DAGSTUHL
REPORTS

Volume 10, Issue 1, January 2020

Dagstuhl Reports, Vol. 10, Issue 1 **ISSN 2192-5283**

ISSN 2192-5283

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at
http://www.dagstuhl.de/dagpub/2192-5283

Publication date
June, 2020

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at `http://dnb.d-nb.de`.

Digital Object Identifier: 10.4230/DagRep.10.1.i

Aims and Scope
The periodical *Dagstuhl Reports* documents the program and the results of Dagstuhl Seminars and Dagstuhl Perspectives Workshops.
In principal, for each Dagstuhl Seminar or Dagstuhl Perspectives Workshop a report is published that contains the following:
- an executive summary of the seminar program and the fundamental results,
- an overview of the talks given during the seminar (summarized as talk abstracts), and
- summaries from working groups (if applicable).

This basic framework can be extended by suitable contributions that are related to the program of the seminar, e. g. summaries from panel discussions or open problem sessions.

Contact
Schloss Dagstuhl – Leibniz-Zentrum für Informatik
Dagstuhl Reports, Editorial Office
Oktavie-Allee, 66687 Wadern, Germany
reports@dagstuhl.de
http://www.dagstuhl.de/dagrep

Report from Dagstuhl Seminar 20021

Spoken Language Interaction with Virtual Agents and Robots (SLIVAR): Towards Effective and Ethical Interaction

Edited by

Laurence Devillers[1], Tatsuya Kawahara[2], Roger K. Moore[3], and Matthias Scheutz[4]

1 **CNRS – Orsay, FR,** devil@limsi.fr
2 **Kyoto University, JP,** kawahara@i.kyoto-u.ac.jp
3 **University of Sheffield, GB,** r.k.moore@sheffield.ac.uk
4 **Tufts University – Medford, US,** matthias.scheutz@tufts.edu

Abstract

This report documents the outcomes of Dagstuhl Seminar 20021 "Spoken Language Interaction with Virtual Agents and Robots (SLIVAR): Towards Effective and Ethical Interaction". Held in January 2020, the seminar brought together world experts on spoken language processing and human-robot interaction. The aims of the seminar were not only to share knowledge and insights across related fields, but also to cultivate a distinct SLIVAR research community. In this report, we present an overview of the seminar program and its outcomes, abstracts from stimulus talks given by prominent researchers, a summary of the 'Show and Tell' demonstrations held during the seminar and open problem statements from participants.

Seminar January 5–10, 2020 – http://www.dagstuhl.de/20021

2012 ACM Subject Classification Computing methodologies → Natural Language Processing, Computer systems organization → Robotics, Computing methodologies → Philosophical/theoretical foundations of artificial intelligence

Keywords and phrases human-robot interaction, spoken language processing, virtual agents

Digital Object Identifier 10.4230/DagRep.10.1.1

Edited in cooperation with Ali Mehenni, Hugues and Skidmore, Lucy

1 Executive Summary

Laurence Devillers (CNRS – Orsay, FR)
Tatsuya Kawahara (Kyoto University, JP)
Roger K. Moore (University of Sheffield, GB)
Matthias Scheutz (Tufts University – Medford, US)

Motivation and aims

Recent times have seen growing interest in spoken language-based interaction between human beings and so-called "intelligent" machines. Presaged by the release of Apple's Siri in 2011, speech-enabled devices – such as Amazon Echo, Google Home, and Apple HomePod – are now becoming a familiar feature in people's homes. Coming years are likely to see the appearance of more embodied social agents (such as robots), but, as yet, there is no clear theoretical basis, nor even practical guidelines, for the optimal integration of spoken language interaction with such entities.

One possible reason for this situation is that the spoken language processing (SLP) and human-robot interaction (HRI) communities are fairly distinct, with only modest overlap. This means that spoken language technologists are often working with arbitrary robots (or limit themselves to conversational agents), and roboticists are typically using off-the-shelf spoken language components without too much regard for their appropriateness. As a consequence, an artefact's visual, vocal, and behavioural affordances are often not aligned (such as providing non-human robots with inappropriate human-like voices), and usability suffers – the human-machine interface is not "habitable".

These usability issues can only be resolved by the establishment of a meaningful dialogue between the SLP and HRI communities. Both would benefit from a deeper understanding of each other's methodologies and research perspectives through an open and flexible discussion. The aim of the seminar was thus to bring together a critical mass of researchers from the SLP and HRI communities in order to (i) provide a timely opportunity to review the critical open research questions, (ii) propose appropriate evaluation protocols for speech-based human-robot interaction, (iii) investigate opportunities to collect and share relevant corpora, and (iv) consider the ethical and societal issues associated with such machines.

Participants

A broad range of expertise was represented by the seminar participants, with a total of 38 attendees including industry experts, PhD students and academics from 14 different countries. The research areas of this interdisciplinary group included SLP, robotics, virtual agents, HRI, dialogue systems, natural language processing, as well as other intersections of SLIVAR.

Seminar overview

The seminar began with short presentations from all attendees, providing them an opportunity to introduce themselves and their research, as well as share their insights on challenges and opportunities in SLIVAR. The presentations were interwoven with four stimulus talks given by leading experts in their respective fields. In light of these presentations, participants formed discussion groups based on the clustering of related topics. The seminar's schedule was intentionally adaptable to allow for discussions to shift and new groups to form over the course of the week. Alongside discussions, "Show and Tell" sessions were organised to provide participants an opportunity to demonstrate their work and further stimulate discussion.

A non-exhaustive list of topics covered are outlined below along with a selection of the questions discussed within groups.

- Adaptability
  - *How do you cope with the frontier between user adaptation and system adaptation?*
  - *Are there representations that better enable adaptivity to users?*
- Architecture
  - *What are the desiderata for a spoken dialogue system-robot architecture?*
- Ethics
  - *What can we do as scientists and engineers to create ethical agents?*
  - *Should a robot be able to pursue goals that you do not know?*
- Evaluation
  - *How do we evaluate HRI systems effectively and efficiently?*
  - *What are the existing evaluation approaches for SLIVAR?*

- Interaction
  - *How do we bridge the gap between dialogue management and interaction management?*
  - *What kind of interaction modules are useful for dialogue and why?*
- Multimodality
  - *What are the minimum representations units for different modalities?*
  - *What is the added value of multimodal features of spoken interaction in HRI?*
- Natural Language Understanding (NLU) Scalability
  - *How should we approach large scale supervised learning for NLU?*
- Speech in Action
  - *How can we create challenging interaction situations where speech performance is coordinated to a partner's action?*
- Usability
  - *What are the use cases for SLIVAR systems?*
  - *What is the role of physical or virtual embodiment?*

Seminar outcomes

The topics and questions outlined above facilitated a stimulating week of discussion and interdisciplinary collaboration, from which several next steps were established. These include participation in a number of workshops, special sessions and conferences, including but not limited to:

- SIGdial 2020 Special Session on Situated Dialogue with Virtual Agents and Robots [1]
- HRI 2020 Second Workshop on Natural Language Generation for HRI [2]
- IJCAI 2020 ROBOTDIAL Workshop on Dialogue Models for HRI [3]
- 29th IEEE International Conference on Robot & Human Interactive Communication [4]
- Interspeech 2020 [5]

Research and position papers were also discussed, specifically focusing on the evaluation and ethics of SLIVAR systems. For the former, suggestions included a survey of existing evaluation approaches, a report paper on issues in SLIVAR and HRI evaluation, and investigations into the automation of SLIVAR system objective evaluation. For the latter, next steps included a survey of existing architectures for embedded ethical competence and a position paper on ethical machine learning and artificial intelligence.

The final, and perhaps most valuable outcome of the seminar was the establishment of a new SLIVAR community. There was a strong enthusiasm for the discussions during the seminar to continue with a second SLIVAR meeting, as well as suggestions for growing the community through the formal establishment of a special interest group. Overall, the seminar provided a unique opportunity to create a foundation for collaborative research in SLIVAR which will no doubt have a positive impact on future work in this field.

[1] https://www.sigdial.org/files/workshops/conference21/
[2] https://hbuschme.github.io/nlg-hri-workshop-2020/
[3] http://sap.ist.i.kyoto-u.ac.jp/ijcai2020/robotdial/
[4] http://ro-man2020.unina.it/
[5] http://www.interspeech2020.org/

2 Table of Contents

3 Overview of Stimulus Talks

3.1 Ethical Issues in SLIVAR

Laurence Devillers, (CNRS – Orsay, FR)

The new uses of affective social robots, conversational virtual agents, and the so-called "intelligent" systems, in fields as diverse as health, education or transport reflect a phase of significant change in human-machine relations which should receive great attention.

What ethical issues arise from the development of spoken language Interaction with Virtual Agents and Robots (SLIVAR)? Human-chatbot/robot interaction raises the crucial issue of trust, especially for conversational agents who assist vulnerable people. Are nudging machines (SDS) using affective computing and cognitive biases ethical?

The Dilemma of the researchers is on the one hand, to achieve the highest performance with conversational virtual agents and robots (close to or even exceeds human capabilities) but on the other hand, to demystify these systems by showing that they are "only machines": on the one hand, the designers of conversational agents seek for many to imitate, simulate the dialogical behaviour of humans, on the other hand, users spontaneously anthropomorphise the conversational agents" capacities and lend them human understanding. The Media Equation [1] explains that people tend to respond to media/computer/robot as they would either to another person by being polite, cooperative, attributing personality characteristics such as aggressiveness, humor, expertise, and even gender depending on the cues they receive from the media. So, an object "which seems to be in the pain", as the robot Atlas of Boston Dynamics, can inspire some empathy. Asking users not to project human traits on machines is not enough, as some reactions may even appear in spite of this knowledge.

At LIMSI-CNRS, we build agents that can recognise, interpret, process and simulate human language and affect (even a kind of machine-humor). With the capacity of interpretation of the emotional state of humans, a robot can adapt his behaviour and give an appropriate response to these emotions. Naturally, it interacts differently with different individuals. The planned scientific work in our chair HUMAAINE focuses on the detection of social emotions in human voice, and on the study of audio and spoken language manipulations (nudges), intended to induce changes in the behaviour of the human interlocutor. A nudge is an indirect suggestion or subtle reminder intended to influence people's behaviour (Richard Thaler: Nobel Prize in Behavioural Economy, Nov 2017). A "nudge" is a tactic of subtly modifying behaviour of a consumer. Nudging mainly operates through the affective system. Nudges work by making use of our cognitive biases and « irrational » way in decision-making our cognitive capacities are limited, we are lacking self-control, we act emotionally, we act by conformity, we act by laziness, etc.

Nudging could be used in a near future in chatbots and social robots: to incentivise purchase, to influence behaviour that may be and may not be desired by users. The first results from an original pre-experiment, conducted by the proposed HUMAAINE Chair's team in June 2019 in partnership with an elementary school, shows that an AI machine (Pepper robot or Google Home) is more efficient at nudging than adults. Our aim is to study these interactions and relationships, in order to audit and measure the potential influence of affective systems on humans, and finally to go towards a conception of "ethical systems by design" and to propose evaluation measures.

The question of liability arises for designers and trainers of virtual conversational agents. There are several design factors that give rise to ethical problems:

1. Specification problem: a complete specification of a virtual agent is impossible. Laws and rules of conduct in society are formulated in natural language. It is illusory to believe that a complete and precise formulation of natural language elements is possible in a computer language (common sense, irony, culture...).
2. Learning bias: Some of the virtual agent models are trained from data selected by a "coach" (human agent in charge of selecting them). The agent can be discriminating, not fair if the data are badly chosen.
3. Learning without understanding: A virtual agent learns from data, but unlike a human being, he does not understand the meaning of the sentences he generates or perceives.
4. Learning instability: Mistakes are inevitable when a learning system classifies data that do not resemble, or falsely resemble, the data contained in the corpus used for learning. The problem of system robustness is important.
5. Impossible to rigorously evaluate a virtual agent: Dialogue is inherently dynamic. It is difficult to reproduce behaviour or results.
6. Confusion of status: The attribution of a name and a personality to the virtual agent. Maintaining such confusion raises ethical issues. The risk is one of decision manipulation (nudging), isolation, and machine addiction.
7. Trust in virtual conversational agents: Human-agent interaction raises the crucial issue of trust, especially for conversational agents who help vulnerable people. They are currently neither transparent nor evaluated.

Ii is important to consider the level of trust in a virtual agent, its capabilities and limits and the capabilities and limits of the pair it forms with the user. Some ethical principles have been proposed by the EU experts:

- Beneficence: promoting well-being, preserving dignity, and sustaining the planet
- Non-maleficence: privacy, security and "capability caution"
- Autonomy: the power to decide
- Justice: promoting prosperity and preserving solidarity
- Transparency and Explicability: enabling the other principles through intelligibility and accountability

For example, the objectives of the IEEE SA – P7008 WG which is a working group with public and private partners are:

- understanding human behaviour , nudging and manipulation of choice with spoken dialogue system
- understanding AI nudging applications with public and private partners,
- discussing ethical solutions that guide people to do what's in their best interest and well-being,
- proposing norms and standards for these ethical solutions.

Conversational virtual agents and robots using autonomous learning systems and affective computing will change the game around ethics. We need to build long-term experimentation to survey Human-Machine Co-evolution and to build ethics by design chatbots and robots.

References

1 Byron Reeves and Clifford Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, Cambridge, U.K., 1996.

3.2 Problems and Questions in SLIVAR

Tatsuya Kawahara (Kyoto University, JP)

While smartphone assistants and smart speakers are prevailing and there are high expectations, spoken language interaction with virtual agents and robots (SLIVAR) is not effectively deployed. It is necessary to analysethe reasons and explore use cases. In this talk, first, the dialogue tasks are categorised based on the service types and goals. Next, a variety of social robots and virtual agents are compared according to the affordance. Then, component technologies including ASR, TTS, SLU, dialogue management, and non-verbal processing are reviewed with the focus on critical points for SLIVAR. Finally, evaluation and ethical issues are addressed. Research Questions:

1. Why social robots/agents are not prevailing in society?
2. What kind of tasks are robots/agents expected to conduct?
3. What kind of robots/agents are suitable (for the task)?
4. Why spoken dialogue (speech input) is not working with robots?
5. What kind of other modalities and interactions are useful?
6. What kind of evaluations should be conducted?

3.3 Grounded Language Acquisition for Robotics

Cynthia Matuszek (University of Maryland, Baltimore County, US)

For this stimulus talk, Cynthia summarised her current research on grounded language acquisition for human-robot interaction, which she defines as "extracting semantically meaningful representations of human language by mapping those representations to the noisy, unpredictable physical world in which robots operate" [1]. In absence of an abstract, see [2] for her research related to the presentation and [3] for her overview of grounded language acquisition, including future directions and challenges that remain.

References

1 Cynthia Matuszek. UMBC, Department of Computer Science and Electrical Engineering: Cynthia Matuszek Bio. https://www.csee.umbc.edu/~cmat/index.html Accessed: 03-04-2020.

2 Nisha Pillai, Cynthia Matuszek and Francis Ferraro. Deep Learning for Category-Free Grounded Language Acquisition. In *Proc. of the NAACL Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics*, NAACL-SpLU-RoboNLP, Minneapolis, MI, USA, June 2019. http://iral.cs.umbc.edu/Pubs/PillaiNAACLws2019.pdf

3 Cynthia Matuszek. Grounded Language Learning: Where Robotics and NLP Meet (early career spotlight). In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, IJCAI, Stockholm, Sweden, July 2018. http://iral.cs.umbc.edu/Pubs/MatuszekIJCAI2018_earlycareer.pdf

3.4 Socially Aware Virtual Interaction Partners

Catherine Pelachaud (Sorbonne University – Paris, FR)

During an interaction, we adapt our behaviours on several levels: we align ourselves linguistically (vocabulary, syntax, level of formality), but also our behaviours (we respond to the smile of our interlocutor, we imitate the posture, the gestural expressiveness...), our conversational strategies (to be perceived more warm or competent), etc. This multi-level adaptation can have several functions: reinforcing engagement in interaction, emphasising our relationship with others, showing empathy, managing the impression we give to others.... The choice of verbal and non-verbal behaviours and their temporal realisation are markers of adaptation. Adaptation, which can take the form of mimicry, synchronisation, alignment, is an important factor in interaction. Several researchers have worked on Embodied Conversational Agents ECAs that can be adapted during an interaction that focuses on imitation, relationship building, empathy [1, 2, 3].

We have conducted several studies to provide the ACA with the capacity for interaction. First, we developed models for agents who were either speakers or interlocutors [4]. Today, we are turning our attention to interaction itself; that is, we are interested in developing agents capable of aligning their behaviours with those of their interlocutors, imitating them, synchronising with them [5]. We are also working to give agents the ability to reason about the expressions they display, to measure their potential impact on their interlocutors. Our current models go beyond our first work on modelling the backchannels of interlocutors [6]. In this first approach, a set of rules specified when a backchannel could be triggered. Then we focused our attention on the ability to equip the virtual agent with the ability to enter into behavioural resonance with his interlocutors. We defined a dynamic coupling model to modulate the level of synchronisation between the agent and his interlocutors [7]. The agent is considered as a dynamic system in constant evolution in real time. The states representing the agent's behaviour can be modified and adapted to allow the emergence of synchronisation between the interlocutors. We have conducted several studies to measure the impact of this motor resonance capacity on the quality of interaction perceived by users. We first evaluated this model between two virtual interactions [8]. We also applied it to the agent capable of laughing and the user listening to funny music (generated according to Peter Schickele's PDQ Bach model) [9]. Virtual agents that are able to synchronise dynamically with other agents or human users are perceived as being more socially involved in the interaction than agents that only send backchannels but do not show any motor resonance.

Now, we focus on the ability to equip the virtual agent with the ability to adapt his behaviour with his interlocutors. We have developed several models that address different aspects of adaptation during an interaction. Over the past two years, we have developed an architecture that allows an ACA to adapt to the non-verbal behaviours of the user during an interaction. We conducted three studies on different levels of adaptation: conversational strategy, nonverbal behaviours, multimodal signals. Each adaptation mechanism has been implemented in the same architecture that includes multimodal analysis of user behaviour using the Eyesweb platform [10], a dialogue manager (Flipper [11]), our virtual agent GRETA. The architecture was adapted to each study. The same scenario was used for the three studies carried out at the Musée des sciences de la Cité des sciences et de l'industrie de Paris. The agent was used as a guide for an exhibition on video games at the Science Museum.

Each of these three studies involved between 70 and 100 participants and followed a similar protocol. Participants first completed a questionnaire based on the NARS questionnaire, used in robotics, to measure their apriori about virtual agents, then interacted with the agent and finally answered other questionnaires on their perception of the agent and interaction. Several hypotheses have been validated, in particular with regard to the competent condition (study 1) and the condition in which the agent adapted his smile to the user's smile (study 3). Study 2 also highlighted the primacy of the warm dimension. In each of the studies, the agent who adapted his or her behaviours to maximise the participants' impression or level of engagement was better perceived. The different coping mechanisms, whether in conversational strategies, non-verbal behaviours or signals, have helped to improve the user experience of the interaction.

References

1 Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. Empathy in virtual agents and robots: A survey. In *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3):11, 2017.
2 Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. Virtual rapport 2.0. In *International Workshop on Intelligent Virtual Agents*, pages 68-79, 2001. Springer.
3 Ran Zhao, Tanmay Sinha, Alan W Black, and Justine Cassell. Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavi In *International conference on intelligent virtual agents*, pages 218-233, 2016. Springer.
4 Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark Ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, et al. Building autonomous sensitive artificial listeners. *IEEE Transactions on Affective Computing*, 3(2):165-183, 2012.
5 Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schroeder. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing*, 3(1):69-87, 2012.
6 Elisabetta Bevacqua, Maurizio Mancini, and Catherine Pelachaud. A listening agent exhibiting variable behaviour. In Helmut Prendinger, James C. Lester, and Mitsuru Ishizuka, editors, *Proceedings of 8th International Conference on Intelligent Virtual Agents*, IVA 2008, Lecture Notes in Computer Science, volume 5208, pages 262-269, Tokyo, Japan, 2008. Springer.
7 Magalie Ochs, Catherine Pelachaud, and Gary Mckeown. A user perception-based approach to create smiling embodied conversational agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1):4, 2017.
8 Ken Prepin, Magalie Ochs, and Catherine Pelachaud. Beyond backchannels: co-construction of dyadic stance by reciprocal reinforcement of smiles between virtual agents. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35, 2013.
9 Maurizio Mancini, Beatrice Biancardi, Florian Pecune, Giovanna Varni, Yu Ding, Catherine Pelachaud, Gualtiero Volpe, and Antonio Camurri. Implementing and evaluating a laughing virtual character. *ACM Transactions on Internet Technology*, 17(1):1-22, 2017.
10 Gualtiero Volpe, Paolo Alborno, Antonio Camurri, Paolo Coletta, Simone Ghisio, Maurizio Mancini, Radoslaw Niewiadomski, and Stefano Piana. Designing multimodal interactive systems using eyesweb xmi. In *SERVE@ AVI*, pages 49-56, 2016.
11 Jelte van Waterschoot, Merijn Bruijnes, Jan Flokstra, Dennis Reidsma, Daniel Davison, Mariët Theune, and Dirk Heylen. Flipper 2.0. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pages 43-50, November 2018.

4 Show and Tell

A number of participants were able to share their work through two "Show and Tell" sessions during the seminar. Participants provided videos of robots and virtual assistants in the lab, previews of prototypes and works-in-progress, as well as live demonstrations. Below are some examples of the work that was presented.

4.1 Android ERICA

Figure 1 The android "ERICA".

Presenter: Tatsuya Kawahara (Kyoto University, JP)
Demonstration: ERICA is an android that can be engaged in human-level conversation including attentive listening and job interview.
Further Information: http://www.sap.ist.i.kyoto-u.ac.jp/erato/index-e.html

4.2 Creating a Voice for the MiRo Biomimetic Robot

Presenter: Roger K. Moore (University of Sheffield, GB)
Demonstration: MiRo is the first commercial biomimetic robot to be based on a hardware and software architecture that is modelled on the biological brain. In particular, MiRo's vocalisation system was designed, not using pre-recorded animal sounds, but based on the implementation of a real-time parametric general-purpose mammalian vocal synthesiser tailored to the specific physical characteristics of the robot. The novel outcome has been the creation of an "appropriate" voice for MiRo that is perfectly aligned to the physical and behavioural affordances of the robot, thereby avoiding the "uncanny valley" effect and contributing strongly to the effectiveness of MiRo as an interactive agent.

Figure 2 Seminar participants interacting with MiRo.

References

1 Roger K. Moore and Ben Mitchinson. Creating a Voice for MiRo, the World's First Commercial Biomimetic Robot. In *Proceedings of INTERSPEECH 2017*, pages 3419–3420, Stockholm, 2017.

2 Roger K. Moore and Ben Mitchinson. A biomimetic vocalisation system for MiRo. In M. Mangan, M. Cutkosky, A. Mura, P. F. M. J. Verschure, T. Prescott, and N. Lepora (Eds.), *Living Machines 2017, LNAI 10384*, pages 363–374, Stanford, CA, 2017. Springer International Publishing.

4.3 Incremental Spoken Dialogue and the Platform for Situated Intelligence

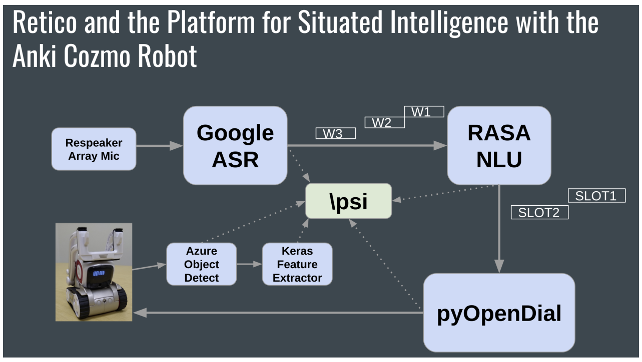


Figure 3 Module architecture for voice-controlled navigation.

Presenter: Casey Kennington (Boise State University, US)
Demonstration: Voice controlled navigation for the Anki Cozmo robot using multimodal, incremental spoken dialogue processing with the Platform for Situated Intelligence and ReTiCo frameworks. Cozmo was able to perform simple navigation commands.
Further Information: https://anki.com/en-us/cozmo.html
https://github.com/microsoft/psi
https://github.com/Uhlo/retico

4.4 Furhat – A Social Robot for Conversational Interaction

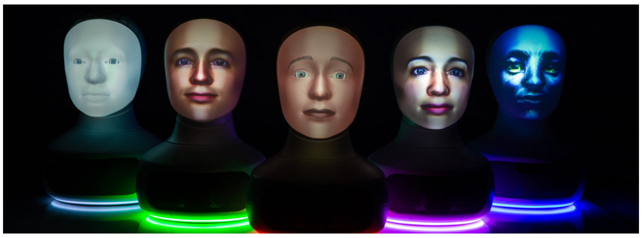


Figure 4 The various personas of Furhat.

Presenter: Gabriel Skantze (KTH Royal Institute of Technology – Stockholm, SE)
Demonstration: Furhat started as a research project at KTH and span off into the company Furhat Robotics in 2014. Furhat is a hardware and software platform that allows researchers and developers to build social robotics applications. Hardware-wise, Furhat has a back-projected face that allows for changing the persona of the robot, as well as expressing subtle facial expressions and accurate lip movement. Software-wise, Furhat provides a platform for building multimodal conversational interactions with Furhat. Furhat is being used by many research groups worldwide and by for real-world applications by companies such as Deutsche Bahn (travel information), Merck (medical screening) and TNG (recruitment interviews).
Further Information: www.furhatrobotics.com

References

1. Al Moubayed, S., Skantze, G., and Beskow, J. The Furhat Back-Projected Humanoid Head – Lip reading, Gaze and Multiparty Interaction. In *International Journal of Humanoid Robotics*, 10(1), 2013.
2. Skantze, G. and Al Moubayed, S. IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *Proceedings of ICMI*. Santa Monica, CA, 2012.
3. Skantze, G. Real-time Coordination in Human-robot Interaction using Face and Voice. *AI Magazine*, 37(4):19-31, 2016.

4.5 VoxHead

Figure 5 VoxHead the humanoid robot.

Presenter: Michael C. Brady (American University of Central Asia, KG)
Demonstration: VoxHead is a humanoid robot, developed to be an "open access" research tool for human-robot [speech] interaction. It is composed of 3D printed parts and off-the-shelf components. The idea is that hobbyists and researchers can build these robots for a fraction of the cost of commercial alternatives. Current work involves developing an operating system for the robot that runs interactive dialogues written in "Fluidscript." This scripting approach is derived from the W3C standard of VoiceXML, and is similar to writing behaviours s for today's Amazon Alexa, except where video input and motor output are both incorporated to specify multimodal interactions.
Further Information: www.fluidbase.com

5 Individual Contributions from Participants

There are several open problems related to SLIVAR still to explore. This section of the report includes statements from attendees providing their perspective on challenges and opportunities within the field.

5.1 Bridging the Habitability Gap

Bruce Balentine (Entreprise Integration Group – Zürich, CH)

Basic design philosophy

Anthropomorphism introduces many challenges, among them ethical, uncanny valley, practical implementation and user mind-reading problems. The anthropomorphic goal of "just like human-to-human conversation" (point **z** in Figure 6) leads to the habitability gap via the "death of a thousand cuts." But what's the alternative?

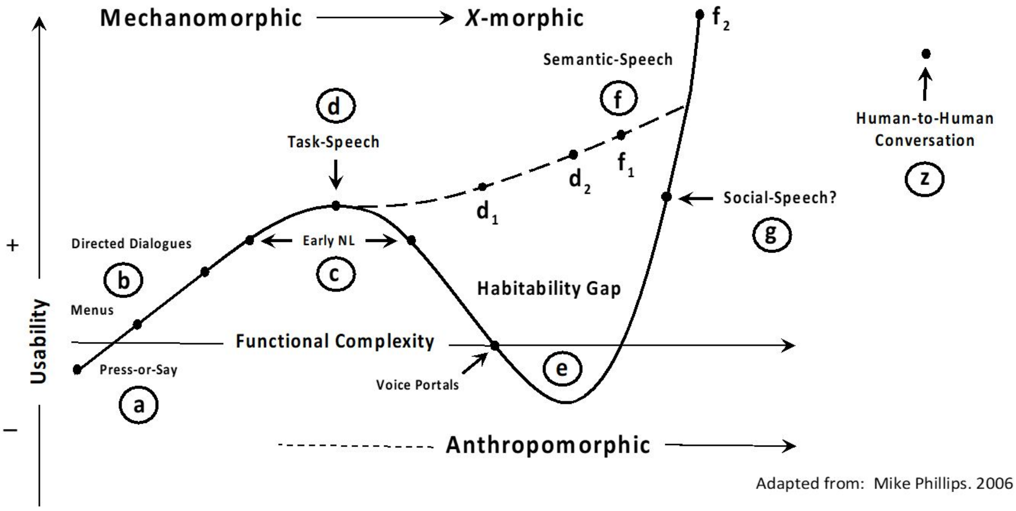


Figure 6 Bridging the Habitability Gap.

Mechanomorphism

A mechanomorphic interface presents itself unabashedly as a machine. Such an entity has no mind. It is emotionally agnostic and socially unaware. It is very skilled within its knowledge and task domain, with affordances that convey stability, learnability and discoverability. Mechanomorphism is achieved through the use of:

- Designative meaning in use of language;
- Repetition and reusability;
- Multimodality (tones, visual display, touch and gesture);
- Consistent use of mechanical UI devices (e.g. signposts); and,
- Well-integrated discoverability features.

TaskSpeech (**d**) is mechanomorphic, and features such affordances. It is built on a smart, half-duplex turn taking foundation. It evolves into (d_1) and (d_2) with full-duplex turn taking, enhanced semantic capabilities and expanded self, user and environment modeling. I am currently involved in a collaboration to develop a TaskSpeech proof-of-concept.

X-Morphism

An X-morphic interface has a mind, allowing (within limits) ostensive-inferential communication and recursive mind-reading skills, including "Theory-of-Mind." But the system is not conscious, sentient nor sapient, making it an "alien mind" – unlike but compatible with human minds. Since we've never built an alien mind before, the X stands for experimental. It is emotionally and socially aware, but not participative. As with mechanomorphism, X-morphic interfaces are task-oriented (their job is to get work done), as measured by traditional HCI usability metrics. The realisation of X-morphism is semantic speech (**f**) featuring all of the capabilities of TaskSpeech plus extensive negotiation skills, meta-knowledge and meta-cognition, and extended world knowledge within specific domains. (f_2) extends the range of domains. Social speech (**g**) is semantic speech with full-blown emotional and social competencies.

My roadmap

My trajectory across the habitability gap is: (**d**) a couple of years, to (**f**) a decade, to (**g**) unknown.

5.2 Ubiquity of Computing and SLIVAR

Timo Baumann (Universität Hamburg, DE)

Human-computer spoken language interaction has come a long way, and audio-only, task-only virtual agents are carried around by many people on their phones. While some level of functionality for such interactions can now be achieved (as measured by "task success"), interactions with today's agents are still far from the natural ideal. This is all the more true for spoken language interaction with robots which fight not only the interdisciplinary issues and additional technical complexity involved but also the the unnaturalness of either modality involved and amplified by the combination of these imperfect modalities.

At the same time, computing has become ubiquitous, and cloud-based computing enables us to access our data from an ever growing multitude of devices, from TV and smart speaker via laptop and tablet to smartphones and connected earphones. The advent of 5G networks will help to virtualise even more computation into the cloud while keeping latencies low enough to not be a nuisance in spoken interaction (and robotics). Internet of Things sensors will provide access to all sorts of sensory data. Thus, ubiquitous computing also has the potential to radically improve and change the way that we interact with machines.

These changes come with numerous challenges and opportunities, some of which I try to summarise below:

1. While moving through an ubiquitous computing environment, e.g., leaving the breakfast table, walking down the stairs and to your car, an agent's realisation should move along and seamlessly transition between devices and modalities, from a possibly embodied agent at the kitchen table to your phone and then your car (or bicycle). The handover poses interesting technical challenges but the more interesting ones are those of availability of modalities (e.g. think twice before reading out e-mails on the subway).
2. Future systems will want to manage and exploit the wealth of data that they acquire about their users from all the modalities and sensors involved. Critical questions here are the users awareness of when she is being observed by the system and how this makes her feel (e.g., does the system observe conversations the user is having with other people?).
3. With the opportunities growing both in system performance and availability, the model of "natural conversation" will limit human-computer interaction. There is no need for a human to be polite to a system, to not interrupt it. Likewise, a system can blend speech, song, signalling noises, etc. in ways that a human never could. It will be interesting to see what forms of "supernatural" sociolect evolve for talking to machines.
4. Human-human interaction, society and culture is easily influenced. Thus, the means of interaction, the assumptions and rules that we design for our future spoken language interaction systems will feed back into human-human language, and from there into society. Already, kids (falsely) ascribe all sorts of properties to Alexa; likewise, female-sounding voices of spoken language systems influence the role model for real(!) females. This influence thus will not only yield an exciting field for research but more importantly requires ethical, societal and cultural far-sightedness from each developer and researcher of spoken language interaction systems.

5.3 SLIVAR Needs Models of Interactional Intelligence

Hendrik Buschmeier (Universität Bielefeld, DE)

Main reference Hendrik Buschmeier: "Attentive Speaking. From Listener Feedback to Interactive Adaptation". PhD thesis, Faculty of Technology, Bielefeld University, Bielefeld, Germany, 2018.
URL https://doi.org/10.4119/unibi/2918295

Language use is successful when we understand the interaction partner and are able to make ourselves understood. Problems in understanding are a major source of frustration in spoken language interaction with virtual agents and robots (SLIVAR), because artificial conversational agents, even in restricted domains, are usually not always able to understand what a human user means – unless users restrict themselves to a specific way of expressing their intention. Although such approaches may work in principle, SLIVAR in this way may feel unnatural and non-spontaneous to users and makes exploration of what a conversational agent can do for the user – discoverability is a general usability-problem in speech-based interfaces – difficult. Both aspects may contribute to the limited acceptance of SLIVAR.

Problems in understanding (non-understanding, partial understanding and misunderstanding) are, however, not limited to SLIVAR. They are prevalent in human communication as well – to an extent that it is argued that "language use is inherently problematic" and that "miscommunication is not a failure but part and parcel of the act of communication" [4, p. 765]. Humans can, however, deal with these problems and repair them interactively through communication.

This insight could be an opportunity for future research on SLIVAR. When problems in understanding the user arise, artificial conversational agents should not give up, but actively try to "come to an understanding" [4, p. 769] with the user by interactively working with the them to make themselves better understood. The ability of human interactive language use is, according to Levinson [3], based on the *human interaction engine*, which provides us with *interactional intelligence*.

In this statement, I want to argue, that artificial conversational agents should be endowed with computational models of interactional intelligence, which would allow them to interactively come to an understanding with their interaction partners – in both the speaking and listening roles – and will thus likely be "better" communicators. In previous work, we have computationally modelled a simple form of interactional intelligence based on the dialogue phenomenon of multimodal communicative feedback [1] and could show that an agent equipped with this model communicates more efficiently and that humans rated it more helpful in resolving their understanding difficulties [2].

References

1 Hendrik Buschmeier. *Attentive Speaking. From Listener Feedback to Interactive Adaptation*. PhD thesis, Faculty of Technology, Bielefeld University, Bielefeld, Germany, 2018. https://doi.org/10.4119/unibi/2918295

2 Hendrik Buschmeier and Stefan Kopp. Communicative listener feedback in human–agent interaction: Artificial speakers need to be attentive and adaptive. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*, pp. 1213–1221, Stockholm, Sweden, 2018.

3 Stephen C. Levinson. On the human "Interaction Engine". In Nick J. Enfield and Stephen C. Levinson, editors, *Roots of Human Sociality: Culture, Cognition and Interaction*, pp. 39–69. Berg, Oxford, UK, 2006.
4 Edda Weigand. Misunderstanding: The standard case. *Journal of Pragmatics*, 31:763–785, 1999. https://doi.org/10.1016/s0378-2166(98)00068-x

5.4 The Role of Social/Moral Norms in SLIVAR

Nigel Crook (Oxford Brookes University, GB)

We must go beyond simply identifying the ethical issues that arise from spoken language interaction with virtual agents if we are to make those interactions more habitable. In my view, an agent's (artificial or human) ability to recognise and observe the moral and social norms that surround spoken interaction goes to the heart of what facilitates habitability. This is because these norms embed some core expectations that people have about their interactions with other agents, guiding what it is morally and socially acceptable to say and do and what is not. For example, in human spoken interaction the relative "status" (for want of a better word) of those engaging in dialogue (adult – adult, adult – child, boss – employee, friend – friend, stranger – stranger, salesperson – customer, etc) will strongly influence both what is said and the manner in which it is spoken. The spatial/physical context of the interaction can also set expectations on what verbal interactions are morally and socially acceptable (a conversation in a child's bedroom, or in a class room, in the office, or at home, etc). More significantly, the cultural context (regional, generational etc) in which the interaction occurs will set the moral and social expectations of the human participants and determine the habitability of that interaction for them. The moment moral or social norms are violated in a spoken interaction, the less habitable that interaction becomes.

Here are some key questions/tasks that I think we need to address here:

1. Identify the role of social/moral norms in spoken language interaction with virtual agents and understand their impact on "habitability".
   This is very challenging as many of these social/moral norms are not directly articulated. They are often acquired by learning through interaction.
2. How can this impact be measured?
   I'm not sure there is a metric here – but it may be possible to evaluate how comfortable people are with certain spoken interactions.
3. Explore how virtual agents and robots can be equipped with sufficient social/moral competence to facilitate habitability?
   Some work has already been done on this – "top-down" (i.e. rule based) and "bottom-up" (machine learning based) approaches are presented in the literature.
4. Determine the functional aspect of these systems that embody/reveal this social/moral competence to users (e.g. choice of vocabulary, tone of voice, posture, bodily gestures)
   Again, difficult to determine how these functional aspects can meet the social/moral expectations of human conversational partners.
5. Accommodating regional/cultural variations in the social/moral norms exhibited through spoken language interaction

It is clear that social/moral norms do not cross cultural boundaries very well. But there are questions here about how these are to be accommodated so that an AI agent can operate in multiple cultural contexts without limiting habitability.

5.5 Human-robot Interactions and Affecting Computing: The Ethical Implications

Laurence Devillers, (CNRS – Orsay, FR)

Social and emotional robotics wants to create companion robots, which are supposed to provide us for example with therapeutic assistance or even monitoring assistance. So, it is necessary to learn how to use these new tools without fear and to understand their usefulness. We need to demystify the artificial intelligence, elaborate ethical rules and put the values of the human being back at the center of the design of these robotic systems. Affective robots and chatbots bring a new dimension to interaction and could become a mean of influencing individuals.

Since the early studies of human behaviour , emotion has attracted the interest of researchers in many disciplines of neuroscience and psychology. Recent advances in neuroscience are highlighting connections between emotion, social functioning, and decision making that have the potential to revolutionise our understanding of the role of affect. Cognitive neuroscience has provided us with new keys to understanding human behaviour , new techniques (such as neuroimaging) and a theoretical framework for their evaluation. The American neuroscientist A. Damasio [1, 2, 3] has suggested that emotions play an essential role in important areas such as learning, memory, motivation, attention, creativity, and decision making. More recently, it is a growing field of research in computer science and machine learning. *Affective Computing* aims at the study and development of systems and devices that use emotion, in particular in human computer and human robot interaction. It is an interdisciplinary field spanning computer science, psychology, and cognitive science. The *affective computing* field of research is related to, arises from, or deliberately influences emotion or other affective phenomena [4]. The three mains technologies are emotion detection and interpretation, dialogue reasoning using emotional information and emotion generation and synthesis.

An affective chatbot or robot is an autonomous system that interacts with humans using affective technologies to detect emotions, decide and to simulate affective answers. It can have an autonomous natural language processing system with at least these components: signal analysis and automatic speech recognition, semantic analysis and dialogue policies, response generation and speech synthesis. The agent can be just a voice assistant, an 2D or 3D on-screen synthetic character or a physically embodied robot. Such artefact has several types of AI modules to develop perceptive, decision-making, and reactive capabilities in real environment for a robot or in virtual world for synthetic character. The robot is a complex object, which can simulate cognitive abilities but without human feelings, nor that desire or "appetite for life' that Spinoza talks as *conatus* (effort to persevere in being) which refers to everything from the mind to the body. Attempts to create machines that behave intelligently often conceptualise intelligence as the ability to achieve goals, leaving unanswered a crucial question: whose goals?

The robotics community is actively creating affective companion robots with the goal of cultivating a lifelong relationship between a human being and an artifact. Enabling autistic children to socialise, helping children at school, encouraging patients to take medications and protecting the elderly within a living space is only few samples of how they could interact with humans. Their seemingly boundless potential stems in part from the fact that they can be physically instantiated, i.e., they are embodied in the real world, unlike many other devices. Social robots will share our space, live in our homes, help us in our work and daily life and also share a certain story with us. Why not give them some machine humour? Humour plays a crucial role in social relationships: it dampens stress, builds confidence and creates complicity between people. If you are alone and unhappy, the robot could joke to comfort you; if you are angry, it could help you to put things into perspective, saying that the situation is not so bad. It could also be self-deprecating if it makes mistakes and realises it!

At Limsi-CNRS, we are working to give robots the ability to recognise emotions and be empathetic, so that they can best help their users. We teach them to dialogue and analyze emotions using verbal and non-verbal cues (acoustic cues, laughter, for example) in order to adapt their responses[5, 6]. How are these "empathetic" robots welcomed? To find out, it is important to conduct perceptual studies on human–machine interaction. Limsi-CNRS has conducted numerous laboratory and Ehpad tests with elderly people, or in rehabilitation centers with the Association Approche [6], as part of the BPI ROMEO2 project, led by Softbank robotics. Created in 1991, the main mission of the Association Approche is to promote new technologies (robotics, electronics, home automation, information and communication technologies, etc.) for the benefit of people in a situation of disability regardless of age and living environment. We are exploring how the expression of emotion is perceived by listeners and how to represent and automatically detect a subject's emotional state in speech6 but also how to simulate emotion answers with a chatbot or robot. Furthermore, in real-life context, we often have mixtures of emotions [7] .We also conducted studies around scenarios of everyday life and games with Professor Anne-Sophie Rigaud's team at the Living Lab of Broca Hospital. All these experiments have shown that robots are quite well-accepted by patients when they have time to experiment with them. Post-experimental discussions also raised a number of legitimate concerns about the lack of transparency and explanation of the behaviour of these machines. The winner of the Nobel Prize in economics, the American Richard Thaler, highlighted in 2008 the concept of nudge, a technique that consists in encouraging individuals to change their behaviour without constraining them using their cognitive biases. The behaviour of human beings is shaped by numerous factors, many of which might not be consciously detected. Thaler and Sunstein [8] advocate "*libertarian paternalism*", which they see as being a form of weak paternalism. From their perspective, "*Libertarian Paternalism is a relatively weak, soft, and non-intrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened*". Numerous types of systems are already beginning to use nudge policies (ex: Carrot, Canada, for health). Assuming for the time being that nudging humans for their own betterment is acceptable in at least some circumstances, then the next logical step is to examine what form these nudges may take. An important distinction to draw attention to is between "positive" and "negative" nudges (sludges) and whether one or both types could be considered ethically acceptable. The LIMSI team in cooperation with a behavioural economist team in France in the Chair AI HUMAAINE HUman-MAchine Affective spoken INteraction & Ethics au

[6] http://www.approche-asso.com/

CNRS (2019-24) will set up experiments with a robot capable of nudges with several types of more or less vulnerable population (children, elderly) to develop nudge assessment tools to show the impact (Project BAD NUDGE BAD ROBOT[7]). The principal focus of this project is to generate discussion about the ethical acceptability of allowing designers to construct companion robots that nudge a user in a particular behavioural direction for different purposes. At the laboratory scale, then in the field, the two teams will study whether fragile people are more sensitive to nudges or not. This research is innovative, it is important to understand the impact of these new tools in the society and to bring this subject on ethics and manipulation by machines internationally[8]. The objects will address us by talking to us. It is necessary to better understand the relationship to these chatty objects without awareness, without emotions and without proper intentions. Users today are not aware of how these systems work, they tend to anthropomorphise them. Designers need to avoid these confusions between life and artifacts to give more transparency and explanation on the capabilities of machines.

Social roboticists are making use of empirical findings from sociologists, psychologists and others to decide their spoken interaction designs, and effectively create conversational robots that elicit strong reactions from users. From a technical perspective, it is clearly feasible that robots could be encoded to shape, at least to some degree, a human companion's behaviour by using verbal and non-verbal cues. But is it ethically appropriate to deliberately design nudging behaviour in a robot?

The imagination of the citizens about robotics and more generally artificial intelligence are mainly founded on science-fiction and myths (*Golem* Myth [9]). To mitigate fantasies that mainly underline gloomy consequences, it is important to demystify the affective computing, robotics and globally-speaking AI science. For example, the expressions used by experts such as "the robots understand emotions", "they make decisions", "the robots will have a consciousness", are not understood as metaphors by those outside the technical research community. The citizens are still not ready to understand the concepts behind these complex AI machines. These emerging interactive and adaptive systems using emotions modify how we will socialise with machines and with humans. These areas inspire critical questions centering on the ethics, the goals and the deployment of innovative products that can change our lives and society [10]. Anthropomorphism is the attribution of human traits, moods, emotions, or intentions to non-human entities. It is considered to be an innate tendency of human psychology. It is clear that the multiple forms of the voice assistants and affective robots already in existence and in the process of being designed will have a profound impact on human life and on human-machine co-adaptation. Human machine co-adaptation is related to how AI is used today to affect people autonomy (in decision, perception, attention, memorisation, ...) by nudging and manipulating them.

Systems have become increasingly capable of mimicking human behaviour through research in affective computing. These systems have provided demonstrated utility, for interactions with vulnerable populations (e.g., the elderly, children with autism). The behaviour of human beings is shaped by several factors, many of which might not be consciously detected. Marketers are aware of this dimension of human psychology as they employ a broad array of tactics to encourage audiences toward a preferred behaviour. One of the main questions in social robotics evaluation is what kind of impact the social robot's appearance has on

[7] https://dataia.eu/en/news/bad-nudge-bad-robot-project-nudge-and-ethics-human-machine-verbal-interaction
[8] https://standards.ieee.org/project/7008.html

the user, and if the robot must have a physical embodiment. The issue is complex, and the Uncanny Valley phenomenon is often cited to show the paradox of increased human likeness and a sudden drop in acceptance. An explanation of this kind of physical or emotional discomfort is based on the perceptual tension that arises from conflicting perceptual cues. When familiar characteristics of the robot are combined with mismatched expectations of its behaviour , the distortion in the category boundary manifests itself as perceptual tension and feelings of creepiness. A solution to avoid the uncanny valley experience might be to match the system's general appearance (robot-like voice, cartoon-like appearance) with its abilities. This can prevent users from expecting behaviour that they will not "see".

Alternatively, users can be exposed to creatures that fall in the uncanny valley (e.g. Geminoids), making the public more used to them. Humans tend to feel greater empathy towards creatures that resemble them, so if the agent can evoke feelings of empathy in the user towards itself, it can enhance the user's natural feeling about the interaction and therefore make communication more effective. Following the reasoning on perceptual categorisation, the robot's appearance as a pleasantly familiar artificial agent and its being perceived as a listening and understanding companion to the user can establish a whole new category for social robots which, in terms of affection and trust, supports natural interaction between the user and the robot.

The research challenge is to build autonomous machines able to learn just by observing the world. For a digital system, autonomy "is the capacity to operate independently from a human operator or from another machine by exhibiting nontrivial behaviour in a complex and changing environment" [11]. In April 2016, Microsoft's Tay chatbot, which had the capacity to learn continuously from its interactions with web users, start racist language after just 24 hours online. Microsoft quickly withdrew Tay. Affective computing and curiosity models will be among the next big research topics. Self-supervised learning systems will extract and use the naturally available relevant context, emotional information and embedded metadata as supervisory signals. Researchers such as A. Bair (MIT lab) created an "Intrinsic Curiosity Model," a self-supervised reinforcement learning system.

The integration of intentionality and human-like creativity is a new area of research. These machines are called "intelligent" because they can also learn. For a robot, the task is extremely difficult because it has neither instinct nor intentions to make decisions. It can only imitate the human being. Giving a robot the ability to learn in interaction with the environment and humans, is the Holy Grail of artificial-intelligence researchers. It is therefore desirable to teach them the common values of life in society. The ability to learn alone constitutes a technological and legal breakthrough, and raises many ethical questions. These robots can be, in a way, creative and autonomous in their decision making, if they are programmed for this. Indeed, according to the American neuroscientist A. Damasio, self-awareness comes from the pleasant or unpleasant feelings generated by the state of homeostasis (mechanisms aimed at the preservation of the individual) of the body. "Consciousness" is a polysemic term: for some, it refers to self-awareness, for others to the consciousness of others, or to phenomenal consciousness, moral consciousness, etc. To be conscious, you need a perception of your body and feelings. The robots would need an artificial body with homeostatic characteristics "similar to ours" to be conscious. The goal of researchers such as K. Man and A. Damasio [12] is to test the conditions that would potentially allow machines to care about what they do or "think". Machines capable of implementing a process resembling homeostasis is possible using soft robotics and multi-sensory abstraction. Homeostatic robots might reap behavioural benefits by acting as if they have feeling. Even if they would never achieve full-blown inner experience in the human sense, their properly motivated behavioural would result in expanded intelligence and better-behaved autonomy. The initial goal of the introduction of physical vulnerability and self-determined self-regulation is not to create robots with authentic feeling, but rather to improve their functionality across a wide range of environments. As a second goal, introducing this new class of machines would constitute a scientific platform for experimentation on robotic brain–body architectures. This platform would open the possibility of investigating important research questions such as "*To what extent is the appearance of feeling and consciousness dependent on a material substrate?*"

How can we assess a system that learns? What decisions can and cannot be delegated to a machine learning system? What information should be given to users on the capacities of machine learning systems? Who is responsible if the machine malfunctions: the designer, the owner of the data, the owner of the system, its user, or perhaps the system itself? What will be the power of manipulation of the voices of these machines? What responsibility is delegated to the creators of these chatbots/robots?

In my book "Robots and Humans: Myths, Fantasies and Reality" [9], I propose to enrich Asimov's laws with commands adapted to life-assistant robots. The foundations of these commandments come in part from feedback from experiences of interactions between elderly people and robots. In my new book, "Emotional robots: health, surveillance, sexuality... and the ethics of it all?" (L'Observatoire, March 2020), I describe these "artificial friends" which will take a growing place in society. Just as an airplane does not flap its wings like a bird to fly, we build machines that can imitate without feeling, speak without understanding, and reason without consciousness. While their role can be extremely positive, particularly in the field of health, the risks of manipulation are also real: emotional dependence, isolation, loss of freedom, amplification of stereotypes (80% of these artefacts have voices, names – Alexa, Sofia – and women's bodies, which turn them into servile assistants or sex robots). Will they be an extension of ourselves? How far will we go to program an emergence of artificial consciousness? Conversational virtual agents and robots using autonomous learning systems and affective computing will change the game around ethics. We need to build long-term experimentation to survey Human-Machine Co-evolution and to build *ethics by design* chatbots and robots.

Developing an interdisciplinary research discipline with computer scientists, doctors and cognitive psychologists to study the effects of co-evolution with these machines in a long-term way is urgent. The machine will learn to adapt to us, but how will we adapt to it? Machines will be increasingly autonomous, talkative and emotionally gifted through sophisticated artificial-intelligence programs?

References

1 A. Damasio. *Descartes' Error: Emotion, Reason, and the Human Brain*. HarperCollins, 1994.
2 A. Damasio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace, New York, 1999.
3 A. Damasio. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Houghton Mifflin Harcourt, New York, 2003.
4 R. W. Picard. *Affective Computing*. MIT Press, Cambridge, MA, 1997.
5 L. Devillers, M. Tahon, M. Sehili, and A. Delaborde. Detection of affective states in spoken interactions: robustness of non-verbal cues, *TAL*, 55(2), 2014.
6 L. Devillers, M. Tahon, M. Sehili, and A. Delaborde. Inference of human beings' emotional states from speech in human-robot interactions. *International Journal of Social Robotics*, 7(4):451–463, 2015.

7 L. Devillers, L. Vidrascu and L. Lamel. Challenges in real-life emotion annotation and machine learning based detection, *Neural Networks*, 18(4):407–422, May 2005.
8 R. H. Thaler and C. R. Sunstein. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, 2008.
9 L. Devillers. *Des robots et des hommes : Mythes, fantasmes et réalité*. (French) [Robots and Humans: myths, fantasies and reality]. Plon, Paris, France, 2017
10 IEEE. Ethically Aligned Design Version 2. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf
11 A. Grinbaum, R. Chatila, L. Devillers, J. G. Ganascia. Ethics in Robotics Research: CERNA Mission and Context. *IEEE Robotics and Automation Magazine*, 24(3):139–145, 2017.
12 K. Man and A. Damasio. Homeostasis and soft robotics in the design of feeling machines, *Nature Machine Intelligence*, 1:446–452, October 2019. https://doi.org/10.1038/s42256-019-0103-7

5.6 SLIVAR in Education

Johanna Dobbriner (TU Dublin, IE)

The role of robots or virtual agents in their interaction with humans can take many forms, e.g. for production of goods, as a personal assistant, companion, guide, story teller, teacher, and more. Frequently, for that interaction work, spoken language will be involved. Appropriate and engaging dialogue needs to be planned and executed by the virtual agent/robot automatically and in real time.

My own research mainly concerns the use of virtual agents in teaching and education and accordingly both the technical and ethical aspects of that use case are of special interest to me.

Ethical concerns:

- When used to teach children, what ethical constraints are there to consider?
- How can these constraints be built into such a system?
- Are the current measures we take for privacy and data protection sufficient?
- How human-like or realistic should a virtual agent be?

Technical aspects:

- Measuring engagement in real time
- Strategies to keep and increase user engagement
- Customisation of dialogue and interaction
- Generation of realistic conversation
- Integrating multimodal aspects of conversation e.g. speech, intonation, gestures and facial expressions
- Modelling context and situational awareness in the interactions
- Specific needs for dialogue systems in teaching

With researchers from different disciplines involved in SLIVAR coming together, some of the above points may be addressed and discussed from multiple perspectives.

5.7 Face-to-face Conversation with Socially Intelligent Robots

Mary Ellen Foster (University of Glasgow, GB)

The overall shape of my research programme is influenced by two well-known facts about interaction:

1. Humans are inherently social creatures who tend to respond socially to any form of technology, whether it is human-like or not [1] This phenomenon is only further enhanced when the technology is embodied in an approximately human-like robot.
2. Face-to-face conversation is the basic form of human communication, and is also arguably the richest possible communications method. Talking to other humans face-to-face permits full, incremental, multimodal communication on all channels – spoken language, prosody, facial displays, proxemics, body posture, body gestures. As Bavelas et al. [2] point out, a useful exercise is to analyse other communication systems by enumerating the ways in which they differ from full face-to-face dialogue.

Taken together, the above facts mean that, when deploying robots into human spaces, it is entirely unavoidable that people will treat those robots as social actors, and will desire and attempt to have a full, multimodal, face-to-face conversation with them. Developers of such robots must acknowledge and understand this phenomenon – and, more importantly, they must also design systems in such a way that the robots are able to detect and respond appropriately to human attempts to engage them in a face-to-face conversation.

More specifically, I am interested in addressing the following research questions in this context:

- What are user expectations of a humanoid robot? How can we understand and manage those expectations (e.g., through appearance, behaviour, or management of the social situation) so that the robots behave in a coherent, predictable way and that users are not disappointed by potentially unrealistic expectations being violated?
- How can a robot allow humans to employ the strategies developed from human-human interaction? More specifically: how can a robot detect and classify the multimodal, conversational social signals of humans in its area? And, how can and should it choose actions to respond appropriately to those signals?
- What are contexts (use cases, scenarios) where a socially intelligent robot can be deployed where it provides an actually useful service for the humans in that space? Many current robot deployment scenarios feel like they were chosen for convenience rather than in response to an actual user demand for a robot to enter those spaces.
- What are the ethical considerations when deploying socially intelligent robots into human spaces? If a robot displays "appropriate" social signals, there is a potential risk of users ascribing intentions and even "feelings" to that robot that are not intended. How can we ensure that these effects are not exploited, especially with vulnerable populations?

References

1 Byron Reeves and Clifford Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, Cambridge, U.K., 1996.
2 Janet Bavelas, Sarah Hutchinson, Christine Kenwood, and Deborah Hunt Matheson. Using Face-to-face Dialogue as a Standard for Other Communication Systems. *Canadian Journal of Communication*, 22(1), 1997. https://doi.org/10.22230/cjc.1997v22n1a973

5.8 Building Casual Conversation

Emer Gilmartin (ADAPT Centre – Dublin, IE)

While every student of natural language processing is familiar with Jelinek's 1988 quote "Whenever I fire a linguist our system performance improves", perhaps not so many know his later[9] remark on the same topic "It is our task to figure out how to make use of the insights of linguists". Insights from linguistics, and pragmatics in particular, into social talk should give engineers a basis on which to model such interaction, from knowledge of the kind of data required to features desirable in such talk. Current interest in companion applications and social bots evoke the well-defined natural phenomenon of human social or casual talk. Such "talk for the sake of talking" ranges from short greeting and small talk routines to longer story swapping or discussion, and serves strong social goals – building and maintaining social bonds, entertainment, informing participants of their interlocutors' personality, values, feelings and affect.

The structure of casual talk involves light chat interleaved with monologic "chunks" of narrative or extended opinion, where one participant dominates, and others contribute feedback. It is clearly not possible to model the entirety of such conversations as a series of adjacency pairs. Chunks are generic, and could be modelled separately, but chat is not easy to specify. Human chat comprises a series of statements and sometimes questions on a topic, interspersed with backchannels and short positive comments with little or no additional information. Contributions often include two elements – a short coda-like acknowledgement of or comment on the previous speaker's utterance, followed by a new statement or question, either related to the current topic or, when a topic is exhausted, shifting or changing topic. At times of topic exhaustion, conversations have been observed to "idle", with participants producing only short and generic comments (the first element of the two part contributions described above), for a number of turns until somebody introduces a new topic. These structures reflect the inherent mixed-initiative nature of real conversation, which is thought to involve a levelling of role, power and status differentials between interlocutors, with equally distributed speaker rights.

An example stretch of casual chat is illustrated below:

1. **A:** I saw the match last night.
2. **B:** Oh. Wasn't Messi great?
3. **A:** Yeah, but Ronaldo really messed up.
4. **B:** Yeah. I guess so.
5. **A:** Yep. He really fluffed that penalty.
6. **B:** Yeah. But a great match overall.
7. **A:** Yeah. It was, wasn't it?
8. **B:** Yeah, good stuff.
9. **A:** Yeah. Are you golfing much these days?
10. **B:** Yeah, a bit, but not as much as I'd like.

In the snippet above, the two-part contributions can be seen in turns 2, 3, 5, 6, 9 and 10, with a short "backward-facing" segment, followed by a "forward-facing" segment advancing the conversation. Turns 4, 7 and 8 are examples of idling, with low-content segments only. In

[9] Antonio Zampolli Prize acceptance speech, LREC 2004

the deep learning dialogue modelling community, a number of features have been identified for use in controlling generated talk. Examples are repetition (undesirable), response- relatedness (desirable), and specificity (desirable). The origin of these features is unclear, and may be evidence of a growing "folk pragmatics" in the domain. Considering these three features in the light of pragmatics research, their desirability or otherwise is more nuanced. Repetition is often noted as a failing in S2S models, with systems generating short high frequency low information responses. However, such responses do occur in talk, but they are only the first half of most turns, (or feedback in chunk phases). The challenge is how to generate both elements for realistic social talk. In addition, human interaction involves interlocutor alignment, often posited to be necessary for efficient grounding and processing, and created through lexical and syntactic priming. While, in very short conversations, this may not manifest, once natural conversation gets longer, the level of repetition of words and phrases grows. Thus, a certain level of repetition is necessary and desirable in casual talk. The same factors apply to response-relatedness, as the information rich second element in responses needs to be related to the previous turn while a topic is in progress. However, the introduction of unrelated content for topic changes and even seeming non-sequiters is also common in natural talk. Specificity, or the provision of varied language has been identified as a challenge, with measures of word rarity such as Inverse Document Frequency (IDF) used to boost the level of low frequency vocabulary in generated content. However, language in conversation is generally quite simple, with lower lexical density than in written text. More diverse language is desirable in longer chunks, or in demonstrations of humour or irony. Again, seeming flaws in current methods actually work well where less specificity is needed, in the first part of contributions, but the more content rich second parts need more specificity. A major challenge is measuring success in casual talk, which is not easily definable at the utterance level or even in a single conversation, as the function of the conversation is often a broad longitudinal building of good relationships. Assessment of success, beyond simple measurement of time users spend chatting to the system, is challenging.

5.9 Architectures for Multimodal Human-Robot Interaction

Manuel Giuliani (University of the West of England – Bristol, GB)

In my work, I am interested in building architectures for multimodal human-robot interaction (HRI). This means architectures that combine verbal and non-verbal input processing components (to understand the human's utterances) with high-level decision mechanisms (for the robot to decide it's next actions based on the input) and multimodal output generation (for the robot to respond to the human in a socially appropriate way). There are a lot of challenges related to building these architectures, but specifically I am interested in the following research questions:

- How can we combine sub-symbolic and symbolic components in a hybrid HRI architecture that is modular so that data-driven and rule-based approaches can both be tested?
- Can we build HRI architectures that can be adapted to different usage contexts?
- Are there basic dialogue acts (e.g., for initiating and ending an interaction) that can be reused in many different HRI usage contexts?
- In terms of software engineering, can we build reusable HRI architectures that are based on commonly used middlewares like ROS (Robot Operating System) to allow the community to build on previous work?

5.10 SLIVAR based on Transparency, Situatedness and Personalisation

Martin Heckmann (Honda Research Institute Europe GmbH, DE)

When two humans engage in an interaction, two independent minds with different experiences and views on the world come together. To make this joint activity a success, they have to work together. They need to align their mental representations to be able to form a common ground and define a joint goal for the interaction [1, 2, 3, 4, 5]. The communicative signals they use for such an alignment are not limited to the words they utter but encompasses a multitude of potentially multimodal signals such as prosodic variations and gestures [6, 7]. These signals also help to coordinate when the partners take their turns [8]. The necessary alignment between the partners affects a multitude of domains, the phonological, syntactic and semantic level as well as the situation model [2, 3]. With situation model, I want to refer to different aspects of the current situation such as the environment in which the interaction takes place and the progression of the interaction so far. During the interaction they also make assumptions about their partner's mental world, her ability to perceive, understand and judge, often referred to as common sense, her prior knowledge on the topic of the interaction, her goals and intentions and her current state, traits, skills and personal preferences [4]. The larger the differences between the partners' mental worlds, the more work the alignment process needs. A true alignment can never be reached but is also not necessary as long as the joint goals of the interaction are achieved.

Unfortunately, the different building blocks fundamental to human-human communication which I have outlined above, i.e. the capabilities to perceive and interpret multimodal communicative signals, build an adequate model of the environment, keep track of the history of the interaction, reason with common sense, recruit domain knowledge, interpret the partner's goals as well as model her state, traits, skills and preferences, are at best manifested in a very different form in an artificial agent. In today's artificial agents they are often not present at all or only in a rudimentary form. Furthermore, if and how these capabilities are implemented in the artificial agent is in most cases opaque to the user, i.e. the system is not transparent to the user. This means that the alignment process will take a lot of effort from the human and in the end will leave a large gap between the partners' mental models, often too large to achieve the goal of the interaction.

A common and logical approach to improve the interaction is to improve the system's capabilities, to make them closer to that of a human. Many people have investigated the role of prosody [9, 6] as well as how to integrate the information from all available modalities [10, 11, 12, 13, 14]. We have extended this by multimodal integration and personalisation for prosodic analysis [15, 16] and reference resolution [17]. In classical dialogue systems, the situation model is often limited to keeping track of the dialogue history [18, 19, 20]. For human robot interaction, mainly models have been developed to visually perceive the world and link these percepts to internal concepts, words in particular. This process is often referred to as grounding, sometimes also as anchoring, and can be based on pre-existing categories [21, 22, 23] or learned in interaction [24, 25]. The domain knowledge of classical dialogue systems is typically very narrow, based on hand-crafted domains such as bus information [18], restaurant reservation [19], or, more recently, also technical support dialogues [20]. For the interaction with robots it is common to establish the domain knowledge by combining information from external ontologies with the learning of new representations in interaction with the user [26, 22]. Common sense reasoning is then implemented by reasoning on these ontologies. We proposed an approach to automatically acquire task-specific domain knowledge from online text-based resources, e.g. in the context of a repair task [27].

Many of the functionalities mentioned above are still nowhere near the level of a human. In my view, the most challenging and promising domains relevant to improve the interaction with artificial agents are better environment models and domain knowledge combined with an adaptation to the individual user.

To really meet human expectations will most likely require an AI with reasoning capabilities on par with a human. Yet such an AI does not seem to be around the corner [28]. Hence, instead of raising the system's perception and reasoning capabilities, I think a more promising approach is to increase its transparency. If the system states, capabilities and intentions are transparent to the human, the alignment process and hence the overall interaction can be much more efficient [29]. Humans use backchannels, facial expressions and emotional displays to give feedback on the interaction and the progress of the alignment [30, 31, 32, 33]. Backchannels, facial expressions and emotions have been also investigated in the context of human-agent interaction [34, 35, 32]. However, the research on emotions in human-agent interaction typically focuses on detecting the emotions of the human to enable emphatic responses of the agent [36, 37]. I consider using emotional displays and other non-verbal cues to provide insights into the capabilities and mental states of the agent an interesting and very relevant topic to help to align the minds of the partners and to improve the interaction. Connected to this is the need to empower the agent to make its reasoning steps transparent, often referred to as explainable AI [38, 39]. We took one step into this direction by suggesting an approach to create annotations with explanations that can be used to train a system to provide explanations for its reasoning [40]. In general, spoken feedback from the agent bears the risk of triggering very high expectations on the human's side with respect to the agent's understanding and reasoning capabilities. I consider finding good models to convey the agent's limitations via its communication an important direction to increase transparency [41].

In short, I am convinced that faster progress can be made if we focus more on making the limitations of the agent transparent then on trying to overcome them.

References

1. H. H. Clark and S. E. Brennan "Grounding in communication.," in *Perspect. Soc. Shar. Cogn.*, pp. 127–149, American Psychological Association, 1991.
2. M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behav. Brain Sci.*, vol. 27, no. 02, 2004.
3. S. Garrod and M. J. Pickering, "Why is conversation so easy?," *Trends Cogn. Sci.*, vol. 8, no. 1, pp. 8–11, 2004.
4. K. Friston and C. Frith, "A Duet for one," *Conscious. Cogn.*, vol. 36, pp. 390–405, nov 2015.
5. T. Scott-Phillips, *Speaking Our Minds: Why human communication is different, and how language evolved to make it special.* Red Globe Press, 2015.
6. J. Hirschberg, "Communication and prosody: Functional aspects of prosody," *Speech Commun.*, vol. 36, no. 1-2, pp. 31–43, 2002.
7. P. Wagner, Z. Malisz, and S. Kopp, "Gesture and speech in interaction: An overview," *Speech Commun.*, vol. 57, pp. 209–232, 2014.
8. H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn-Taking for Conversation," *Language (Baltim).*, vol. 50, no. 4, p. 696, 1974.
9. Y. Sagisaka, N. Campbell, and N. Higuchi, eds., *Computing Prosody.* New York, NY: Springer US, 1997.

10 M. Turk, "Multimodal interaction: A review," *Pattern Recognit. Lett.*, vol. 36, no. 1, pp. 189–195, 2014.
11 S. Oviatt, B. Schuller, P. R. Cohen, D. Sonntag, G. Potamianos, and A. Krüger, eds., *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1.* ACM Books, 2018.
12 R. Stiefelhagen, C. Fügen, P. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel, "Natural human-robot interaction using speech, head pose and gestures," in *2004 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, vol. 3, pp. 2422–2427, 2004.
13 K. Funakoshi, M. Nakano, T. Tokunaga, and R. Iida, "A unified probabilistic approach to referring expressions," in *Proc. 13th Annu. Meet. Spec. Interes. Gr. Discourse Dialogue*, pp. 237–246, 2012.
14 C. Kennington and D. Schlangen, "A simple generative model of incremental reference resolution for situated dialogue," *Comput. Speech Lang.*, vol. 41, pp. 43–67, 2017.
15 M. Heckmann, "Audio-visual word prominence detection from clean and noisy speech," *Comput. Speech Lang.*, vol. 48, pp. 15–30, 2018.
16 A. Schnall and M. Heckmann, "Feature-space SVM adaptation for speaker adapted word prominence detection," *Comput. Speech Lang.*, vol. 53, pp. 198–216, jun 2019.
17 D. Kleingarn, N. Nabizadeh, M. Heckmann, and D. Kolossa, "Speaker-adapted neural-network-based fusion for multimodal reference resolution," in *Proc. 20th Annu. SIGdial Meet. Discourse Dialogue*, (Stockholm, Sweden), pp. 210–214, Association for Computational Linguistics, 2019.
18 J. Williams, A. Raux, D. Ramachandran, and A. Black, "The dialogue state tracking challenge," in *Proc. SIGDIAL 2013 Conf.*, pp. 404–413, Association for Computational Linguistics, 2013.
19 M. Henderson, B. Thomson, and S. Young, "Word-based dialogue state tracking with recurrent neural networks," in *Proc. 15th Annu. Meet. Spec. Interes. Gr. Discourse Dialogue*, pp. 292–299, Association for Computational Linguistics (ACL), 2014.
20 K. Yoshino, C. Hori, J. Perez, L. F. D'Haro, L. Polymenakos, C. Gunasekara, W. S. Lasecki, J. K. Kummerfeld, M. Galley, C. Brockett, J. Gao, B. Dolan, X. Gao, H. Alamari, T. K. Marks, D. Parikh, and D. Batra, "dialogue System Technology Challenge 7," in *Proc. NIPS2018 2nd Conversational AI Work.*, 2018.
21 D. K. Misra, J. Sung, K. Lee, and A. Saxena, "Tell me Dave: Context-sensitive grounding of natural language to manipulation instructions," *Int. J. Rob. Res.*, 2016.
22 S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human–robot interaction: An implementation," *Artif. Intell.*, vol. 247, pp. 45–69, jun 2017.
23 L. Fischer, S. Hasler, J. Deigmöller, T. Schnürer, M. Redert, U. Pluntke, K. Nagel, C. Senzel, A. Richter, and J. Eggert, "Where is the tool? – grounded reasoning in everyday environment with a robot," in *Proc. Int. Cogn. Robot. Work.*, CEUR Workshop Proceedings, 2018.
24 D. K. Roy and A. P. Pentland, "Learning words from sights and sounds: a computational model," *Cogn. Sci.*, vol. 26, pp. 113–146, jan 2002.
25 C. Matuszek, "Grounded language learning: Where robotics and NLP meet," in *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2018-July, pp. 5687–5691, 2018.
26 M. Tenorth and M. Beetz, "KnowRob: A knowledge processing infrastructure for cognition-enabled robots," *Int. J. Rob. Res.*, vol. 32, no. 5, pp. 566–590, 2013.
27 N. Nabizadeh, D. Kolossa, and M. Heckmann, "MyFixit : An Annotated Dataset , Annotation Tool , and Baseline Methods for Information Extraction from Repair Manuals," in *Proc. Twelth Int. Conf. Lang. Resour. Eval.*, 2020.

28 K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, "Viewpoint: When will ai exceed human performance? Evidence from ai experts," *J. Artif. Intell. Res.*, vol. 62, pp. 729–754, 2018.

29 R. K. Moore and M. Nicolao, "Toward a needs-based architecture for "intelligent" communicative agents: Speaking with intention," *Front. Robot. AI*, vol. 4, p. 66, dec 2017.

30 A. Eshghi, C. Howes, E. Gregoromichelaki, J. Hough, and M. Purver, "Feedback in conversation as incremental semantic update," in *IWCS 2015 – Proc. 11th Int. Conf. Comput. Semant.*, pp. 261–271, 2015.

31 K. P. Truong, R. Poppe, I. De Kok, and D. Heylen, "A multimodal analysis of vocal and visual backchannels in spontaneous dialogs," in *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 2973–2976, 2011.

32 C. Lang, S. Wachsmuth, M. Hanheide, and H. Wersing, "Facial Communicative Signals – Valence Recognition in Task-Oriented Human-Robot Interaction," *Int. J. Soc. Robot.*, 2012.

33 P. Ekman, *Emotions revealed: recognizing faces and feelings to improve communication and emotional life.* Times Books, 2003.

34 G. Skantze, A. Hjalmarsson, and C. Oertel, "Turn-taking, feedback and joint attention in situated human-robot interaction," *Speech Commun.*, vol. 65, pp. 50–66, 2014.

35 C. Becker, S. Kopp, and I. Wachsmuth, "Why Emotions should be Integrated into Conversational Agents," in *Conversational Informatics An Eng. Approach* (T. Nishida, ed.), pp. 49–67, John Wiley & Sons, 2007.

36 H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," in *2011 IEEE Int. Conf. Autom. Face Gesture Recognit. Work. FG 2011*, pp. 827–834, 2011.

37 L. Devillers, M. Tahon, M. A. Sehili, and A. Delaborde, "Inference of Human Beings' Emotional States from Speech in Human–Robot Interactions," *Int. J. Soc. Robot.*, vol. 7, pp. 451–463, aug 2015.

38 A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, "Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda," in *Conf. Hum. Factors Comput. Syst. – Proc.*, vol. 2018-April, 2018.

39 A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

40 N. Attari, M. Heckmann, and D. Schlangen, "From Explainability to Explanation: Using a Dialogue Setting to Elicit Annotations with Justifications," in *Proc. 20th Annu. SIGdial Meet. Discourse Dialogue*, (Stockholm, Sweden), pp. 331–335, Association for Computational Linguistics, 2019.

41 R. K. Moore, "Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction," in *Lect. Notes Electr. Eng.*, vol. 999 LNEE, pp. 281–291, 2017.

5.11 On Boundary-Crossing Robots

Kristiina Jokinen (AIST – Tokyo Waterfront, JP)

My work has dealt with communicative competence and enablements of communication that are crucial in modelling natural interaction between humans and between humans and intelligent agents. Within the cascaded dialogue modelling framework, based on the cycle of contact, perception, understanding, and reaction, such enablements are seen as preconditions for the interaction to proceed in a smooth manner. In particular, my research interests have focused on situational awareness and its signaling by eye-gaze and gesturing. When interacting with social robots, the same enablements are assumed to be valid, albeit with slightly different behaviour patterns, and social robots, enabling spoken interaction with humans, are not only sophisticated computers with a capability to quickly process huge amounts of data, but they are also perceived as interactive agents with an ability to communicate with human partners using natural language. Given the robot's dual characteristics as a computer and as an interactive agent, the main issues in HRI are related to technological enablements to support natural language interaction and to the modelling of complicated issues in the interaction between humans and robots.

Interesting issues in the SLIVAR context concern situational awareness of the agent and the aspects of communication enablements that can be learnt from the human-animal interaction in this respect. Such issues do not only concern speech and vocalisation, not even affect and emotion, but becoming aware of the partner and of the partner being ready for communication. Such recognition of the partner's communicative intention is an essential part of the basic enablements of communication (contact with the partner), and monitoring of the partner is crucial for the dialogue dynamics.

The current HRI and dialogue systems in general lack this kind of situational awareness. They are designed to answer certain factual questions and do some preliminary inferencing, but not to monitor their environment. In my research I have focused on eye-gaze studies and checking the gaze-patterns in various situations from which we can infer the partner's emotional and affective state, as well as comparing gaze patterns between humans and between humans and robots.

From this view point, one of the challenges that I see crucial for the SLIVAR workshop concern timing of the robot's reaction and its coordination with the human partner. How can the dialogue model take into account such immediate reaction and then, deliberate on the higher-level communicative aspects as cooperation, collaboration and planning together with human agents. What kind of dialogue model and interaction architecture would support this kind of functionality and processing of observations.

Another question concerns symbiotic relation between humans and robots. Can robots and robot interaction be as natural as that with pet animals, different but still accepted as one type of the many interactions humans can have with their environment and the world in general. I have introduced the concept of "Boundary Crossing Robot' which refers to robotic agents capable of interacting with humans in everyday life situations, and thus gradually shifting the boundaries of what are typical interactive agents in our environment that we need to take into account and which can take us into account when dealing with everyday tasks. The boundary crossing refers to conceptual boundaries that humans have of the structure of the world, e.g. what are the agents we can communicate with, and thus crossing of the boundary refers to accepting robots as agents which can interact. Another boundary concerns acceptance of robots as partners and co-workers. BCRs are to facilitate interaction and mutual intelligibility between different perspectives among humans and HRI, and especially pave way towards the views of Society 5.0 where the robotic agents and human agents can have symbiotic relation and co-habit the world.

5.12 Human-level Spoken Dialogue Processing for Multimodal Human-Robot Interaction

Tatsuya Kawahara (Kyoto University, JP)

Following the success of spoken dialogue systems (SDS) in smartphone assistants and smart speakers, a number of communicative robots are being developed and commercialised. Since robots have a face and a body, the interaction is essentially multimodal. I have investigated spoken dialogue with robots in the context of multimodal interaction. Compared with the conventional SDSs, people tend to talk to a robot in a closer manner to talking to a human (or a pet?) because of the anthropomorphism and physical presence. This poses fundamental changes in the design and methodology of dialogue and interaction, since the conventional SDSs are designed as a human-machine interface. For example, you don't need a robot just to ask for weather information or news. And a robot should detect when you speak even without pressing a button or saying a magic word.

We first need to explore desirable tasks and interactions conducted by humanoid robots engaged in spoken dialogue. These obviously depend on the character design of the robots, and I focus on long and deep interactions such as counseling and interview, which have a definite task but do not have observable goals. They will expand the potential of communicative robots. Then, we need to enhance the methodology of spoken dialogue processing including speech recognition and synthesis for human-robot interaction. Moreover, non-verbal processing also needs to be incorporated. In particular, smooth turn-taking and real-time feedback including backchannels are critically important for keeping the user engaged in the dialogue, so the interaction will be duplex consisting of not only speaking but also attentive listening.

We are investigating human-level dialogue and developing an autonomous android ERICA. Our ultimate goal is to make the android fully autonomous, passing a "total Turing test", but we also hope we can make clear what is essential and missing in the current technology for natural dialogue and interaction through this grand challenge.

5.13 Dialogue and Embodiment as Requirements for Understanding

Casey Kennington (Boise State University, US)

Symbol grounding for holistic semantics

If machines have any chance of fully understanding humans, they need to be able to learn how to interact with humans *on human terms*; that is, using the most natural and effective communication medium that humans use with each other: speech. Learning speech requires a learning the semantic meaning of language, yet words are not just symbols or high-dimensional vectors, they have connections with the physical (*symbol grounding*) and social (*conversational grounding*) world. A holistic word meaning would, therefore, need to ground into physical modalities such vision, proprioperception, interoception, etc., requiring physical embodiment and the ability to interact and manipulate physical objects. Moreover, the *setting* for grounding into the physical modalities is co-located, social interaction with other agents. This physical embodiment in a co-located, interactive speech setting for holistic semantics requires spoken dialogue with robots.

Anthropomorphisation

However, when humans are confronted with robots, they tend to anthropomorphise those robots based on their physical characteristics (e.g., they assign gender and age) and their interactive abilities (e.g., intelligence, or even sincerity). Humans assign affect and valence to robotic behaviours (e.g., a robot with certain face configurations is perceived as *angry*, or a robot that takes a long time to respond is perceived as *confused*). This has implications for the kinds of settings in which they hope to use robots. For example, a robot with a certain color, shape, and facial features will be perceived as friendly or unfriendly in a setting of hospital assistive care. Robots need to have physical characteristics that are amenable to setting in which they are in and what they are tasked with. If the task is to learn semantic concepts (or if learning semantic concepts is a byproduct of another task) the physical characteristics of the robot play a role in the dynamics of the interaction.

Concrete vs. abstract concepts

Concrete linguistics concepts denote physical things; abstract concepts do not, though concrete concepts and be used abstractly (e.g., one can talk about a *dog* without a physical dog being present). Concrete concepts are learned in physical environments, such as referring to physical objects. Abstract concepts are only learned in more social contexts–language building upon language–concepts that are required for holistic semantics, but since concrete concepts are holistically learned in embodied agents, and abstract concepts build on concrete concepts, abstract concepts are likely best learned also in embodied robots.

Social vs. business

Robots and dialogue systems are often task-based with pre-defined goals. Robots and dialogue systems that are optimised for those tasks tend to only have language and affordances (respectively) to complete those tasks. I refer to these as *all business* in that they aren't concerned with socially acceptable (or sometimes required) behaviour. This is of course acceptable because they are machines. Conversely, some systems–robots or dialogue systems–focus more on social aspects, such as social robots or chatbots that don't accomplish any task beyond socialising. I would call these *social* systems. There is a continuum between *social* and *all business* robots and dialogue systems. The point between the two extremes where one decides to set the business/social side depends on the task. Both are required for holistic semantics: building concrete concepts can happen in all business scenarios, but abstract concepts are more likely to come through more socially-driven systems. Moreover, abstract concepts have been shown to be grounded into elements of social dynamics such as valence and affect more that concrete concepts do.

5.14 Challenges in Processing Disaster Response Team Communication

Ivana Kruijff-Korbayová (DFKI – Saarbrücken, DE)

Abstract

Our current work on team communication processing and teamwork support for disaster response missions provides insights on the communication capabilities robots should ultimately have as team members contributing to such missions. This includes references to complex partially damaged/destroyed dynamic environments with potentially unusual objects, which all pose a challenge for object detection; descriptions of activities and states and knowledge integration over time; it may involve co-present interaction, although more typical is remote interaction, which poses a challenge for orientation and achieving common ground; the teams are complex and teamwork involves multiparty communication, another challenge for common ground modeling; to support shared situation awareness such interaction is often (but not always) multimodal, using a graphical/visual interface, such as a map and/or a video feed.

Emergency first response teams operate in high risk situations and make critical decisions despite partial and uncertain information. In order for technology, such as robots or assistive software agents, to provide optimal support for mission execution, it needs mission knowledge, i.e., run-time awareness and understanding of the current mission goals, team composition, resource allocation, the tasks of the team(s), how and by whom they are being carried out, the state of their execution, etc. Since first responders typically operate under high cognitive load and time pressure, it is paramount to keep the burden of entering mission knowledge into the system at a minimum. It is our goal to develop methods for interpreting the verbal communication in the response team and extracting run-time mission knowledge from it. In [1] we have addressed one particular sub-problem: the recognition of dialogue acts in the communication among the human members of a robot-assisted emergency response team. The acquired mission knowledge is then used to assist the first responders during or after the mission, for example, by supporting the real-time coordination of human and robot actions or by mission documentation generation [2, 3].

The goal we are pursuing is very challenging and requires progress beyond state of the art in many aspects. I describe some of the challenges below.

Noisy speech input

Obviously, one challenge is to deal with noisy speech input. The team members use walkie-talkies and move around in a noisy environment. They may be wearing personal protective equipment. This could have built-in microphones, but currently it does not.

Low-resource domain

There do not exist large amounts of transcribed, let alone otherwise annotated data for disaster response. Recording data in exercises and real missions is only becoming possible with spreading use of digital radio equipment. Obtaining transcriptions and annotations remains a challenge. Obtaining realistic data from robot-assisted missions is even more difficult. Moreover, the content of the team communication varies with the nature of robot involvement in the mission, i.e., the tasks, the degree of robot autonomy and other aspects of the technical system realisation, such as user interface functionality.

Grounding in complex (remote) physical situation and “mixed reality”

The speech refers to the physical environment which is complex, dynamic and nonstandard, e.g., partially destroyed. The team is distributed, which means that they do not fully share the physical/visual context. Moreover, when we deal with robot-assisted disaster response, the human team members are not themselves in the environment, they use robots for remote operation in the danger zone. They share situation awareness in a multimodal way by a combination of speech and a graphical user interface, such as a map with annotations. This means that they refer to objects if the physical world and in the map. We have observed that a kind of mixed or bent reality emerges. Sometimes it is necessary to distinguish the two for proper interpretation, e.g., moving an icon on the map is something else than moving an object in the real world; but sometimes the speech is vague in this respect and the distinction is even not important. BTW, we did not use virtual reality in the experiment we have done so far, but I would expect the worlds to blend even more in that case.

Integrating verbal communication with sensor input

First responders use various sensors for “measuring” the environment, and even more visual and sensor data is available when robots are used. So the verbal communication needs to be together with the visual/sensor input.

Extended/complex “multistep” activities

The activities of the team consist of multiple/many tasks and steps that are related to one another. A point of interest, such as a victim, a hazard source or a fire may be detected by one team and further handled by another team.

Dynamic situation modeling

The physical situation is changing as the mission progresses: resources move around, victims are being extracted, hazard sources neutralised, fires extinguished, etc. The temporal aspect cannot be disregarded in the interpretation of the communication and in the modelling of the situation.

Common ground modelling in multiparty communication

The communication involves multiple team members. Although they normally share the radio communication channel, they explicitly establish pairwise or group threads. Modeling common ground for shared situation awareness needs to take these threads into account and not assume that the team members all have the same mission knowledge.

Overhearing vs. active involvement in communication

Our software agent for interpreting the team communication, as we have conceived it so far, overhears the team communication, it does not itself engage in it as a participant. Nevertheless, there needs to be a possibility to correct or at least reject misunderstandings, i.e., wrong interpretations created by the system, e.g., a wrong or wrongly assigned task. This needs to be done by a non-intrusive way, e.g., by easy to handle text editing in a task management interface, rather than complex clarification dialogues between the system and the user. The functionality we develop for the overhearing agent is a subset of the functionality that will be required when this agent is to actively engage in the team communication.

Communication extended over multiple sessions

Disaster response missions may stretch over extended periods involving multiple sessions with the system or prolonged sessions with team switches. The system needs to handle the integration of knowledge over extended time and support continuing, interrupting and resuming sessions.

References

1 T. Anakina and I. A. Kruijff-Korbayov. Dialogue act classification in team communication for robot assisted disaster response. In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue*, Stockholm, Sweden, September 11-13, 2019.

2 W. Kasper. Team monitoring and reporting for robot-assisted USAR missions. In *International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, IEEE, pp. 246–251, 2016.

3 C. Willms, C. Houy, J.-R. Rehse, P. Fettke and I. A, Kruijff-Korbayov. Team communication processing and process analytics for supporting robot-assisted emergency response. In *International Conference on Safety, Security, and Rescue Robotics (SSRR)*, 2019.

5.15 Personal Statement on SLIVAR

Joseph J. Mariani (CNRS – Orsay, FR)

Background

My interest in participating is to have a better view of the state-of-the-art in those fields, and to encourage the two communities of specialists in robotics and human-machine (spoken language) interaction to work more closely, using best practices, and especially evaluation.

State-of-the-art

Where are we now regarding (spoken) interaction with robots? I recently asked Roger Moore at the LT4All conference, who was to say “stone age”?

Mutual understanding

For many years, researchers thought that spoken interaction with robots was simply adding speech input/output to robots. Things have now slightly progressed, but the two communities still have to better understand each other, their research agenda and their constraints.

HRI evaluation metrics and protocols

Objective and quantitative evaluations have been decisive in the development of workable language technologies, starting in the 80s. It seems that the situation is not comparable in the robotics area. How could the evaluation paradigm be also used in robotics? What would be the evaluation metrics and possible protocols? Conducting evaluation tests with embodied robots appears difficult and costly. Can we conduct them in simulated environments, including virtual agents? To what degree is it similar to using actual robots? Similarly, conducting spoken dialogue systems evaluation with human subjects is costly and time consuming. How can we automatise this process with proper data, metrics and protocols?

Multi-party human(s)-robot(s) interaction

Humans are often able to understand whether they are concerned by a statement issued by other humans. How can we handle this when one or several humans are communicating with one or several robots?

Language portability

Most of the developments concern the English language, while robots will be used by humans who want to keep on speaking their native language. What is the size of the effort to port a spoken dialogue system to another language? Do we have to devote the same effort, or is there a way to adapt the system to a different language? It is usually necessary to benefit from large quantities of speech to develop an application, while they are difficult to gather in spoken dialogue situations. How can we address this difficulty?

Ethics of robots

Who is responsible for the robots behaviour, especially given that it is based on machine learning? Is it the designer, the programmer, the trainer, the seller, the owner?

5.16 The Importance of Aligning Visual, Vocal, Behavioural and Cognitive Affordances

Roger K. Moore (University of Sheffield, GB)

Recent years have witnessed astonishing progress in the development of speech-enabled artefacts. Indeed, the appearance of such "intelligent" personal assistants is often hailed as a significant step along the road towards more natural interaction between human beings and future autonomous social agents (such as robots). However, studies into the usage of such technology suggest that, far from engaging in a promised "natural" conversational interaction, users tend to resort to formulaic language and focus on a handful of niche applications which work for them. Given the pace of technological development, it might be expected that the capabilities of such devices will improve steadily. However, evidence suggests that there is a "habitability gap" in which usability drops as flexibility increases. I have hypothesised that the habitability gap is a manifestation of the "uncanny valley" effect whereby a near human-looking artefact (such as a humanoid robot) can trigger feelings of eeriness and repulsion. In particular, I developed a Bayesian model of the effect which reveals that it can be caused by misaligned perceptual cues. So, for example, a device with an inappropriate (e.g. humanlike) voice can create unnecessary confusion in a user. The Bayesian model suggests that the habitability gap can only be avoided if the visual, vocal, behavioural and cognitive affordances of an artefact are aligned. However, given that the state-of-the-art in these areas varies significantly, this means that the capabilities of an artificial agent should be determined by the affordance with the lowest capability, which probably points to an agent's cognitive abilities as limiting factor. My work in this area suggests that future progress depends on designers taking a whole-system perspective, and that emulating a human being is a recipe for failure. However, I have also raised the possibility that there may be a fundamental limit to the interaction that can take place between mismatched interlocutors (such as humans and machines), and on-going research is looking into the implications for future speech-based human-robot interaction particularly by studying vocal interactivity in-and-between humans, animals and robots. In summary, I have identified two open challenges in the field of spoken language interaction with virtual agents and robots: (a) how to optimise the multimodal coupling between intentional agents and their environments (including other agents), and (b) how to optimise spoken language understanding between mismatched interlocutors.

References

1. Roger K. Moore. A Bayesian explanation of the "Uncanny Valley" effect and related psychological phenomena. *Nature Scientific Reports*, 2(864), 2012.
2. Roger K. Moore. From talking and listening robots to intelligent communicative machines. In J. Markowitz (Ed.), *Robots That Talk and Listen* (pp. 317–335). Boston, MA: De Gruyter, 2015.
3. Roger K. Moore, Ricard Marxer, and Serge Thill. (2016). Vocal interactivity in-and-between humans, animals and robots. *Frontiers in Robotics and AI*, 3(61).
4. Roger K. Moore. Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In K. Jokinen and G. Wilcock (Eds.), *Dialogues with Social Robots – Enablements, Analyses, and Evaluation*, (pp. 281–291). Springer Lecture Notes in Electrical Engineering (LNEE), 2016.
5. Roger K. Moore and Ben Mitchinson. Creating a Voice for MiRo, the World's First Commercial Biomimetic Robot. In *INTERSPEECH 2017* (pp. 3419–3420). Stockholm, 2017.
6. Roger K. Moore. Appropriate voices for artefacts: some key insights. In *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR-2017)*, (pp. 7–11). Skovde, Sweden: VIHAR, 2017.
7. Sarah Wilson and Roger K. Moore. Robot, alien and cartoon voices: implications for speech-enabled systems. In *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR-2017)*, (pp. 40–44). Skovde, Sweden: VIHAR, 2017.
8. Roger K. Moore and Mauro Nicolao. Towards a Needs-Based Architecture for "Intelligent" Communicative Agents: Speaking with Intention. *Frontiers in Robotics and AI*, 4(66), 2017.
9. Roger K. Moore. A "Canny" Approach to Spoken Language Interfaces. In *CHI-19 Workshop on Mapping Theoretical and Methodological Perspectives for Understanding Speech Interface Interactions.* Glasgow: ACM, 2019.
10. Roger K. Moore. Vocal interactivity in crowds, flocks and swarms: implications for voice user interfaces. In *2nd International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR-2019).* London, 2019.
11. Roger K. Moore. Talking with Robots: Opportunities and Challenges. In *International Conference Language Technologies for All (LT4All)*, Paris, France: Unesco, 2019.

5.17 Chat, Personal Information Acquisition, and Turn-Taking in Multi-Party, Multimodal Dialogues

Mikio Nakano (Honda Research Institute Japan – Wako, JP)

Spoken dialogue is one of the promising media for human-machine interfaces. Human users and machines can exchange complicated information through multi-turn dialogues. Recent advances in speech and language processing technologies have enabled us to develop dialogue systems to engage in a variety of tasks. However, only performing task is not enough for many people to repeatedly use such dialogue systems; it is also crucial to give the user a good impression and establish good a relationship between the user and the system. These are also effective to alleviate the problems caused by the system's intention understanding errors.

Virtual agents and robots (agents hereafter) are expected to play a crucial role in those issues, since they can exploit non-verbal behaviour s such as eye gaze, facial expressions, gestures, and postures. However, although there have been many studies on what verbal and physical behaviours s of agents can improve the users' impressions, more investigations are needed to establish guidelines for designing agents' behaviours s. Below are the challenges that I think are important.

The first is to engage in multimodal chat dialogues, or non-task-oriented dialogues. Chat dialogues have been found effective in giving a good impression [1] and building rapport with users [3, 4], and thus combining task-oriented dialogues and chat dialogues would be effective. However, what kind of non-verbal behaviours s should be generated during chat dialogues is yet to be investigated.

Second, acquiring the users' personal information, such as interests, habits, and experiences, through dialogues is one of the important functions for agents, because tailoring dialogues using acquired personal information would contribute to strengthen the relationship between the user and the agent. Not only the linguistic contents of user utterances but also prosody and non-verbal information would be useful for estimating the users' intentions and attitudes during dialogues which leads to precisely acquiring personal information [2].

The third challenge is to achieve smooth turn-taking in multi-party multimodal dialogues. Agents often need to interact with multiple users and sometimes there are multiple agents, but turn-taking in multi-party dialogues is far complicated than that in two-party dialogues [5]. A user speech might be directed to the agent or another user. When a user finishes speaking, the system should start speaking or should wait for another user to start speaking. Non-verbal behaviours s of the users and the agents are considered useful for sophisticated conversational floor management.

To address these challenges, recognising and generating social signals in language, prosody, and non-verbal information are important. In addition to investigating what kinds of social signals play crucial roles in these challenges, we also need to make effort for better recognising and generating social signals.

References

1 Takahiro Kobori, Mikio Nakano, and Tomoaki Nakamura. Small Talk Improves User Impressions of Interview Dialogue Systems. In *Proc. SIGDIAL-2016*, pp. 370-380, 2016.

2 Masahiro Araki, Sayaka Tomimasu, Mikio Nakano, Kazunori Komatani, Shogo Okada, Shinya Fujie, Hiroaki Sugiyama. Collection of Multimodal dialogue Data and Analysis of the Result of Annotation of Users' Interest Level. In *Proc. LREC 2018*, 2018.

3 Gale M. Lucas, Jill Boberg, David R. Traum, Ron Artstein, Jonathan Gratch, Alesia Gainer, Emmanuel Johnson, Anton Leuski and Mikio Nakano. Getting to Know Each Other: The Role of Social Dialogue in Recovery from Errors in Social Robots. In *Proc. HRI-2018*, pp. 344-351, 2018.

4 Gale M. Lucas, Jill Boberg, David R. Traum, Ron Artstein, Jonathan Gratch, Alesia Gainer, Emmanuel Johnson, Anton Leuski and Mikio Nakano. Culture, Errors, and Rapport-building Dialogue in Social Agents. In *Proc. IVA 2018*, pp. 51-58, 2018.

5 Takaaki Sugiyama, Kotaro Funakoshi, Mikio Nakano, and Kazunori Komatani. Estimating response obligation in multi-party human-robot dialogues. In *Proc. Humanoids 2015*: pp. 166-172, 2016.

5.18 Natural Dialogue Interaction with Autonomous Robots

Matthias Scheutz (Tufts University – Medford, US)

URL https://hrilab.tufts.edu/publications/

Natural language is often viewed as an appendix or add-on to a robotic control architecture in the robotics community and there is currently still little interest in deep natural language integration. For one, the reason is that robotics tasks are difficult enough as is, and enabling natural language interactions on robots requires not only expertise in various areas of computational linguistics (such as parsing and dialogue systems), but also an understanding of how language interacts with perceptions and actions, and how humans use spoken language in dialogue interactions (which is very different from processing written text). But more importantly, we still do not have good architectural theories of how language needs to be integrated into a cognitive robotic architecture: What roles should prosody and disfluencies play in speech recognition and subsequent parsing? How are referential expressions resolved in open worlds when the robot does not even know the referent? What kind of inferences and common sense knowledge are required for robots to understand indirect requests (expressed as indirect speech acts)? When should the robot hold the floor in dyadic interactions and how should it do that (forget the dynamics of multi-user dialogues with different overlaps among the speakers' utterances)? And how are the natural language processing components in the architecture connected to the rest to allow for information exchanges, the sharing of mutual constraints, the coordination among different parallel processes, and the overall time-sensitive processing required by human speakers? These are just a few of the open questions that need to be tackled if we want to build robots that are reasonably language competent and can be instructed and taught naturally. Critically, we need to revisit component algorithms for all subsystems, from the ASR, to the syntactic and semantic parsers, the pragmatic reasoner, the dialogue manager, the text generation component, and the speech synthesiser and determine the extent to which they can work incrementally in a time-sensitive manner; for that is what humans expect from interlocutors, and this includes non-linguistic aspects such as attention shifts, search and exploration actions (e.g., turn the head, purposefully looking in a particular direction, etc.), carefully timed backchannel feedback while the interlocutor is still speaking, as well as the initiation of actions , including speech actions, while an interlocutor's utterance is still processed.

5.19 Some Open Questions

David Schlangen (Universität Potsdam, DE)

What can deep (end-to-end) learning offer to dialogue research?

Deep learning success in many areas (MT, ASR) has so far not translated into success in dialogue modelling. There are some advances in NLU (intent classification), but also, arguably, much success that is only apparent (natural language inference) and mostly due to the very large datasets that are now available. Also, arguably, a lot of resources have been spent on something that seems more like a regression (deep learning chatbots, where it took a long time to rediscover basic notions like coherence).

Is dialogue modelling perhaps not a task that is usefully approached end-to-end? There are many interesting questions here about how to approach language science and engineering in general. (E.g., relation btw. machine learning and human learning; modularity; ... I've touched upon some of these in a recent manuscript on ArXiv, "language tasks and language games", [5].)

Is building conversational agents AI complete?

Unrestricted conversation clearly is AI complete (ie., requires the full realisation of intelligence): Any problem-solving behaviour of a person can be "simulate" in conversation (by imagining and describing it), so if any behaviour of humans is intelligent, conversation is also it. But conversation is also more narrowly and directly very challenging, not just because of what can be talked about, as it involves planning, reasoning, and acting under time-pressure.

The question is whether restricted conversation is not AI complete, and whether restricted conversation exists. Allen et al. [1] assume that it is, and does: "The Practical Dialogue Hypothesis: The conversational competence required for practical dialogues, while still complex, is significantly simpler to achieve than general human conversational competence" (see also [6, 3]).

The jury is still out on this, I would say. When things work, the illusion can be created that conversation is happening. When things don't anymore, this illusion breaks down very quickly. It seems that a) recognising misunderstandings, even in restricted tasks / domains and b) recovering from them might be AI complete.

In "The Symbolic Species", Terrence Deacon [2] makes the startling observation that there are no simple languages. But maybe there are simpler language users (human first language learners), or simpler language addressees? (See References.)

Social agents that don't use language

There is a model organism for social interaction, though, which is human/dog interaction. There is a strong impression that some form of communication is happening. What are its limitations? Can you achieve reference with a non-linguistic agent, or only joint attention? What are its mechanisms? Modelling this won't tell us about the role of language in this, but it might tell us about the role of paralinguistic signals and of interaction management (monitoring, turn-taking).

(Exploring this idea with a roomba-type robot has been on my "grant proposal ideas" list for a long time.)

Language-using agents that aren't social

Is it possible to build language interfaces that avoid giving the impression that they are more capable than they are? Interfaces that don't say "I", and don't pretend to be an "I"? Is that a useful goal?

It seems to me that a lot of the frustration that users have comes from them expecting "normal" language behaviour from systems, which they can only provide within a very narrow corridor of choices. Maybe that is a problem that goes away with exposure (perhaps Siri, Alexa, etc. have by now trained their users better?). Or is there space for designing interfaces that avoid giving the impression of having capabilities that aren't really there? It works for command line interfaces, but these are only for experts.

In [4], we tried to explore how using a non-speech modality could keep some aspects of conversational behaviour (quick feedback; prediction; adaptation / common-ground) while otherwise exposing limitations (restricted set of expectations). Many more design cues could be explored (e.g., "robotic" voices).

References

1 James F. Allen, Donna Byron, M. Dzikovska, George Ferguson, L. Galescu, and A. Stent. An architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3), 2000.
2 Terrence Deacon. *The Symbolic Species*. Norton & Co, 1997.
3 Jens Edlund, Joakim Gustafson, Mattias Heldner, and Anna Hjalmarsson. Towards human-line spoken dialogue systems. *Speech Communication*, 50:630–645, 2008.
4 Casey Kennington and David Schlangen. Supporting Spoken Assistant Systems with a Graphical User Interface that Signals Incremental Understanding and Prediction State. In *Proceedings of the 17th Annual SIGdial Meeting on Discourse and Dialogue*, 2016. http://clp.ling.uni-potsdam.de/publications/Kennington-2016.pdf
5 David Schlangen. Language Tasks and Language Games: On Methodology in Current Natural Language Processing Research; *CoRR 2019*, 2019. http://clp.ling.uni-potsdam.de/publications/Schlangen-2019-1.pdf
6 David Schlangen. What we can learn from Dialogue Systems that don't work: On Dialogue Systems as Cognitive Models. In *Proceedings of DiaHolmia, the 13th International Workshop on the Semantics and Pragmatics of Dialogue*, SEMDIAL 2009, pp. 51–58, Stockholm, Sweden, 2009.

5.20 Interaction Model for SLIVAR

Abhishek Shrivastava (Indian Institute of Technology – Guwahati, IN)

Interaction models or Conceptual models are well defined in the area of Human-Computer Interaction (HCI). To quote Preece et al. ([2], p. 40), Interaction model is

> *a description of the proposed system in terms of a set of integrated ideas and concepts about what it should do, behave and look like, that will be understandable by the users in the manner intended.*

These models are evaluated across three dimensions: (a) descriptive, (b) generative and (c) evaluative [1]. This means that an interaction model can help designers describe a range of possible interactions between the human and the computer, generate newer interactions within a specific conceptual framework, and (c) evaluate interactions across a range of design alternatives. I argue that a similar conceptual understanding of interactions between the humans and the virtual agents and robots is yet to substantiate. The research and design community are still required to arrive at a common understanding of conceptual models which drive the design and developments of SLIVAR. Not only that, we need to evolve a common understanding of these models, we need to find relevant candidates which can be evaluated across descriptive, generative and evaluative dimensions. During SLIVAR Dagstuhl seminar, it was seen that the community was hugely concerned about articulating these models. Anthropomorphism and, at times, animal-like did come across as existing conceptual models. However, there were open questions on evaluating interactions born out of these models. I suspect that unless we find methods to evaluate these interactions, we may be designing only limited scope point-designs. In my proposal, here, I believe that the answer lies in finding relevant interaction models for SLIVAR. This may be an interdisciplinary research where we may as well utilise methodologies involved in evolving Interactions models (as in HCI).

References

1 M. Beaudouin-Lafon. Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces – AVI '04*, p. 15, New York, New York, USA, 2004. ACM Press. https://doi.org/10.1145/989863.989865

2 J. Preece, Y. Rogers and H. Sharp. Interaction Design: Beyond Human-Computer Interaction. *Design*, 18, 2007. https://doi.org/10.1016/S0010-4485(86)80021-5

5.21 Personal Statement on Spoken Language Interaction with Virtual Agents and Robots

Gabriel Skantze (KTH Royal Institute of Technology – Stockholm, SE)

Long-term benefits of human-robot interaction

Social robots are often perceived as more engaging than other speech interfaces (if we may call them that), such as avatars or smart speakers. To what extent is this only a novelty factor? What happens when this wears off? What are the long-lasting benefits of human-robot interaction, compared to other forms of spoken interaction? Since robots are much more expensive, the advantages they provide must be very clear. In theory, and intuitively, it is clear what these advantages are. I often make the analogy that we are very reluctant to have important meetings (and would never take a Friday beer) over Skype or over telephone. So, there is clearly something special about physical face-to-face meetings and situated interaction (even if we do not actually interact physically). But the scientific evidence for the benefits of human-robot interaction are not abundant. Several studies have shown benefits of robots in for example educational settings [1]. But recent large-scale studies have not found these effects [4]. How can we go about understanding these phenomena better?

Anthropomorphism

Given that we think that social robots provide an added value compared to other speech interfaces, another important question is whether they should look like humans? Most social robots of today (NAO, Pepper, etc) are not very human-like, and their faces are not very expressive. Other robots (such as Jibo) are certainly expressive, but not human-like. Two of the most common arguments against human-likeness are: (1) the Uncanny Valley, and (2) the increased expectations from the user (which cannot be met). I would argue that (1) is certainly a problem for robots like Sophia, and is essentially a problem of mismatch between behaviour and appearance. However, we do not typically find human-like animated agents (like those found in Pixar movies) uncanny, so there is no reason why that would have to be the case. Regarding (2), I think robots situated in a specific setting (such as a reception) can help to limit the expectations (you would not ask a human receptionist for the meaning of life). I think that one of the strongest argument for human-likeness is that the face helps to coordinate the interaction, as it carries a large set of social signals that we already know how to interpret and can relate to [2].

Deep learning for conversational agents

While there have been enormous benefits of deep learning in other areas of speech and language processing (ASR, TTS, MT, etc.), this is not really the case for dialogue systems. I think there are at least four explanations for this: (1) The mapping between input and output is much more indirect for a dialogue system as a whole, compared to ASR, TTS, MT, etc. Simply put, there are many potential answers to the same user utterance or dialogue context. Standard measures like BLEU, etc., used for other technologies do not apply very well. (2) Dialogue is inherently interactive, which makes dialogue systems inadequate to evaluate using fixed datasets. Thus, the ultimate test for a dialogue system is interactive challenges such as the Alexa challenge, but these are very expensive to perform (and hard to define in a meaningful way). (3) Dialogue systems operate in specific domains where there is often not much data to train on. (4) End-to-end training of dialogue systems goes against the idea of modularisation. Thus, it is not clear how such a system could be updated with new vocabulary, database items, etc., without retraining the whole system and complement the dataset with new examples used in the specific contexts, etc. So, the question is how we can make use of deep learning for dialogue systems (beyond the individual components)? Personally, I think representation and transfer learning for dialogue is an interesting topic that should be investigated more. At KTH, we are currently looking into how models of turn-taking can be learned from human-human data and transferred to human-computer dialogue, using deep learning [3].

References

1 D. Leyzberg, S. Spaulding, M. Toneva & B. Scassellati. The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains. In *34th Annual Conference of the Cognitive Science Society*, 2012.

2 G. Skantze. Real-Time Coordination in Human-Robot Interaction Using Face and Voice. *AI Magazine*, 37(4):19, 2016.

3 G. Skantze. Towards a General, Continuous Model of Turn-taking in Spoken Dialogue using LSTM Recurrent Neural Networks. In *Proceedings of SIGdial*, 220–230, 2017.

4 P. Vogt, R. Van den Berghe, M. de Haas, L. Hoffman, J. Kanero, E. Mamus et al. Second Language Tutoring using Social Robots: A Large-Scale Study. In *Proceedings of HRI*, 2019.

5.22 SLIVAR and Language Learning

Lucy Skidmore (University of Sheffield, GB)

The study of SLIVAR in relation to second language acquisition is a rich topic which has been explored at length in the field of computer-assisted language learning (CALL), continuously diversifying as new technology emerges. The evolution of speech technology in particular has created new ground for exploration in dialogue-based computer-assisted language applications. With this new territory comes challenges and opportunities, some of which are highlighted below.

Technical challenges

Despite vast improvements in accuracy of automatic speech recognition (ASR) for non-native language, it can still fall short when used in language learning applications. Accommodating these scenarios is a fundamental challenge for researchers in CALL and imaginative steps need to be taken to both minimise the occurrence of errors but also navigate inaccurate recognition in a constructive way for learners.

Choice of platform

With the continuous accommodation of new technologies comes the vast array of platform choice for language learning applications. Naturally accessibility plays an important role in this decision – both robot-assisted language learning and mobile-assisted language learning are well-established sub-fields within CALL, however mobile-assisted language learning research has more direct impact on learners due to the accessibility of smartphones. With the increasing ownership of products such as Alexa and Google Assistant, voice-assisted smart devices are more accessible compared to robots. This is one of the important factors to consider when making a choice between platforms.

Role of anthropomorphism

The fact that this research is concerned with non-native speaker-computer dialogue rather than native speaker-computer dialogue raises interesting questions about the role of anthropomorphism in human-computer dialogue for language learning. How important is it for learners to hear human speech? Is synthesised speech sufficient for language practice? Are virtual agents and robots appropriate communication partners for learners?

What learners want

Motivation to learn has been shown to be a strongly influential factor in successful language acquisition. It is therefore essential for any research in this area to take learners' opinions into account in order to understand how learners want to use these systems. This in turn may provide interesting opportunities for applications not previously imagined.

An interdisciplinary approach is key

Interdisciplinary in its nature, any successful research into the applicability of SLIVAR to language learning will face the challenge of understanding the topic from multiple perspectives. These include but are not limited to second language acquisition, speech technology (ASR, speech synthesis and dialogue modelling), human-robot interaction and psycholinguistics. However, with this challenge comes the chance for collaboration amongst research communities, which is perhaps the most exciting opportunity of all.

5.23 SLIVAR and the Role of the Body

Serge Thill (Radboud University Nijmegen, NL)

Human signal interpretation is multimodal

Embodied cognition has long pushed the idea that sensorimtor areas of the brain are involved much more in all aspects of cognition than classic theories allow f Relatedly, it has also become understood that human perception is multimodal, and thus not just focused on one sense at a time.

In particular, studies have demonstrated that humans are very attuned to perceiving biological motion even in the context of HRI: when a robot moves with biological kinematic profiles, humans imitate both the action and the speed at which the robot moves, but if the robot uses other types of kinematics, humans are no longer sensitive to the speed of the action, imitating only the goal [2]. This demonstrates that information from the visual modality may be augmented by motor information, provided that the observation can be parsed in human motor terms.

The first question is therefore whether similar processes might apply to sound processing: are we better at understanding speech when the speech is comprised of sounds for which we have a motor programme? If so, can HRI profit, and if so, how? For example, would a detailed morphological model of human sound production help a robot and are there robot vocalisations that are easier to understand for humans because they map onto human motor programs?

Human language is grounded

Genuine vocal interaction requires, in particular on the part of the machine, an understanding of the meaning of the concepts used. From a computational linguistics perspective, it has been clear for a while that purely statistical approaches on the textual modality is insufficient [5]. While this has been debated for quite a while already, the core question remains: to what degree would a robot need to "understand" the sensorimotor experience underlying human language, what exactly are the mechanisms of grounding, and does the fact that humans and robots have different bodies impose any limitations on the degree we can communicate [6]?

Do we even need sophisticated vocal interaction?

Most, if not all of present-day HRI operates in relatively constrained scenario and we are pretty far away from idealised generic "robot companions". This raises the question whether there is a genuine added benefit to providing vocal interactions in realistic scenarios. Even in situations where robots are specifically meant as companions, for example as companion robots for the elderly, there is evidence to suggest that end users are satisfied with animal-like command interactions [3].

An interesting question to explore is therefore where exactly the added benefits of vocal interaction lie. It is clear, for example, that sophisticated vocal interaction may reduce the need to interpret other types of social signals, which is also a non-trivial problem [1] that taps into embodiment and would appear to require advanced models of Theory of Mind [4]. But what shape does this trade off take exactly?

References

1 M. E. Bartlett, C. E. R. Edmunds, T. Belpaeme, S. Thill and S. Lemaignan. What can you see? identifying cues on internal states from the movements of natural social interactions. *Frontiers in Robotics and AI*, 6:49, 2019.
2 A. Bisio, A. Sciutti, F. Nori, G. Metta, L.Fadiga, G. Sandini and T. Pozzo. Motor contagion during human-human and human-robot interaction. *PLOS ONE*, 9(8):1–10, 2014.
3 H. L Bradwell, K. J. Edwards, R. Winnington, S. Thill, and R. B. Jones. Companion robots for older people: importance of user-centred design demonstrated through observations and focus groups comparing preferences of older people and roboticists in South West England. *BMJ Open*, 9, 2019.
4 H. Svensson and S. Thill. Beyond bodily anticipation: internal simulations in social interaction. *Cognitive Systems Research*, 40:161–171, 2016.
5 S. Thill, S. Padó and T. Ziemke. On the importance of a rich embodiment in the grounding of concepts: perspectives from embodied cognitive science and computational linguistics. *Topics in Cognitive Science*, 6(3):545–558, 2014.
6 S. Thill and K. Twomey. What's on the inside counts: A grounded account of concept acquisition and development. *Frontiers in Psychology: Cognition*, 7:402, 2016.

5.24 What should an agent's identity be?

David R. Traum (USC – Playa Vista, US)

In our conceptual structure, mirrored also in grammatical categories in many languages, we have different types of entities we consider and engage with. Agents, exemplified by people, have cognitive structure (e.g. beliefs, desires, intentions, emotions) and can be the moral and physical causer of actions. Objects can be acted upon (moved, modified, created, destroyed), and tools or instruments can be used by an agent to facilitate an action. Instruments thus can play a role in action, but usually without the conceptual structure and with another causer (agent) as taking ultimate responsibility. Interfaces and even human languages themselves could be seen as a kind of tool that allows humans to communicate with each other and with the physical and virtual worlds. When we come to robots and "virtual agents", the question immediately arises as to whether they better fit the agent or instrument categories? Despite the name "agent", many researchers and users take the instrument view – that the entity is there primarily to serve a human user or enhance their abilities, by allowing them to focus on a higher level of problem. Taking this point of view, the main goal of interaction should be efficient task completion, with a minimum of time or cognitive overload spent. On the other hand, if the agent perspective is taken, we would anticipate more effort spent at social-relationship building, empathy and perspective-taking and more symmetric interactions. While the tool view seems too limiting for many of the uses people desire to put robots and virtual agents, few, if any, are comfortable seeing these artificial entities as fully competent humans. While we don't have many examples of intermediate types of entities, we do have some. Examples of entities that are seen as having some cognitive and moral agentive capacity but not fully competent members of human society include animals, children, and at least for some foreigners (who don't fully grasp the language or culture), slaves, mentally ill or incapacitated, and criminals. While these roles vary from culture to culture (and culture views change across time), what is common is that members of these classes are generally seen as having some of the conceptual structure and moral responsibility of full persons, but not all of it. These patterns may be a good launching point for how to construct and think about robot and virtual agent identity, as they often have been in fiction. A problem is that we often have conflicting intuitions about the source of rights to person and agenthood, e.g. whether it is based on biological relatedness or physical and cognitive abilities, or potential for these. There is often also a disconnect between how an entity portrays itself, how it is perceived, and the actual abilities. We are often willing to take quite superficial displays as signals of a host of abilities and attitudes, whether these displays come from other people or other types of entities. There is thus a potential for deception, which might be either benign, neutral or harmful to the specific interaction or for trust and expectations of future interactions. My position is that the identity and supporting displays for an artificial agent should match the role it is expected to play in desired interactions but also its capabilities as required by that interaction. Human-like activities require human-like identities. Some capabilities could be assumed while others must be demonstrated. Key also will be strategies for maintaining identity as well as the fluidity of interaction across communication errors and problems, especially as these can be seen as opportunities for building social relationships rather than just places where the system seems to be malfunctioning.

5.25 Are we building thinking machines or are we illusionists?

Preben Wik (Furhat Robotics – Stockholm, SE)

When we talk about using Spoken Language to Interact with Virtual Agents or Robots it is a very intuitive thing for the non-expert to understand. Because people are so adept at conversing with each other, it is fairly easy to understand how it should be working on a high level of abstraction. A lot easier than it is to actually implement it. Lots of really smart people are working really hard on it, yet a 5 years old kid will often do better than today's state-of-the-art conversational systems. People wonder what is so hard?

Of course I don't have a solution or quick fix for how to make it work well, but I am hoping that we will spend some time together these days to look at some of the bigger, structural challenges from new angles, and ask ourselves some new questions, and that way perhaps come up with some fresh ideas.

We may ask ourselves: What is it that the kid is so much better at? How come we are not able to do that aspect well with machines yet? Do we even know what these difficulties and challenges are? Let's consider the possibility that we might be barking up the wrong tree.

Most AI systems today are doing "Neo-cortex kind of stuff". Can we take a closer look at what cognitive tasks are done in the reptilian brain, or the limbic system and how they relate to our task? Perhaps there are some lower level features needed to make a system more responsive, and feel more "alive"?

I have worked in the chatbot industry, and the social robotics industry, and I have built systems for language learning (CALL), and for human-animal interaction. Although they are all about spoken interaction between different agents, I find it interesting that the questions asked in the various contexts are so different from each other.

While working with Human-Dolphin interaction some linguists say "They Can't do it because animals don't have a language acquisition device -LAD!" That statement is never heard in a SLIVAR context. We have had long academic debates about the origin of language and the source of our ability. Is it a divine quality? Do we have a language instinct? Do we have a LAD in our brains? If so, where does it reside? And what exactly does it do? Could we make an artificial LAD?

What should the building blocks for creating conversational AI be? Could there be something missing in our toolbox? Human-robot interactions are today typically built by writing scripts with some tools such as FurhatSDK, Watson, DialogFlow, LOUIS, Chatscript, Teneo etc. using building blocks such as Entities, Intents, and Topics. We need units on several different levels of abstraction. If we look at "a body" as an analogy, we can talk about it on different levels: atoms, molecules, amino acids, cells, joints, tendons, muscles, organs, arms legs etc. But we cannot understand a kidney from a molecular-level. It will just be a big mess of a bunch of molecules. Similarly, how can we deal with implicatures in a "Gricean" sense for example? Are we able to capture the cooperative principles described in his "logic and conversation"? What about replicators such as "memes"? And what about metaphors and analogies?

As we stand now, we should not forget that we are in the illusion business. Today our job is to create the illusion that you are talking with something that understands you and cares about what you are saying. Which it doesn't. There is nothing wrong with that, but it is a distinctly different discipline from the engineering of building thinking machines. Is that where we see ourselves heading? Is there another path where we are building sentient machines?

Participants

- Hugues Ali Mehenni
CNRS – Orsay, FR
- Gérard Bailly
University Grenoble Alpes, FR
- Bruce Balentine
Entreprise Integration Group – Zürich, CH
- Roberto Basili
University of Rome "Tor Vergata", IT
- Timo Baumann
Universität Hamburg, DE
- Michael C. Brady
American University of Central Asia, KG
- Hendrik Buschmeier
Universität Bielefeld, DE
- Nick Campbell
Trinity College Dublin, IE
- Nigel Crook
Oxford Brookes University, GB
- Laurence Devillers
CNRS – Orsay, FR
- Johanna Dobbriner
TU Dublin, IE
- Jens Edlund
KTH Royal Institute of Technology – Stockholm, SE
- Mary Ellen Foster
University of Glasgow, GB
- Emer Gilmartin
ADAPT Centre – Dublin, IE
- Manuel Giuliani
University of the West of England – Bristol, GB
- Martin Heckmann
Honda Research Institute Europe GmbH, DE
- Kristiina Jokinen
AIST – Tokyo Waterfront, JP
- Tatsuya Kawahara
Kyoto University, JP
- Casey Kennington
Boise State University, US
- Evangelia Kordoni
HU Berlin, DE
- Ivana Kruijff-Korbayová
DFKI – Saarbrücken, DE
- Pierre Lison
Norwegian Computing Center, NO
- Joseph J. Mariani
CNRS – Orsay, FR
- Cynthia Matuszek
University of Maryland, Baltimore County, US
- Roger K. Moore
University of Sheffield, GB
- Mikio Nakano
Honda Research Institute Japan – Wako, JP
- Catherine Pelachaud
Sorbonne University – Paris, FR
- Roberto Pieraccini
Google Switzerland – Zürich, CH
- Matthias Scheutz
Tufts University – Medford, US
- David Schlangen
Universität Potsdam, DE
- Abhishek Shrivastava
Indian Institute of Technology – Guwahati, IN
- Gabriel Skantze
KTH Royal Institute of Technology – Stockholm, SE
- Lucy Skidmore
University of Sheffield, GB
- Serge Thill
Radboud University Nijmegen, NL
- David R. Traum
USC – Playa Vista, US
- Matthew Walter
TTIC – Chicago, US
- Lun Wang
Sapienza University of Rome, IT
- Preben Wik
Furhat Robotics – Stockholm, SE

Report from Dagstuhl Seminar 20031

Scalability in Multiobjective Optimization

Edited by

Carlos M. Fonseca[1], Kathrin Klamroth[2], Günter Rudolph[3], and Margaret M. Wiecek[4]

1 **University of Coimbra, PT,** cmfonsec@dei.uc.pt
2 **Universität Wuppertal, DE,** klamroth@math.uni-wuppertal.de
3 **TU Dortmund, DE,** guenter.rudolph@tu-dortmund.de
4 **Clemson University, US,** wmalgor@clemson.edu

Abstract

The Dagstuhl Seminar 20031 Scalability in Multiobjective Optimization carried on a series of six previous Dagstuhl Seminars (04461, 06501, 09041, 12041, 15031 and 18031) that were focused on Multiobjective Optimization. The continuing goal of this series is to strengthen the links between the Evolutionary Multiobjective Optimization (EMO) and the Multiple Criteria Decision Making (MCDM) communities, two of the largest communities concerned with multiobjective optimization today.

This report documents the program and the outcomes of Dagstuhl Seminar 20031 "Scalability in Multiobjective Optimization". The seminar focused on three main aspects of scalability in multiobjective optimization (MO) and their interplay, namely (1) MO with many objective functions, (2) MO with many decision makers, and (3) MO with many variables and large amounts of data.

Seminar January 12–17, 2020 – http://www.dagstuhl.de/20031
2012 ACM Subject Classification Mathematics of computing → Mathematical optimization, Applied computing → Operations research, Computing methodologies → Artificial intelligence
Keywords and phrases multiple criteria decision making, evolutionary multiobjective optimization, scalability
Digital Object Identifier 10.4230/DagRep.10.1.52
Edited in cooperation with Allmendinger, Richard

1 Executive Summary

Carlos M. Fonseca (University of Coimbra, PT)
Kathrin Klamroth (Universität Wuppertal, DE)
Günter Rudolph (TU Dortmund, DE)
Margaret M. Wiecek (Clemson University, US)

To continue being useful to society, MO has to address new challenges brought to science and engineering by big data that are continuously being produced and stored with a much lower cost than ever in the past. Since massive production of data takes place in the areas of human activity that have traditionally benefited from MCDM (e.g., social media analysis, retail sales, or high-frequency finance), MO needs to enter a new stage of development to be able to handle the high-dimensional data. Driven by this increasing availability of data and also motivated by an unprecedented demand for efficient, reliable and robust optimization methods, research in MO has to focus on the particular difficulties arising in large-scale problems. This requires from MCDM researchers new statistical thinking and leads to an increasing demand for efficient solution methods for large-scale problems, involving *many* objective functions and/or constraints, *many* decision makers, *many* variables and large amounts of data.

In this spirit, the focus in the seminar was on *scalability* which has become a universal challenge for mathematical optimization, and for EMO *and* MCDM in particular. *Scalability* is a characteristic of a system that describes its capability to cope and perform under an increased or expanding workload. A system that scales well will be able to maintain or even increase its level of performance or efficiency when tested by larger operational demands. In an economic context, a company's scalability implies that the underlying business model offers the potential for economic growth within the company. Therefore the main goals of the seminar were the exploration and elucidation of scalability in three fundamental domains: MO with many objective functions, MO with many decision makers, and MO with many variables.

While single objective optimization problems possess (at most) one optimal objective value, biobjective optimization problems are already intractable in many cases, i.e., combinatorial problems such as, for example, shortest path and spanning tree problems, may have an exponential number of nondominated solutions. Going from two to three objectives is another major step in difficulty since there does no longer exist a complete ordering of nondominated solutions. Problems with *many objective functions* pose even greater challenges. Since the number of nondominated solutions generally grows exponentially with the number of objective functions (as long as these are conflicting), efficient strategies for the detection of redundancies, for model reduction and for metamodelling are crucial for the scalability of existing methods. In the domain of MO with many objective functions the following specific topics were addressed:

- *Model building* and the derivation of technical properties are crucial for understanding the specific challenges in many-objective optimization. The following topics were undertaken: (i) Identification of interdependencies between objective functions as compared to real conflict; (ii) Relevance of many objective functions to a given real-life decision-making situation; (iii) Exploration of mathematical or statistical tools that can compress information while retaining the important problem features. Methodological approaches in this context included data analysis, metamodelling, partial and full scalarization, and a new concept for approximation schemes with quality guarantees.
- Concise *representations* are indispensable for the development of efficient algorithms, particularly EMO algorithms, interactive approaches, and decision support tools. The scalability of quality measures and associated representations, including hypervolume, uniformity, coverage, and ε-indicator were discussed and novel representation and visualization paradigms suitable for many-objective optimization were proposed.
- *Efficient solution algorithms* that scale well with an increasing number of objective functions or computationally expensive objective functions are needed. The shortcomings of existing methods were discussed and new strategies that are specifically designed for large-scale problems were derived.
- *Scalable test cases* are needed for the evaluation of the developed approaches. This has been a concern of the EMO community to some extent. The difficulties pertaining to construction of the test cases were identified and future work in this direction was proposed.

The discussion of MO with many decision makers considered the inherent changes in the decision process as soon as there is not just a single decision maker. The focus was on building a formal framework that guarantees a fair decision respecting the preferences of all decision makers with the least total loss.

The domain of MO with many variables was discussed jointly with the domain of MO with many objective functions because large-scale optimization problems involving many variables and large amounts of data often also involve many objective functions. However, an emphasis was put on the required adaptations of EMO and MCDM approaches to handle problems with a high-dimensional decision space. While EMO algorithms often scale relatively well with an increasing dimension of the decision space (at least as long as the number of objective functions remains relatively small), this is in general not the case for MCDM approaches. In particular, the most commonly used exact solution approaches, such as dynamic programming and branch and bound, suffer from the curse of dimensionality. Complexity theoretic aspects were discussed and the use of approximation paradigms, metamodelling, hybridization, or parallelization in this situation was investigated.

During the seminar the schedule was updated on a daily basis to maintain flexibility in balancing time slots for talks, discussions, and working groups, and to integrate in the program the talks whose authors were inspired by the ongoing seminar. The working groups were established on the first day in an interactive fashion. Starting with three large working groups focused around the three central topics of the seminar (MO with many objectives, MO with many decision makers, and MO with many variables), each participant was requested to formulate her/his favorite topics and most important challenges.

The three groups then rejoined and the prevailing topics were put into groups through a natural clustering process while the participants made up initial five working groups. During the week the participants were allowed to change the working groups while some groups split. Overall, the teams remained fairly stable throughout to eventually form eight groups by the end of the seminar. Abstracts of the talks and extended abstracts of the eight working groups can be found in subsequent chapters of this report.

Further notable events during the week included: (i) an invitation to the opening of the art exhibition with paintings of the artist Lola Sprenger, (ii) a hike during a time period with the best (!) weather conditions in the entire week, (iii) a presentation session allowing the participants to share details of upcoming events in the research community, and (iv) a wine and cheese party (see Fig. 1) made possible by a donation of *Fraunhofer-Institut für Techno- und Wirtschaftsmathematik (ITWM)* represented by Karl-Heinz Küfer. The participants are pleased to announce that they made a donation to Schloss Dagstuhl to make a painting by Lola Sprenger entitled "Berg und Tal" part of the permanent art display.

Offers and Needs Market

A major innovation to this seminar was the *Offers & Needs Market* open for the entire week. The participants could write their research offers and needs regarding MO on notepads in different colors and post on pin boards (see fig. 2) to attract or find a possible collaborator. The idea was well received and the participants desired its repetition in future events.

Figure 1 Traditional wine & cheese party.

Figure 2 Offers and needs market.

Outcomes

The outcomes of each of the working groups can be seen in the sequel. Extended versions of their findings will be submitted to a Special Issue of *Computers and Operations Research* entitled "Modern Trends in Multiobjective Optimization" and guest-edited by the organizers of this Dagstuhl seminar.

This seminar resulted in a very insightful, productive and enjoyable week. It has already led to first new results and formed new cooperation, research teams and topics.

Acknowledgements

The organizers would like to express their appreciation to the Dagstuhl office and its helpful and patient staff for their professional and smooth cooperation; huge thanks to the organizers of the previous seminars in this series for setting us up for success; and thanks to all the participants, who worked hard and were amiable company all week. In a later section, we also give special thanks to Kathrin Klamroth and Günter Rudolph as they step down from the organizer role.

2 Table of Contents

3 Overview of Talks

3.1 Scaling up Multi-Objective Bayesian Optimization

Mickaël Binois (INRIA – Valbonne, FR)

Joint work of Mickaël Binois, Abderrahmane Habbal, Victor Picheny, Stefan M. Wild (Argonne), Nathan Wycoff

Bayesian optimization (BO) aims at efficiently optimizing expensive black-box functions, such as hyperparameter tuning problems in machine learning. Scaling up BO to many variables relies on structural assumptions about the underlying black-box, to alleviate the curse of dimensionality. In this talk, we review several options to tackle this challenge. We also discuss the use of the Kalai-Smorodinski solution when the number of objectives increases, for which a stepwise uncertainty reduction infill criterion is detailed.

References

1 Mickaël Binois, Victor Picheny, Patrick Taillandier, Abderrahmane Habbal. *The Kalai-Smorodinski solution for many-objective Bayesian optimization*. ArXiv preprint 1902.06565. 2019
2 Nathan Wycoff, Mickaël Binois, Stefan Wild, *Sequential Learning of Active Subspaces*. ArXiv preprint 1907.11572. 2019

3.2 Output-sensitive Complexity in Multiobjective Optimization

Fritz Bökler (Universität Osnabrück, DE)

Main reference Fritz Bökler: "Output-sensitive Complexity of Multiobjective Combinatorial Optimization Problems with an Application to the Multiobjective Shortest Path Problem", PhD Thesis, TU Dortmund University, 2018
URL http://dx.doi.org/10.17877/DE290R-19130

In this talk, I summarize my core findings on output-sensitive complexity of multiobjective optimization problems from the past years. I contrast the results with current open problems in the respective areas. I also present new open problems on counting and approximation of nondominated sets.

3.3 On Set-Indicator-Based Search: Using Single-Objective Solvers for Multiobjective Problems

Dimo Brockhoff (INRIA Saclay – Palaiseau, FR)

Joint work of Dimo Brockhoff, Cheikh Touré, Anne Auger, Nikolaus Hansen
Main reference Cheikh Touré, Nikolaus Hansen, Anne Auger, Dimo Brockhoff: "Uncrowded hypervolume improvement: COMO-CMA-ES and the sofomore framework", in Proc. of the Genetic and Evolutionary Computation Conference, GECCO 2019, Prague, Czech Republic, July 13-17, 2019, pp. 638–646, ACM, 2019.
URL https://doi.org/10.1145/3321707.3321852

One approach to solve multiobjective optimization problems is to formulate them as single-objective set problems via indicators: the goal is then to find the set of solutions (of a given size) that maximizes a certain quality. The hypervolume indicator has been regularly used in this context because it has favorable theoretical properties. The "classical" definition of the hypervolume indicator and how it is used in multiobjective solvers, however, has some disadvantages: (i) the resulting single-objective set problem is of high dimension and the gradient of the hypervolume indicator is zero in dominated areas of the search space – not giving a solver enough information about where to search for good solutions.

In this talk, I discussed and visualized these disadvantages and presented a set quality criterion which is based on the hypervolume indicator but solves the mentioned disadvantages (joint work with Cheikh Touré, Anne Auger, and Nikolaus Hansen). The implementation of this idea can be combined with any existing single-objective solver with an ask&tell interface, in particular solvers that can handle expensive or large-scale problems.

3.4 A Multiobjective Trust Region Method for Expensive and Cheap Functions

Gabriele Eichfelder (TU Ilmenau, DE)

Joint work of Gabriele Eichfelder, Jana Thomann
Main reference Jana Thomann, Gabriele Eichfelder: "A Trust-Region Algorithm for Heterogeneous Multiobjective Optimization", SIAM Journal on Optimization, Vol. 29(2), pp. 1017–1047, 2019.
URL https://doi.org/10.1137/18M1173277

The talk is about multiobjective optimization problems where one or more of the objective functions are expensive, i.e. computationally heavy, see [1, 3]. In case just some of the functions are expensive while the others are cheap we speak of heterogeneous problems [3]. Such problems occur in applications for instance in the context of Lorentz force velocimetry, when the task is to find an optimal design of a magnet which minimizes the weight of the magnet and maximizes the induced Lorentz force. The latter might be computable only by a time-consuming simulation. We discuss the reasons why classical methods as scalarizations, descent methods and so on are not a suitable approach here. We present the basic concepts of trust region approaches which are often used in case of expensive functions. The main idea is to restrict the computations in every iteration to a local area and to replace the objective functions by suitable models. The number of function evaluations also depends on the models chosen (linear vs. quadratic), and we give recommendations for the model choice based on numerical experiments [3]. Moreover, we discuss acceptation criteria for the iteration steps which are suitable for an application, the electromagnetic mixing in a liquid metal [4].

References

1 Jana Thomann and Gabriele Eichfelder. *A Trust-Region Algorithm for Heterogeneous Multiobjective Optimization*. SIAM Journal on Optimization, 29(2), 1017 – 1047, 2019.
2 Jana Thomann and Gabriele Eichfelder. *Numerical Results for the Multi-Objective Trust Region Algorithm MHT*. Data in Briefs, 25 , 2019.
3 Jana Thomann and Gabriele Eichfelder. *Representation of the Pareto front for heterogeneous multi-objective optimization*. Journal of Applied and Numerical Optimization, 1(3), 293-323, 2019.
4 Sebastian Prinz, Jana Thomann, Gabriele Eichfelder, Thomas Boeck, Jörg Schumacher. *Expensive multi-objective optimization of electromagnetic mixing in a liquid metal* . OptimizationOnline, 2019

3.5 Multi-Objective Multi-Scale Optimization with Massively Large Number of Variables: Design of Graded Components

Georges Fadel (Clemson University – Clemson, US)

Joint work of Georges Fadel, Anthony Garland

Additive Manufacturing processes that deposit material on a point-by-point basis are able to control the material type that is placed at every location within a single object, which enables customizing the object's internal material composition on a region-by-region or cell-by-cell basis. Furthermore, Additive Manufacturing can create complex shapes that cannot be manufactured using other manufacturing techniques. The optimal object design should then seek to find both the optimal part topology and an optimal internal material composition. In particular, the design for many parts requires considering both thermal and elastic loading. This presentation describes how an object's optimal topology and optimal internal composition can be generated using multi-scale optimization, and how the emphasis on one objective versus the other affects the final solution. Since the design process and the associated finite element analyses must be carried out at the macro-cell level, the discretization of the object results in a number of variables in the order of thousands. Additionally, each macro-cell is again discretized into a large number of smaller micro level cells, again in the order of thousands each, and a multi-level coordination method (ATC) is required to obtain an optimal solution for each level yielding an optimal solution for the two scales. The optimization problem is solved as a weighted bi-objective problem at the macro level with the topological objective converted to a constraint. It is solved as a single objective problem at the micro-level using parallel computers. Example designed structures in 2D are shown.

References

1 Anthony Garland. *Optimal Design of Gradient Materials and Bi-Level Optimization of Topology Using Targets (BOTT)*. Ph.D. dissertation, Department of Mechanical Engineering, Clemson University, August 2017
2 Anthony Garland and Georges M. Fadel. *Optimizing Topology and Gradient Orthotropic Material Properties Under multiple Loads*. J. Comput. Inf. Sci. Eng. Jun 2019, 19(2): 021007

3.6 Multi-Objective Simulation Optimization: Theory and Practice

Susan R. Hunter (Purdue University, US)

Main reference Susan R. Hunter, Eric A. Applegate, Viplove Arora, Bryan Chong, Kyle Cooper, Oscar Rincón-Guevara, Carolina Vivas-Valencia: "An Introduction to Multiobjective Simulation Optimization", ACM Trans. Model. Comput. Simul., Vol. 29(1), Association for Computing Machinery, 2019.
URL https://doi.org/10.1145/3299872

The multi-objective simulation optimization (MOSO) problem is a nonlinear multi-objective optimization problem in which the objective functions can only be observed with stochastic error, e.g., through a Monte Carlo simulation oracle. To date, these difficult problems have seen little theoretical development in the literature. We discuss the MOSO problem statement, existing MOSO methods, and promising directions for the future development of MOSO theory and practice.

References

1 Susan R. Hunter, Eric A. Applegate, Viplove Arora, Bryan Chong, Kyle Cooper, Oscar Rincón-Guevara, and Carolina Vivas-Valencia. An introduction to multi-objective simulation optimization. *ACM Transactions on Modeling and Computer Simulation*, 29 (1): 7:1–7:36, January 2019. 10.1145/3299872.

2 Kyle Cooper, Susan R. Hunter, and Kalyani Nagaraj. Bi-objective simulation optimization on integer lattices using the epsilon-constraint method in a retrospective approximation framework. *INFORMS Journal on Computing*, 2020. 10.1287/ijoc.2019.0918.

3 Kyle Cooper and Susan R. Hunter. PyMOSO: Software for multi-objective simulation optimization with R-PERLE and R-MinRLE. *INFORMS Journal on Computing*, 2020. 10.1287/ijoc.2019.0902.

3.7 Chances and Challenges of Multimodality in Multi-Objective Continuous Optimization Problems

Pascal Kerschke (Universität Münster, DE)

Joint work of Pascal Kerschke, Christian Grimme, Heike Trautmann, Michael Emmerich, Hao Wang, Andre Deutz, Mike Preuss
Main reference Christian Grimme, Pascal Kerschke, Heike Trautmann: "Multimodality in Multi-objective Optimization – More Boon than Bane?", in Proc. of the Evolutionary Multi-Criterion Optimization, pp. 126–138, Springer International Publishing, 2019.
URL https://doi.org/10.1007/978-3-030-12598-1_11

As the quality of multi-objective (MO) optimization algorithms needs to be assessed, they are usually tested on a variety of MO benchmark problems (DTLZ, ZDT, bi-objective BBOB, etc.). Those problems have usually been designed with some expectations regarding their difficulty for a MO algorithm, which is usually inferred from our understanding of structures in the single-objective domain. As such, structural properties like multimodality are assumed to be rather challenging – especially for local search algorithms. However, recent developments on the visualization of MO test problems revealed interesting insights, which question those assumptions [3, 2]. Therefore – in order to get away from fine-tuning algorithms and/or designing problems without actually understanding the challenges (and chances) of MO optimization – we should look for new approaches that actually improve our understanding of such problems [1]. Some possible ideas would be:

(1) How can we visualize more than 2-3 search and/or objective variables (ideally simultaneously)?

(2) How to scale MO benchmarks w.r.t. structural properties (e.g., how does multimodality change with increasing dimensionality in search and/or objective space)?

(3) What would be (scalable) features to characterize continuous MO problems? (ideally those features should of course be informative and efficiently computable in high-dimensional spaces)

References

1 Christian Grimme and Pascal Kerschke and Heike Trautmann. *Multimodality in Multi-Objective Optimization–More Boon than Bane?*. International Conference on Evolutionary Multi-Criterion Optimization, pp. 126–138, East Lansing, MI, USA, Springer, 2019.

2 Pascal Kerschke and Hao Wang and Mike Preuss and Christian Grimme and André H. Deutz and Heike Trautmann and Michael T. M. Emmerich. *Search Dynamics on Multimodal Multiobjective Problems*. Evolutionary Computation, 27(4), pp. 577–609, MIT Press, 2019.

3 Pascal Kerschke and Christian Grimme. *An Expedition to Multimodal Multi-Objective Optimization Landscapes*. International Conference on Evolutionary Multi-Criterion Optimization, pp. 329–343, Münster, Germany, Springer.

3.8 On the Difficulty of Multiobjective Combinatorial Optimization Problems

Arnaud Liefooghe (University of Lille, FR)

Joint work of Arnaud Liefooghe, Fabio Daolio, Sébastien Verel, Bilel Derbel, Hernán Aguirre, Kiyoshi Tanaka, Manuel López-Ibánez, Luís Paquete

We first analyze in [1] the impact of the number of variables (n) and of the number of objectives (m) on the difficulty of multiobjective combinatorial optimization problems. Based on extensive experiments conducted on multiobjective NK landscapes, a general family of multiobjective pseudo-boolean functions, we measure the Pareto set approximation quality reached by multiobjective search algorithms with respect to n and m. Additionally, we discuss the relative importance of n and m compared against other facets of problem difficulty. Based on landscape analysis, a sound and concise summary of features characterizing the structure of a multiobjective combinatorial optimization problem are identified, including objective correlation, ruggedness and multimodality. We then expose and contrast the relation between these properties and algorithm performance, thus enhancing our understanding about why and when a multiobjective optimization algorithm is actually successful, and about the main challenges that such methods have to face.

Secondly, we report in [2] an in-depth experimental analysis on local optimal set (LO-set) under given definitions of neighborhood and preference relation among subsets of solutions, such as set-based dominance relation, hypervolume or epsilon indicator. Our results reveal that, whatever the preference relation, the number of LO-sets typically increases with the problem non-linearity, and decreases with the number of objectives. We observe that strict LO-sets of bounded cardinality under set-dominance are LO-sets under both epsilon and hypervolume, and that LO-sets under hyper-volume are LO-sets under set-dominance, whereas LO-sets under epsilon are not. Nonetheless, LO-sets under set-dominance are more similar to LO-sets under epsilon than under hypervolume. These findings have important implications for multi-objective local search. For instance, a dominance-based approach with bounded archive gets more easily trapped and might experience difficulty to identify an LO-set under epsilon or hypervolume. On the contrary, a hypervolume-based approach is expected to perform more steps before converging to better approximations.

References

1 Arnaud Liefooghe, Fabio Daolio, Sébastien Verel, Bilel Derbel, Hernán Aguirre, Kiyoshi Tanaka. *Landscape-aware performance prediction for evolutionary multi-objective optimization.* IEEE Transactions on Evolutionary Computation, 2019 (early access)

2 Arnaud Liefooghe, Manuel López-Ibánez, Luís Paquete, Sébastien Verel. *Dominance, epsilon, and hypervolume local optimal sets in multi-objective optimization, and how to tell the difference.* Genetic and Evolutionary Computation Conference (GECCO 2018), pp 324-331, Kyoto, Japan, 2018

3.9 Robust Multiobjective Optimization Problems and an Approach for Solving them

Anita Schöbel (Fraunhofer ITWM – Kaiserslautern, DE)

Joint work of Marco Botte, Anita Schöbel
Main reference Marco Botte, Anita Schöbel: "Dominance for multi-objective robust optimization concepts", Eur. J. Oper. Res., Vol. 273(2), pp. 430–440, 2019.
URL https://doi.org/10.1016/j.ejor.2018.08.020

Multiobjective optimization problems often face uncertainties. Instead of efficient solutions there is hence a need for robust efficient solutions in many practical settings. Concepts for defining such robust efficient solutions exist, but unfortunately, they are usually hard to find.

One promising algorithmic idea is to reduce an uncertain multi-objective optimization problem to a deterministic multiobjective optimization problem with very many objective functions. The approach has been studied for single-objective optimization before (among others in Klamroth et al, 2017) and is motivated by an idea of Wiecek et al (2009). The talk is based on the paper by Botte and Schöbel (2019) mentioned above.

The talk shows in particular the need for solving multi-objective problems with very many objective functions.

4 Working groups

4.1 Many Objectives: Characterization and Structure (WG2)

Richard Allmendinger (University of Manchester, GB), Andrzej Jaszkiewicz (Poznan University of Technology, PL), Arnaud Liefooghe (University of Lille, FR), and Christiane Tammer (Martin-Luther-Universität Halle-Wittenberg, DE)

4.1.1 Topics covered

This working group has covered a number of topics related to impact that many objectives (in an optimization problem) have on algorithm design and theoretical properties of tools. The availability of suitable many-objective optimization test problems has been touched upon as well. What follows is a summary of topics covered.

Influence of the number of objectives on problem characteristics
- Types of characteristics
- Theoretical results
- Existing experimental results

Considerations on the effect of the number of objectives on the complexity of multiobjective procedures and algorithms
- Updating the archive of non-dominated solutions
- Solving scalarizing problem(s)
- Computing and approximating hypervolume
- EMO algorithms

Exemplary problems
- Location problems [11, 3]
- Distance problems [18, 19, 9]
- ρMNK-landscapes [27]
- Instance generators

4.1.2 Effect of the Number of Objectives on Problem Characteristics

In this section, we study the influence of the number of objectives on different problem characteristics such as the number of Pareto optimal solutions or of preference parameters.

Number of Pareto Optimal Solutions (Combinatorial Case)

In the combinatorial case, the number of Pareto optimal solutions grows exponentially with the number of objectives in the worst case, that is $O(c^{m-1})$, where c is a constant [5]. Furthermore, as shown in [5], this bound is tight for many classical multiobjective combinatorial optimization problems, such as selection, knapsack, shortest path, spanning tree, traveling salesperson, and s–t cut problems. Obviously, the number of Pareto optimal solutions is also bounded by the size of the whole feasible set.

In Figure 3, we report the proportion of Pareto optimal solutions in the solution space with respect to the number of objectives (from 2 to 20 objectives) for ρMNK-landscapes [27] (see Section 4.1.4). Let us focus on independent objectives ($\rho = 0$). We see that less than 5%

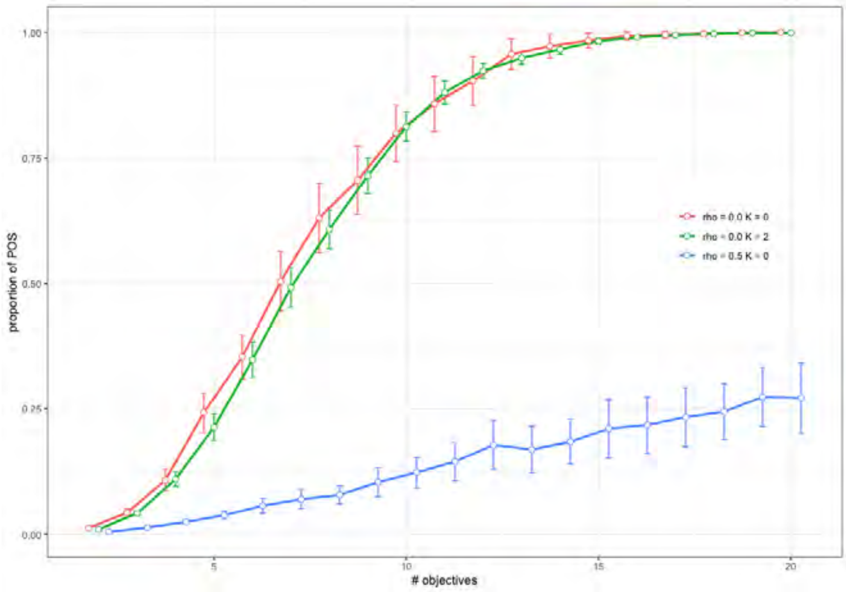


Figure 3 Proportional number of Pareto optimal solutions with respect to the number of objectives for ρMNK-landscapes with different degree of non-linearity (k) and correlations among the objectives (ρ).

of solutions are Pareto optimal for 2-objective problems ($m = 2$), whereas this proportion grows to about 50% for $m = 7$ objectives. For $m = 20$ objectives, more than 99% are Pareto optimal solutions.

Discriminative Power of the Dominance Relation

With a growing number of objectives, the dominance relation becomes less discriminative. Let us consider the comparison between two arbitrary solutions x^1 and x^2 on m objectives. Assume that the probability of equal objective values can be neglected, and that the comparison with respect to each objective is independent. For each objective there is a 1/2 probability that x^1 has a better value for this objective, and a 1/2 probability of the opposite situation. As such, the probability that one solution dominates the other one is $1/2^{(m-1)}$. Thus, it becomes more and more likely that two arbitrary solutions are mutually non-dominated. If the objectives are positively correlated, this probability increases, and if they are negatively correlated this probability decreases.

Probability for a Solution to be Non-Dominated in a Population

As a consequence of the reduced discriminative power of the dominance relation, the probability that a given solution is Pareto optimal increases. Consider a population of μ random solutions. The probability that a given solution is not dominated by another one is

$$1 - \frac{1}{2^m}$$

and thus the probability that one of them is not dominated by any other solution in the population is

$$\prod_{i=1}^{\mu-1}\left(1-\frac{1}{2^m}\right)=\left(1-\frac{1}{2^m}\right)^{\mu-1},$$

and the expected number of non-dominated solutions in this population is then

$$\mu\left(1-\frac{1}{2^m}\right)^{\mu-1}.$$

Dimensionality of the Objective Space

In the continuous case, the size of the set of Pareto optimal solutions formally does not grow, because it is infinite, $\mathfrak{c}$ to be precise, already for two objectives. However, the dimensionality of the objective space and the Pareto front grows. This means that more points or directions are typically required to approximately cover the Pareto front or possible search directions, respectively.

Number of Preference Parameters

In many multiobjective optimization methods, the decision maker (DM) is expected to express his/her preferences e.g. in the form of weighting coefficients or reference levels (aspiration levels/goals) specified for each objective. The number of such parameters grows just linearly with the number of objectives. In the case of some methods like AHP [26], preference parameters are expressed with respect to each pair of objectives; their number then grows quadratically.

Probability of Having Heterogeneous Objectives

By heterogeneous objectives, we here mean objectives that differ for example in their mathematical form (e.g. linear vs. non-linear), cost and/or time of evaluation (e.g. analytical form vs. simulation vs. real physical experiment). Intuitively, the higher the number of objectives, the higher the chance that some of them will differ from other objectives; i.e. that some of them will be more multimodal or costly/time-consuming in evaluation, an issue that has been investigated e.g. in [2, 6].

4.1.3 Effect of the Number of Objectives on the Complexity of Multiobjective Procedures and Algorithms

Although most studies are based on a number of pairwise comparisons of solutions, it is important to notice that the elementary operation for complexity results reported below is a pairwise comparison *per objective*. This choice is motivated by the fact that we want to highlight the effect of the number of objectives (m) on different multiobjective optimization tools and methods.

Updating the Pareto Archive

The Pareto archive is a structure used to store the set of points in the objective space (and corresponding solutions) generated by a multiobjective optimization method, in particular solutions being non-dominated with respect to all solutions generated so far. Updating the Pareto archive A with a new solution x means that all solutions dominated by x are removed from A and x is added to A if it is not dominated by any solution in A.

Updating the Pareto archive can be performed efficiently with the ND-Tree data structure [16]. The complexity of this process is as follows:

- Worst case: $O(m\ \mu)$;
- Best case: $\Theta(m\ \log(\mu))$;
- Average case: $\Theta(m\ \mu^b)$, where μ is the size of the archive, and $b \in [0, 1]$ is the probability of branching.

Note that sublinear time complexity in average case could also be obtained with another recently proposed data structure BSP Tree [12].

An interesting question is how b changes with respect to the number of objectives. In ND-Tree, the archive is recursively divided into subsets of points being close to each other in the objective space. For each subset, a local ideal point and a local nadir point are maintained. A new solution is compared first to the two points and only if the decision with respect to this subset cannot be made, then the corresponding branch of the tree is expanded. The latter situation occurs when the new solution is dominated by the local ideal point and/or when the new solution dominates the local nadir point. Given that, as discussed above, the probability that the dominance relation holds for two solutions (points) decreases with the number of objectives, these situations may become even less likely with a growing number of objectives.

The Pareto archive may be either bounded in size or unbounded, i.e. contain only some or all non-dominated solutions generated so far [10]. In the latter case, the size of the Pareto archive may, in general, grow exponentially with the number of objectives (see Section 4.1.2). Assuming that $\mu = O(c^{m-1})$, the complexity of the update process becomes:

- Worst case: $O(m\ \mu) = O(m\ c^{m-1})$;
- Best case: $\Theta(m\ \log(\mu)) = \Theta(m\ \log(c^{m-1})) = \Theta(m^2\ \log(c))$;
- Average case: $\Theta(m\ \mu^b) = \Theta(m\ c^{(m-1)b})$.

In other words, in the average case the time grows exponentially with the number of objectives, however, with a relatively low exponent, assuming that probability of branching is $b \ll 1$.

Dominance Test

In this section, we consider the process of testing if a solution x is non-dominated or dominated by a Pareto archive. The complexity analysis of this process with the use of ND-tree is the same as for updating the Pareto archive, since the dominance test is the bottleneck part of the updating process. The same holds for BSP Tree [12].

Solving Scalarizing Problem(s)

Let $c(n, m)$ be the complexity for solving one scalar problem. Assuming that the complexity grows linearly with m, then we have $c(n, m) = O(c(n)\ m)$. As such, when multiple scalar sub-problems are to be solved, as in decomposition-based evolutionary multiobjective optimization (e.g., [25]), the complexity is $O(c(n)\ m\ \mu)$, with μ being the number of sub-problems. However, notice that it is often assumed that μ increases with the number objectives m in order obtain a good approximation of the Pareto set [25].

Computing and Approximating Hypervolume

When assessing the performance of multiobjective optimization algorithms, or in indicator-based evolutionary multiobjective optimization, the indicator-value of a solutions-set of size μ is to be computed multiple times. One of the recommended and most-often used indicator is the hypervolume [32]. Unfortunately, the exact hypervolume computation is known to grow exponentially with the number of objectives, more particularly: $O(\mu^{m-2}\ \log(\mu)\ m)$ [4].

Alternatively, the hypervolume can be approximated by Monte Carlo sampling [4]. In this case, the complexity is $\Theta(s\ m\ \mu^b)$, where s is the number of sampling points (see above). Since, Monte Carlo sampling is just a sequence of s independent experiments, each asking a yes/no question (dominated/non-dominated), the confidence interval can be derived from a binomial distribution and does not depend on the number of objectives. On the other hand, the question remains if the size of the confidence intervals should be reduced with growing number of objectives or growing μ. Furthermore, as discussed above, it is often assumed that μ increases with m.

EMO Algorithms

This section tries to understand the impact of many objectives on the working principles of different types of evolutionary multiobjective optimization (EMO) algorithms.

For *all* EMO algorithms that use a constant population size and no external archive:

- Algorithm performance should be evaluated in terms of representation quality.
- The distance between solutions in the objective space increases, so that the quality of representation likely decreases, with poorer coverage.
- The distance between solutions in the decision space increases, so that (blind) recombination likely becomes less effective.

For *dominance-based EMO algorithms* (e.g. NSGA-II [7]):

- The dominance relation becomes less discriminative (see Section 4.1.2). Because of this, we expect a lower selection pressure based on dominance, and then a lower quality in terms of Pareto front approximation.
- Since most individuals are mutually non-dominated, the EMO selection pressure is mostly guided by the diversity preservation mechanisms.
- Convergence is potentially affected.
- The complexity of non-dominated sorting is potentially affected.

For *decomposition/scalarization-based EMO algorithms* (e.g. MOEA/D [29]):

- Assuming a constant number of weight vectors (and population size), all issues mentioned above for constant population size hold. Moreover, the distance between weight vectors increases.
- Assuming the number of weight vectors increases with m (in order to maintain the same level of coverage), the algorithm complexity increases with the number of weight vectors. It remains unclear how the number of weight vectors shall increase, e.g. polynomially or exponentially.

For *indicator-based EMO algorithms* (e.g. IBEA [31]):

- For exact hypervolume computation, the complexity is exponential with m (see above).
- For Monte Carlo approximation, and assuming a constant population size, all issues mentioned above for constant population size hold. By contrast, assuming an population size that increases with m, it remains unclear at this stage how the number of sampling points shall be changed (or not) to reach the same level of hypervolume approximation quality.

4.1.4 Case Studies on Real and Artificial Problems

Multiobjective NK Landscapes (Artificial)

ρMNK-landscapes [27] are a problem-independent model used for constructing multiobjective multimodal combinatorial problems with objective correlation. They extend single-objective NK-landscapes [17] and multiobjective NK-landscapes with independent objectives [1]. Candidate solutions are binary strings of size n. The objective function vector $f = (f_1, \dots, f_i, \dots, f_m)$ is defined as $f\colon \{0,1\}^n \mapsto [0,1]^m$ such that each objective f_i is to be maximized. As in the single-objective case, the objective value $f_i(x)$ of a solution $x = (x_1, \dots, x_j, \dots, x_n)$ is an average value of the individual contributions associated with each variable x_j. Given objective f_i, $i \in 1, \dots, m$, and each variable x_j, $j \in 1, \dots, n$, a component function $f_{ij}\colon \{0,1\}^{k+1} \mapsto [0,1]$ assigns a real-valued contribution for every combination of x_j and its k *epistatic interactions* $x_{j_1}, \dots, x_{j_k}$. These f_{ij}-values are uniformly distributed in $[0,1]$. Thus, the individual contribution of a variable x_j depends on its value and on the values of $k < n$ variables $x_{j_1}, \dots, x_{j_k}$ other than x_j. The problem can be formalized as follows:

$$\begin{aligned} \max \quad & f_i(x) = \frac{1}{n}\sum_{j=1}^{n} f_{ij}(x_j, x_{j_1}, \dots, x_{j_k}) && i \in 1, \dots, m \\ \text{s.t.} \quad & x_j \in \{0,1\} && j \in 1, \dots, n \end{aligned}$$

The epistatic interactions, i.e. the k variables that influence the contribution of x_j, are typically set uniformly at random among the $(n-1)$ variables other than x_j, following the random neighborhood model from [17]. By increasing the number of epistatic interactions k from 0 to $(n-1)$, problem instances can be gradually tuned from smooth to rugged. In ρMNK-landscapes, f_{ij}-values additionally follow a multivariate uniform distribution of dimension m, defined by an $m \times m$ positive-definite symmetric covariance matrix (c_{pq}) such that $c_{pp} = 1$ and $c_{pq} = \rho$ for all $p, q \in 1, \dots, m$ with $p \neq q$, where $\rho > \frac{-1}{m-1}$ defines the correlation among the objectives. The positive (respectively, negative) objective correlation ρ decreases (respectively, increases) the degree of conflict between the different objective function values. Interestingly, ρMNK-landscapes exhibit different characteristics and different degrees of difficulty for multiobjective optimization methods [20]. The source code of the ρMNK-landscapes generator (and other problem classes) as well as a set of multiobjective combinatorial benchmark instances are available at the following URL: http://mocobench.sf.net.

Multiobjective Distance-based Problems (Artificial)

Distance-based optimization problems have been developed with the aim to visualize the movement of a population through the design space over time in order to understand, e.g. search bias. This is achieved by assuming a 2D design space with arbitrarily many (m) objective dimensions. More formally, in a standard visualisable distance-based test problem, the ith objective is calculated as

$$f_i(x) = \min_{v \in V_i}(\texttt{dist}(x, \mathbf{v})),$$

where $\mathcal{X}$ is a feasible design space, and $x \in \mathcal{X} \subseteq \mathbb{R}^2$ a 2D point (solution) in the design space. There are m sets of vectors defined, where the ith set, $V_i = \{\mathbf{v}_1, \dots, \mathbf{v}_{s_i}\}$, determines the quality of a putative design vector $\mathbf{x} \in \mathcal{X}$, on the ith objective. Note as s_i is the number of elements of V_i, which depends on i, it is legal for $|V_p| \neq |V_q|$, but $|V_p| \geq 1$ for all p. The function $\texttt{dist}(x, \mathbf{v})$ typically returns the Euclidean distance between x and $\mathbf{v}$.

Figure 4 illustrates the simplest distance-based problem formulation using points, where $|V_i| = 1$ for all i. This means that there is a single connected Pareto set, and no additional locally Pareto optimal regions.

There are two approaches to setting the set the elements of V_i: (i) setting the elements of V_i directly by fixing $2 \times m$ parameters or (ii) using a centre (in a 2D space), a circle radius and an angle to each objective minimizing vector resulting in $3 + m$ parameters to fix. The

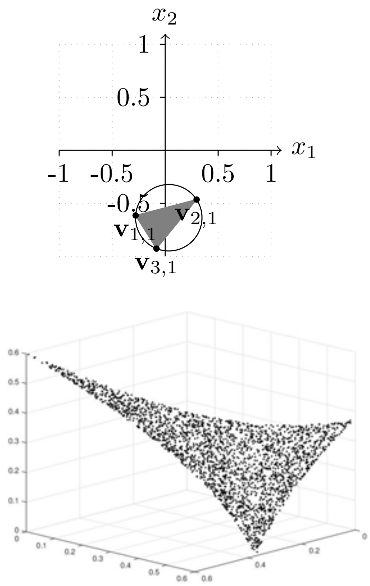


Figure 4 A problem with three objectives, $V_i = \{\mathbf{v}_{i,j}\}$, $|V_i| = 1$; figure taken from [9]. *Left*: The three locations in $\mathcal{X}$, which lie on the circumference of the black circle, determine the objective value minima. They describe a three-sided polygonal Pareto set (coloured grey). *Right*: Samples on the corresponding Pareto front generated by Monte Carlo sampling the Pareto set.

latter approach is more convenient as it requires fixing fewer (or same number) parameters for problems with $m > 2$ objectives compared to the former approach. That is, in the context of many-objective optimization, it is advisable to adopt this approach.

Distance-based problems have been proposed initially in 2005 [18, 19]. Since then the community has suggested a number of extensions to the problem formulation to replicate different complex problem characteristics, such as arbitrarily large decision spaces that could be projected back to the 2D visualization space [22], disconnected Pareto sets of the same [15] or different shapes [14], non-identical disconnected Pareto sets [14], alternative distance metrics, e.g. the Manhattan distance [30, 28], dominance resistance regions [8], local fronts [21], and variants of real-world constraints [24]. To automate the design of feature rich distance-based problems and thus increase the uptake of these problems within the community, Fieldsend et al. [9] have also proposed a configurable test problem generator. The source code of the generator is available at the following URL: https://github.com/fieldsend/DBMOPP_generator.

Multiobjective Location Problems (Real)

The aim of this section is to discuss our investigations for multiobjective optimization problems with a special structure. Especially, we consider point-objective location problems. Using the special special structure of these multiobjective optimization problems, [3] and [11] derived suitable duality assertions and corresponding algorithms for generating the whole set of Pareto optimal solutions in the decision space.

Let m points $a^1, \dots, a^m \in \mathbb{R}^n$ be a priori given. The distance from the new facility $x \in \mathbb{R}^n$ to a given existing facility $a^i \in \mathbb{R}^n$ will be measured by the metric induced by a norm $|| \cdot ||$, especially by the Euclidean norm $|| \cdot ||_2 : \mathbb{R}^n \to \mathbb{R}$, i.e., $||x||_2 := (\sum_{j=1}^n x_j^2)^{\frac{1}{2}}$ or by the Manhattan norm $|| \cdot ||_1 : \mathbb{R}^n \to \mathbb{R}$, i.e., $||x||_1 := |x_1| + \cdots + |x_n|$ or by the maximum norm $|| \cdot ||_\infty : \mathbb{R}^n \to \mathbb{R}$, i.e., $||x||_\infty := \max\{|x_1|, \cdots, |x_n|\}$.

The constrained point-objective location problem involving a certain norm $|| \cdot ||$ is formulated as:

$$\begin{cases} f(x) = (||x - -a^1||, \dots, ||x - -a^m||) \to \min \quad \text{w.r.t. } \mathbb{R}^m_+ \\ x \in X, \end{cases} \tag{POLP$_m$}$$

where the feasible set X is a nonempty and closed set in $\mathbb{R}^n$.

Pareto Optimal Solutions: A point $x \in X$ is called Pareto optimal solution for (POLP$_m$) if

$$\nexists\, x' \in X \ \text{ s.t. } \begin{cases} \forall\, i \in I_m : \ ||x' - -a^i|| \leq ||x - -a^i||, \\ \exists\, j \in I_m : \ ||x' - -a^j|| < ||x - -a^j||, \end{cases}$$

where $I_m := \{1, 2, \cdots, m\}$.

The set of all Pareto optimal solutions is denoted by PO$(X \mid f)$. We have

$$\text{PO}(X \mid f) = \{x \in X \mid f[X] \cap (f(x) - -\mathbb{R}^m_+ \setminus \{0\}) = \emptyset\}.$$

For generating the set of all Pareto optimal solutions PO$(X \mid f)$ of (POLP$_m$), we can use algorithms based on duality statements [3, 11]. These algorithms and many other algorithms for solving scalar as well as multiobjective location problems are implemented in the software *FLO* freely available at https://project-flo.de; see [13].

Generating the Solution Set: Using the software *FLO*, we generate in Figure 5 the set PO$(X \mid f)$ of (POLP$_7$) with seven existing facilities a^i $(i \in I_7)$, $X = \mathbb{R}^2$ and $|| \cdot || = || \cdot ||_\infty$.

If we consider one additional existing facility a^8, i.e., we consider a multiobjective location problem with one more objective function, we generate the set of all Pareto optimal solutions PO$(X \mid f)$ of (POLP$_8$) using the duality based algorithm included in *FLO*. The set of all Pareto optimal solutions PO$(X \mid f)$ of (POLP$_8$) is given in Figure 6. It is possible to see that the solution set of (POLP$_8$) is very different from the solution set of (POLP$_7$).

Furthermore, if we consider the multiobjective location problem (POLP$_8$) where the *Manhattan norm* is involved, we get the set of all Pareto optimal solutions reported in Figure 7.

At last, Figure 8 shows PO$(X \mid f)$ of (POLP$_{11}$) with different norms.

Summary: Using the special structure of the multiobjective location problem, it is of interest to study the influence of the number of objectives (i.e., the number of existing facilities) from the theoretical as well as numerical point of view. Experimental results could be derived

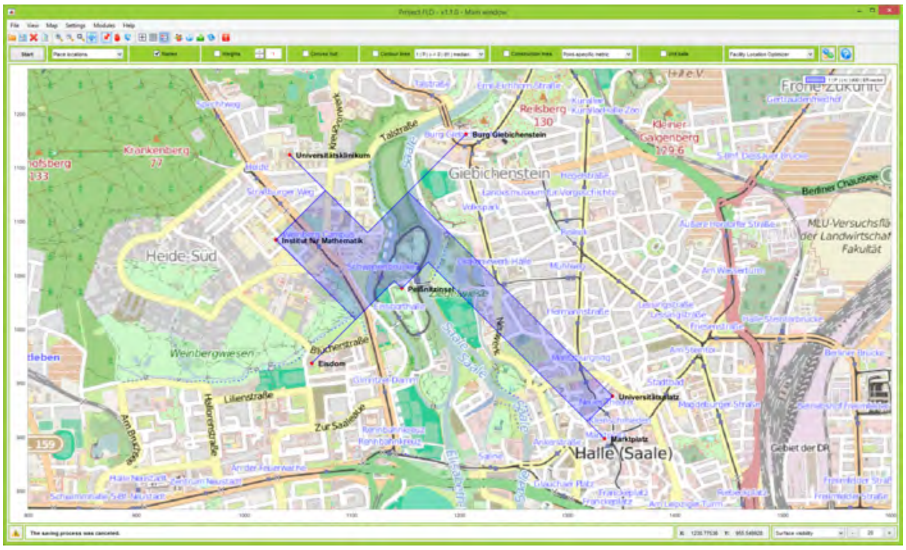


Figure 5 $\mathrm{PO}(X \mid f)$ of (POLP_7) where the maximum norm is involved.

for multiobjective location problems using the software *FLO*. Especially, the influence of the number of existing facilities on the algorithm is of interest. Decomposition methods and scalarization based algorithms for generating the set $\mathrm{PO}(X \mid f)$ of (POLP_m) should be derived. Approaches based on multiple scalar subproblems could be developed for solving multiobjective location problems.

Furthermore, it would be interesting to consider variable domination structures in (POLP_m).

The results can be used in several fields of applications:

- **Location theory**.
- **Economics:** Considering models in utility theory (Cobb-Douglas-function) and production theory.
- **Bioinformatics:** Considering entropy maximization models (based on entropies by [23] for DNA sequence analysis.

4.1.5 Conclusions

Summary

This working group has analyzed the impact an increase in the number of objectives has on (i) problem characteristics and the (ii) complexity of multiobjective procedures and algorithms. Several existing feature-rich test instance generators for problems with many objectives where covered as well. These generators can be used, for example, to validate the observations made but also test the performance of many-objective optimization algorithms on problems with different properties. Our main findings in terms of scaling efficiencies can be summarized as:

- *Good* scaling behavior (i.e., polynomially):
  - Single scalarizations,
  - Approximate hypervolume (though the approximation quality itself may be affected).
- *Relatively good* scaling behavior:
  - Updating the archive of non-dominated solutions.

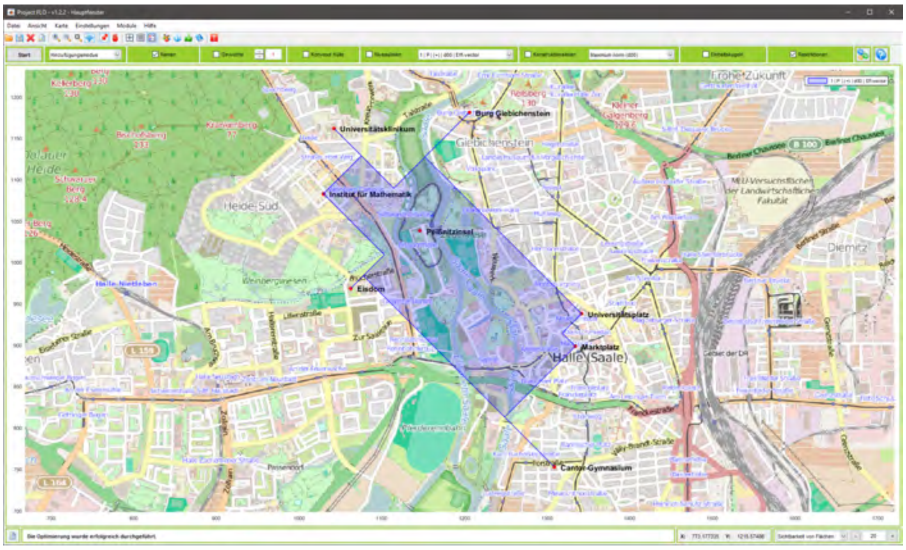


Figure 6 PO($X \mid f$) of (POLP$_8$) where the maximum norm is involved.

- *Poor* scaling behavior (i.e., exponential complexity, decreased quality):
  - Exact hypervolume computation,
  - Approximating the whole PF with guaranteed quality.

Future/Ongoing work

The following directions for future work have been identified:
- Verify theoretical properties on artificial and real problems, problems with special structures.
- Investigate a proper setting for the population size or the number of weights (approximation quality) w.r.t the number of objectives and the correlation among them.
- Investigate the accuracy of Monte Carlo estimation for the hypervolume, how many sampling points w.r.t the number of objectives?
- Changing DM preferences for generating preferred solutions.

References

1 Hernan. Aguirre and Kiyoshi Tanaka. Working principles, behavior, and performance of MOEAs on MNK-landscapes. *European Journal of Operational Research*, 181(3):1670–1690, 2007.

2 Richard Allmendinger, Julia Handl, and Joshua Knowles. Multiobjective optimization: When objectives exhibit non-uniform latencies. *European Journal of Operational Research*, 243(2):497–513, 2015.

3 Shaghaf Alzorba, Christian Günther, Nicolae Popovici, and Christiane Tammer. A new algorithm for solving planar multiobjective location problems involving the manhattan norm. *European Journal of Operational Research*, 258(1):35–46, 2017.

4 Johannes Bader and Eckart Zitzler. HypE: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, 2011.

5 Cristina Bazgan, Florian Jamain, and Daniel Vanderpooten. On the number of non-dominated points of a multicriteria optimization problem. *Discrete Applied Mathematics*, 161(18):2841 – 2850, 2013.

6 Tinkle Chugh, Richard Allmendinger, Vesa Ojalehto, and Kaisa Miettinen. Surrogate-assisted evolutionary biobjective optimization for objectives with non-uniform latencies.

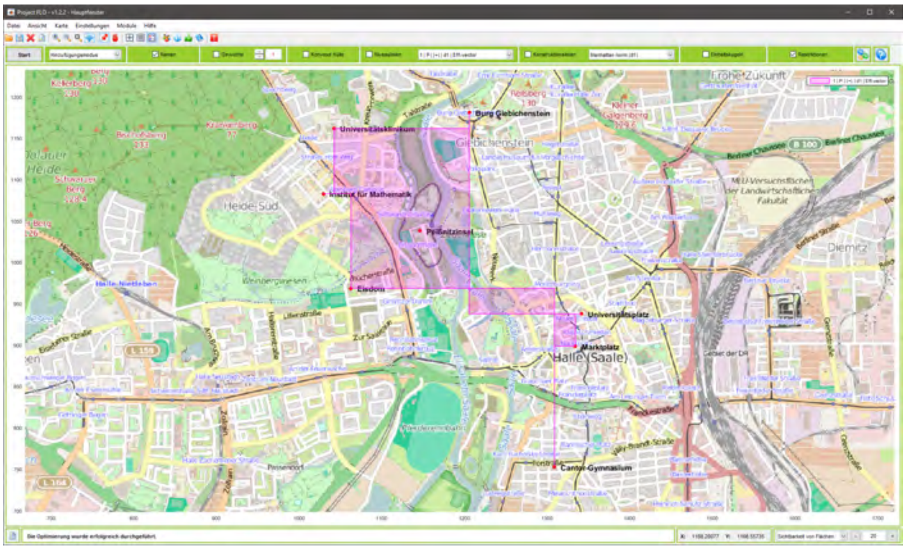


Figure 7 PO($X \mid f$) of (POLP$_8$) where the Manhattan norm is involved.

In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 609–616, 2018.

7 Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

8 Jonathan E. Fieldsend. Enabling dominance resistance in visualisable distance-based many-objective problems. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, pages 1429–1436. ACM, 2016.

9 Jonathan E. Fieldsend, Tinkle Chugh, Richard Allmendinger, and Kaisa Miettinen. A feature rich distance-based many-objective visualisable test problem generator. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO ,äô19, page 541,äì549, New York, NY, USA, 2019. Association for Computing Machinery.

10 Jonathan E. Fieldsend, Richard. M. Everson, and Sameer Singh. Using unconstrained elite archives for multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 7(3):305–323, June 2003.

11 Christiane Gerth and Karin Pöhler. Dualität und algorithmische anwendung beim vektoriellen standortproblem. *Optimization*, 19(4):491–512, 1988.

12 Tobias Glasmachers. A fast incremental bsp tree archive for non-dominated points. In Heike Trautmann, Günter Rudolph, Kathrin Klamroth, Oliver Schütze, Margaret Wiecek, Yaochu Jin, and Christian Grimme, editors, *Evolutionary Multi-Criterion Optimization*, pages 252–266, Cham, 2017. Springer International Publishing.

13 Christian Günther and Christiane Tammer. Relationships between constrained and unconstrained multi-objective optimization and application in location theory. *Mathematical Methods of Operations Research*, 84(2):359–387, 2016.

14 Hisao Ishibuchi, Naoya Akedo, and Yusuke Nojima. A many-objective test problem for visually examining diversity maintenance behavior in a decision space. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, pages 649–656. ACM, 2011.

15 Hisao Ishibuchi, Yasuhiro Hitotsuyanagi, Noritaka Tsukamoto, and Yusuke Nojima. Many-objective test problems to visually examine the behavior of multiobjective evolution in a decision space. In R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph, editors, *Parallel*

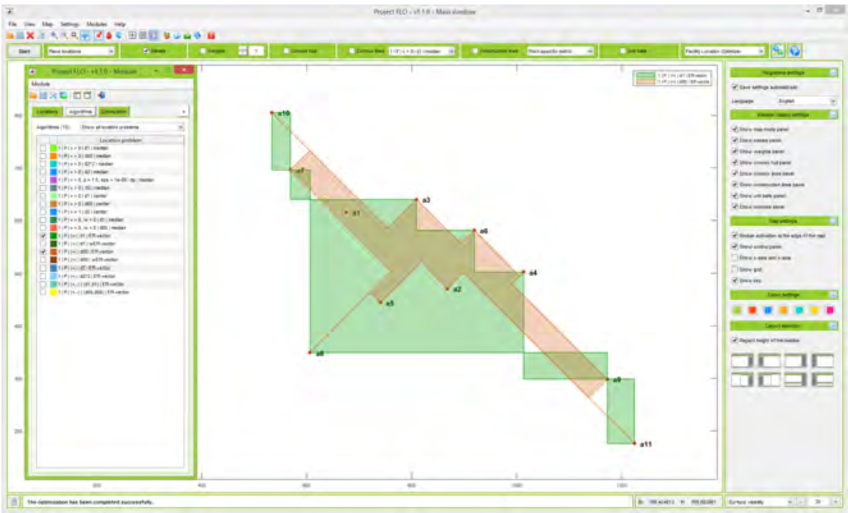


Figure 8 $\mathrm{PO}(X \mid f)$ of (POLP_{11}) where the Manhattan norm (green) and the maximum norm (red) are involved (generated using *FLO*).

Problem Solving from Nature, PPSN XI, 11th International Conference, Proceedings, Part II, pages 91–100, Berlin, Heidelberg, 2010. Springer.

16 Andrzej Jaszkiewicz and Thibaut Lust. ND-Tree-based update: A fast algorithm for the dynamic nondominance problem. *IEEE Transactions on Evolutionary Computation*, 22(5):778–791, Oct 2018.

17 Stuart A. Kauffman. *The Origins of Order*. Oxford University Press, 1993.

18 Mario Köppen, Raul Vicente-Garcia, and Bertram Nickolay. Fuzzy-Pareto-dominance and its application in evolutionary multi-objective optimization. In C. A. Coello Coello, A. H. Aguirre, and E. Zitzler, editors, *Evolutionary Multi-Criterion Optimization, Third International Conference, Proceedings*, pages 399–412, Berlin, Heidelberg, 2005. Springer.

19 Mario Köppen and Kaori Yoshida. Visualization of Pareto-sets in evolutionary multi-objective optimization. In *Proceedings of the 7th International Conference on Hybrid Intelligent Systems*, pages 156–161, 2007.

20 Aarnaud Liefooghe, Fabio Daolio, Sebastien Verel, Bilel Derbel, Hernan Aguirre, and Kiyoshi Tanaka. Landscape-aware performance prediction for evolutionary multi-objective optimization. *IEEE Transactions on Evolutionary Computation*, page (early access), 2019.

21 Yiping Liu, Hisao Ishibuchi, Yusuke Nojima, Naoki Masuyama, and Ke Shang. A double-niched evolutionary algorithm and its behavior on polygon-based problems. In *International Conference on Parallel Problem Solving from Nature*, pages 262–273. Springer, 2018.

22 Hiroyuki Masuda, Yusuke Nojima, and Hisao Ishibuchi. Visual examination of the behavior of EMO algorithms for many-objective optimization with many decision variables. In *IEEE Congress on Evolutionary Computation (CEC), Proceedings*, pages 2633–2640. IEEE, 2014.

23 Tomasz Maszczyk and Włodzisław Duch. Comparison of shannon, renyi and tsallis entropy used in decision trees. In *International Conference on Artificial Intelligence and Soft Computing*, pages 643–651. Springer, 2008.

24 Yusuke Nojima, Yiping Liu Takafumi Fukase, Naoki Masuyama, and Hisao Ishibuchi. Constrained multiobjective distance minimization problems. In *Proceedings of the 2019 on Genetic and Evolutionary Computation Conference Companion*, pages 1–1. ACM, 2019.

25 Qingfu Zhang and Hui Li. MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 2007.

26 Roseanna W Saaty. The analytic hierarchy process‚äîwhat it is and how it is used. *Mathematical modelling*, 9(3-5):161–176, 1987.
27 Sébastien Verel, Arnaud Liefooghe, Laetitia Jourdan, and Clarisse Dhaenens. On the structure of multiobjective combinatorial search space: MNK-landscapes with correlated objectives. *European Journal of Operational Research*, 227(2):331–342, 2013.
28 Justin Xu, Kalyanmoy Deb, and Abhinav Gaur. Identifying the Pareto-optimal solutions for multi-point distance minimization problem in Manhattan space. Technical Report COIN Report Number 2015018, Michigan State University, 2015.
29 Qingfu Zhang and Hui Li. Moea/d: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on evolutionary computation*, 11(6):712–731, 2007.
30 Heiner Zille and Sanaz Mostaghim. Properties of scalable distance minimization problems using the Manhattan metric. In *2015 IEEE Congress on Evolutionary Computation (CEC), Proceedings*, pages 2875–2882. IEEE, 2015.
31 Eckart Zitzler and Simon Künzli. Indicator-based selection in multiobjective search. In *International conference on parallel problem solving from nature*, pages 832–842. Springer, 2004.
32 Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and Viviane Grunert da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, 2003.

4.2 The Output-sensitive Complexity of the BUCO Problem

Fritz Bökler (Universität Osnabrück, DE), Matthias Ehrgott (Lancaster University, GB), José Rui Figueira (IST – Lisbon, PT), Andreia P. Guerreiro (IST – Lisbon, PT), Kathrin Klamroth (Universität Wuppertal, DE), Britta Schulze (Universität Wuppertal, DE), and Daniel Vanderpooten (University Paris-Dauphine, FR)

The question how the running-times of algorithms scale with respect to certain parameters of the input lies at the heart of computational complexity theory. Traditionally, we are interested in how the worst-case running-time of a given algorithm for a given problem scales in the size of the input. Especially, whether this function scales polynomially. For multiobjective combinatorial optimization (MOCO) problems, we often know a negative answer. In particular, for multiobjective variants of the shortest path, spanning tree, assignment, and many more problems, there is no algorithm with a running-time scaling polynomially in the input size [3].

But what happens, if we investigate the running-time of an algorithm as a function of more than just the input size? A new approach in multiobjective optimization is to investigate the running-time as a function of the input size and the output size. Other possible viewpoints include fixed-parameter tractability, wherein the running-time is studied in the input-size and a set of parameters that are assumed to be fixed (cf., e.g., [2]).

In our group, we studied a problem which is easy to describe, but a negative or positive answer has far-reaching consequences. Our object of study was the biobjective unconstrained combinatorial optimization (BUCO) problem given in the following mathematical form:

$$\begin{aligned} \max\ & {c^1}^\top x \\ \min\ & {c^2}^\top x \\ \text{s.t.}\ & x \in \{0,1\}^n \end{aligned}$$

Where $c^1, c^2 \in \mathbb{N}^n$ and $n \in \mathbb{N}$. We are interested in the set of *nondominated points* of this problem, i.e., $\mathcal{Y}_N \min \left\{(-c^1, c^2)^x \mid x \in \{0,1\}^n\right\}$, where min denotes the set of minimal elements with respect to the component-wise less-or-equal order on vectors.

An intuitive description of the problem is the following: We are given n items. Item $i \in \{1, \ldots, n\}$ has profit c_i^1 and weight c_i^2. A solution is a filling of a knapsack with a subset of these items. The profit of a filling is the sum of the profits of its items and the weight of a filling is the sum of the weights of its items. We want to maximize the profit and minimize the weight but these objectives may be conflicting. Consequently, we want to find the set of best compromises between the profit and weight of those knapsack fillings.

In [1] it is proven that if there is no output-sensitive algorithm for the BUCO problem then there is also none for the multiobjective spanning tree problem. Thus, a negative result is especially interesting, since the output-sensitive complexity of the multiobjective spanning-tree problem is open. On the other hand, as this problem is of very general nature, a positive result for the BUCO problem may result in new solution methods for MOCO problems in general. We investigated several angles to attack the problem:

The Nemhauser-Ullmann Algorithm

With a given BUCO instance (c^1, c^2), we associate a set of knapsack (KP) instances parameterized by $k \in \mathbb{N}$:

$$\begin{aligned} &\max\ {c^1}^\mathsf{T} x \\ &\text{s.t.}\ \ {c^2}^\mathsf{T} m x \leqslant k \\ &\qquad\quad x \in \{0,1\}^n \end{aligned}$$

One algorithm to solve KP is the Nemhauser-Ullmann (NU) algorithm [4, 5]. The basic idea is to iteratively consider a growing subset of items. We start with the empty set that only allows for the empty solution and the cost-vector $(0,0)^\mathsf{T}$. Iteratively, we consider one more item in the order given by the input, add its weight and profit to all previous vectors, join these new vectors with the old ones, and delete the dominated vectors. This computes the nondominated set of the BUCO problem. Moreover, for each given k, we can find the maximum value of the corresponding KP instance among this nondominated set. It is unknown whether the NU algorithm is output-sensitive for the BUCO problem.

The main difficulty arises as follows: For each index $i \in \{0, \ldots, n\}$, we define $T(i)$ as the set of solutions after the ith iteration. Thus, $T(0) = \left\{(0,0)^\mathsf{T}\right\}$ and $T(n) = \mathcal{Y}_N$. Can it happen that one of the $T(i)$ for $i \in \{1, \ldots, n-1\}$ is much larger, say more than any polynomial in n and $|\mathcal{Y}_N|$ larger, than $T(n)$? Moreover, while changing the ordering of the items does not change $T(n)$, the sets $T(i)$ for $i \in \{1, \ldots n-1\}$ can change significantly. We thus discussed the behavior of the algorithm with respect to several orderings, including

- ordering by $\frac{c_i^1}{c_i^2}$,
- lexicographic ordering, and
- ordering by $\left\| \begin{pmatrix} c_i^1 \\ c_i^2 \end{pmatrix} \right\|$.

Parameter: Number of large items. The NU algorithm solves the BUCO problem in pseudo-polynomial time. An implication of pseudo-polynomiality is the following: If all numbers in the input are polynomially bounded by the input size, these instances are polynomial-time solvable. We asked the question: What happens if we allow only a small number of super-polynomial items? More formally: Let $p \colon \mathbb{N} \to \mathbb{N}$ be a polynomial. For a given instance, let n be the number of items, and let k be the number of items that have at least one of c_i^1, c_i^2 with value larger than $p(n)$. Is there an algorithm solving BUCO with a running-time bounded by $\mathcal{O}(f(k) \cdot \text{poly}(n))$ for any computable function f? We answered this question affirmatively.

Computing supported nondominated points. The *weighted-sum scalarization* for a given $\lambda \in \mathbb{R}^2_>$ of the BUCO problem is the following problem:

$$\begin{aligned} 2\max \quad & \lambda_1 {c^1}^\top x - -\lambda_2 {c^2}^\top x \\ \text{s.t.} \quad & x \in \{0,1\}^n \end{aligned}$$

A point $y \in \mathcal{Y}_N$ is called a *supported nondominated point*, if there is a $\lambda \in \mathbb{R}^2_>$ such that there is an optimal solution x to the weighted-sum-scalarization with weight λ and ${c^1}^\top x = y_1$ and ${c^2}^\top x = y_2$. A point $y \in \mathcal{Y}_N$ is called an *extreme nondominated point*, if there is a $\lambda \in \mathbb{R}^2_>$ with the above property that leads to no other point $y' \in \mathcal{Y}_N \backslash \{y\}$.

In [6] it is proven that the set of extreme nondominated points of BUCO can be found in polynomial time. However, the number of supported nondominated points can be exponential in the input size. As a third question we thus asked, if we can find all of the supported nondominated points in output-sensitive running-time. While the set of Pareto-optimal supported solutions can be computed in an output-sensitive way by a local search scheme, this is not clear for its image. It can happen that many of the Pareto-optimal supported solutions map to the same point in the objective space, prohibiting an output-sensitive running-time.

We also answered this question in the affirmative. The general idea is to use the characterization in [6] and enumerate solutions comprised of objects with the same profit-to-weight ratios in an output-sensitive way. We also discussed further ideas for the generalized problem, where more than two objectives are present and we aim to continue this line of research in the future.

Acknowledgements

Andreia P. Guerreiro and José Rui Figueira acknowledge the support from DOME (Discrete Optimization Methods for Energy management) FCT Research Project (Ref: PTDC/CCI-COM/31198/2017).

References

1 Bökler, Fritz Output-sensitive Complexity of Multiobjective Combinatorial Optimization Problems with an Application to the Multiobjective Shortest Path Problem. PhD Thesis, Department of Computer Science, TU Dortmund University, 2018.
2 Downey, Rodney G. and Michael R. Fellows. Parameterized Complexity. Springer, 1999.
3 Ehrgott, Matthias. Multicriteria Optimization. Springer, 2005.
4 Kellerer, Hans, Ulrich Pferschy, and David Pisinger. Knapsack Problems. Springer, 2004.
5 Nemhauser, Georg L., Zev Ullmann. Discrete dynamic programming and capital allocation. Management Science 15, pp. 494–263, 1994.
6 Schulze, Britta, Michael Stiglmayr, and Kathrin Klamroth. Multi-objective unconstrained combinatorial optimization: a polynomial bound on the number of extreme supported solutions. Journal of Global Optimization 74(3), pp. 495-–522, 2019.

4.3 Computationally Expensive Functions and Large Scale Test Instances

Dimo Brockhoff (INRIA Saclay – Palaiseau, FR), Gabriele Eichfelder (TU Ilmenau, DE), Carlos M. Fonseca (University of Coimbra, PT), Susan R. Hunter (Purdue University, US), Enrico Rigoni (ESTECO SpA – Trieste, IT), Michael Stiglmayr (Universität Wuppertal, DE)

4.3.1 Introduction

The group first discussed general issues and current challenges related to scaling up to solve large-scale problems. The group identified a number of issues and open research questions on which we give an overview in Section 4.3.2. The group examined in more detail the issue of computationally heavy problems, i.e. problems where function evaluations are computationally demanding, see Section 4.3.3. Further, when developing new algorithms for large-scale problems, the group emphasized the importance of being able to test the algorithms well. The group identified a lack of test instances which are not just simple extensions from the single-objective to the multiobjective setting but which really transfer single-objective challenges to the mutliobjective case or highlight challenges inherent to the multiobjective setting that may not exist in the single-objective setting. We summarize this discussion and provide some solution approaches in Section 4.3.4.

4.3.2 Large Scale Issues and Current Challenges

In application problems and thus in optimization problems of interest there can be many factors that imply a multiobjective problem to be of "large scale." These factors include a large number of variables, a large amount of computation time to evaluate the objective functions which may be outputs of a black box [16], a large number of processors that must evaluate the objectives, the existence of a large Pareto front (especially in discrete problems), or the solution approach which must navigate a large number of local Pareto fronts.

The group identified several open research topics that relate to the identified factors implying a multiobjective problem is large-scale. These research questions apply to either very specific problem formulations or to a more general setting. Some of these open research topics include the following:

- *Many decision variables:* If the multiobjective optimization problem has a large number of decision variables, and particularly if there are many integer variables, constructive methods such as Branch-and-Bound may be too computationally burdensome. For instance, in [3] the proposed Branch-and-Bound procedure could only solve mixed-integer convex multiobjective problems up to 30 integer variables. New theory and methods are needed to handle this scenario. In particular, the group identified column generation for multiobjective combinatorial problems (MOCO) and complexity theory for multiobjective unconstrained combinatorial optimization (MUCO) as possible research topics.
- *Computationally heavy problems:* If the objective functions are the output of a black-box, such as a deterministic black-box oracle [17] or a *Monte-Carlo* simulation oracle [13], one or more of the objective or constraint function evaluations may be time consuming. Thus, only a limited number of function evaluations may be possible. Furthermore, in the case of a Monte-Carlo simulation oracle, more than one processor may be required to obtain a sufficiently accurate estimator of the objective function value [10, 8, 9, 14]. How should

Table 1 Comparison of trust region methods with surrogate methods.

Trust Region Methods	Surrogate Methods
uses a local model, determined only with information along a "path" in the area you trust	uses a global model
basically a local solver, i.e. aims to find locally optimal solutions	aims to find a globally optimal solution
refined when coming close to an optimum	refined in promising areas where you do not trust the model (high variance)
requires model assumptions, e.g., trust the model	requires model assumptions, e.g. Gaussian process

such problems be approached? Trust-region [24, 23] or surrogate modeling may be useful tools in this context; how can they be modified to ensure efficiency?

- *Many local Pareto sets:* In many single-objective settings, local methods are used in the search for global optima (e.g.[18, 19, 15]). Does a similar framework make sense in a multiobjective setting, and if so, how should solvers handle many local Pareto sets? How can one avoid calculating unnecessary Pareto sets, and how might one go about storing the data generated when solving such problems? In particular, mixed-integer nonlinear multiobjective optimization problems might have a huge number of local Pareto sets, one for each fixed setting of the integer variables. Then, the overall Pareto set is the non-dominated points among them, which can be identified by comparing sets. How can such calculations be conducted efficiently?
- *Test instances:* While a large number of test instances exist for multiobjective optimization (e.g. [2, 1, 12, 6, 5, 4]), there is a lack of scalable test instances that truly extend the challenges observed in a single-objective setting to the multiobjective setting. Such test instances are required to test algorithms developed for large-scale problems.

4.3.3 Computationally Heavy Problems

The group decided first to focus specifically on computationally heavy problems, i.e., on optimization problems where the functions are large-scale in terms of computational time. Furthermore, the group envisaged to identify how such problems might be approached using surrogate or trust-region methods. First, we summarize similarities and differences between these methods, which we display in Table 1.

Especially the idea of a local model which is refined close to the optimum can be used in many fields. There are also different possibilities for the model itself: one can built a model for each objective individually, as done in the trust region approaches [22, 24], or one can define a global one based on scalarization, as for instance for the hypervolume or for a multiplicative reformulation [22].

When such a model approach should be chosen for computationally heavy objectives, then one has also to keep in mind the following aspects. First, the question is whether all function evaluations are expensive, or whether it will be cheaper to calculate some value $f(y)$ for y close to some x in case one has already calculated $f(x)$. An example for that might be mesh adaptation or just slight changes in a mesh. It is still an open research question how this can be used for efficient algorithms, for instance for the numerical approximation of derivatives. Of course, this issue is also a task for single-objective expensive optimization.

Another aspect is that the objective functions might be of different type (heterogeneous), like, e.g., different simulation times (3h vs 20 minutes; 2h vs 1 second). See for example [7, 24].

Moreover, in the case that there are different levels of granularities or accuracies of the model an efficient algorithm should choose adaptively. Such approaches already exist, but theoretical proofs are always a true challenge.

Hence, when dealing with computationally heavy problems, several aspects have to be taken into account which are not directly related to the presence of multiple objectives. Moreover, there are many different problem types (heterogeneous, stochastic, locally cheaper, ...) and methods have to be developed for each problem type individually.

4.3.4 Consistent and scalable extensions of test instances for large scale

There is already a large number of test instances available for multiobjective solvers. Nevertheless, we see a lack in instances which truly transfer the single-objective challenges to the multiobjective setting. Often, the difficulties known from single-objective optimization are just transferred to one of the objective functions, while the other objective functions are just chosen to be of a quite simple structure to obtain a simple extension.

An example in this direction is the test instance known as DTLZ7 [5], denoted MaF7 in [2], with

$$\min_{x\in\mathbb{R}^n} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{m-1} \\ f_m(x) \end{pmatrix}$$

with $1 \leq m \leq n$ and functions $f_m, g\colon \mathbb{R}^n \to \mathbb{R}$,

$$f_m(x) = m - \sum_{i=1}^{m-1} \left(\frac{x_i}{1+g(x)} (1 + \sin(3\pi x_i)) \right)$$

and $g(x) = 1 + \frac{9}{n-m+1} \sum_{i=m}^{n} x_i$. Other instances with many objectives are for instance obtained by taking sinus and cosinus values of the components x_i and by a multiplication of them.

Hence, we suggest to construct new multiobjective test instances which extend the classical test functions as the *Rosenbrock* [20] or the *Himmelblau function* [11]. Thereby, we want to use a consistent extension, i.e., all objective functions should be of the same type. Moreover, to be usable for large scale issues, the test instances should be scalable w.r.t. the number of objective functions and the number of variables. Of course, as expected from test instances, they should have predefined mathematical properties as, e.g., a known Pareto set.

Another aspect is that real-world problems often exhibit a complex correlation structure between objectives: conflict and agreement among objective functions are in general local properties. For example, Figure 9 shows the correlation structure of the crankshaft problem [21] on the Pareto front. Available multiobjective test instances consider in general only problems with independent objectives, missing the complexity of real-world scenarios.

Consequently, test instances should have the property that when adding functions the problem should keep the structure in the following sense:

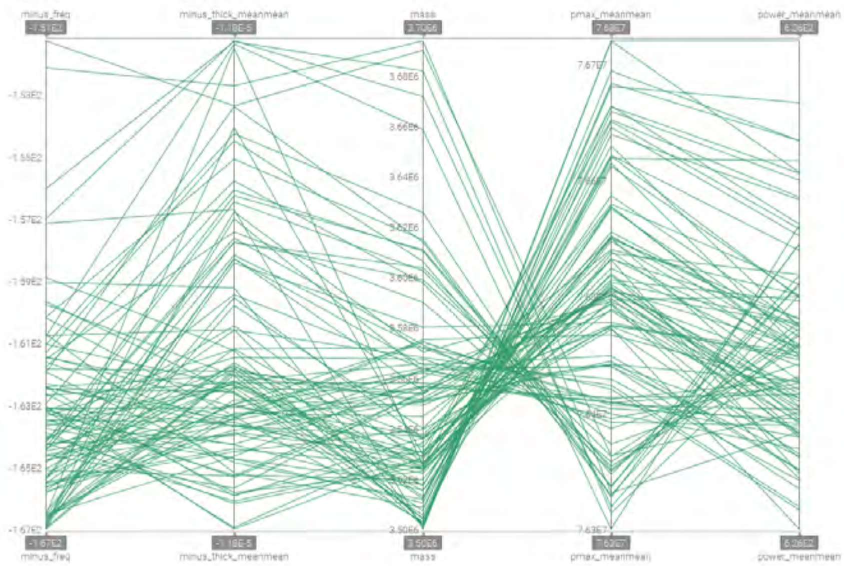


Figure 9 Parallel coordinates chart of an empirical Pareto set of the crankshaft problem. Signs have been changed so that all five objectives are to be minimized. Both conflict and agreement among different objectives are visible.

Assume there are k objectives $f_1, \ldots, f_k \colon \mathbb{R}^n \to \mathbb{R}$. Then these define some kind of ordering in the pre-image space in the sense that we say

$$x \leq_f \tilde{x} \;:\Longleftrightarrow\; (f_1(x), \ldots, f_k(x)) \leq (f_1(\tilde{x}), \ldots, f_k(\tilde{x})). \tag{1}$$

Now we aim on defining for the test instances a new objective $f_{k+1} \colon \mathbb{R}^n \to \mathbb{R}$, which is not just a copy of the previous objective functions, but that also keeps this ordering, meaning

$$x \leq_f \tilde{x} \iff (f_1(x), \ldots, f_k(x), f_{k+1}(x)) \leq (f_1(\tilde{x}), \ldots, f_k(\tilde{x}), f_{k+1}(\tilde{x})).$$

The overall idea is that if we want to study scalability with respect to the number of objectives, we should keep everything else fixed. Furthermore, this requirement introduces a rich correlation structure between objectives as a fringe benefit.

We illustrate this with a biobjective and a triobjective example based on the Rosenbrock function [20]. First, recall that the Rosenbrock function $f \colon \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$f(x, y) = (a - x)^2 + b(y - x^2)^2$$

and it has a unique global minimum at $(x, y) = (a, a^2)$, where $f(x, y) = 0$. Usually the parameters are set such that $a = 1$ and $b = 100$. Then the global minimum, located at $(1, 1)$, is inside a narrow, parabolic shaped valley, as shown in Figure 10.

▶ **Example 1.** Let $a, c, d \in \mathbb{R}$ and $b > 0$ be scalars and define $f_a \colon \mathbb{R}^2 \to \mathbb{R}$ by

$$f_a(x, y) = (a - x)^2 + b(y - x^2)^2.$$

We define a second function by $f_d \colon \mathbb{R}^2 \to \mathbb{R}$,

$$f_d(x, y) = (d - x)^2 + b(y - x^2)^2.$$

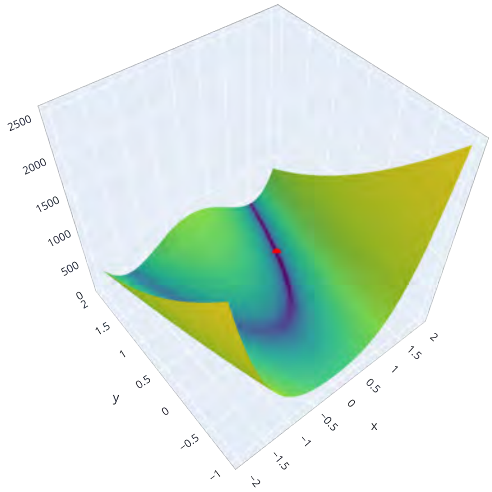


Figure 10 Plot of the Rosenbrock function with $a = 1$ and $b = 100$. The unique global minimum, outlined with a red point, is at $(1, 1)$.

Figure 11 shows that the shape of the function and the position of the global minimum are affected by the value of a (or d).

Then a simple calculation yields that the set of efficient points of the biobjective optimization problem

$$\min_{(x,y)\in\mathbb{R}^2} (f_a(x,y), f_d(x,y))$$

is

$$\mathbb{S} := \{(x,y) \in \mathbb{R}^2 \mid x = (1-\lambda)a + \lambda d,\ y = x^2,\ \lambda \in [0,1]\}.$$

An example of optimal solution set $\mathbb{S}$ is shown in Figure 12, where the two Rosenbrock functions presented in Figure 11 are used as objectives for a biobjective optimization problem. The relevant Pareto front is presented in Figure 13.

These two objectives f_a, f_d define now an ordering as given in (1). Now we can add a third new objective $f_c\colon \mathbb{R}^2 \to \mathbb{R}$ with the same structure, i.e.

$$f_c(x,y) = (c-x)^2 + b(y-x^2)^2$$

and as long as $c \in [a,d]$ the optimal solution set of the three objective problem

$$\min_{(x,y)\in\mathbb{R}^2} (f_a(x,y), f_d(x,y), f_c(x,y))$$

is still equal to $\mathbb{S}$.

Note that the discussion in the above example would cover the quadratic case for $b = 0$. This can be seen even better with the forthcoming Example 3. For the moment, let us show with the next example that also with quadratic functions $f\colon \mathbb{R}^n \to \mathbb{R}$ of the form

$$f(x) = \sum_{i=1}^{n} (a_i - x_i)^2$$

with parameters $a_i \in \mathbb{R}$, $i = 1, \ldots, n$ we get a similar structure (which is of course less challenging from an algorithmic point of view).

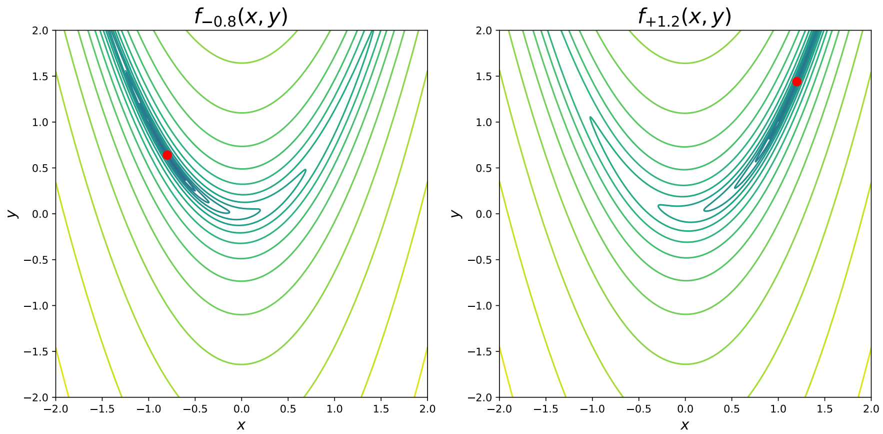


Figure 11 Contour plots of two Rosenbrock functions (in logarithmic scale) with different values for the parameter a (or d) and fixed $b = 100$: the position of the unique global minimum, outlined with a red point, changes accordingly to the value of a.

▶ **Example 2.** We consider a triobjective optimization problem with three objective functions of the type $f_j : \mathbb{R}^2 \to \mathbb{R}$ with

$$f_j(x, y) = (x - \alpha_j)^2 + (y - \alpha_j)^2$$

with $\alpha_j \in \mathbb{R}$, $j = 1, 2, 3$. Each individual objective function has the unique minimal solution $(x, y) = (\alpha_j, \alpha_j)$. Now let $A = (\alpha_1, \alpha_1)$, $B = (\alpha_2, \alpha_2)$, $C = (\alpha_3, \alpha_3)$. Figure 14 shows the optimal solution set for such a triobjective optimization problem. The optimal solution set is

$$\mathbb{S} := \Big\{(x, y) \in \mathbb{R}^2 \mid \alpha_{\min} := \min_{j=1,2,3} \alpha_j \le x \le \max_{j=1,2,3} \alpha_j =: \alpha_{\max},\ y = x\Big\}.$$

The Pareto front is presented in Figure 15. The solution set would not change in case one adds objectives of the above type as long as for the additional functions f_j we have $\alpha_{\min} \le \alpha_j \le \alpha_{\max}$. Figure 16 shows the correlation structure between objectives on the Pareto front.

The test instances should also be scalable w.r.t. the number of variables. This can be easily achieved by following the approach from the following example:

▶ **Example 3.** Let a function $g : \mathbb{R}^n \to \mathbb{R}$ be defined by

$$g(x) = \sum_{i=1}^{k} (a_i - x_i)^2 + \sum_{j=k+1}^{n} b_j \Big(x_j - \sum_{i \in S_j} x_i^2\Big)^2$$

where $a_i \in \mathbb{R}$, $i = 1, \ldots, k$, and $b_j > 0$, $j = k+1, \ldots, n$, are scalars and $S_j \subseteq \{1, \ldots, k\}$ are index sets with $S_j \neq \emptyset$ for all $j = k+1, \ldots, n$. Then the single-objective unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} g(x)$$

has the optimal solution set

$$\Big\{x \in \mathbb{R}^n \mid x_i = a_i,\ i = 1, \ldots, k,\ x_j = \sum_{i \in S_j} x_i^2\Big\}$$

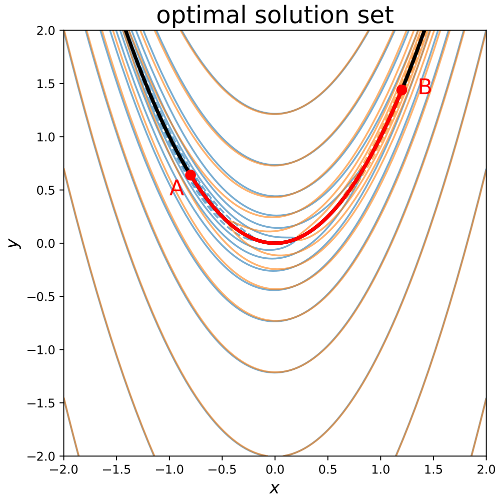


Figure 12 Optimal solution set (outlined with a red parabolic arch) for the biobjective optimization problem defined by the two objectives $f_{-0.8}(x, y)$ and $f_{+1.2}(x, y)$ shown in Figure 11. The contour plots of two objectives functions are represented in blue and in orange, respectively. The minima of the objective functions (shown as red points) are, respectively, $A = \arg\min f_{-0.8}(x, y)$ and $B = \arg\min f_{+1.2}(x, y)$.

4.3.5 Conclusion

Future work contains the formulation of clear test instances with full information on the parameters to choose and on the Pareto set. It is also of interest to discuss the above procedures on how they can be extended to other functions as to construct a multiobjective version of the Himmelblau function [11] $h\colon \mathbb{R}^2 \to \mathbb{R}$

$$h(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

and also to see how general the proposed approach is to generate other types of level sets and shapes of the Pareto set. Most of all on how the dimensionality of the Pareto front can be influenced and how the choice of the parameters influences the conditioning of the problem.

References

1 D. Brockhoff, T. Tusar, A. Auger, and N. Hansen. Using well-understood single-objective functions in multiobjective black-box optimization test suites. *arXiv preprint arXiv:1604.00359*, 2016.

2 R. Cheng, M. Li, Y. Tian, X. Zhang, S. Yang, Y. Jin, and X. Yao. A benchmark test suite for evolutionary many-objective optimization. *Complex & Intelligent Systems*, 3(1):67–81, 2017.

3 M. de Santis, G. Eichfelder, J. Niebling, and S. Rocktäschel. Solving multiobjective mixed integer convex optimization problems. *Optimization Online*, 2019.

4 K. Deb. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7(3):205–230, 1999.

5 K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable test problems for evolutionary multi-objective optimization. Technical report, Institut für Technische Informatik und Kommunikationsnetze, ETH Zürich, 2001. TIK-Technical Report No. 112.

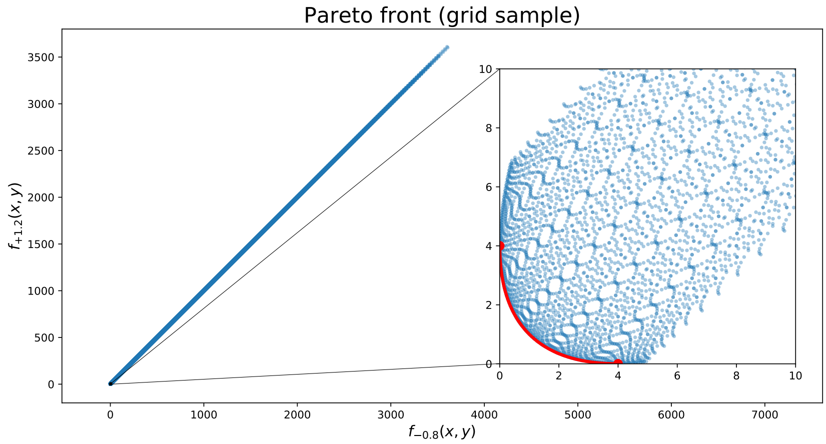


Figure 13 Objective space of the biobjective optimization problem presented in Figure 12: the Pareto front is outlined with a red curve. The blue points represent the image of the objective functions when the domain is sampled on a regular grid. It is evident that the two objectives are highly correlated.

6 K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. Scalable multi-objective optimization test problems. In *Proceedings of the 2002 Congress on Evolutionary Computation*, volume 1, pages 825–830, Piscataway, NJ, May 2002. IEEE.

7 G. Eichfelder, X. Gandibleux, M.J. Geiger, J. Jahn, A. Jaszkiewicz, J. Knowles, P.K. Shukla, H. Trautmann, and S. Wessing. Heterogeneous functions. *In: Understanding Complexity in Multiobjective Optimization, Dagstuhl Seminar 15031*, pages 121–129, 2015.

8 P. W. Glynn and P. Heidelberger. Bias properties of budget constrained simulations. *Operations Research*, 38(5):801–814, 1990.

9 P. W. Glynn and P. Heidelberger. Analysis of parallel replicated simulations under a completion time constraint. *ACM Transactions on Modeling and Computer Simulations*, 1(1):3–23, 1991.

10 P. Heidelberger. Discrete event simulations and parallel processing: statistical properties. *Siam J. Stat. Comput.*, 9(6):1114–1132, 1988.

11 D. M. Himmelblau. *Applied nonlinear programming.* McGraw-Hill, 1972.

12 S. Huband, P. Hingston, L. Barone, and L. While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Transactions on Evolutionary Computation*, 10(5):477–506, 2006.

13 S. R. Hunter, E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia. An introduction to multi-objective simulation optimization. *ACM Transactions on Modeling and Computer Simulation*, 29(1):7:1–7:36, January 2019.

14 S. R. Hunter and B. L. Nelson. Parallel ranking and selection. In A. Tolk, J. Fowler, G. Shao, and E. Yücesan, editors, *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences*, Simulation Foundations, Methods and Applications, chapter 12, pages 249–275. Springer International, Switzerland, 2017.

15 J. Larson and S. M. Wild. A batch, derivative-free algorithm for finding multiple local minima. *Optimization and Engineering*, 17(1):205–228, March 2016.

16 Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87 of *Applied Optimization.* Springer, 2004.

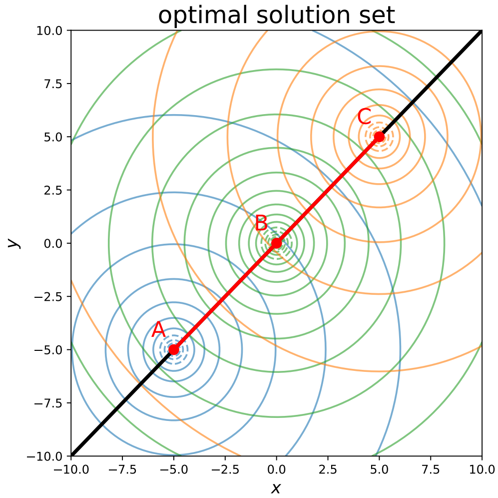


Figure 14 Optimal solution set (outlined with a red segment) for a triobjective optimization problem defined by three quadratic functions as objectives. The contour plots of the tree quadratic functions are represented in blue, green, and orange. The minima of the objective functions (shown as red points) are, respectively, A, B, and C.

17 S. Prinz, J. Thomann, G. Eichfelder, T. Boeck, and J. Schumacher. Expensive multi-objective optimization of electromagnetic mixing in a liquid metal. *Optimization Online*, 2019.

18 A. H. G. Rinnooy Kan and G. T. Timmer. Stochastic global optimization methods part I: Clustering methods. *Mathematical Programming*, 39:27–56, 1987.

19 A. H. G. Rinnooy Kan and G. T. Timmer. Stochastic global optimization methods part II: Multi level methods. *Mathematical Programming*, 39:57–78, 1987.

20 H. H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, 01 1960.

21 Rosario Russo, Alberto Clarich, and Marco Carriglio. A multi-objective optimization of engine crankshaft design using modeFRONTIER. *International Review of Mechanical Engineering*, 6(3), 2012.

22 J.-H. Ryu and S. Kim. A derivative-free trust-region method for biobjective optimization. *SIAM Journal on Optimization*, 24:334,äì362, 2014.

23 J. Thomann and G. Eichfelder. Representation of the pareto front for heterogeneous multi-objective optimization. *Journal of Applied and Numerical Optimization*, 1(2):293–323, 2019.

24 J. Thomann and G. Eichfelder. A trust-region algorithm for heterogeneous multiobjective optimization. *SIAM Journal on Optimization*, 29(2):1017–1047, 2019.

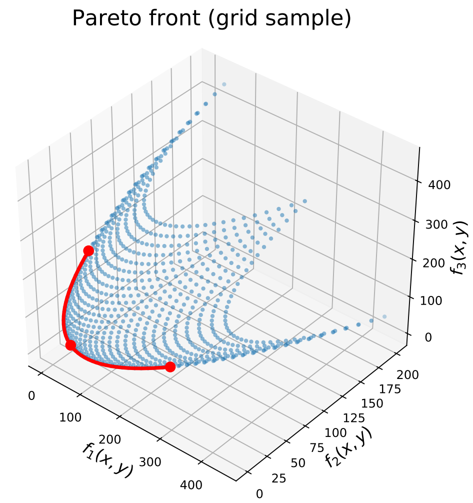


Figure 15 Objective space of the triobjective optimization problem presented in Figure 14: The Pareto front is outlined with a red curve. The blue points represent the image of the objective functions when the domain is sampled on a regular grid.

4.4 Performance Indicators

Carlos A. Coello Coello (CINVESTAV – Mexico, MX), Hisao Ishibuchi (Southern Univ. of Science and Technology – Shenzen, CN), Pascal Kerschke (Universität Münster, DE), Boris Naujoks (TH Köln, DE), and Tea Tušar (Jožef Stefan Institute – Ljubljana, SI)

4.4.1 Introduction

Within any aspect of scalability in multi-objective optimisation, performance indicators play a critical role. These indicators are important when comparing solutions and evaluating the performance of different algorithmic approaches.

Within many-objective optimisation, one special aspect of scalability in multi-objective optimisation, performance indicators recently received a lot of interest. With increasing objective space dimension, this aspect becomes more and more complex since such indicators need to fulfil different properties. Such indicators are wanted to be[1]

- Fast to compute
- Independent from the optimal Pareto set/front
- Independent from the shape of the Pareto front
- Pareto-compliant
- Not emphasizing boundary points
- Measures spread in the decision and objective space
- Scalable in the number of objectives
- Interpretable
- Featuring only a few parameters (that are easy to understand)
- Sharing the ability to represent preference of the DM
- Sharing biases of the indicator that can be understood (are known)

[1] List compiled by a working group on the topic "Performance Indicators" at the Lorentz Center workshop MACODA (MAny Criteria Optimization and Decision Analysis, 16. – 20. September 2020, http://lorentzcenter.nl/lc/web/2019/1160/info.php3?wsid=1160&venue=Oort)

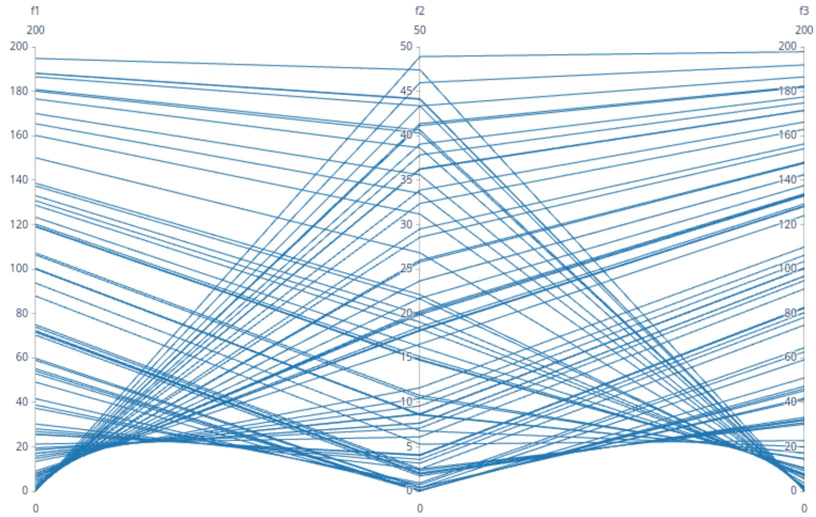


Figure 16 Parallel coordinates chart of a random sample of the Pareto front of the triobjective optimization problem presented in Figure 14. This chart complements the information already shown in Figure 15. There is a rich correlation structure between objectives: Both conflict and agreement among objectives are visible. This pattern closely resembles the features observed in the real-world problem shown in Figure 9.

The idea to discuss the aspect of performance indicators at the Dagstuhl seminar origins in the understanding that no single indicator exists that fulfils all these properties. Moreover, recent work indicates that optimally distributed points generated in the sense of different indicators are not optimal in any case [28]. Since the distribution of points is a critical aspect, the working group started with the idea to use statistical method to generate these points and build a new performance indicator based on the corresponding methodology.

The remainder of the text is organised as follows: The following part briefly summarises the history of performance indicators in multi- and many-objective optimisation. It thus summarises the preliminary work that lead to the one at hand. This is followed by a detailled description of involved methods and the implementation derived during the Dagstuhl seminar. Finally, first result are presented and a short overview on conclusion and future work is provided.

4.4.2 Preliminary Work

In the early days of multi-objective evolutionary algorithms (MOEAs), no performance indicators were adopted to assess performance, and comparisons of results were based purely on graphical representations of the approximations generated by two or more MOEAs. The first performance indicators were proposed in the mid-1990s. Some examples are: Distributed Spacing (ι) [27], Attainment Functions [12] and Efficient Set Spacing [24]. However, it was in the late 1990s when a wide variety of performance indicators were introduced. For example, David Van Veldhuizen [29] proposed: Generational Distance, Error Ratio, Maximum Pareto Front Error, Average Pareto Front Error, Overall Nondominated Vector Generation, Overall Nondominated Vector Generation Ratio and Generational Nondominated Vector Generation. Zitzler et al. [36, 34] proposed: Relative Coverage Comparison of Two Sets, Size of the Space Covered and Hypervolume. Later on, Zitzler also proposed the ϵ-indicator [35].

During this period (late 1990s and early 2000s), there were also several concerns regarding the appropriate methodology to assess the performance of a MOEA (see for example [16]). But a more important issue that soon arose was Pareto compliance. In 2003, Zitzler et al. [37] showed that most of the performance indicators that were in common use at that time were Pareto non-compliant, which meant that their results were unreliable. The hypervolume was identified as the only unary indicator which is Pareto compliant, but its high computational cost when dealing with problems having a high number of objectives triggered a significant amount of research in the last 15 years [31, 2, 23, 15].

Throughout the years, many other performance indicators have been proposed (see for example [10, 11, 6, 26, 18, 3, 7, 25, 20, 14]). Also, a number of surveys on performance indicators are currently available (see for example [21, 5, 22, 17]). Also, some researchers have proposed other interesting ideas such as the use of an ensemble of performance indicators [32].

Although the development of performance measures for assessing convergence and diversity of the approximations generated by a MOEA is a fundamental topic in evolutionary multi-objective optimization, in recent years, there have been few papers focusing on the development of new performance measures. The emphasis in recent years has been on the development of performance measures to assess performance in problems having a large number of objectives. In this case, diversity is of particular interest, and some interesting proposals have been made in that regard (see for example [13, 30]).

While the work of Harding and Saff [13] is more theory driven, there are some recent publications that address the construction of weights vectors [33] and the easy creation of any arbitrary number of uniformly distributed reference points [8]. Finally, [28] analyzes nine quality indicators using their approximated optimal distributions for $p = 3$. The analysis demonstrates that uniformly-distributed objective vectors over the entire Pareto front are not optimal in many cases. Each quality indicator has its own optimal distribution for each Pareto front.

4.4.3 Methods and Implementation

We propose a performance indicator based on reference vectors. The advantages of the latter are manifold: for instance, they (i) allow to incorporate a decision maker's preferences, (ii) are independent from the shape of the true Pareto front, and (iii) are scalable in the number of objectives.

Reference vectors $\mathbf{r} = (w_1, \ldots, w_p) \in [0,1]^p$ are p-dimensional vectors (in the objective space) and their elements (or weights) w_i indicate how strong the respective vector is affected by each of the p underlying objectives $f_1, \ldots, f_p$. W.l.o.g. we further assume $\sum_{i=1}^{p} w_i = 1$. In most applications, those reference vectors are uniformly distributed across the $(p-1)$-dimensional Pareto front (see, e.g., [4]). However, a variety of research on efficient sampling strategies has shown that evenly spaced structures (like a grid layout in ≥ 2 dimensions) are suboptimal. Therefore, we decided to utilize two sophisticated sampling strategies – *Latin Hypercube Sampling (LHS)* [19] and *Maximally Sparse Selection (MSS)* [9] – for finding promising weight configurations (which in turn will hopefully result in a suitable alignment of the reference vectors across the Pareto front).

In the first strategy, we use LHS to generate a set of K coefficient vectors $\boldsymbol{\theta}^{(k)} = \left(\theta_1^{(k)}, \theta_2^{(k)}, \ldots, \theta_{p-1}^{(k)}\right) \in \left[0, \frac{\pi}{2}\right]^{p-1}$ with $k = 1, \ldots, K$ in the $(p-1)$-dimensional box with lower bounds 0 and upper bounds $\pi/2$. We treat those vectors as angles in the $(p-1)$-dimensional

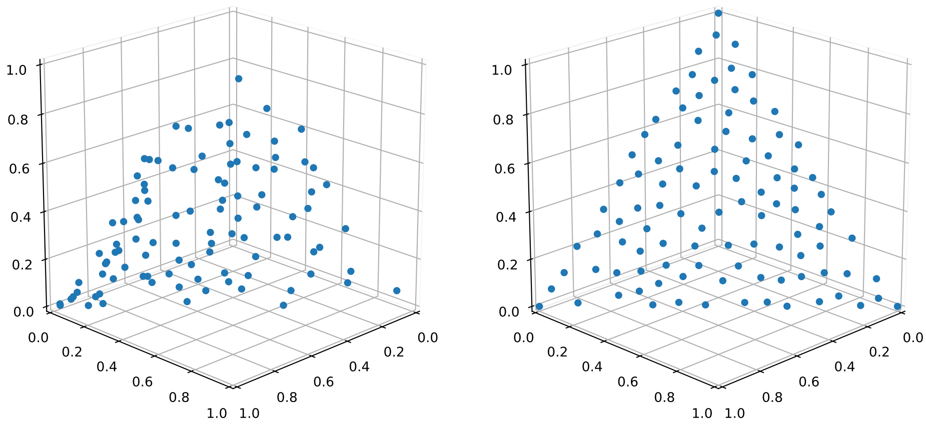


Figure 17 Schematic example of our proposed approach for generating 100 weight vectors in a 3-objective problem. The left image shows the results using LHS, and the right one is based on MSS.

space and use them to compute the corresponding p-dimensional (polar) coordinates $\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_p^{(k)})$ (in the so-called hyperspherical coordinate system [1]):

$$\begin{aligned} x_1^{(k)} &= \cos(\theta_1^{(k)}) \\ x_j^{(k)} &= \cos(\theta_j^{(k)}) \cdot \Pi_{i=1}^{j-1} \sin(\theta_i^{(k)}), \qquad j = 2, \dots, p-1 \\ x_p^{(k)} &= \Pi_{i=1}^{p-1} \sin(\theta_i^{(k)}). \end{aligned}$$

The resulting vectors are located on the $(p-1)$-dimensional unit sphere and thus contradict the desired property of all weights summing up to one. Thus, we project the coordinates onto the $(p-1)$-dimensional simplex by normalizing the polar coordinates (with the L1-norm of the respective vector's distance to the origin), i.e., $w_i^{(k)} := x_i^{(k)} / \sum_{j=1}^{p} x_j^{(k)}$.

On the other hand, the MSS sampling strategy already starts with points on the $(p-1)$-dimensional simplex. Each of the initial $\overline{K}$ simplex points $\mathbf{s}^{(k)} = \left(s_1^{(k)}, \dots, s_p^{(k)}\right) \in [0,1]^p$, where $\sum_{i=1}^{p} s_i^{(k)} = 1$ and $k = 1, \dots, \overline{K}$, is constructed by first sorting $p - -1$ uniformly distributed random numbers $a_i^{(k)} \in [0,1]$, $i = 1, \dots, p-1$, so that $0 \leq a_1^{(k)} \leq a_2^{(k)} \leq \cdots \leq a_{p-1}^{(k)} \leq 1$. We interpret the numbers 0 and 1 as $a_0^{(k)}$ and $a_p^{(k)}$, respectively. Then, the coordinates of the simplex points are computed as $s_i^{(k)} = a_i^{(k)} - -a_{i-1}^{(k)}$, for $i = 1, \dots, p$.

The final set of K simplex points, where $\overline{K} \gg K$, is constructed iteratively according to the maximally sparse selection. In a first step, all p unit vectors are added to this set. Then, until the size of the set reaches K, the point $\mathbf{s}^{(k)}$ that has the greatest sum of distances to all already chosen points is added to the set. By increasing the number of initial points $\overline{K}$, the resulting set has a more uniform distribution of points, but it also takes longer to compute.

4.4.4 Results

Figure 17 depicts two exemplary samples of reference vectors using our proposed approach for a 3-objective problem. The left image shows 100 reference vectors generated using the LHS approach, and the right image depicts the 100 reference vectors created using MSS as a sampling strategy.

In the depicted scenario, the distribution of the reference vectors generated by the MSS approach (i.e., right image) looks more promising than the one based on LHS (left). Therefore, we will consider only reference vectors generated by the MSS approach in the further study. Therein, we will use our reference vectors in combination with the R2 indicator (called our R2-MSS indicator in the following) in order to assess the quality of a selection of approximated optimal distribution sets that were obtained in [28]. The assessment will be performed both from a qualitative (visual) and quantitative (indicator-based) perspective.

We will use the approximated optimal distribution sets for nine different performance indicators (HV, IGD, IGD+, R2, NR2, $\epsilon+$, SE, Δ, PD), across MOO problems with six types of Pareto fronts (linear, concave, convex, and their inverted versions), and three, five and eight objectives, generated by Tanabe and Ishibuchi [28]. The performance of these sets will be evaluated by the R2 indicator where 100 reference vectors were chosen by the MSS approach from an initial selection of 1000 random reference vectors.

Tables 2 and 3 show the performance of our R2-MSS indicator that uses 100 reference vectors sampled with the MSS approach (see Section 4.4.3) on the approximated optimal distribution sets from [28]. The columns contain sets of the same shape and the rows those that correspond to the same indicator. For each set we show our R2-MSS indicator value and its rank among all sets of the same shape and dimension. In addition, we visualize the approximated optimal distribution sets for the 3-objective case. We have summed up the ranks for sets optimizing each indicator (across the three considered dimensions and all six shapes) and ordered the sets in the tables from the best (top) to the worst (bottom) indicator performance according to this sum.

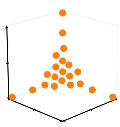
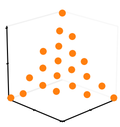
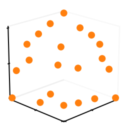
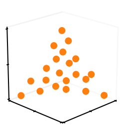
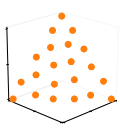
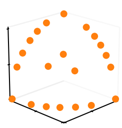
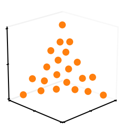
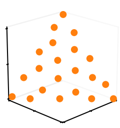
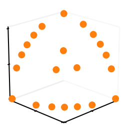
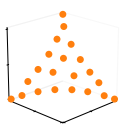
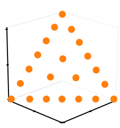
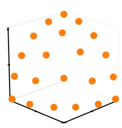
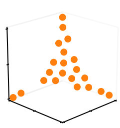
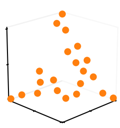
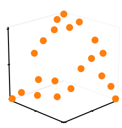
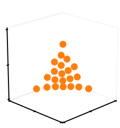
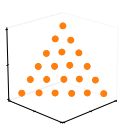
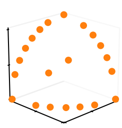
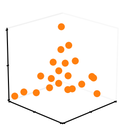
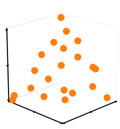
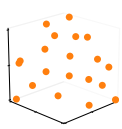
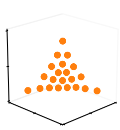
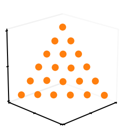
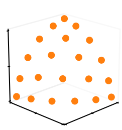
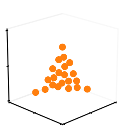
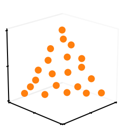
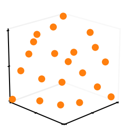
We can see that, unsurprisingly, the best results are almost always achieved on sets optimizing the regular R2 indicator, which are closely followed by those optimizing the NR2 indicator. As indicated by the plots, the sets for the HV indicator resemble those for the NR2 indicator and generally yield similar values. The sets optimizing the SE indicator perform really well on linear shapes and poorly on the others and are therefore scored fourth overall. The sets for indicators Δ, PD and $\epsilon+$ look unevenly distributed and are appropriately given poor scores. On the other hand, sets corresponding to the IGD and IGD+ indicators often have an even distribution of points, which is not in line with their performance as evaluated by our indicator. A further inspection suggests that the main reason for the poor performance of these sets is that they are missing points at the location of the unit vectors on non-inverted fronts (see Table 2), which are always included in our 100 selected reference vectors and are well approximated by some of the other sets. In fact, a short experiment has shown that if we add such extreme points to all sets from Table 2, the order of the indicators changes and sets optimizing the IGD+ and IGD indicators are scored more favourably.

4.4.5 Conclusions and Future Work

First of all we have to admit that the definition of a performance indicator needs much more than a neat idea. All aspects that needed to be considered to finally derive a proper performance indicator was more work and needed more discussion than expected. In addition, more compromises than expected with respect to different aspects like sampling strategy, reference-based or not etc., needed to define such an indicator had to be made.

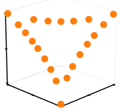
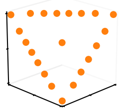
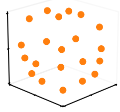
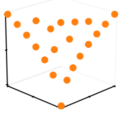
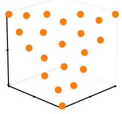
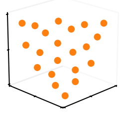
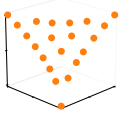
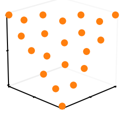
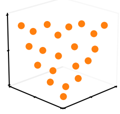
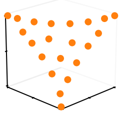
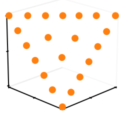
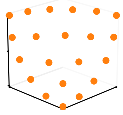
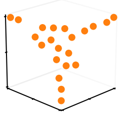
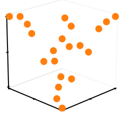
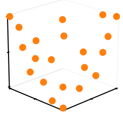
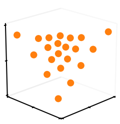
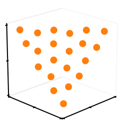
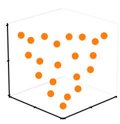
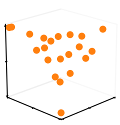
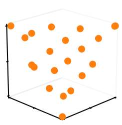
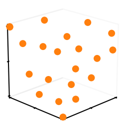
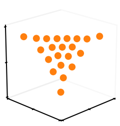
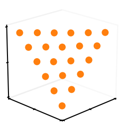
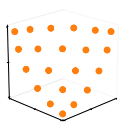
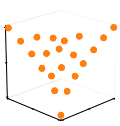
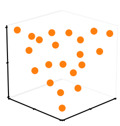
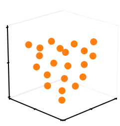
Nevertheless, we managed to define a new indicator based on statistical methods, namely LHS and MSS, to generate equally distributed points in high-dimensional spaces. Two such methods were involved in the implementation of a new indicator and first results supported the decision to continue investigations using MSS. Incorporating this method into an R2 like indicator led to our new indicator R2-MSS. This was tested on already existing approximately optimal distributed sets yielding promising results.

Table 2 The values (and their ranks) of our R2-MSS indicator that uses 100 MSS sampled reference vectors on the approximated optimal distribution sets for indicator I with convex, linear and concave shapes in dimensions $p = 3, 5, 8$ (see textual description for more details).

	convex				linear				concave			
I	$p = 3$	p	value	rank	$p = 3$	p	value	rank	$p = 3$	p	value	rank
R2		3	0.03497	1		3	0.06009	1		3	0.08219	4
		5	0.00814	1		5	0.01559	1		5	0.01872	3
		8	0.00269	2		8	0.00890	2		8	0.00980	3
NR2		3	0.03672	2		3	0.06169	4		3	0.08183	1
		5	0.00849	2		5	0.01807	4		5	0.01882	4
		8	0.00259	1		8	0.00953	3		8	0.01138	4
HV		3	0.03682	3		3	0.06159	3		3	0.08210	3
		5	0.00863	3		5	0.01956	5		5	0.01858	2
		8	0.00323	3		8	0.01002	4		8	0.00835	1
SE		3	0.03907	6		3	0.06186	5		3	0.08230	5
		5	0.01082	7		5	0.01596	2		5	0.01979	5
		8	0.00604	9		8	0.00784	1		8	0.01703	6
Δ		3	0.03803	5		3	0.06121	2		3	0.08329	6
		5	0.01049	6		5	0.01789	3		5	0.02113	6
		8	0.00358	5		8	0.01100	5		8	0.01393	5
IGD+		3	0.04144	8		3	0.07305	8		3	0.08193	2
		5	0.01135	8		5	0.04167	9		5	0.01854	1
		8	0.00368	6		8	0.02460	9		8	0.00843	2
PD		3	0.03718	4		3	0.06262	6		3	0.08926	8
		5	0.00990	4		5	0.02514	6		5	0.04626	7
		8	0.00348	4		8	0.01633	6		8	0.03326	7
IGD		3	0.04048	7		3	0.07284	7		3	0.10066	9
		5	0.01039	5		5	0.03929	7		5	0.06322	9
		8	0.00389	7		8	0.02457	7		8	0.05234	9
$\epsilon+$		3	0.04192	9		3	0.07699	9		3	0.08896	7
		5	0.01268	9		5	0.04149	8		5	0.04834	8
		8	0.00423	8		8	0.02457	7		8	0.03820	8

It seems clear that this approach needs to be examined on different solutions sets, in particular steaming from algorithm runs in the future. Different scenarios need to be investigated using different algorithms, different benchmarking functions yielding different Pareto front shapes, and compared to other performance indicators for different objective space dimensions. Finally, the new indicator should be investigated with respect to all of the

Table 3 The values (and their ranks) of our R2-MSS indicator that uses 100 MSS sampled reference vectors on the approximated optimal distribution sets for indicator I with inverted convex, linear and concave shapes in dimensions $p = 3, 5, 8$ (see textual description for more details).

	inverted convex				inverted linear				inverted concave			
I	$p = 3$	p	value	rank	$p = 3$	p	value	rank	$p = 3$	p	value	rank
R2		3	0.22936	1		3	0.19256	1		3	0.12929	1
		5	0.23385	3		5	0.20172	2		5	0.15192	1
		8	0.23359	2		8	0.21235	2		8	0.16376	1
NR2		3	0.23084	2		3	0.19670	4		3	0.13240	4
		5	0.23140	1		5	0.21995	5		5	0.16292	8
		8	0.24153	4		8	0.26378	7		8	0.17817	7
HV		3	0.23172	3		3	0.19625	3		3	0.13174	3
		5	0.23332	2		5	0.23400	7		5	0.16039	6
		8	0.23175	1		8	0.26654	8		8	0.18751	9
SE		3	0.23285	4		3	0.19450	2		3	0.13281	7
		5	0.23534	4		5	0.20003	1		5	0.15728	5
		8	0.23975	3		8	0.20727	1		8	0.17949	8
Δ		3	0.24822	7		3	0.21063	9		3	0.13613	8
		5	0.29475	7		5	0.21630	3		5	0.16178	7
		8	0.36511	8		8	0.22674	3		8	0.16612	3
IGD+		3	0.24816	6		3	0.20900	8		3	0.13274	5
		5	0.27227	6		5	0.23800	8		5	0.15403	3
		8	0.26911	5		8	0.25465	5		8	0.17334	5
PD		3	0.24967	8		3	0.20092	5		3	0.13280	6
		5	0.35168	9		5	0.23380	6		5	0.15457	4
		8	0.38488	9		8	0.27272	9		8	0.16828	4
IGD		3	0.28437	9		3	0.20827	7		3	0.13146	2
		5	0.33947	8		5	0.24139	9		5	0.15319	2
		8	0.33546	7		8	0.26131	6		8	0.16597	2
$\epsilon+$		3	0.23383	5		3	0.20549	6		3	0.13654	9
		5	0.24714	5		5	0.21832	4		5	0.16309	9
		8	0.27915	6		8	0.22893	4		8	0.17647	6

properties listed above. This investigation is expected to prove which properties are owned by our indicator and which are not. Based on all these results we expect some hints on how to improve our algorithm further and, thus, to hopefully improve performance measurement and comparison in many-objective optimisation.

References

1 L. E. Blumenson. A Derivation of n-Dimensional Spherical Coordinates. *The American Mathematical Monthly*, 67(1):63 – 66, 1960.

2 Lucas Bradstreet, Lyndon While, and Luigi Barone. A Fast Incremental Hypervolume Algorithm. *IEEE Transactions on Evolutionary Computation*, 12(6):714–723, December 2008.

3 Dimo Brockhoff, Tobias Wagner, and Heike Trautmann. On the Properties of the R2 Indicator. In *2012 Genetic and Evolutionary Computation Conference (GECCO'2012)*, pages 465–472, Philadelphia, USA, July 2012. ACM Press. ISBN: 978-1-4503-1177-9.

4 Ran Cheng, Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions on Evolutionary Computation*, 20(5):773 – 791, 2016.

5 Shi Cheng, Yuhui Shi, and Quande Qin. On the Performance Metrics of Multiobjective Optimization. In Ying Tan, Yuhui Shi, and Zhen Ji, editors, *Advances in Swarm Intelligence, Third International Conference, ICSI 2012*, pages 504–512, Shenzhen, China, June 17-20 2012. Springer. Lecture Notes in Computer Science Vol. 7331.

6 Carlos A. Coello Coello and Nareli Cruz Cortés. Solving Multiobjective Optimization Problems using an Artificial Immune System. *Genetic Programming and Evolvable Machines*, 6(2):163–190, June 2005.

7 Dilip Datta and Jose Rui Figueira. Some convergence-based M-ary cardinal metrics for comparing performances of multi-objective optimizers. *Computers & Operations Research*, 39(7):1754–1762, July 2012.

8 Kalyanmoy Deb, Sunith Bandaru, and Haitham Seada. Generating uniformly distributed points on a unit simplex for evolutionary many-objective optimization. In Kalyanmoy Deb, Erik Goodman, Carlos A. Coello Coello, Kathrin Klamroth, Kaisa Miettinen, Sanaz Mostaghim, and Patrick Reed, editors, *Evolutionary Multi-Criterion Optimization*, pages 179–190, Cham, 2019. Springer International Publishing.

9 Kalyanmoy Deb, Sunith Bandaru, and Haitham Seada. Generating Uniformly Distributed Points on a Unit Simplex for Evolutionary Many-Objective Optimization. In Kalyanmoy Deb, Erik Goodman, Carlos A. Coello Coello, Kathrin Klamroth, Kaisa Miettinen, Sanaz Mostaghim, and Patrick Reed, editors, *Evolutionary Multi-Criterion Optimization*, pages 179 – 190. Springer, 2019.

10 Ali Farhang-Mehr and Shapour Azarm. Diversity Assessment of Pareto Optimal Solution Sets: An Entropy Approach. In *Congress on Evolutionary Computation (CEC'2002)*, volume 1, pages 723–728, Piscataway, New Jersey, May 2002. IEEE Service Center.

11 Ali Farhang-Mehr and Shapour Azarm. Entropy-Based Multi-Objective Genetic Algorithm for Design Optimization. *Structural and Multidisciplinary Optimization*, 24(5):351–361, November 2002.

12 Carlos M. Fonseca and Peter J. Fleming. On the Performance Assessment and Comparison of Stochastic Multiobjective Optimizers. In *Parallel Problem Solving from Nature,äîPPSN IV*, Lecture Notes in Computer Science, pages 584–593, Berlin, Germany, September 1996. Springer-Verlag.

13 Douglas P. Hardin and Edward B. Saff. Discretizing Manifolds via Minimum Energy Points. *Notices of the American Mathematical Society*, 51(10):1186 – 1194, 2004.

14 Hisao Ishibuchi, Hiroyuki Masuda, Yuki Tanigaki, and Yusuke Nojima. Modified Distance Calculation in Generational Distance and Inverted Generational Distance. In António Gaspar-Cunha, Carlos Henggeler Antunes, and Carlos A. Coello Coello, editors, *Evolutionary Multi-Criterion Optimization, 8th International Conference, EMO 2015*, pages 110–125. Springer. Lecture Notes in Computer Science Vol. 9019, Guimarães, Portugal, March 29 – April 1 2015.

15 Andrzej Jaszkiewicz. Improved Quick Hypervolume Algorithm. *Computers & Operations Research*, 90:72–83, February 2018.

16 Joshua Knowles, Lothar Thiele, and Eckart Zitzler. A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland, 2006. revised version.

17 Miqin Li and Xin Yao. Quality Evaluation of Solution Sets in Multiobjective Optimisation: A Survey. *ACM Computing Surveys*, 52(2), May 2019. Article number: 26.

18 Giovanni Lizárraga Lizárraga, Arturo Hernández Aguirre, and Salvador Botello Rionda. G-Metric: an M-ary Quality Indicator for the Evaluation of Non-dominated Sets. In *2008 Genetic and Evolutionary Computation Conference (GECCO'2008)*, pages 665–672, Atlanta, USA, July 2008. ACM Press. ISBN 978-1-60558-131-6.

19 Michael D. McKay, Richard J. Beckman, and William J. Conover. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 42(1):55 – 61, 2000.

20 Seyedali Mirjalili and Andrew Lewis. Novel Performance Metrics for Robust Multi-Objective Optimization Algorithms. *Swarm and Evolutionary Computation*, 21:1–23, April 2015.

21 Tatsuya Okabe, Yaochu Jin, and Bernhard Sendhoff. A Critical Survey of Performance Indices for Multi-Objective Optimization. In *Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003)*, volume 2, pages 878–885, Canberra, Australia, December 2003. IEEE Press.

22 Nery Riquelme, Christian Von Lücken, and Benjamin Baran. Performance Metrics in Multi-Objective Optimization. In *2015 Latin American Computing Conference (CLEI)*, pages 1 – 11. IEEE Press, October 2015.

23 Luis M. S. Russo and Alexandre P. Francisco. Quick Hypervolume. *IEEE Transactions on Evolutionary Computation*, 18(4):481–502, August 2014.

24 Jason R. Schott. Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization. Master's thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Massachusetts, May 1995.

25 Oliver Schütze, Xavier Esquivel, Adriana Lara, and Carlos A. Coello Coello. Using the Averaged Hausdorff Distance as a Performance Measure in Evolutionary Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation*, 16(4):504–522, August 2012.

26 Vinícius L. S. Silva, Elizabeth F. Wanner, Sérgio A. A. G. Cerqueira, and Ricardo H. C. Takahashi. A New Performance Metric for Multiobjective Optimization: The Integrated Sphere Counting. In *2007 IEEE Congress on Evolutionary Computation (CEC'2007)*, pages 3625 – 3630, Singapore, September 2007. IEEE Press.

27 N. Srinivas and Kalyanmoy Deb. Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms. *Evolutionary Computation*, 2(3):221–248, Fall 1994.

28 Ryoji Tanabe and Hisao Ishibuchi. An Analysis of Quality Indicators Using Approximated Optimal Distributions in a Three-dimensional Objective Space. *IEEE Transactions on Evolutionary Computation*, pages 1–15, 2020.

29 David A. Van Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Department of Electrical and Computer Engineering. Graduate School of Engineering. Air Force Institute of Technology, Wright-Patterson AFB, Ohio, USA, May 1999.

30 Handing Wang, Yaochu Jin, and Xin Yao. Diversity Assessment in Many-Objective Optimization. *IEEE Transactions on Cybernetics*, 47(6):1510–1522, June 2017.

31 Lyndon While, Phil Hingston, Luigi Barone, and Simon Huband. A Faster Algorithm for Calculating Hypervolume. *IEEE Transactions on Evolutionary Computation*, 10(1):29–38, February 2006.

32 Gary G. Yen and Zhenan He. Performance Metric Ensemble for Multiobjective Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 18(1):131–144, February 2014.

33 Saul Zapotecas-Martínez, Hernán E. Aguirre, Kiyoshi Tanaka, and Carlos A. Coello Coello. On the Low-Discrepancy Sequences and Their Use in MOEA/D for High-Dimensional Objective Spaces. In *2015 IEEE Congress on Evolutionary Computation (CEC)*, pages 2835–2842, May 2015.

34 Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195, 2000.

35 Eckart Zitzler, Marco Laumanns, and Stefan Bleuler. A Tutorial on Evolutionary Multiobjective Optimization. In Xavier Gandibleux, Marc Sevaux, Kenneth Sörensen, and Vincent T'kindt, editors, *Metaheuristics for Multiobjective Optimisation*, Lecture Notes in Economics and Mathematical Systems, pages 3 – 37, Berlin, 2004. Springer.

36 Eckart Zitzler and Lothar Thiele. Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, November 1999.

37 Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M. Fonseca, and Viviane Grunert da Fonseca. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation*, 7(2):117–132, April 2003.

4.5 KaKaRaKe – User-Friendly Visualization for Multiobjective Optimization with High-Dimensional Objective Vectors

Kerstin Dächert (Fraunhofer ITWM – Kaiserslautern, DE), Kathrin Klamroth (Universität Wuppertal, DE), Kaisa Miettinen (University of Jyväskylä, FI), and Ralph E. Steuer (University of Georgia, US)

4.5.1 Introduction

Multiobjective optimization problems involve multiple conflicting objective functions to be optimized simultaneously. When solving such problems, we generate Pareto optimal solutions reflecting different trade-offs among the objectives. Typically, some type of preference information coming from a decision maker, a domain expert, is needed to identify the most preferred Pareto optimal solution as the final one.

By studying different Pareto optimal solutions and providing preference information, the decision maker can learn about interdependencies among the objectives. This task can be supported by visualizations. Visualizations can also be used to represent the progress of the solution process.

Surveys of various visualization techniques to represent a set of Pareto optimal objective vectors include [9, 12, 13, 18]. They discuss widely-used visualization techniques like parallel coordinate plots (sometimes known as value paths), spider web charts, petal diagrams, star coordinate plots, and glyphs. Examples of more recently proposed visualization methods for multiobjective optimization are heatmaps [6], knowCube [16], interactive decision maps [11], the prosection method [17], and 3d-radvis [7]. Visualization aspects in multiobjective optimization are also discussed in [5].

Questions involved in developing and applying visualizations include how to scale visualizations with increasing numbers of objectives; how to support the decision maker to gain understanding of the progress of the solution process; how to conduct "sensitivity analysis" by which the decision maker can understand the consequences of actions; can visualizations point to directions where small sacrifices in some objective can provide good improvement in some other objectives; how to find solutions with "robustness" properties; and how to communicate this to the decision maker. Overall, the decision maker should be able to affect the appearance of the visualization.

Our motivation is to develop visualizations to provide support to decision makers by identifying interesting aspects of Pareto optimal objective vectors as solution candidates and of the solution process itself. We aim at handling larger numbers of objectives (say between five and nine) in a comprehensible way without cognitively burdening the DM too much.

When using traditional visualization techniques, e.g., radar plots or parallel coordinates plots, the decision maker can conveniently grasp important interrelationships among the objectives when they are placed close to each other yet finding the best order in which to present the objectives may require some trial and error. Our goal is to assist the decision maker in identifying aspects of interest by pre-processing the set of solution candidates. We have two goals: to detect which objectives are strongly correlated or uncorrelated with one another, and to reduce the number of solutions presented when desirable by detecting similarities among them. The latter can be understood as finding good representative solutions.

One can detect correlations among objectives, for example, to reduce the computational burden by decreasing the number of objectives (see, e.g. [1, 2]). However, we are not aware of any such approach that is applied in a combined way to reduce the visual effort of the decision maker by clustering both objectives and solution candidates simultaneously.

We propose a new visualization technique by applying bi- or co-clustering to the set of Pareto optimal objective vectors. In this way, we can simultaneously visualize similarities and differences among objective functions and solution candidates. To communicate the information visually, we modify the idea of parallel coordinate plots so that the distances between correlated objective functions are shorter than otherwise and similar solutions are given the same color. Thanks to this kind of visualization, the decision maker can handle higher numbers of objectives and solution candidates and concentrate on aspects that are of greatest interest.

As for the following, in Section 4.5.2 we describe how data for a study like this can be obtained. In Section 4.5.3 we survey clustering techniques applicable to our needs, and in Section 4.5.4 the new visualization technique of this study is proposed. Finally, we conclude in Section 4.5.5.

4.5.2 Data generation

As mentioned above, we want to support the decision maker in handling larger numbers of objectives than is usually the case and by this we mean problems with 5 or more objectives. Of course, if one thinks long enough, one can probably imagine problems with almost any number of objectives, but because of the almost total lack of work in the area, we will in this paper only focus on problems with up to 9 objectives to get things started. However, there is a kind of "Catch-22" with problems in this area. On one hand, people are reluctant to attempt applications with many more than 5 objectives as there are essentially no tools for processing points from such high-dimensional Pareto fronts, and on the other hand, people have been slow to begin work on tools for processing high-dimensional solution vectors because of the lack of data from problems upon which to test such tools.

To get around this we describe two methods for randomly generating nondominated vectors in objective space. The first method is for the generation of Pareto optimal solutions as if coming from a problem in which all of the objectives have no especial correlations with one another. This method uses the random multiple objective linear program (MOLP) generator described in the documentation for Adbase [15] and then uses that code to solve the resulting MOLPs for all nondominated vertices of the problem's feasible region in the objective space. Data sets of any size for any number of objectives can be generated in this way by adjusting the parameters (numbers of objectives, constraints, variables) of the generator. Once a data set of Pareto solution vectors is generated, it is immaterial how it was generated for testing.

The second method is for the generation of Pareto solutions as if coming from a problem in which there are groups of objectives that have within group correlations with one another. For instance, consider a problem with 9 objectives such that 4 of the objectives are clustered in one group, 3 are in another, and the last two are in a third. In this method, the feasible region of a problem is generated in the same way as the first method, but instead of letting the MOLP generator generate the gradients of the objectives, the gradients of the objectives are randomly generated by a special routine that assures the clusters desired and their within group correlations. In this way, we can develop as many data sets of the types desired as needed. As for the example of this report, only the second method was necessary to generate data for it.

4.5.3 Clustering techniques

We now assume a given two-dimensional data set where each row corresponds to a Pareto optimal objective vector (solution candidate) and each column corresponds to an objective function. Our aim is to aggregate the given data matrix into solution-objective clusters according to the following rules.

1. Objectives are clustered if the values in any solution are "similar".
2. Solutions are clustered if the values in any objective are "similar".

"Similar" means as close as possible with respect to a specified distance, e.g., a Manhattan or Euclidean distance. The clustering is to be done simultaneously in both dimensions, rows and columns.

In the literature, there exist the concepts of biclustering and co-clustering which seem to denote the same, but are used in different communities and/or applications. A helpful survey is presented in [10].

Biclustering was first applied to bioinformatics, in particular to identify co-expressed genes under a subset of all conditions/samples, see [19] for a recent overview. A mathematical review of successful biclustering techniques is given in [3]. The authors particularly show that most of them are based on singular value decomposition (SVD). However, algorithms differ in the definition of biclusters. Some assume constant values in the data on rows within one cluster, some on columns, some on both. Others are more flexible and allow coherent values along rows and/or columns.

There are also "spectral" variants, i.e., "spectral biclustering" and "spectral co-clustering", that make use of the underlying graph structure of the problem. Spectral algorithms are common in graph partitioning problems and refer to algorithms that compute eigenvalues, eigenvectors, and singular values to solve the underlying graph problem, see e.g., the lecture notes https://courses.cs.washington.edu/courses/cse521/16sp/521-lecture-11.pdf. There, we also find an illustrative example in which common clustering methods like k-means fail while a spectral algorithm detects a meaningful cluster. In [4], a spectral co-clustering algorithm is proposed for a text mining problem.

In the next section, we use a Python implementation of the biclustering algorithm of [8]. One should note that even though the terms bi- and co-clustering are often regarded as synonyms, they refer to different algorithms in the scikit-learn package employed. More details are given below.

4.5.4 Preliminary results

We present tentative numerical results for a data set containing 88 mutually nondominated objective vectors with nine objective functions. The data were generated by the second method described in Section 4.5.2, aiming at three objective clusters. The tests were implemented in Python 3.7 using the scikit-learn package for data clustering [14] and plotly.express to generate the parallel coordinate plots. The data were clustered w.r.t. objective vectors (rows) and objective functions (columns) using spectral biclustering [8]. This includes an automatic data normalization which was implemented as log-normalization, see the documentation on https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralBiclustering.html for further details.

The original, unclustered data are shown in a parallel coordinates plot in Figure 18a, while Figures 18b to 18d show the results obtained with spectral biclustering using log-normalization for different numbers of solution clusters (S) and objective clusters (O). Each solution cluster is identified by a specific color, and the objective clusters are distinguished by larger distances between the coordinate lines of the different clusters.

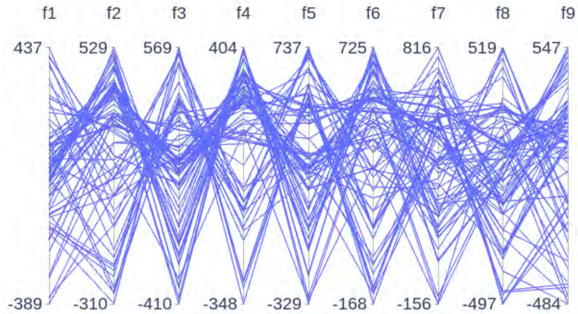
The visualizations nicely demonstrate that biclustering is a valuable tool to reveal tendencies among the solutions and correlations between the objective functions. For appropriately chosen values of S and O, the three objective clusters that are present in the data are retrieved. However, we note that the outcome largely depends on the parameter settings, in particular on the choices of S and O, but also on the employed normalization method.

In the future, we would like to use a measure to evaluate the quality of the visualization. Moreover, it would be interesting to automatically test different sizes of row and column clusters and present the “best” (with respect to a quality measure) result or results to the decision maker. Another open question is the way the columns are ordered. So far, we simply use the output of the biclustering algorithm. In the figures presented above we manually inserted gaps between the different clusters. By varying the sizes of these gaps, they could also serve as a source of information for the decision maker, e.g. by linking larger distances to a lower correlations between clusters. This and other technical issues like an automation are left for further improvements in the future.

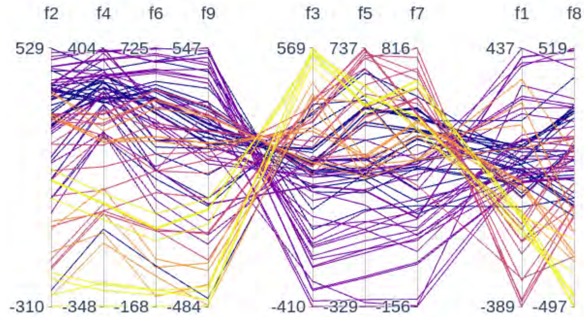
4.5.5 Conclusions

We have proposed a novel way of visualizing sets of Pareto optimal objective vectors by applying bi- or co-clustering and modifying parallel coordinate plots. Thanks to these visualizations, the decision maker can gain insight in the correlations among both objective functions and objective vectors simultaneously.

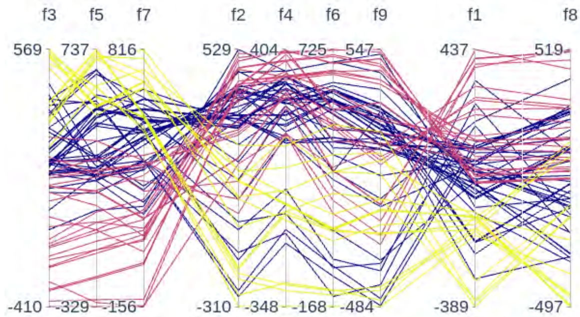
The novel visualizations can be applied to analyze any set of objective vectors. They can also be applied as a part of an interactive solution process. Our future research direction is to apply the findings and develop visualization assistance for solving multiobjective optimization problems with more than three objective functions.



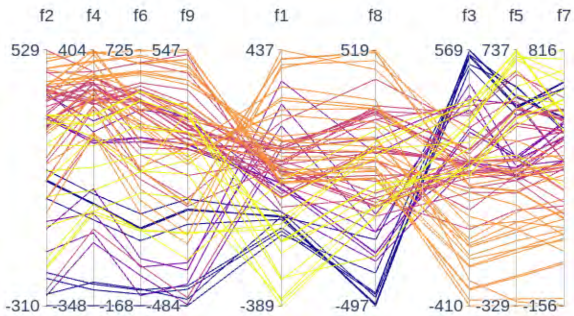
(a) Original data set



(b) BC with $S = 5$ and $O = 3$



(c) BC with $S = 3$ and $O = 4$



(d) BC with $S = 5$ and $O = 4$

Figure 18 Spectral biclustering applied to a data set containing 88 nondominated points (a). Subfigures (b)-(d) show biclustering results for S solution clusters and O objective clusters.

Acknowledgements

We would like to warmly thank Bhupinder Singh Saini for his valuable implementation support. This research is related to the thematic research area DEMO (Decision Analytics utilizing Causal Models and Multiobjective Optimization, jyu.fi/demo) of the University of Jyvaskyla.

References

1 D. Brockhoff and E. Zitzler. Are all objectives necessary? On dimensionality reduction in evolutionary multiobjective optimization. In T. P. Runarsson, H. G. Beyer, E. Burke, J. J. Merelo-Guervós, L. D. Whitley, and X. Yao, editors, *Parallel Problem Solving from Nature, PPSN IX, Proceedings*, pages 533–542, Berlin, Heidelberg, 2006. Springer.

2 D. Brockhoff and E. Zitzler. Dimensionality reduction in multiobjective optimization: The minimum objective subset problem. In K. H. Waldmann and U. M. Stocker, editors, *Operations Research Proceedings 2006*, pages 423–429, Berlin, Heidelberg, 2007. Springer.

3 S. Busygin, O. Prokopyev, and P. M. Pardalos. Biclustering in data mining. *Computers & Operations Research*, 35(9):2964–2987, 2008.

4 I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 269–274, 2001.

5 C. M. Fonseca, C. A. Antunes, R. Lacour, K. Miettinen, P. M. Reed, and T. Tusar. Visualization in multiobjective optimization. In S. Greco, K. Klamroth, J. D. Knowles, and G. Rudolph, editors, *Understanding Complexity in Multiobjective Optimization, Report from Dagstuhl Seminar 15031*, pages 129–139, Dagstuhl, 2015.

6 J. Hettenhausen, A. Lewis, and S. Mostaghim. Interactive multi-objective particle swarm optimization with heatmap-visualization-based user interface. *Engineering Optimization*, 42(2):119–139, 2010.

7 A. Ibrahim, S. Rahnamayan, M. V. Martin, and K. Deb. 3d-radvis: Visualization of Pareto front in many-objective optimization. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 736–745. IEEE, 2016.

8 Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein. Spectral biclustering of microarray data: Coclustering genes and conditions. *Genome Research*, 13:703–716, 2003.

9 P. Korhonen and J. Wallenius. Visualization in the multiple objective decision-making framework. In J. Branke, K. Deb, K. Miettinen, and R. Slowinski, editors, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 195–212. Springer, Berlin, 2008.

10 H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data*, 3(1):1–58, 2009.

11 A. V. Lotov, V. A. Bushenkov, and G. K. Kamenev. *Interactive Decision Maps: Approximation and Visualization of Pareto Frontier*. Kluwer Academic Publishers, Boston, 2004.

12 A. V. Lotov and K. Miettinen. Visualizing the Pareto frontier. In J. Branke, K. Deb, K. Miettinen, and R. Slowinski, editors, *Multiobjective Optimization: Interactive and Evolutionary Approaches*, pages 213–243. Springer, Berlin, 2008.

13 K. Miettinen. Survey of methods to visualize alternatives in multiple criteria decision making problems. *OR Spectrum*, 36(1):3–37, 2014.

14 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

15 R. E. Steuer. "ADBASE: A multiple objective linear programming solver for all efficient extreme points and unbounded efficient edges," Version 17.1, Terry College of Business, University of Georgia, Athens, Georgia, 2017.

16 H. L. Trinkaus and T. Hanne. knowCube: A visual and interactive support for multicriteria decision making. *Computers and Operations Research*, 32(5):1289–1309, 2005.

17 T. Tusar and B. Filipic. Visualization of Pareto front approximations in evolutionary multiobjective optimization: A critical review and the prosection method. *IEEE Transactions on Evolutionary Computation*, 19(2):225–245, 2015.

18 M. J. Woodruff, P. M. Reed, and T. W. Simpson. Many objective visual analytics: Rethinking the design of complex engineered systems. *Structural and Multidisciplinary Optimization*, 48(1):201–219, 2013.

19 J. Xie, A. Ma, A. Fennell, Q. Ma, and J. Zhao. It is time to apply biclustering: A comprehensive review of biclustering applications in biological and biomedical data. *Briefings in Bioinformatics*, 20(4):1450–1465, 2018.

4.6 Supporting Problem Solving with Many Decision Makers in Multi-objective Optimization

Michael Emmerich (Leiden University, NL), Jussi Hakanen (University of Jyväskylä, FI), Patrick Reed (Cornell University, US), Pradyumn Kumar Shukla (KIT – Karlsruher Institut fürTechnologie, DE), Jürgen Branke (University of Warwick, GB), Günter Rudolph (TU Dortmund, DE), Sanaz Mostaghim (Universität Magdeburg, DE), Heike Trautmann (Universität Münster, DE)

4.6.1 Introduction

The problem of multiobjective optimization changes qualitatively as soon as many decision makers are involved. All problems that occur when a single decision maker is involved are inherited, but new problem aspects are added. Three main additional aspects are:

1. The different decision makers can differ on the constraints and objective functions that are relevant for the problem, and the way they are computed.
2. They may have different preferences for the different objective functions and this way the problem of fairness arises, that is the problem of considering different preferences in a balanced manner.
3. There is a possibility of negotiations and group dynamics that should be considered in designing decision making processes. Moreover, decision makers might form coalitions and there might be different types of (power) relations between decision makers and hidden objectives/agendas.

In this report we consider the somewhat ideal situation of a group of equal decision makers that are able and willing to express their preferences. In such cases computer systems can be used to find out solutions that are non-dominated with respect to all objectives considered by the decision makers and among them present solutions that achieve a high performance in fairness on the one side, and total gain in terms as being close in average to the decision makers' preferred solutions or reference points on the other side.

4.6.2 Problem formulation

General decision making problem

There are three main steps for finding a consensus solution when multiple decision makers are involved:

1. Agreement on the model/goal formulation(s)
2. Selecting a subset of interesting solutions from a large set of alternatives, consensus on solvers
3. Selecting a single solution in mediated negotiation

To start with, all the DMs need to agree on the optimization problem formulation which is not a trivial task as such since different DMs may have different opinions on what is important to take into account. Therefore, this first step is usually an iterative process where alternative formulations are considered and some illustrative results are then shown to the DMs in a facilitated manner. One alternative approach to consider here is the *value focused thinking* approach by Prof. Ralph Keeney [3]. There the idea is to start from the values that the DM actually cares and, then, identify suitable criteria to measure them and define their relative importance. In case of multiple decision makers in all of these steps consensus needs to be achieved among DMs, making the value focused thinking process particularly challenging. In this report, we will not consider the first step in more detail.

When the problem formulation has been agreed, the aim of the second step is to provide solution candidates for the DMs based on their preferences. There exist different ways of providing preferences but, in this report for simplicity, we assume that every DM specifies preferences in the form of a reference point consisting of aspiration levels for each objective. In the case of multiple DMs, it can be assumed that the reference points can vary a lot meaning that the preferences have clear conflicts. In such a case, it is not clear what kind of solution candidates should be presented to the DMs. We feel that this step can be supported by computational tools and that will be our focus in this report. Our idea is based on using measures for fairness and gain when evaluating the solution candidates.

Keeping in mind that the overall goal is to find a single solution for the problem such that all the DMs can agree to that, the last step requires also a facilitated process. Finding the consensus can be difficult especially in cases where the DMs have very conflicting preferences.

Fairness and gain measures

As opposed to multi-objective optimization with a single DM, in the case of many-decision makers a solution can be unfair in the sense that it is closer to one decision maker's preference than to that of another decision maker.

But how can we measure the preference of a decision maker? Examples of preference modeling techniques are

- the specification of a preference by reference points, that are ideal points for each of the decision makers. There apply certain constraints of setting such ideal points. For instance, in the examples in this paper (*NSGA-II for teams*) we suggest to choose the reference point from the Pareto front approximation. It may also be possible to allow arbitrary reference points that our approach would then map onto the Pareto front approximation.
- the assessment of preferences by means of achievement scalarization functions [5]
- the definition of *desirability functions* for each objective and decision maker and aggregating *desirablility indices*, that are products of desirability functions over all objectives considered by the decision makers (cf. [2]).

In the following we illustrate fairness and gain measures by a relative straightforward, yet useful approach of preference formulation. The approach can be transferred easily to other, potentially more sophisticated means of preference modelling.

The basis of modeling gain and fairness can be the notion of losses with respect to the ideal solution a DM can obtain. This loss is also termed *Pareto regret* ([4]). In the case that we define the loss by means of distance to reference vectors the Pareto regret can be computed as:

$$PR_j(x) = \sum_{i=1}^{m} (f_i(x) - r_{ij})$$

where j is the index of the decision maker, $j = 1, ..., d$, m is the number of objectives, and x is the final solution selected. In the following we will introduce some new definitions that are related to the notion of Pareto regret w.r.t. reference points which we have described above. These notions generalise easily to other notions of Pareto regret.

▶ **Definition 1.** Pareto Regret, Average Pareto Regret, Inequality in Pareto Regret Let PR_i denote the Pareto regret for each decision maker. Then the *average Pareto regret* of a solution x is defined as the average of the Pareto regrets of all DMs, i.e.,

$$\mathrm{APR}(x) = \frac{1}{d} \sum_{i=1}^{d} \mathrm{PR}_i(x)$$

and the *Inequality in Pareto Regret* (IPR) of a solution x is defined as the Gini index[2] of the Pareto regret.

$$\mathrm{IPR}(x) = \sum_{i=1}^{d} |\mathrm{PR}_i(x) - \mathrm{APR}(x)|$$

In the following, we will often denote the APR as *gain*, and the IPR as *fairness*.[3]

4.6.3 Matlab code

In order to illustrate our ideas, we wrote a small Matlab script. An example having two objectives is used which means that the objective space and the fairness-gain space have the same dimensionality. The reason for this choice is the ease of visualization and interpretation but the ideas work for problems with more objectives. Further, we have three DMs in our example but, again, the ideas work in case we have more.

The code produces plots in two different spaces: the original objective space and the fairness-gain space. We use a mesh grid to discretize the objective space and we then compute values for fairness and gain in all the points. In the objective space, we visualize the level curves of both the fairness and gain to have some idea of their behaviour with respect to each other. In addition, the reference points of all the DMs are shown.

[2] the average absolute deviation from the mean

[3] Mind that, following the convention in mathematical programming, we aim to minimise both objectives which strictly speaking would require using the more abstract terms IPR and APR. From that perspective fairness and gain are thus improved by means of minimisation.

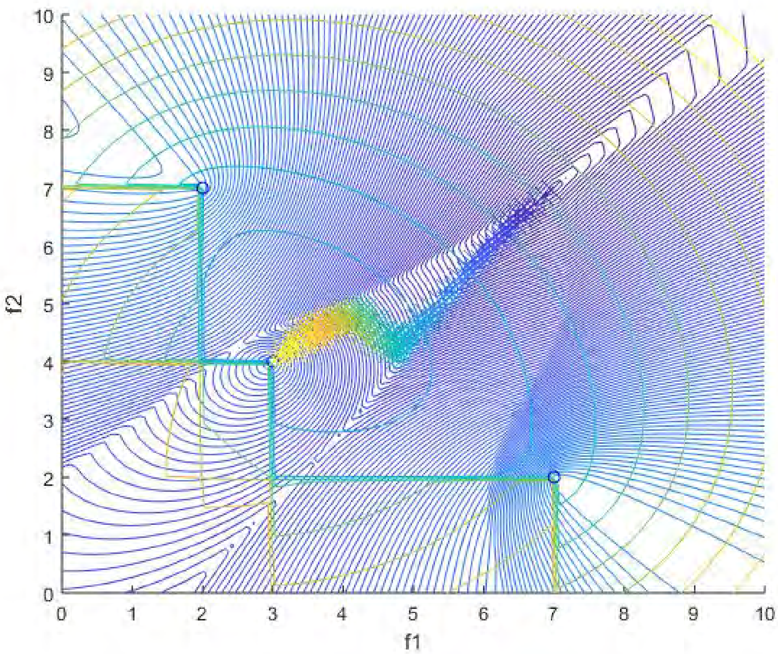


Figure 19 Reference points for DMs (blue circles) and level curves for fairness (denser) and gain (sparse) in the objective space.

Then, we do non-dominated sorting[1] in the fairness-gain space and obtain set of non-dominated points. We visualize these points in both the fairness-gain space and the original objective space. In order to be able to identify same non-dominated points in different plots, we color code them according to increasing values of fairness. An example of the visualizations in the objective space and the fairness-gain space are shown in Figures 19 and 20, respectively.

Figures 21 and 22 show the same trade-offs but this time for a particular example problem where there is a Pareto-optimal front that cannot be improved upon in the objective space (Figure 21. As can be seen, the solutions that belong to the Pareto front in the objective space form a rather peculiar line in the Gain/Fairness space and vice versa. We would be looking for solutions that perform well in both spaces as good compromise solutions.

4.6.4 NSGA-II for teams

It seems relatively straightforward to integrate the above ideas into the NSGA-II framework and focus the search towards solutions that are Pareto-optimal in the objective space, but also Pareto-optimal in the Fairness/Gain space. The proposed NSGA-II for teams algorithm first partitions a given set of solutions (population) into different layers of equal dominance ranks in the objective space. Then, for solutions of the same dominance rank, a non-dominated sorting procedure is executed w.r.t. the objectives fairness and gain.

This is the basic idea for the construction of the proposed algorithm termed *NSGA II for teams*. We leave the discussion of its performance, parameter, scalability studies to the future work and it shall also be remarked that NSGA-II can probably be replaced by other Pareto-based evolutionary multi-objective optimziation algorithms.

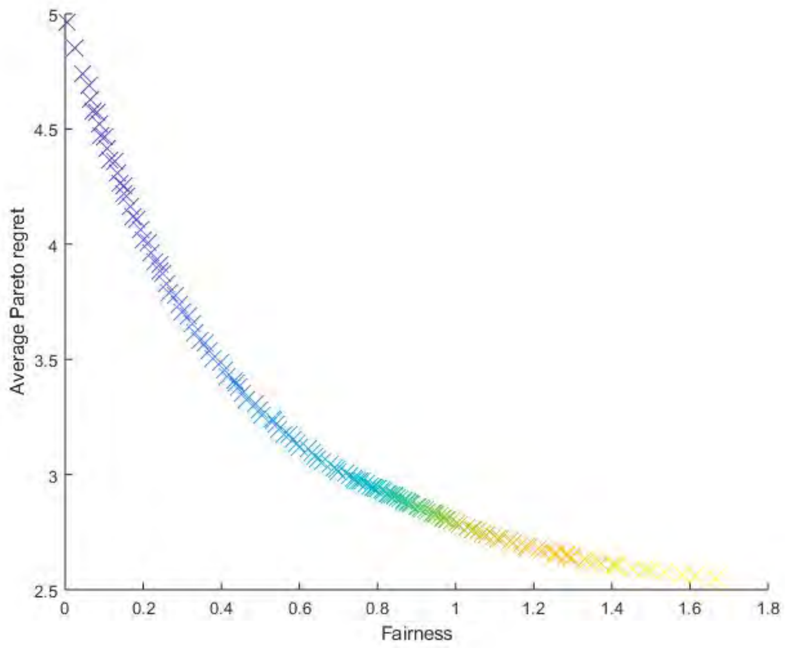


Figure 20 Non-dominated points in the fairness-gain space.

In summary our proposed idea of NSGA-II for teams proposes a straightforward way to integrate an important aspect of problems with many decision makers into EMOA frameworks, namely the trade-off between fairness and gain.

4.6.5 Relation to constraint satisfaction

This part of the report focuses on another aspect that distinguishes problems with many DMs those with a single DM which is the problem of constraint satisfaction in terms of conflicting views of the problem. It shall be noted that constraints and objective functions can coincide, if there is a minimum threshold for an objective function to be met for a solution to be feasible.

Let there be $d \in \mathbb{N}$ decision makers trying to arrive at a solution to an $m \in \mathbb{N}$ objective optimization problem. The (vector-valued) objective function $f(x) := (f_1(x), \ldots, f_m(x))$ is assumed to be a function from a non-empty set $X \subseteq \mathbb{R}^n$, where $n \in \mathbb{N}$ is the dimensionality of the search space. The function $f(x)$ is assumed to be the same for all decision makers (this assumption could, however, be relaxed). They may have their own set of constraints, or have a common set. We assume that decision-maker i has a constraint set $X_i \subseteq \mathbb{R}^n$, for all $i \in \{1, \ldots, d\}$. In the general form, the sets X_i's are can be assumed to be defined using k_i inequality and ℓ_i equality constrains in the following way:

$$X_i = \{x \in \mathbb{R}^n | g_1^i(x) \leq 0, \ldots, g_{k_i}^i(x) \leq 0, h_1^i(x) = 0, \ldots, h_{\ell_i}^i(x) = 0\}. \quad (2)$$

All the functions used in the description (2) above are assumed to be known.

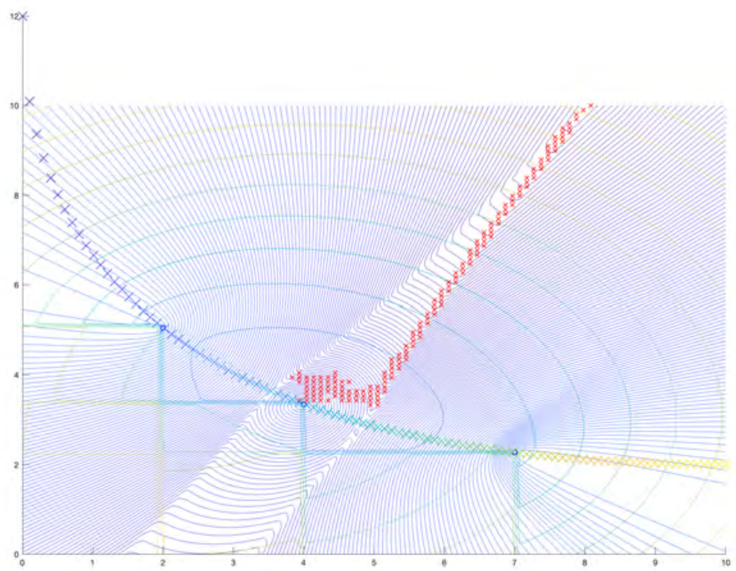


Figure 21 The coloured crosses are the Pareto front approximation in objective space, the red crosses are the solutions that form the Pareto front approximation in the Gain/Fairness space.

Detecting intra-inconsistencies within the constraint sets

The aim here is to detect whether the constraints X_i, defined by (2), are consistent or not. This would be done for all the decision-makers. To do this, for every decision-maker i, we consider the following optimization problem:

$$\begin{aligned} \max_{(u,x)\in\mathbb{B}^{k_i+\ell_i}\times X} \quad & \sum_{i=1}^{k_i+\ell_i} u_i \\ \text{s.t.} \quad & u_j g_j^i(x) \leq 0,\ \forall j = 1,\dots,k_i, \\ & u_l h_j^l(x) = 0,\ \forall l = 1,\dots,\ell_i, \end{aligned} \tag{Opt(i)}$$

where $\mathbb{B}$ denotes the set $\{0,1\}$.

▶ **Proposition 2.** *The constraint set of Opt(i) is non-empty.*

Proof. We assumed that X is nonempty. Let $x \in X$ be a feasible point. It is easy to see that the point $(0,x) \in \mathbb{B}^{k_i+\ell_i} \times X$ is feasible to Opt(i). ◀

Let n_i denote the (global) optimal value of Opt(i).

▶ **Proposition 3.** *If n_i equals $k_i + \ell_i$, then X_i is consistent.*

Proof. If the optimal value n_i equal to $k_i+\ell_i$, then $u_{\tilde{j}} = 1$ for all $\tilde{j} \in \{1,2,\dots,k_i+\ell_i\}$. From the constraints in Opt(i), this implies that there is an $\tilde{x} \in X$ that satisfies the constrains in the description of (2). Therefore, $\tilde{x} \in X_i$ and $X_i \neq \emptyset$. ◀

▶ **Proposition 4.** *If $n_i < k_i + \ell_i$, then X_i is inconsistent. Furthermore, the set of constraints in Opt(i) that correspond to $u_{\tilde{j}} = 1$ for $\tilde{j} \in \{1,2,\dots,k_i+\ell_i\}$ form the largest subset of constraints in X_i that is consistent.*

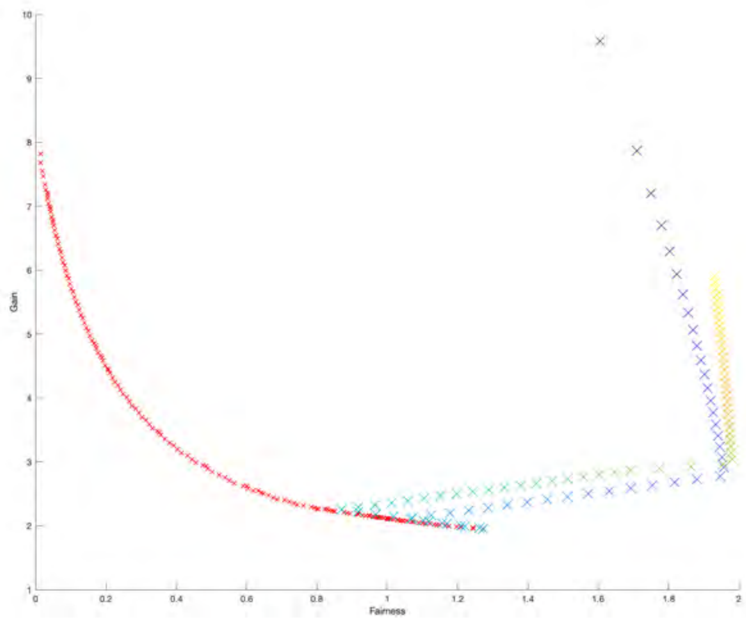


Figure 22 The red crosses are the Pareto front approximation in the Gain/Fairness space, the coloured crosses are the solutions that form the Pareto front approximation in the objective space.

Proof. If $n_i < k_i + \ell_i$, then $u_{\tilde{j}} = 0$ for exactly $k_i + \ell_i - n_i$ indices $\tilde{j}$'s. These are the constrains for which $u_{\tilde{j}} \neq 1$, meaning they cannot be satisfied. Those $\tilde{j}$'s for which $u_{\tilde{j}} = 1$ correspond to the constraint that can be simultaneously satisfied, and therefore the statement of the proposition follows. ◀

Solving Opt(i), therefore, is informative to the decision maker i. Depending on the solution to Opt(i), this either shows that the constraints formulated by him/her are consistent, or, in the inconsistent case, gives information about the constraints that are causing inconsistency. It is clear that all the X_i's need to be non-empty, before any optimization problem or preference information could be formulated.

▶ Remark. is to be noted that for checking whether a set is consistent or not can often be realized by minimizing a penalty function like $\sum_j \max\{0, g_j^i(x)\} + \sum_l |h_j^l(x)|$ or its differentiable version $\sum_j (\max\{0, g_j^i(x)\})^2 + \sum_l |h_j^l(x)|^2$. A minimal point x with minimal value zero is feasible, while a positive minimal value proves inconsistency. The latter can sometimes also be verified by a lower bounding procedure (like interval arithmetic) if this yields a positive lower bound for the penalty function. If only inequality constraints are present, a common feasibility test also is the minimization of some variable t subject to the constraints $g_j(x) \leq t$ for all j and $t \geq -1$ (to guarantee boundedness). A positive optimal value proves inconsistency, otherwise a feasible point x is generated. The advantage of our approach based on Opt(i) is that it explicitly gives the constraints that are causing inconsistency.

Detecting inter-inconsistencies across the constraint sets

In this section we assume that $X_i \neq \emptyset$ for all $i \in \{1, \dots, d\}$. The individual constraint sets of the decision-makers are, therefore, non-empty (i.e., the intra-inconsistencies, if any, are assumed to be removed). The aim now is to check if $\bigcap_{i=1}^d X_i \neq \emptyset$. Obviously, $\bigcap_{i=1}^d X_i = \emptyset$ would mean that there exist at least two decision-makers i and j such that $X_i \cap X_j = \emptyset$, meaning that these decision makers *cannot* be simultaneously satisfied. To check this, we consider the following optimization problem:

$$\begin{aligned} \max_{(v,x) \in \mathbb{B}^d \times X} \quad & \sum_{i=1}^{d} v_i \\ \text{s.t.} \quad & v_i g_j^i(x) \leq 0,\ \forall j = 1, \dots, k_i, \forall i = 1, \dots, d \\ & v_i h_l^i(x) = 0,\ \forall l = 1, \dots, \ell_i, \forall i = 1, \dots, d. \end{aligned} \tag{Opt}$$

▶ **Proposition 5.** *The constraint set of Opt is non-empty.*

Proof. We assumed that X is nonempty. Let $x \in X$ be a feasible point. It is easy to see that the point $(0, x) \in \mathbb{B}^d \times X$ is feasible to Opt(i). ◀

Let $\tilde{n}$ denote the (global) optimal value of Opt.

▶ **Proposition 6.** *Let $X_i \neq \emptyset$ for all i. If $\tilde{n}$ equals d, then $\bigcap_{i=1}^d X_i \neq \emptyset$.*

Proof. As $X_i \neq \emptyset$ for all $i \in \{1, \dots, d\}$ we obtain from Proposition 6 that all the components of u in Opt(i) equal 1, for every i. This means that a common value v_i can be used for all the inequalities and equalities that define X_j. Therefore, for every $i \in \{1, \dots, d\}$, it is sufficient to use just one boolean variable v_i to represent all the k_i inequality and ℓ_i equality constrainsts.

If the optimal value $\tilde{n}$ equal to d, then $v_i = 1$ for all $i \in \{1, \dots, d\}$. From the constraints in Opt, this implies that there is an $\tilde{x} \in X$ that satisfies the constrains in the description of (2) for every index i. Therefore, $\bigcap_{i=1}^d X_i \neq \emptyset$. ◀

▶ **Proposition 7.** *If $\tilde{n} < d$, then there are at least two decision makers having inconsistent constraint sets. Furthermore, the set of constraints in Opt that correspond to $v_i = 1$ for $i \in \{1, 2, \dots, d\}$ form the largest number of decision makers that have common feasible sets.*

Proof. If $\tilde{n} < d$, then $v_i = 0$ for exactly $d - \tilde{n}$ indices i's. These are the constraint sets of decision makers for which $v_i \neq 1$, meaning they cannot be satisfied simultaneously. Those i's for which $v_i = 1$ correspond to the constraint sets (or decision makers) that can be simultaneously satisfied, and therefore the statement of the proposition follows. ◀

Solving Opt, therefore, is informative to all the decision makers. Depending on the solution to Opt, this either shows that the constraints formulated by the decision makers are consistent globally, or, in the inconsistent case, gives information about the decision makers that are causing inconsistency.

4.6.6 Conclusions and future directions

This report discussed in general problems that arise when moving from problems with a single DM to problems with multiple or many DMs. We discussed the general process of solving such problems, and then focused on two more specific aspects, namely:

1. integrating fairness and gain in problems with conflicting preferences among the decision makers. For this we identified NSGA-II for teams and introduced the notion of the fairness-gain space in addition to the objective and decision space for problems with many DMs.
2. solving and discussing problems with conflicting constraints among different decision makers using a formal framework.

Future work will have to be done to further refine these ideas and test them on benchmarks or real world problem solving scenarios. Moreover, it might become necessary to develop additional techniques in case the number of DMs is large (in accordance to the term 'many objective optimization one might term this 'many decision makers' scenario). In this context the following idea might become very relevant:

Apart from including all DM preferences individually into the optimization and decision making process, an alternative and promising perspective for future research lies in reducing complexity of decision preferences prior to optimization. Thus, e.g. by means of clustering techniques, reference points which are representative for subgroups of decision makers could be generated which in combination then reduce the amount of different decision maker preferences and potentially reduces the complexity of agreeing on final compromise solutions. Ideally, cluster centers reflecting preference compromises of the subgroups should be presented to the decision makers prior to optimization.

In general such frameworks should also be integrated further with methodologies used in the MCDM community that have been developed for team decision making and consensus finding given only a small number of alternative solutions. There is a rich literature to be explored here and it would extend the scope of this report to provide a comprehensive overview.

Source code

Matlab-code is available on: http://moda.liacs.nl.

Acknowledgements

We are grateful to Prof. José Figueira about his input on the definitions of fairness/equity, Prof. Hisao Ishibuchi for contributions in the initial discussions, and Prof. Oliver Stein for his remarks on constraint feasibility.

References

1 Kalyanmoy Deb. Multi-objective optimisation using evolutionary algorithms: an introduction. In *Multi-objective evolutionary optimisation for product design and manufacturing*, pages 3–34. Springer, 2011.

2 Michael Emmerich, André Deutz, Longmei Li, Asep Maulana, and Iryna Yevseyeva. Maximizing consensus in portfolio selection in multicriteria group decision making. *Procedia Computer Science*, 100:848–855, 2016.

3 Ralph L. Keeney. Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, 92(3):537–549, 1996.

4 Wouter M Koolen. The pareto regret frontier. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 863–871. Curran Associates, Inc., 2013.

5 Kaisa Miettinen and Marko M Mäkelä. On scalarizing functions in multiobjective optimization. *OR spectrum*, 24(2):193–213, 2002.

4.7 Turning Objective Functions into Constraints? Or Vice Versa?

Georges Fadel (Clemson University – Clemson, US), Karl-Heinz Küfer (Fraunhofer ITWM – Kaiserslautern, DE), Manuel López-Ibáñez (University of Manchester, GB), Luís Paquete (University of Coimbra, PT), Stefan Ruzika (TU Kaiserslautern, DE), and Anita Schöbel (Fraunhofer ITWM – Kaiserslautern, DE)

4.7.1 Introduction and Motivation

Multiobjective mathematical programming problems are characterized by the presence of several, typically incommensurable objective functions. This circumstance is a fundamental difference to single objective optimization. As a consequence, the notion of optimality has to be revised. One common way to define optimality in the presence of multiple objective functions is the notion of Pareto-optimality. An efficient (or Pareto-optimal) solution to a mathematical programming problem is characterized by the fact that there does not exist another solution that is at least as good in all objectives and strictly better in at least one objective. The images of these efficient solutions are referred to as nondominated points. The goal of multiobjective optimization is to compute the set of nondominated points and at least one efficient solution for each nondominated point. For a rigorous introduction to multiobjective programming, we refer to the book [3].

It is well-known that there is an increase in difficulty when considering an arbitrary number of objective functions compared to the biobjective case [5]. This can be explained by the kind of trade-off which stems from the notion of Pareto optimality: In biobjective problems, two different nondominated points are characterized by the fact that one solution is better in one objective and worse in the other. As a consequence, there is a monotonicity among the nondominated points, i.e., they can be sorted by increasing values of the first objective and, as a consequence, the values of the second objective will turn out to be decreasingly sorted. This is not the case when having three or more objectives.

A research direction that has lately attracted quite some interest, is multiobjective problems with *many* objective. In this context, the term 'many' is not precisely defined. It is generally understood as pointing towards the fact that the mere number of objectives imposes some difficulty, e.g., in the theoretical analysis of structural properties, of the numerical solution of the problem, or in the decision-making process.

This situation is, in principle, not new to multiobjective optimization, since the presence of more than one objective inherently already imposes an increase in difficulty compared to the single-objective situation. One way of copying with this fact is as old as multiobjective optimization and widely spread: the idea of reducing the number of objective functions. This technique is called *scalarization* and there is a huge body of literature on scalarization [3, 4].

In this article, we address the question of whether it is beneficial to reformulate some given multiobjective optimization problem by converting one or more objectives to constraints. The resulting problem would have fewer objectives and should be easier to solve. However, only a subset of Pareto-optimal solutions is found and, therefore, this process of formulating and solving a smaller problem has to be iteratively repeated to guarantee finding all or a good representation of the Pareto optimal solutions. In a sense, this technique can be referred to as *partial scalarization.*

4.7.2 A First Theoretical Framework

We consider a multiobjective optimization problem with k objectives and m constraints. On top of these constraints, we denote additional constraints by X. This set X subsumes other restrictions, e.g., sign constraints, bounds on variables, or other side constraints that are not in the focus of the following consideration. We denote the encoding length of X by n. A multiobjective optimization problem can thus be referred to by $MO(k,m)$ and concisely stated as

$$\begin{aligned} \min \quad & \begin{pmatrix} f_1(x) \\ \vdots \\ f_{k-1}(x) \\ f_k(x) \end{pmatrix} \\ \text{s.t.} \quad & g_1(x) \leq 0 \\ & \quad \vdots \\ & g_m(x) \leq 0 \end{aligned} \tag{3}$$

One way of studying and balancing the trade-off between objectives and constraints is by means of the so-called *ε-constraint method* [7]. In this scalarization technique, one of the objective functions is minimized while all other are reformulated as constraints. In the following, we slightly adapt this idea: we choose one objective and reformulate it as a constraint while keeping all remaining objectives. In view of this partial scalarization, the above $MO(k,m)$ problem (Eq. 3) would be reformulated as the following (multiobjective) mathematical programming problem:

$$\begin{aligned} \min \quad & \begin{pmatrix} f_1(x) \\ \vdots \\ f_{k-1}(x) \end{pmatrix} \\ \text{s.t.} \quad & g_1(x) \leq 0 \\ & \quad \vdots \\ & g_m(x) \leq 0 \\ & f_k(x) \leq \varepsilon \end{aligned} \tag{4}$$

For some given ε, we refer to the above problem as $MO(k-1, m+1)$. Comparing the two problems $MO(k,m)$ and $MO(k-1, m+1)$, the sets of efficient solutions and nondominated points obviously differ (see [2, 6] on the effect of adding/deleting objective functions). The step of reducing the objectives by one and adding one more constraint can of course be iteratively applied until, at the end, the resulting problem is equivalent to the ε-constraint problem. This technique of reducing the number of objectives and increasing the number of constraints is key to our analysis.

To find all nondominated points of $MO(k,m)$ by means of solving problems of the kind $MO(k-1, m+1)$, one can use an iterative algorithm that varies the chosen bound ε and solves these smaller problems. Let I denote the number of relevant values for the bound ε. We refer to such an iterative algorithm by $\varepsilon ALG(k,m)$. In other words, $\varepsilon ALG(k,m)$ is an iterative algorithm based on partial ε-constraint scalarization calling I-times an algorithm $ALG(k-1, m+1)$ that solves some specific problem with k objectives and m constraints. Figure 23 illustrates this idea.

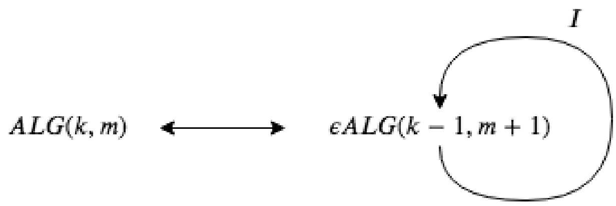


Figure 23 Illustration of partial ε-constraint scalarization used in an iterative manner.

The following result establishes an obvious connection between $\varepsilon ALG(k, m)$ and $ALG(k, m)$.

▶ **Theorem 1.** *$ALG(k, m)$ is faster than $\varepsilon ALG(k, m)$ if*

$$time[ALG(k, m)] < I \cdot time[ALG(k-1, m+1)]$$

Therefore, the algorithm $ALG(k, m)$ is computationally preferable if $I > \frac{time[ALG(k,m)]}{time[ALG(k-1,m+1)]}$. Note that I is determined by the number of nondominated solutions of $MO(k, m)$.

4.7.3 Some Examples

The Multiobjective Unconstrained (Integer) Quadratic Minimization Problem

The problem

$$\min(x - a)^2$$

for a given parameter $a \in \mathbb{R}^n$ and a (maybe integer) variable $x \in \mathbb{R}^n$ is a nice problem for visualization for $n = 2$. For example, Given $a_1, \ldots, a_k$ points in the plane, we may formulate k objective functions, namely $f_k(x) = (x - a_k)^2$ describing the distance from x to a_k. One can visualize what reducing the number of objective functions means: Instead of minimizing f_k we restrict the solution space to a ball with radius ε centered at a_k. It can be easily seen that due to the nonlinear constraint, the problem becomes much more complicated. While it can be used for visualization, it seems not to be suited for an analytical analysis.

Multiobjective Linear Programming

Multiobjective linear programming seems to be a suitable problem since

a) the multiobjective simplex algorithm is available for experiments;

b) the ε-constraint problem is again a multiobjective linear program that can be solved by the multiobjective simplex algorithm.

It remains open to use this setting for numerical experiments.

4.7.4 The Multiobjective Shortest Path Problem

In the following, we illustrate the application of these concepts for a particular case of the shortest path problem with k objectives in a directed acyclic graph. We consider an extension of the *pulling* algorithm [1], which is the fastest approach for finding the single-source single-sink shortest path in acyclic and topologically ordered networks. This extended pulling algorithm processes the nodes in the topological order. At each iteration i $(i = 1, \ldots, n)$, it

Table 4 Computational results.

k	objectives	constraints	CPU-time
3	3	0	0.00
	2	1	0.35
	1	2	1.45
4	4	0	0.00
	3	1	1.77
	2	2	6.84
	1	3	28.47

calculates the nondominated shortest paths from node 1 to node v_i by considering only the distances from each node v_j that is incident to j, $j < i$. Since the network is topologically ordered, the nondominated shortest paths to each node v_j were already computed.

The following pseudo-code implements the algorithm above, where $ND_{1,\dots,k}(\cdot)$ is a procedure that extracts the nondominated paths with respect to the k objectives and L_i is a list that stores the nondominated paths at each node i.

Input: An acyclic graph $G = (V, A)$ with $n = |V|$ nodes, $V = \{v_1, \dots, v_n\}$, $p = |A|$ arcs and k costs at each arc

Initialization: set $L_1 = \{(0, \dots, 0)\}$ and $L_i = \{\}$, for $i = 2, \dots, n$
for $i = 1, \dots, n-1$ **do**
    for all v_j, $(v_i, v_j) \in A$ **do**
        $L_j = ND_{1,\dots,k}\left(L_j \cup \{\ell + c_{i,j} \mid \ell \in L_i\}\right)$
Return L_n

In the following, we present the pseudo-code of an algorithm that solves the mutiobjective shortest path problem with $k-1$ objectives and a resource constraint, which corresponds to one step of the ϵ-constraint for this problem. This algorithm is similar to the previous one, except that only nondominated paths with respect to the $k-1$ objectives and the cost, and that satisfy the resource constraint, are stored on each list L_i, $i = 1, \dots, n$. Note that is easy to adapt it to handle more resource constraints.

Input: An acyclic graph $G = (V, A)$ with $n = |V|$ nodes, $V = \{v_1, \dots, v_n\}$, $p = |A|$ arcs, k costs at each arc and a resource constraint ϵ on the k-th cost

Initialization: set $L_1 = \{(0, \dots, 0)\}$ and $L_i = \{\}$, for $i = 2, \dots, n$
for $i = 1, \dots, n-1$ **do**
    for all v_j, $(v_i, v_j) \in A$ **do**
        $L_j = ND_{1,\dots,k}\left(L_j \cup \{\ell + c_{i,j} \mid \ell^k + c^k_{i,j} \le \epsilon, \ell \in L_i\}\right)$
Return L_n

We performed an experimental analysis of these algorithms on several randomly generated graphs for size 50 with 3 and 4 objectives. Table 4 presents the running time in seconds for each combination on the number of objectives and number of resource constraints. The results clearly indicate that the CPU-time increases quite strongly with the increase on the number of constraints.

4.7.5 Summary and Open Questions

Our claim is that, despite solving a sequence of supposedly easier problems with fewer objectives, $\varepsilon ALG(k, m)$ is (typically) inferior to $ALG(k, m)$ under reasonable assumptions. However, more precise analysis of run time necessary to compare the two algorithmic approaches.

Furthermore, if a specific algorithm is available for the many-objective problem, it may be even more efficient to solve it than equally structured fewer objective problem. Also, the many-objective problem may yield more information potentially relevant for the decision-maker.

References

1 Ahuja, Ravindra K., Magnanti, Thomas, Orlin, James B.: Network Flows: Theory, Algorithms and Applications. Prentice-Hall (1993)
2 Dempe, Stephan, Eichfelder, Gabriele, Fliege, Joerg: On the effects of combining objectives in multi-objective optimization. Mathematical Methods of Operations Research **82**(1), 1–18 (2015)
3 Ehrgott, Matthias: Multicriteria Optimization. Springer, Berlin, Germany, 2nd edn. (2005). 10.1007/3-540-27659-9
4 Ehrgott, Matthias: A discussion of scalarization techniques for multiple objective integer programming. Annals of Operations Research **147**(1), 343–360 (2006)
5 Figueira, Jose R., Fonseca, Carlos M., Halffmann, Pascal, Klamroth, Kathrin, Paquete, Luis, Ruzika, Stefan, Schulze, Britta, Stiglmayr, Michael, Willems, David: Easy to say they are hard, but hard to see they are easy—towards a categorization of tractable multiobjective combinatorial optimization problems. Journal of Multi-Criteria Decision Analysis **24**(1-2), 82–98 (2017). 10.1002/mcda.1574
6 Fliege, Joerg: The effects of adding objectives to an optimisation problem on the solution set. Operations Research Letters **35**(6), 782–790 (2007)
7 Haimes, Yacov, Lasdon, Leon, Da Wismer, Da: On a bicriterion formation of the problems of integrated system identification and system optimization. IEEE Transactions on Systems, Man, and Cybernetics **1**(3), 296–297 (1971). 10.1109/TSMC.1971.4308298

4.8 Data and Preference Driven Objective Space Reduction in Multiobjective Optimization

Serpil Sayin (Koc University – Istanbul, TR), Mickaël Binois (INRIA – Valbonne, FR), and Margaret M. Wiecek (Clemson University, US)

The complexity of multiobjective optimization problems (MOPs) is addressed from the perspective of a large number of objective functions. Two approaches to reduce the number of objectives are proposed. In contrast to prior studies that have concentrated on mathematical features of the MOP allowing for decrease of the number of objectives, the developed approaches exploit the real-life context in which the MOP is being solved. The first approach relies on the classical Principal Component Analysis and is based on the data that is carried by an instance of the MOP, while the other is the Active Subspace Approach, additionally making use of the decision maker's desirability function. Numerical examples are included.

4.8.1 Introduction

This paper contains a report of the work performed by the Working Group "Objective reduction for many-objective problems" at the Dagstuhl Seminar 20031 "Scalability in Multiobjective Optimization" that took place in Schloss Dagstuhl – Leibniz Center for Informatics – on January 12-17, 2020.

Solving multiobjective optimization problems (MOPs) has been well studied and is known to be difficult for some classes of problems (see [6] and many others). The computational difficulty grows with an increase in the number of objective functions. However, in the presence of many objectives, not all functions may be of interest to the decision maker (DM) or not all objectives may be in conflict with each other. It is of interest to make the original MOP simpler by removing unnecessary objective functions, or by building combinations of objectives, while the solution set remains unchanged.

The concept of redundant (or, also called later, *nonessential*) objective functions is first introduced in [7]. An objective function is said to be *redundant* if the efficient set is unchanged when that function is removed. For multiobjective linear programs (MOLPs), sufficient conditions are given for determining the redundancy of an objective function in [7] and later extended to necessary and sufficient conditions in [13, 11]. Similar results for MOPs with quasiconvex objective functions under the assumptions of a compact and convex feasible set or injective objective functions are developed in [12]. The weakness of this approach is that only one objective at a time may be tested for redundancy and that this testing may have a high computational cost if there is a large or infinite number of objectives. The concept of representative objective functions for MOLPs is introduced in [21] and algebraic properties of objectives are used to reduce their number but maintain the same or an equivalent efficient set. A collection of objectives is called *representative* provided all objectives not in the collection can be represented as a conical combination of the criteria in the collection. It is shown that the efficient set to the original MOLP is equal to the efficient set of the problem reduced to a representative family of objectives.

Another way of reducing a large number of objective functions is to construct subproblems with a smaller number of criteria. In [14] it is shown that the weakly efficient set of the convex MOP with n variables is equal to the union of the efficient sets of subproblems obtained from the original MOP by selecting at most $n + 1$ criteria [14]. In this context, the MOP is said to be *Pareto reducible* if its weakly efficient solutions are the efficient solutions for this MOP itself and also for a subproblem obtained from it by selecting certain objectives [17, 18].

The aim of this paper is to further study MOPs with a large number of objectives, possibly non-linear, and develop other rational and practical possibilities for reducing the number of objectives. In contrast to prior studies as above that have concentrated on mathematical features of the MOP allowing for decrease of the number of objectives, we propose two reduction approaches exploiting the real-life context in which the MOP is being solved. The first approach is Principal Component Analysis (PCA), which is based on the data that is carried by an instance of the MOP. PCA has been incorporated into Evolutionary Multiobjective Optimization (EMO) methods that solve problems with *many* objectives. In [5], PCA is incorporated into the EMO algorithm NSGA-II to reduce the number of objectives throughout the solution process as populations evolve. This has been reported to improve performance of EMO algorithms. Other PCA variations, including kernel PCA, have been implemented in different settings such as in [19]. In addition, PCA has been used to improve computational performance of the hypervolume based approaches [2] or to provide visualizations as in [4]. The analysis presented in this report differs from those studies mainly in its focus on identifying a mapping that would support the user's understanding of the set of Pareto solutions rather than solely using those methods as tools to improve computational efficiency of existing algorithms. The second approach relies on the cost (utility) function used by the DM who is involved in a decision-making process of choosing the most preferred solution from among all Pareto solutions of the MOP. This approach is referred to as the Active Subspace Approach (ASA) since it builds on the active subspace methodology for dimension reduction in parameter studies [3].

We envision an environment where some efficient solutions and thus Pareto points have been obtained by utilizing an optimization algorithm. The available Pareto points are given as input data that is used to explore the relationships among the objective functions using the two techniques, PCA and ASA. If the techniques identify a reduction of objectives, we anticipate that the reduced MOP would be solved instead of the original more complex one.

In this preliminary work, we study the two reduction approaches in more detail to gain insight into how they might contribute to the effort of MOP complexity reduction in terms of reducing the number of objective functions. The report is structured as follows. We provide a mathematical definition of the problem under study and introduce two problem instances of an MOP we experiment with. In Sections 4.8.2 and 4.8.3 we present the PCA and ASA approaches respectively, while in Section 4.8.4 we compare the obtained results and discuss future research directions.

Problem Statement

The MOP of interest is defined as

$$\begin{array}{lll} (\mathbf{MOP}) & \min & f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_p(\mathbf{x})) \\ & \text{s.t.} & \mathbf{x} \in \mathcal{X} \end{array}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ denotes the feasible set and $f_i : \mathbb{R}^n \to \mathbb{R},\ i = 1, \dots, p$ are the objective functions.

We assume the MOP is difficult to solve due to the large number of objectives it contains. Therefore a goal could be to eliminate some objectives from the MOP and be left with the same MOP but with a smaller number of objectives, which would call for identifying redundant objectives, if the MOP contains any. Alternatively, a goal could be to build another MOP, which is equivalent to the original one in some sense, but having a smaller number of objectives that are not all a subset of the original ones but perhaps a combination of them.

Examples

We demonstrate the ideas and work on two instances of the multiobjective knapsack problem (MOKP). In the MOKP, there are n objects each of which has a positive integer weight w_r and p non-negative integer profits v_r. The decision variable x_r denotes whether object r is selected for the knapsack or not. The total weight of selected objects should be within an integer capacity $W > 0$.

$$(\mathbf{MOKP}) \quad \max \quad \sum_{r=1}^{n} v_r^j x_r \quad j = \{1, \dots, p\} \tag{5}$$

$$\text{s.t.} \quad \sum_{r=1}^{n} w_r x_r \leq W \tag{6}$$

$$x_r \in \{0, 1\} \quad r = 1, \dots, n \tag{7}$$

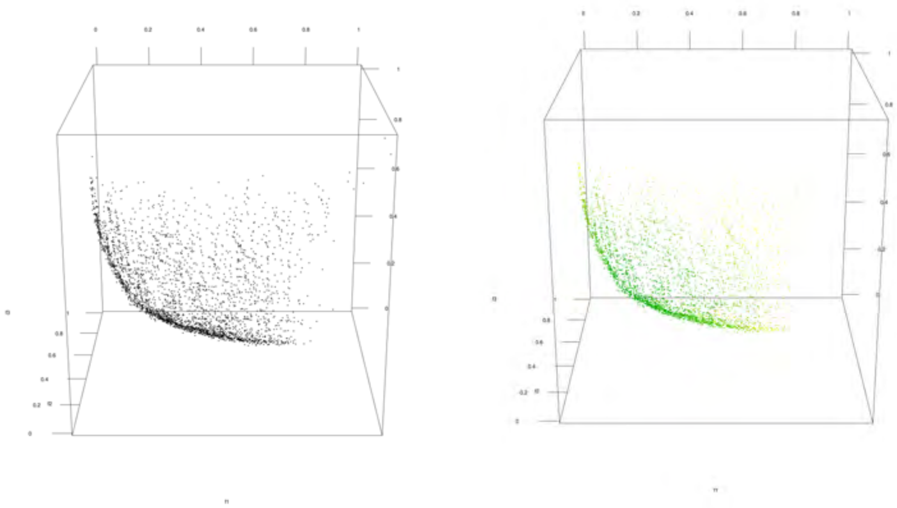


Figure 24 Left: Pareto points for the 3D case. Right: Pareto points for the 3D case with desirability values in color scale (yellow is low, green is high).

The vector of p objective functions where each objective function denotes the total profit of chosen objects is given in (5). Inequality (6) models the capacity constraint meaning that the total weight of selected objects has to be less than or equal to the knapsack's capacity. The binary constraints complete the model in (7) .

Two instances of the MOKP used in the computational experiments in [10] are selected. The first instance has $p = 3,\ n = 100$ and the second instance has $p = 5,\ n = 20$. These are instances $KP\_p - 3\_n - 100\_ins - 10$ and $KP\_p - 5\_n - 20\_ins - 10$ in [9] and from here on they are referred to as the 3D case and the 5D case, respectively. The Pareto set of the 3D case contains about 3200 points and of the 5D case about 400 points.

The PCA-based approach does not require any additional information from the DM and relies on problem data. The ASA can be used when some valuations emulating decision maker's preferences are available. In the two cases, this preference/desirability function is taken as $F : \mathbf{x} \in [0, 1]^p \rightarrow \sum_{i=1}^{p} (x_i - -1)^2$, operating on normalized data. The Pareto points of the 3D case with and without desirability information are depicted in Figure 24.

4.8.2 Principal Component Analysis

In this section the PCA is reviewed, and the resulting approach is presented and applied to the two cases of the MOKP.

Data Driven Reduction of the Objective Space

PCA [16] is a statistical procedure that is designed to reduce the number of *attributes* in a data set without losing its essence. A reduction is achieved by building new *components* that are linear combinations of the original attributes. Therefore the main idea is to build a transformation of the original data set that can be used for exploratory or predictive analysis. In the Pearson sense, this transformation is built by seeking the sum of least squares that optimizes the projection of the original data set onto a lower dimensional subspace [20]. In the Hotelling version, components are built iteratively in a way to carry the maximum

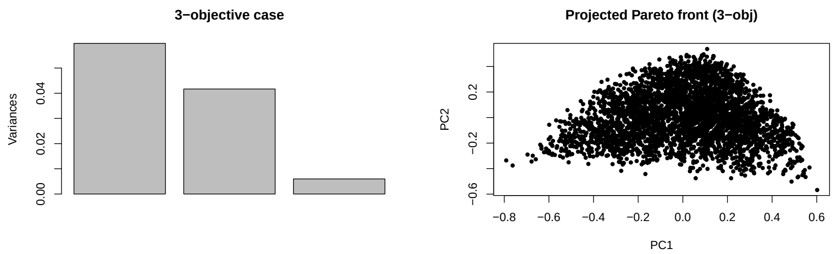


Figure 25 PCA results for the 3D case. Left: variances. Right: projection of the Pareto set on the first two principal components.

variance and remain orthonormal to the previously built ones [1]. There are various methods with different computational aspects to implement PCA. In general, PCA requires computing the covariance matrix and then finding its eigenvalues and eigenvectors. The magnitude of the eigenvalues are indicators of the importance, and the associated eigenvectors indicate the components. Truncation is generally performed, relying on an arbitrary threshold on the variance explained or on probabilistic interpretations, see, e.g., [15]. PCA is used extensively in engineering and machine learning for dimensionality reduction.

In Algorithm 1 we give a pseudo code of the procedure we apply to the test cases.

Algorithm 1 Pseudo-code for the PCA-based procedure.

Require: A set of uniformly spread Pareto points $\mathbf{y}^1, \ldots, \mathbf{y}^K$ in the objective space, represented by the $K \times p$ matrix Y.
1: Normalize Y. Compute the covariance matrix of Y, denoted as C.
2: Obtain the eigen-decomposition of C.
3: Sort the eigenvectors in the decreasing order of eigenvalues and determine the number of principal components, k, that suffice to represent Y.
4: Map Y using the first k eigenvectors.

The 3D Example

Algorithm 1 applied to the 3D case produces the matrix of eigenvectors given below, where every column is referred to as a principal component.

$$\begin{bmatrix} PC1 & PC2 & PC3 \\ -0.199 & -0.891 & -0.408 \\ 0.710 & 0.156 & -0.687 \\ -0.675 & 0.426 & -0.602 \end{bmatrix}$$

The results given in Figure 25 show that the first two principal components (objectives) carry most of the variance, and that the projection on the first two components keeps most of the shape of the 3D Pareto set. This result suggests that the 3D problem can be reformulated as an MOP with two objective functions using the first two component vectors as weights that combine the three objective functions with minimal loss of information.

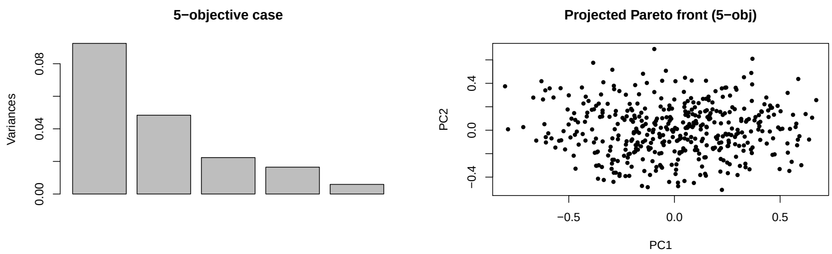


Figure 26 PCA results for the 5D case. Left: variances. Right: projection of the Pareto set on the first two principal components.

The 5D Example

When we apply Algorithm 1 to the 5D example, we obtain the following matrix of five principal components:

$$\begin{bmatrix} PC1 & PC2 & PC3 & PC4 & PC5 \\ -0.140 & 0.753 & -0.077 & 0.547 & -0.330 \\ 0.498 & -0.279 & 0.315 & 0.102 & -0.7511 \\ -0.554 & 0.082 & 0.800 & -0.198 & -0.089 \\ 0.413 & 0.582 & 0.052 & -0.698 & -0.014 \\ -0.505 & -0.098 & -0.503 & -0.406 & -0.563 \end{bmatrix}$$

The results depicted in Figure 26 indicate that the fifth component carries very little variance and therefore can be dropped from consideration in the transformed MOP. In the same figure we also observe the projection of the points onto two-dimensional subspace determined by the first two principal components. We note that it may be possible to drop the fourth and even the third principal component and still not lose the essence of the Pareto set of the original problem. However, the implications of all these reformulations must be studied and understood carefully.

4.8.3 Active Subspace Approach

In this section the ASA is reviewed, and the resulting algorithm is presented and applied to the two cases of the MOKP.

Preference Driven Reduction of the Objective Space

When the decision maker is able to provide a desirability or cost function, which can be only approximate or estimated by another method, then this additional information can be taken into account and used in the ASA. We study the transposition of the active subspace methodology, see [3], to the context of interest in this report.

In parameter studies with a function $F : \mathbb{R}^p \to \mathbb{R}$, ASA is applied to reduce the dimension on the parameter space $\mathbb{R}^p$ to avoid the curse of dimensionality. The principle is to estimate the $p \times p$ matrix $\mathbf{C} = \int_{\mathbf{y} \in D \subset \mathbb{R}^p} \nabla F(\mathbf{y}) \nabla F(\mathbf{y})^\top \mu(d\mathbf{y})$, where D is the domain of F and μ is an appropriate measure on D, usually uniform for bounded domains and Gaussian for unbounded ones. The eigen-decomposition of $\mathbf{C}$ is then computed in the form $\mathbf{C} = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^\top$, where $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues with its columns sorted in the decreasing order of the eigenvalues, and $\mathbf{W}$ is the matrix of eigenvectors.

The first eigenvector is then the direction along which most of the variations of F occur on average. The last eigenvector determines the direction along which least variations of F

occur on average. The active subspaces are identified by gaps in the eigenvalues: the larger the gap between two subsequent eigenvalues, the more important is the subspace defined by the eigenvectors preceding the gap. The eigenvectors define a rotation of the original space $\mathbb{R}^p$ and consequently the domain of F. The $\mathbb{R}^p$ space is reduced by dropping the directions associated with the smallest eigenvalues [3, 8]. A multi-objective version of the method is proposed by [22] and could be further explored.

In the many objective context we consider, the Pareto points in $\mathbb{R}^p$ are the designs, the desirability function is F, and a natural choice for μ is the uniform measure on the Pareto set. In practice, the estimation of $\mathbf{C}$ is detailed in Algorithm 2.

Algorithm 2 Pseudo-code for the ASA procedure.

Require: A set of uniformly spread Pareto points $\mathbf{y}^1, \ldots, \mathbf{y}^K$ in the objective space, and a cost/desirability function $F : \mathbb{R}^p \to \mathbb{R}$

1: Compute the gradient of F at each point of the sample: $\nabla F(\mathbf{y}^1), \ldots, \nabla F(\mathbf{y}^K)$

2: Compute $\hat{\mathbf{C}} = \frac{1}{K} \sum_{i=1}^{K} \nabla F(\mathbf{y}^i) \nabla F(\mathbf{y}^i)^\top$

3: Obtain the eigen-decomposition of $\hat{\mathbf{C}}$

4: Look for gaps in the eigenvalues of $\hat{\mathbf{C}}$

We identify two potential applications of the ASA approach. First, when the desirability function is not known precisely, and second, to drive an optimization algorithm. These circumstances are further discussed in the following examples.

The 3D Example

For comparison with the components produced by PCA, we give $\hat{\mathbf{C}}$, the matrix of eigenvectors computed by Algorithm 2.

$$\begin{bmatrix} AS1 & AS2 & AS3 \\ 0.576 & 0.132 & 0.806 \\ 0.588 & -0.752 & -0.297 \\ 0.567 & 0.645 & -0.511 \end{bmatrix}$$

Similar to the PCA example, in the top of Figure 27 we give the weights of each component in the form of the (log) eigenvalues (left) and the projection of the Pareto points along the first two directions (right). In this case the first eigenvalue is much larger than the other two, hence a gap is formed. The projection on the first two components also resembles the shape of the Pareto set of the original problem. With the additional information carried in F, with ASA it is also possible to show the variability of F with respect to the active directions. In Figure 27 bottom left we observe that the F values almost linearly increase along the first eigenvector. On the contrary, as depicted in Figure 27 bottom right, representing the F values along the last eigenvector is much less informative, since they vary significantly when fixing the abscissa. In any case, this analysis provides information on the directions along which the DM's desirability varies less, which may be helpful when DM explores solutions of equivalent desirability when the desirability is not well known.

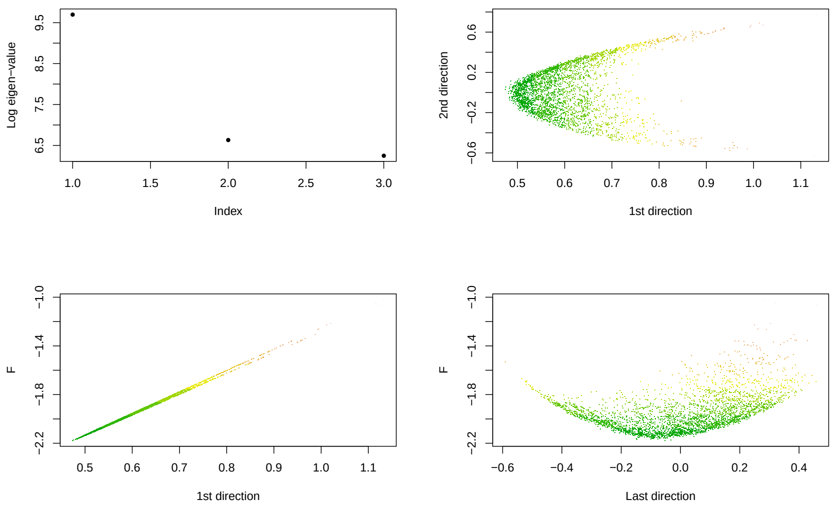


Figure 27 ASA results for the 3D case. Top left: log eigenvalues. Top right: projection of the Pareto set on the first two active directions. Bottom left: values of F along the first active direction. Bottom right: values of F along the second active direction. The color scale is the same as in Figure 24 (right).

The 5D Example

Again, for comparison with the components produced by PCA, we give $\hat{\mathbf{C}}$, the matrix of eigenvectors computed by Algorithm 2.

$$\begin{bmatrix} AS1 & AS2 & AS3 & AS4 & AS5 \\ -0.423 & 0.070 & 0.678 & -0.079 & 0.591 \\ -0.485 & -0.540 & -0.576 & -0.093 & 0.365 \\ -0.469 & 0.499 & -0.159 & -0.644 & -0.300 \\ -0.423 & -0.485 & 0.372 & 0.145 & -0.653 \\ -0.432 & 0.469 & -0.211 & 0.741 & -0.024 \end{bmatrix}$$

From the eigenvalues given in Figure 28 top left, there is a gap between the first and second eigenvalues, indicating that the first eigenvector is mostly sufficient to represent the F values, as shown in the bottom left panel in this figure. The projection on the last eigenvector is again not informative, since the F values vary greatly for the same abscissa. This is probably related to the simple structure of the F function used in this example, as its values depend on the radius of the sphere, which is recovered by ASA. Also the projection on the first two eigenvectors (in Figure 28 top right) is this time different from the one obtained in Figure 26 top right, which has a V shape. The reasons for this difference and the implications in terms of visualisation need further research.

4.8.4 Conclusion

Both PCA and ASA are descriptive tools giving insight into the problem at hand. Because PCA uses the information embedded in the data of the problem, while ASA additionally uses information external to the problem, the two methods provide different results.

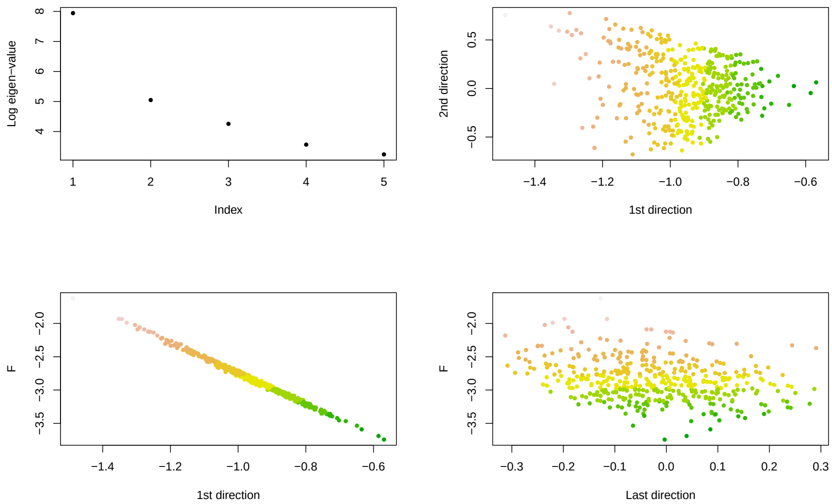


Figure 28 ASA results for the 5D case. Top left: log eigenvalues. Top right: projection of the Pareto set on the first two active directions. Bottom left: values of F along the first active direction. Bottom right: values of F along the second active direction. The color scale is the same as in Figure 24 (right).

Based on this preliminary study, we pose the following tasks for further research:

- Investigate the sensitivity of the PCA approach with respect to the number and/or distribution of the initial Pareto points supplied;
- Investigate the sensitivity of ASA with respect to the number and/or distribution of the initial Pareto points supplied **and** their costs/desirability;
- Establish the relationship between the Pareto set of the reduced problem and the Pareto set of the original problem using numerical instances;
- Study the effect of changing the Pareto cone into another polyhedral cone;
- Improve visualization of the projection of the basis directions of the original space as it is done for PCA with biplots, see e.g., [4].

We believe these investigations would be helpful in defining *reducibility* of MOPs and assessing the applicability of the two approaches presented in this report.

References

1 Rasmus Bro and Age K Smilde. Principal component analysis. *Analytical Methods*, 6(9):2812–2831, 2014.
2 Dimo Brockhoff and Eckart Zitzler. Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods. In *2007 IEEE congress on evolutionary computation*, pages 2086–2093. IEEE, 2007.
3 Paul G Constantine. *Active subspaces: Emerging ideas for dimension reduction in parameter studies*, volume 2. SIAM, 2015.
4 Lino Costa and Pedro Oliveira. Biplots in offline multiobjective reduction. In *IEEE Congress on Evolutionary Computation*, pages 1–8. IEEE, 2010.
5 Kalyanmoy Deb and D Saxena. Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. In *Proceedings of the World Congress on Computational Intelligence (WCCI-2006)*, pages 3352–3360, 2006.

6 Matthias Ehrgott. *Multicriteria Optimization*. Springer, New York, 2005.

7 Tomas Gal and Heiner Leberling. Redundant objective functions in linear vector maximum problems and their determination. *European Journal of Operational Research*, 1(3):176–184, 5 1977.

8 John T Holodnak, Ilse CF Ipsen, and Ralph C Smith. A probabilistic subspace bound with application to active subspaces. *SIAM Journal on Matrix Analysis and Applications*, 39(3):1208–1220, 2018.

9 Gokhan Kirlik. Test instances for multiobjective discrete optimization problems, 2014.

10 Gokhan Kirlik and Serpil Sayın. A new algorithm for generating all nondominated solutions of multiobjective discrete optimization problems. *European Journal of Operational Research*, 232(3):479–488, 2014.

11 Agnieszka B Malinowska and Delfim FM Torres. Computational approach to essential and nonessential objective functions in linear multicriteria optimization. *Journal of Optimization Theory and Applications*, 139(3):577–590, 2008.

12 Agnieszka Barbara Malinowska. Changes of the set of efficient solutions by extending the number of objectives and its evaluation. *Control and Cybernetics*, 31(4):965–974, 2002.

13 Agnieszka Barbara Malinowska. Nonessential objective functions in linear multiobjective optimization problems. *Control and Cybernetics*, 35(4):873–880, 2006.

14 C Malivert and N Boissard. Structure of efficient sets for strictly quasi convex objectives. *Journal of Convex Analysis*, 1(2):143–150, 1994.

15 Thomas P Minka. Automatic choice of dimensionality for PCA. In *Advances in Neural Information Processing Systems*, pages 598–604, 2001.

16 Karl Pearson. Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2):559, 1901.

17 Nicolae Popovici. Pareto reducible multicriteria optimization problems. *Optimization*, 54(3):253–263, 2005.

18 Nicolae Popovici. Involving the helly number in Pareto reducibility. *Operations Research Letters*, 36(2):173–176, 2008.

19 Dhish Kumar Saxena and Kalyanmoy Deb. Non-linear dimensionality reduction procedures for certain large-dimensional multi-objective optimization problems: Employing correntropy and a novel maximum variance unfolding. In *International Conference on Evolutionary Multi-Criterion Optimization*, pages 772–787. Springer, 2007.

20 Jos MF Ten Berge and Henk AL Kiers. Optimality criteria for principal component analysis and generalizations. *British Journal of Mathematical and Statistical Psychology*, 49(2):335–345, 1996.

21 Nguyen V Thoai. Criteria and dimension reduction of linear multiple criteria optimization problems. *Journal of Global Optimization*, 52(3):499–508, 2012.

22 Olivier Zahm, Paul Constantine, Clementine Prieur, and Youssef Marzouk. Gradient-based dimension reduction of multivariate vector-valued functions. *arXiv preprint arXiv:1801.07922*, 2018.

5 Seminar schedule

Monday, January 13, 2020

09:00 – 10:30: Welcome Session

- Welcome and Introduction
- Short presentation of all participants (2 minutes each!)

Coffee Break

11:00 – 12:00: Challenges in Models

- Dimo Brockhoff, Michael Emmerich, Boris Naujoks & Robin Purshouse: MACODA – A Lorentz Center Workshop on "MAny Criteria Optimization and Decision Analysis"
- Arnaud Liefooghe: On the Difficulty of Multiobjective Combinatorial Optimization Problems

Lunch

14:00 – 15:00: Challenges in Methodology

- Mickael Binois: Scaling up Multi-Objective Bayesian Optimization
- Sanaz Mostaghim: Recent Advances in Multi-Objective Large Scale Optimisation

15:00 – 15:30: Group Discussion

Coffee Break

16:00 – 16:30: Working Group Formation

16:30 – 18:00: Working Groups

Dinner

19:30: Opening of the art exhibit "Das Loch das von der anderen Seite kam" by the German artist Lola Sprenger

Tuesday, January 14, 2020

09:00 – 09:30: Many Objectives in Stochastic Settings Chair: Heike Trautmann

- Susan Hunter: Multi-Objective Simulation Optimization: Theory and Practice

09:30 – 10:30: Reporting from Working Groups and Splitting into Smaller Groups

Coffee Break

11:00 – 12:00: Many Objectives in Stochastic Settings Chair: Christiane Tammer

- Anita Schoebel: Robust Multiobjective Optimization Problems – An Approach with Very Many Objective Functions
- Gabriele Eichfelder: A Multiobjective Trust Region Method for Expensive and Cheap Functions

Lunch

14:00 – 15:30: Small Working Groups

Coffee Break

16:00 – 17:00: Small Working Groups

17:00 – 18:00: Reporting from Small Working Groups
- General discussion and working group adaptations

Wednesday, January 15, 2020

09:00 – 09:30: Problems with Many Variables Chair: Patrick M. Reed
- Georges Fadel: Multi-Objective Topology Design of Functionally Graded Components

09:30 – 10:30: Small Working Groups

Coffee Break

11:00 – 12:00: Small Working Groups

Lunch

14:00: Group Foto (Outside)

14:05 – 15:30: Hiking Trip

Coffee Break

16:00 – 17:00: Small Working Groups

17:00 – 18:00: Participant Announcements

Thursday, January 16, 2020

09:00 – 09:30: Future Directions Chair: Juergen Branke
- Fritz Boekler: Complexity in Multiobjective Optimization
- Dimo Brockhoff: On Set-Indicator-Based Search: Using Single-Objective Solvers for Multiobjective Problems
- Pascal Kerschke: Chances and Challenges of Multimodality in Multi-Objective Continuous Optimization Problems

10:30 – 11:00: Coffee Break

11:00 – 12:00: State of Play

Lunch

Time for Individual Discussions (e.g. Offers-and-Needs Market)

15:30 – 16:00: Coffee Break

16:00 – 18:00: Small Working Groups

Dinner

20:00: Wine-and-Cheese Party (Music room)

Friday, January 17, 2020

09:00 – 10:30: Final Reporting from Working Groups

10:30 – 11:00: Coffee Break

11:00 – 12:00: Closing Session

Lunch

6 Topics of interest for participants for next Dagstuhl seminar

It has evolved as a tradition to jointly discuss future challenges and topics of particular interest for the EMO and MCDM community during the closing session on Friday. During this discussion the participants identified the following prevalent topics: Machine learning and data science (ML & DS) for multiobjective optimization and multiobjective optimization for ML & DS, neuro sciences, dynamic and adaptive systems, mixed models, ethical issues, social choice theory, and communicating multiobjective optimization. The organizers will use these suggestions as the basis for their discussion about possible topics for the next edition of this seminar series and for the preparation of a proposal for a continuation of the series.

7 Changes in the seminar organization body

7.1 Kathrin Klamroth and Günter Rudolph step down as co-organizers

As part of a continuing effort to renew the organizing board of this series of Dagstuhl seminars, Kathrin Klamroth and Günter Rudolph step down from the team of organizers, a role that they have held for three terms of office.

On behalf of all the participants of the seminar, Carlos Fonseca and Margaret Wiecek would like to express appreciation to Kathrin and Günter for their contributions and leadership that have been fundamental for the series success.

7.2 Welcome to Richard Allmendinger and Serpil Sayin

We are pleased to announce that our esteemed colleagues, Richard Allmendinger and Serpil Sayin, have agreed to serve as co-organizers for future editions of this Dagstuhl seminar series on Multiobjective Optimization. We look forward to collaborating with them in the near future.

Participants

- Richard Allmendinger
University of Manchester, GB
- Mickaël Binois
INRIA – Valbonne, FR
- Fritz Bökler
Universität Osnabrück, DE
- Jürgen Branke
University of Warwick, GB
- Dimo Brockhoff
INRIA Saclay – Palaiseau, FR
- Carlos A. Coello Coello
CINVESTAV – Mexico, MX
- Kerstin Dächert
Fraunhofer ITWM – Kaiserslautern, DE
- Matthias Ehrgott
Lancaster University, GB
- Gabriele Eichfelder
TU Ilmenau, DE
- Michael Emmerich
Leiden University, NL
- Georges Fadel
Clemson University – Clemson, US
- José Rui Figueira
IST – Lisbon, PT
- Carlos M. Fonseca
University of Coimbra, PT
- Andreia P. Guerreiro
IST – Lisbon, PT
- Jussi Hakanen
University of Jyväskylä, FI
- Susan R. Hunter
Purdue University, US
- Hisao Ishibuchi
Southern Univ. of Science and Technology – Shenzen, CN
- Andrzej Jaszkiewicz
Poznan University of Technology, PL
- Pascal Kerschke
Universität Münster, DE
- Kathrin Klamroth
Universität Wuppertal, DE
- Karl Heinz Küfer
Fraunhofer ITWM – Kaiserslautern, DE
- Arnaud Liefooghe
University of Lille, FR
- Manuel López-Ibánez
University of Manchester, GB
- Kaisa Miettinen
University of Jyväskylä, FI
- Sanaz Mostaghim
Universität Magdeburg, DE
- Boris Naujoks
TH Köln, DE
- Luís Paquete
University of Coimbra, PT
- Patrick M. Reed
Cornell University – Ithaca, US
- Enrico Rigoni
ESTECO SpA – Trieste, IT
- Günter Rudolph
TU Dortmund, DE
- Stefan Ruzika
TU Kaiserslautern, DE
- Serpil Sayin
Koc University – Istanbul, TR
- Anita Schöbel
Fraunhofer ITWM – Kaiserslautern, DE
- Britta Schulze
Universität Wuppertal, DE
- Pradyumn Kumar Shukla
KIT – Karlsruher Institut für Technologie, DE
- Ralph E. Steuer
University of Georgia, US
- Michael Stiglmayr
Universität Wuppertal, DE
- Christiane Tammer
Martin-Luther-Universität Halle-Wittenberg, DE
- Heike Trautmann
Universität Münster, DE
- Tea Tusar
Jozef Stefan Institute – Ljubljana, SI
- Daniel Vanderpooten
University Paris-Dauphine, FR
- Margaret M. Wiecek
Clemson University, US

Report from Dagstuhl Seminar 20041

Symmetric Cryptography

Edited by

Nils Gregor Leander[1], Bart Mennink[2], Kaisa Nyberg[3], and Kan Yasuda[4]

1 **Ruhr-Universität Bochum, DE,** gregor.leander@rub.de
2 **Radboud University Nijmegen, NL,** b.mennink@cs.ru.nl
3 **Aalto University, FI,** kaisa.nyberg@aalto.fi
4 **NTT – Tokyo, JP,** kan.yasuda.hy@hco.ntt.co.jp

Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 20041 "Symmetric Cryptography". The seminar was held on January 19-24, 2020 in Schloss Dagstuhl – Leibniz Center for Informatics. This was the seventh seminar in the series "Symmetric Cryptography". Previous editions were held in 2007, 2009, 2012, 2014, 2016, and 2018.

Participants of the seminar presented their ongoing work and new results on topics of (quantum) cryptanalysis and provable security of symmetric cryptographic primitives. In this report, a brief summary of the seminar is given followed by the abstracts of given talks.

Seminar January 19–24, 2020 – http://www.dagstuhl.de/20041
2012 ACM Subject Classification Security and privacy → Cryptanalysis and other attacks, Security and privacy → Symmetric cryptography and hash functions
Keywords and phrases (quantum) cryptanalysis, constrained platforms, symmetric cryptography
Digital Object Identifier 10.4230/DagRep.10.1.130
Edited in cooperation with Aleksei Udovenko

1 Executive Summary

Nils Gregor Leander (Ruhr-Universität Bochum, DE)
Bart Mennink (Radboud University Nijmegen, NL)
Kaisa Nyberg (Aalto University, FI)
Kan Yasuda (NTT – Tokyo, JP)

IT Security plays a crucial role in everyday life and business. Virtually all modern security solutions are based on cryptographic primitives. *Symmetric* cryptography deals with the case that both the sender and the receiver of a message are using the same key and is highly relevant not only for academia, but also for industrial research and applications.

We identified the following areas as among the most important topics for future research.

Cryptography in the presence of strong constraints. This area deals with the development of symmetric cryptographic primitives and modes that must operate under strong constraints. The area, often indicated by the misleading term lightweight cryptography, has become a very active research field in recent years.

Symmetric Cryptography, *Dagstuhl Reports*, Vol. 10, Issue 1, pp. 130–143
Editors: Nils Gregor Leander, Bart Mennink, Kaisa Nyberg, and Kan Yasuda
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Proving relevant bounds for permutations and (tweakable) block ciphers. Security arguments for symmetric cryptographic primitives often rely on simplifying assumptions and unproven heuristics. Moreover, not only are they often limited by those simplifications, but more fundamentally by the resulting statements.

Development of modes for dedicated functionality or robustness. A cryptographic primitive, e.g., a cryptographic permutation or a (tweakable) block cipher, is of little use without being embedded in a suitable mode of operation. Traditional modes turn such a primitive into an (authenticated) encryption scheme, a message authentication code or a hash function. However, modes of operations could provide more advanced functionalities on the one hand and advanced security features on the other hand.

Quantum cryptanalysis. The threat that one would be able to build a sufficiently large quantum computer has a major impact on the security of many cryptographic schemes we are using today. In particular, the seminal work of Shor showed that such computers would allow to factor large integers and compute discrete logs over large groups in practical time. In the case of symmetric cryptography, the situation seems less critical – but is also significantly less studied. For almost 20 years, it was believed that the only advantage an attacker would have by using a quantum computer when attacking symmetric cryptography is due to Grover's algorithm for speeding up brute force search. Only recently researchers have started to investigate in more detail how the security of symmetric primitives would be affected by attackers equipped with quantum computers.

Seminar Program

The seminar program consisted of short presentations and group meetings. Presentations were about the above topics and other relevant areas of symmetric cryptography, including state-of-the-art cryptanalytic techniques and new designs. Below one can find the list of abstracts for talks given during the seminar. Also, participants met in smaller groups and spent a significant portion of the week, each group intensively discussing a specific research topic. There were eight research groups: 1) Design and analyze ciphers over prime fields, 2) Bounds on the degree of Feistel ciphers with round functions with low univariate degree, 3) Forkcipher, 4) Time-space tradeoffs, 5) Quantum cryptanalysis of hash functions, 6) NIST LWC, 7) Cryptanalysis of the Russian standards, and 8) Security of ProMACs. On the last day of the week the leaders of each group gave brief summaries of achievements. Some teams continued working on the topic after the seminar and started new research collaborations.

2 Table of Contents

3 Overview of Talks

3.1 A MAC Construction for Continuous Message Streams

Frederik Armknecht (Universität Mannheim, DE)

Joint work of Frederik Armknech, Paul Walther, Thorsten Strufe, Gene Tsudik, Martin Beck

Efficiently ensuring integrity of received data requires message authentication code (MAC) tags. The dominating factor determining their security is their length, measured in bits: Short tags are easy to guess, and improving security corresponds to expanding tags. High security constraints hence require sufficiently long tags, which in turn can entail prohibitive cost. This becomes particularly apparent in the context of increasingly common scenarios with typically small payload sizes but strict delay requirements, like robot- or drone control. It is of similar importance in scenarios that suffer from resource scarcity, like LoRaWAN networks with limited battery capacities, or memory protection in Intel SGX with a limitation on the number of costly, additional cells that can be used for integrity protection.

Prior techniques suggested truncation of tags, thus achieving linear performance gain at exponential loss of security. To guarantee security identical to full MAC schemes at the performance of truncated MACs, we suggest a new construction. It introduces internal state to facilitate gradually increasing security upon reception of subsequent messages. We define such schemes as Progressive MACs, provide a formal security framework, prove their security, and evaluate their applicability in several realistic scenarios.

3.2 Security of the STARK-friendly hash functions

Anne Canteaut (INRIA – Paris, FR)

Joint work of Tim Beyne, Anne Canteaut, Itai Dinur, Maria Eichlseder, Gregor Leander, Gaëtan Leurent, María Naya-Plasencia, Léo Perrin, Yu Sasaki, Yosuke Todo, Friedrich Wiemer
Main reference Tim Beyne, Anne Canteaut, Itai Dinur, Maria Eichlseder, Gregor Leander, Gaëtan Leurent, María Naya-Plasencia, Léo Perrin, Yu Sasaki, Yosuke Todo, Friedrich Wiemer: "Out of Oddity – New Cryptanalytic Techniques against Symmetric Primitives Optimized for Integrity Proof Systems", IACR Cryptol. ePrint Arch., Vol. 2020, p. 188, 2020.
URL https://eprint.iacr.org/2020/188

The security and performance of many integrity proof systems like SNARKs, STARKs and Bulletproofs highly depend on the underlying hash function. For this reason several new proposals have recently been developed. These primitives obviously require an in-depth security evaluation, especially since their implementation constraints have led to less standard design approaches. This work compares the security levels offered by three recent families of such primitives, namely GMiMC, Hades-MiMC and Vision/Rescue. We exhibit low-complexity distinguishers against the GMiMC and Hades-MiMC permutations for most parameters proposed in recently launched public challenges for STARK-friendly hash functions. To achieve those results, we adapt and generalize several cryptographic techniques to fields of odd characteristic.

3.3 Tight Time-Space Lower Bounds for Finding Multiple Collision Pairs (and Applications)

Itai Dinur (Ben Gurion University – Beer Sheva, IL)

Main reference Itai Dinur: "Tight Time-Space Lower Bounds for Finding Multiple Collision Pairs and Their Applications", in Proc. of the Advances in Cryptology – EUROCRYPT 2020 – 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 12105, pp. 405–434, Springer, 2020.
URL https://doi.org/10.1007/978-3-030-45721-1_15

We consider a *collision search problem* (CSP), where given a parameter C, the goal is to find C collision pairs in a random function $f : [N] \to [N]$ (where $[N] = \{0, 1, \ldots, N-1\}$) using S bits of memory. Algorithms for CSP have numerous cryptanalytic applications such as space-efficient attacks on double and triple encryption. The best known algorithm for CSP is *parallel collision search* (PCS) published by van Oorschot and Wiener, which achieves the time-space tradeoff $T^2 \cdot S = \tilde{O}(C^2 \cdot N)$.

In this talk, I will prove that any algorithm for CSP satisfies $T^2 \cdot S = \tilde{\Omega}(C^2 \cdot N)$, hence the best known time-space tradeoff is optimal. On the other hand, I give strong evidence that proving similar unconditional time-space tradeoff lower bounds on CSP applications (such as breaking double and triple encryption) may be very difficult, and would imply a breakthrough in complexity theory. Hence, I propose a new restricted model of computation and prove that under this model, the best known time-space tradeoff attack on double encryption is optimal.

3.4 Analyzing the Linear Keystream Biases in AEGIS

Maria Eichlseder (TU Graz, AT)

Joint work of Maria Eichlseder, Marcel Nageler, Robert Primas
Main reference Maria Eichlseder, Marcel Nageler, Robert Primas: "Analyzing the Linear Keystream Biases in AEGIS", IACR Cryptol. ePrint Arch., Vol. 2019, p. 1372, 2019.
URL https://eprint.iacr.org/2019/1372

AEGIS is one of the authenticated encryption designs selected for the final portfolio of the CAESAR competition [2, 3]. It combines the AES round function and simple Boolean operations to update its large state and extract a keystream to achieve an excellent software performance. In 2014, Minaud discovered slight biases in the keystream based on linear characteristics [1]. For family member AEGIS-256, these could be exploited to undermine the confidentiality faster than generic attacks, but this still requires very large amounts of data. For final portfolio member AEGIS-128, these attacks are currently less efficient than generic attacks.

We search for better linear characteristics, as well as upper bounds on the best possible correlation. We observe that straightforward truncated models of linear characteristics of AEGIS only produce very weak bounds since they fail to capture connections and constraints that follow from dependencies in the AEGIS state update function. We briefly discuss several examples of such linear incompatibilities from the related literature, where they have primarily been identified in the context of linear key or tweakey schedules. To obtain tighter bounds and consistent solutions, we identify additional constraints on the differences and higher-order differences of the linear masks and propose an improved truncated model. This model yields much better results, including consistent solutions for AEGIS-128, but still shows a significant gap between the bounds and the best found characteristics, mainly due to the Boolean output function. We propose a partially bitwise model to close this gap. As a result, for all AEGIS family members, we derive upper bounds below 2^{-128} for the squared correlation contribution of any single suitable linear characteristic. This supports AEGIS' security with realistic amounts of data. Finally, we apply Constraint Programming (CP) to find consistent characteristics and obtain improved attacks for all members.

References

1. Brice Minaud. Linear biases in AEGIS keystream. In Antoine Joux and Amr M. Youssef, editors, *Selected Areas in Cryptography – SAC 2014*, volume 8781 of *LNCS*, pages 290–305. Springer, 2014.
2. Hongjun Wu and Bart Preneel. AEGIS: A fast authenticated encryption algorithm. In Tanja Lange, Kristin E. Lauter, and Petr Lisonek, editors, *Selected Areas in Cryptography – SAC 2013*, volume 8282 of *LNCS*, pages 185–201. Springer, 2013.
3. Hongjun Wu and Bart Preneel. AEGIS: A fast authenticated encryption algorithm (v1.1). Submission to CAESAR: Competition for Authenticated Encryption. Security, Applicability, and Robustness (Round 3 and Final Portfolio), September 2016. http://competitions.cr.yp.to/round3/aegisv11.pdf.

3.5 Higher-Order Differential Attacks on Ciphers with Low-Degree Polynomial S-Boxes in $GF(2^n)$: Open Problems

Lorenzo Grassi (TU Graz, AT)

Joint work of Lorenzo Grassi, Carlos Cid, Maria Eichlseder, Reinhard Lüftenegger, Christian Rechberger, Markus Schofnegger, Qingju Wang

Higher-order differential attacks are among the most powerful attacks against low-degree ciphers and hash functions. Predicting the evolution of the degree of the cipher (as a function of the number of rounds) is the main issue in such attacks. Given an SPN cipher over a field $\mathbb{F}$, where each round has algebraic degree δ, it is a common belief that the degree grows essentially exponentially in δ. Several analyses made in the literature confirm this belief, with the only exception of the case in which the algebraic degree of the function is close to its maximum. As a result, the number of rounds necessary for security against higher-order differential attacks grows logarithmic in the size of F.

In this presentation, we show that surprisingly, if the round function/S-Box can be described as an invertible (low-degree) polynomial function in $\mathbb{F}_{2^n}$, then the algebraic degree grows linearly with the number of rounds, and not exponentially. In particular, we present several examples of this, including iterated Even-Mansour and SPN ciphers with (low-degree) polynomial round functions/S-Boxes.

3.6 Analysis on Adiantum

Tetsu Iwata (Nagoya University, JP)

Joint work of Habu Makoto, Tetsu Iwata

Adiantum is a disk sector encryption scheme designed by Google [1]. It can be seen as a tweakable, variable-input-length strong pseudorandom permutation, and has an indistinguishability security proof. In this talk, we first present a distinguishing attack with the birthday complexity. We then present plaintext recovery and forgery attacks, with almost the same complexity as the distinguishing attack. These results do not violate the security proof.

References

1 Paul Crowley and Eric Biggers. Adiantum: length-preserving encryption for entry-level processors. *IACR Transactions on Symmetric Cryptology*, 2018, Issue 4:39–61, 2018.

3.7 Some Thoughts on Boomerang Switches

Virginie Lallemand (LORIA – Nancy, FR)

Joint work of Hamid Boukerrou, Paul Huynh, Virginie Lallemand, Bimal Mandal, Marine Minier

Boomerang distinguishers were introduced at FSE 1999 by David Wagner. It is a variant of differential cryptanalysis that works on quartets of messages and studies if a difference "comes back". Namely, it looks at the probability that:

$$E^{-1}(E(M_1) + b) + E^{-1}(E(M_1 + a) + b) = a.$$

In practice, this type of distinguisher is built by splitting the cipher in three parts:

$$E = E_1 \circ E_m \circ E_0,$$

where E_m is a middle part that contains the boomerang switch. With such a framework, the probability of the distinguisher is evaluated to be: p^2q^2r where p is the probability of the differential used over E_0, q the one used over E_1 and r is the probability of the boomerang switch.

At Eurocrypt 2018, Cid et al. introduced the Boomerang Connectivity Table (BCT), a tool to easily compute the value of r for the case where the cipher E is a substitution-permutation network and where E_m covers one round.

In this talk, we introduce the FBCT, the counterpart of the BCT for the case where the cipher follows a Feistel construction. We show that the value of an FBCT coefficient is related to the second order derivative of the Sbox at play and study its properties.

3.8 The First Chosen-Prefix Collision on SHA-1

Gaëtan Leurent (INRIA – Paris, FR) and Thomas Peyrin

Main reference Gaëtan Leurent, Thomas Peyrin: "SHA-1 is a Shambles – First Chosen-Prefix Collision on SHA-1 and Application to the PGP Web of Trust", IACR Cryptol. ePrint Arch., Vol. 2020, p. 14, 2020.
URL https://eprint.iacr.org/2020/014

The SHA-1 hash function was designed in 1995 and has been widely used during two decades. A theoretical collision attack was first proposed in 2004 [3], but due to its high complexity it was only implemented in practice in 2017, using a large GPU cluster [2]. More recently, an almost practical *chosen-prefix* collision attack against SHA-1 has been proposed [1]. This more powerful attack allows to build colliding messages with two arbitrary prefixes, which is much more threatening for real protocols.

In this talk, we reported the first practical implementation of this attack, and its impact on real-world security with a PGP/GnuPG impersonation attack. We managed to significantly reduce the complexity of collisions attack against SHA-1: on an Nvidia GTX 970, identical-prefix collisions can now be computed with a complexity of $2^{61.2}$ rather than $2^{64.7}$, and chosen-prefix collisions with a complexity of $2^{63.4}$ rather than $2^{67.1}$. When renting cheap GPUs, this translates to a cost of 11k US$ for a collision, and 45k US$ for a chosen-prefix collision, within the means of academic researchers. Our actual attack required two months of computations using 900 Nvidia GTX 1060 GPUs (we paid 75k US$ because GPU prices were higher, and we wasted some time preparing the attack).

Therefore, the same attacks that have been practical on MD-5 since 2009 are now practical on SHA-1. In particular, chosen-prefix collisions can break signature schemes and handshake security in secure channel protocols (TLS, SSH). We strongly advise to remove SHA-1 from those type of applications as soon as possible.

We exemplify our cryptanalysis by creating a pair of PGP/GnuPG keys with different identities, but colliding SHA-1 certificates. A SHA-1 certification of the first key can therefore be transferred to the second key, leading to an impersonation attack. This proves that SHA-1 signatures now offers virtually no security in practice. The legacy branch of GnuPG still uses SHA-1 by default for identity certifications, but after notifying the authors, the modern branch now rejects SHA-1 signatures (the issue is tracked as CVE-2019-14855).

References

1 Gaëtan Leurent and Thomas Peyrin. From collisions to chosen-prefix collisions application to full SHA-1. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part III*, volume 11478 of *LNCS*, pages 527–555. Springer, Heidelberg, May 2019.
2 Marc Stevens, Elie Bursztein, Pierre Karpman, Ange Albertini, and Yarik Markov. The first collision for full SHA-1. In Jonathan Katz and Hovav Shacham, editors, *CRYPTO 2017, Part I*, volume 10401 of *LNCS*, pages 570–596. Springer, Heidelberg, August 2017.
3 Xiaoyun Wang, Yiqun Lisa Yin, and Hongbo Yu. Finding collisions in the full SHA-1. In Victor Shoup, editor, *CRYPTO 2005*, volume 3621 of *LNCS*, pages 17–36. Springer, Heidelberg, August 2005.

3.9 Conditional Cube Attack on Keccak Keyed Modes

Willi Meier (FH Nordwestschweiz – Windisch, CH)

Joint work of Zheng Li, Xiaoyang Dong, Wenquan Bi, Keting Jia, Xiaoyun Wang, Willi Meier
Main reference Zheng Li, Xiaoyang Dong, Wenquan Bi, Keting Jia, Xiaoyun Wang, Willi Meier: "New Conditional Cube Attack on Keccak Keyed Modes", IACR Trans. Symmetric Cryptol., Vol. 2019(2), pp. 94–124, 2019.
URL https://doi.org/10.13154/tosc.v2019.i2.94-124

The conditional cube attack on round-reduced Keccak keyed modes was proposed by Huang et al. at Eurocrypt 2017. A new conditional cube attack on Keccak is proposed by removing some limitations of previous attacks. As a result, the time complexity of key recovery attacks on 7-round Keccak-MAC-512 can be reduced from 2^{111} to 2^{72}, and similarly, the time complexity of key recovery on KMAC256 can be reduced from 2^{147} to 2^{139}.

3.10 Accelerating MRAE

Kazuhiko Minematsu (NEC – Kawasaki, JP)

Since nonce-based AE (NAE) schemes are generally fragile to a misuse of nonce, MRAE has received significant attention from the initial proposal by Rogaway and Shrimpton at Eurocrypt 2006. They showed a generic MRAE construction called SIV. SIV has become a de-facto scheme for MRAE, however, one notable drawback is its two-pass operation for both encryption and decryption. This implies that MRAE built on SIV is slower than the integrated nonce-based AE schemes, such as OCB.

In this talk, we propose a new method to improve this situation. Particularly, our MRAE proposal (decryption-integrated SIV or DI-SIV) allows to decrypt as fast as a plain decryption, hence theoretically doubles its speed from the original SIV, while keeping the encryption speed equivalent to SIV.

We show three generic compositions for DI-SIV, called DI-SIV1, DI-SIV2 and DI-SIV3, and prove their security bounds that are comparable to the original SIV. We also provide several concrete instantiations to show their effectiveness compare to the existing MRAE schemes, namely the same encryption speed but decryption is ideally fast.

3.11 Bits and Pieces

Orr Dunkelman (University of Haifa, IL)

Joint work of Orr Dunkelman, Nathan Keller, Abhishek Kumar, Eran Lambooij, Somitra Sandhya, Ariel Weizman

This talk presented a few ideas in the context of block cipher's cryptanalysis.

1. The partition of plaintext pairs according whether they satisfied the differential characteristic in the first round or not. This allows improving the probability of boomerang attacks

and the bias in differential-linear attacks, as in each partition the probability/biased is increased significantly. (joint work with Nathan Keller and Ariel Weizman)

2. We showed how to use multiple differential-linear approximations to recover the decorrelation module keys (joint work with Nathan Keller and Ariel Weizman)
3. We showed that counting the number of active S-boxes is not always a good measure for security estimation. We showed a 4-round Feistel cipher with a round function composed of many S-box/MDS layers, but with very high probability one could build a decent differential characteristic for the scheme. (joint work with Eran Lambooij, Abhishek Kumar, Somitra Sandhya).
4. Finally, we used ideas related to the above idea to attack the Korean FPE standard FEA-1.

3.12 Update on the ISO Standardization of Kuznyechik

Léo Perrin (INRIA – Paris, FR)

Joint work of Xavier Bonnetain, Léo Perrin, Shizhu Tian
Main reference Xavier Bonnetain, Léo Perrin, Shizhu Tian: "Anomalies and Vector Space Search: Tools for S-Box Analysis", in Proc. of the Advances in Cryptology – ASIACRYPT 2019 – 25th International Conference on the Theory and Application of Cryptology and Information Security, Kobe, Japan, December 8-12, 2019, Proceedings, Part I, Lecture Notes in Computer Science, Vol. 11921, pp. 196–223, Springer, 2019.
URL https://doi.org/10.1007/978-3-030-34578-5_8

In this talk, I presented the latest results on a specific S-box, how they disprove verifiable claims by its designers, and what their consequences were at ISO.

A year ago, we established that the S-box of Kuznyechik [2] (the block cipher recently standardized in Russia) is more structured than previously thought [3]: it can be written as a so-called TKlog. Yet, at ISO/IEC meetings, the Russian delegation was still pushing for the standardization of this block cipher, insisting that the S-box was generated by picking permutations uniformly at random until some properties were met.

To figure out if this claim could be true, we investigated the properties of random permutations (both in terms of cryptographic properties and in terms of structure) [1]. We found that a C implementation of this S-box exists that fits in 1155 bits. As there are $256! \approx 2^{1684}$ distinct 8-bit permutations, the probability that a C-implementation at least this short exists is at most $2^{1155+1-1684} = 2^{-528}$. This bound would be tight if all 2^{1155+1} bit strings of length at most 1155 were valid ASCII encoded C programs implementing 8-bit permutations; we thus expect it to be an extremely loose upper bound. As a consequence, we have to conclude that the designers of Kuznyechik are lying about the design process of a key component of their cipher: the probability of obtaining such a structured S-box using the process they disclosed is negligible.

At an ISO meeting held in Paris in October, the Russian delegation thus tried to convince the audience that all permutations are in fact structured (in spite of the facts highlighted above), the aim being to argue that their claims of randomness are true. Unsurprisingly, they failed to convince other countries representatives. As a consequence, the standardization of Kuznyechik has been stopped.

References

1 Bonnetain X., Perrin L., Tian S. *Anomalies and Vector Space Search: Tools for S-Box Analysis*. In: Galbraith S., Moriai S. (eds) Advances in Cryptology – ASIACRYPT 2019. Lecture Notes in Computer Science, vol 11921, pp 196–223. Springer, Cham.
2 Federal Agency on Technical Regulation and Metrology. *Information technology – data security: Block ciphers*. 2015. English version available at http://wwwold.tc26.ru/en/standard/gost/GOST_R_34_12_2015_ENG.pdf
3 Perrin, L. (2019). *Partitions in the S-Box of Streebog and Kuznyechik*. IACR Transactions on Symmetric Cryptology, 2019(1), 302-329.

3.13 On generating collisions in blinded keyed hashing

Yann Rotella (University of Versailles, FR)

Joint work of Yann Rotella, Joan Daemen, Jonathan Fuchs

In this talk, we analyze keyed-hashing modes with respect to collision resistance in a blinded keyed hashing model for the attacker in both serial and parallel constructions to do compression functions in cryptography.

The serial construction is used in CBC-MAC for blockcipher-based or DonkeySponge for Permutation-based, while the parallel one is used in P-MAC (blockcipher-based) or Farfalle (Permutation-based).

We try to obtain collisions in this setting by using differential trails existing in the inner permutation (or underlying blockcipher). Eventually, we mount two different attack strategies for both constructions, by using a single trail core. Our attack takes use of a huge set of trails, all sharing the same trail core.

More precisely, the expected number of inputs that we need to take into account for finding a collision is 2^W where W is defined as the sum of the weights of the round differentials starting from the 2nd round and where the weight of the last round is divided by 2. Also, in the case of the parallel construction, W is twice as large as in the case of the serial construction.

So in the case of a collision attack based on a single trail core, under reasonable assumptions the parallel construction offers twice the security level than the serial construction.

3.14 Improved Differential-Linear Attacks with Applications to ARX Ciphers

Yosuke Todo (NTT – Tokyo, JP)

Joint work of Christof Beierle, Gregor Leander, Yosuke Todo

Differential cryptanalysis and linear cryptanalysis are ones of the most common cryptanalysis techniques. The differential-linear attack is an extension of their techniques, and it used both in the same time: the differential characteristic for the first part and the linear trail for the second part. Usually, when the differential probability is p and the linear correlation is q, the required data complexity is $p^{-2}q^{-4}$. We proposed several new techniques for the differential-linear attack, in particular, the main focus of the application is ARX design.

On the differential part, we propose a new technique, where many "right pairs" are generated for free once we find only one "right pair". This technique allows us to distinguish the ciphers with data complexity of $p^{-1}q^{-4}$.

On the linear part, we propose a new partition technique using multiple linear trails. ARX ciphers have many multiple linear trails with particular structure, and these linear trails can be evaluated by guessing the same key bits. Moreover, we propose a new key-recovery algorithm, where the involved key bits are decomposed into two parts and only guessing the first part is enough to recover the whole of keys.

3.15 Attacks on the Legendre PRF

Aleksei Udovenko (CryptoExperts – Paris, FR)

Joint work of Aleksei Udovenko, Ward Beullens, Tim Beyne, Giuseppe Vitto
Main reference Ward Beullens, Tim Beyne, Aleksei Udovenko, Giuseppe Vitto: "Cryptanalysis of the Legendre PRF and generalizations", IACR Cryptol. ePrint Arch., Vol. 2019, p. 1357, 2019.
URL https://eprint.iacr.org/2019/1357

The Legendre PRF relies on the conjectured pseudorandomness properties of the Legendre symbol with a hidden shift. Originally proposed as a PRG by Damgård at CRYPTO 1988 [1], it was recently suggested as an efficient PRF for multiparty computation purposes by Grassi et al. at CCS 2016. Moreover, the Legendre PRF is being considered for usage in the Ethereum 2.0 blockchain.

In the talk, I describe a birthday-bound attack on the Legendre PRF with reduced query complexity compared to previous attacks due to Khovratovich [2]. Furthermore, I study a higher-degree generalization of the PRF and point out a large class of weak keys for this construction.

References

1 Ivan Damgard. *On the randomness of Legendre and Jacobi sequences.* In Shafi Goldwasser, editor, CRYPTO'88, volume 403 of LNCS, pages 163–172. Springer, Heidelberg, August 1990
2 Dmitry Khovratovich. Key recovery attacks on the Legendre PRFs within the birthday bound. Cryptology ePrint Archive, Report 2019/862, 2019

3.16 Forkciphers and Provable Security

Damian Vizár (CSEM – Neuchatel, CH)

Joint work of Elena Andreeva, Virginie Lallemand, Antoon Purnal, Reza Reyhanitabar, Arnab Roy, Damian Vizár

We report updates on the security of NIST Lightweight Cryptography candidate algorithm SAEF [1]. SAEF is a mode of operation of a forkcipher for authenticated encryption. SAEF was proposed with security up to $\approx 2^{n/2}$ processed bytes in the nonce-based AE security model in the original submission. The new result says that SAEF has online-AE (OAE) security up to $2^{n/2}$ processed bytes. This means that SAEF can be safely used when plaintext or ciphertext arrives in blocks, and does not crumble if nonces accidentally repeat, while being more efficient than many existing constructions.

We then propose several directions of interest. Firstly we point out the similarity of DECK function and multi-forkcipher security notions, and propose to study their relation. We remark that a recent DECK construction Farfalle can never achieve quantitatively optimal DECK security, and suggest that a notion between MFC and DECK would model it more closely.

References

1 Andreeva, E., Lallemand, V., Purnal, A., Reyhanitabar, R., Roy, A., Vizár, D.: ForkAE v1. Submission to NIST Lightweight Cryptography Project (2019)

Participants

- Elena Andreeva
Technical University of Denmark – Lyngby, DK
- Frederik Armknecht
Universität Mannheim, DE
- Christof Beierle
Ruhr-Universität Bochum, DE
- Daniel J. Bernstein
University of Illinois – Chicago, US
- Eli Biham
Technion – Haifa, IL
- Christina Boura
University of Versailles, FR
- Anne Canteaut
INRIA – Paris, FR
- Joo Yeon Cho
ADVA Optical Networking – Martinsried, DE
- Itai Dinur
Ben Gurion University – Beer Sheva, IL
- Christoph Dobraunig
Radboud University Nijmegen, NL
- Orr Dunkelman
University of Haifa, IL
- Maria Eichlseder
TU Graz, AT
- Patrick Felke
FH Emden, DE
- Henri Gilbert
ANSSI – Paris, FR
- Lorenzo Grassi
TU Graz, AT
- Tetsu Iwata
Nagoya University, JP
- Pierre Karpman
Université Grenoble Alpes – Saint Martin d'Hères, FR
- Dmitry Khovratovich
Ethereum – Luxembourg, LU
- Virginie Lallemand
LORIA – Nancy, FR
- Tanja Lange
TU Eindhoven, NL
- Nils Gregor Leander
Ruhr-Universität Bochum, DE
- Gaëtan Leurent
INRIA – Paris, FR
- Stefan Lucks
Bauhaus-Universität Weimar, DE
- Atul Luykx
Swirlds – San Francisco, US
- Willi Meier
FH Nordwestschweiz – Windisch, CH
- Florian Mendel
Infineon Technologies AG – Neubiberg, DE
- Bart Mennink
Radboud University Nijmegen, NL
- Kazuhiko Minematsu
NEC – Kawasaki, JP
- Maria Naya-Plasencia
INRIA – Paris, FR
- Kaisa Nyberg
Aalto University, FI
- Léo Perrin
INRIA – Paris, FR
- Bart Preneel
KU Leuven, BE
- Yann Rotella
University of Versailles, FR
- Arnab Roy
University of Bristol, GB
- Yu Sasaki
NTT – Tokyo, JP
- Ling Song
Chinese Academy of Sciences – Beijing, CN
- Meltem Sonmez Turan
NIST – Gaithersburg, US
- Marc Stevens
CWI – Amsterdam, NL
- Stefano Tessaro
University of Washington – Seattle, US
- Emmanuel Thomé
INRIA Nancy – Grand Est, FR
- Yosuke Todo
NTT – Tokyo, JP
- Aleksei Udovenko
CryptoExperts – Paris, FR
- Damian Vizár
CSEM – Neuchatel, CH
- Kan Yasuda
NTT – Tokyo, JP

Report from Dagstuhl Seminar 20051

Computational Metabolomics: From Cheminformatics to Machine Learning

Edited by

Sebastian Böcker[1], Corey Broeckling[2], Emma Schymanski[3], and Nicola Zamboni[4]

1 **Friedrich-Schiller-Universität Jena, DE,** sebastian.boecker@uni-jena.de
2 **Colorado State University, Fort Collins, CO, US,** corey.broeckling@colostate.edu
3 **University of Luxembourg, LU,** emma.schymanski@uni.lu
4 **ETH Zürich, CH,** zamboni@imsb.biol.ethz.ch

Abstract

Dagstuhl Seminar 20051 on Computational Metabolomics is the third edition of seminars on this topic and focused on Cheminformatics and Machine Learning. With the advent of higher precision instrumentation, application of metabolomics to a wider variety of small molecules, and ever increasing amounts of raw and processed data available, developments in cheminformatics and machine learning are sorely needed to facilitate interoperability and leverage further insights from these data. Following on from Seminars 17491 and 15492, this edition convened both experimental and computational experts, many of whom had attended the previous sessions and brought much-valued perspective to the week's proceedings and discussions. Throughout the week, participants first debated on what topics to discuss in detail, before dispersing into smaller, focused working groups for more in-depth discussions. This dynamic format was found to be most productive and ensured active engagement amongst the participants. The abstracts in this report reflect these working group discussions, in addition to summarising several informal evening sessions. Action points to follow-up on after the seminar were also discussed, including future workshops and possibly another Dagstuhl seminar in late 2021 or 2022.

Seminar January 26–31, 2020 – http://www.dagstuhl.de/20051
2012 ACM Subject Classification Applied computing → Life and medical sciences
Keywords and phrases bioinformatics, chemoinformatics, computational mass spectrometry, computational metabolomics, machine learning
Digital Object Identifier 10.4230/DagRep.10.1.144
Edited in cooperation with Adelene Lai

1 Executive Summary

Sebastian Böcker (Friedrich-Schiller-Universität Jena, DE)
Corey Broeckling (Colorado State University – Fort Collins, CO, US)
Emma Schymanski (University of Luxembourg, LU)
Nicola Zamboni (ETH Zürich, CH)

Mass spectrometry is the predominant analytical technique for detection, identification, and quantification in metabolomics experiments. Technological advances in mass spectrometry and experimental workflows during the last decade enabled novel investigations of biological systems on the metabolite level. Metabolomics started as the study of all metabolites in a living cell or organism; in comparison to transcriptome and proteome, the metabolome is a better proxy of metabolic activity. Emerging fields including personalized medicine and exposomics have expanded the scope of metabolomics to "all" small molecules, including those of non-biological origin. Advances in instrumentation plus rapid increase in popularity, throughput and desired compound coverage has resulted in vast amounts of both raw and processed data; the field is in desperate need for further developments in computational methods. Methods established in other -omics fields are frequently not transferable to metabolomics due to the structural diversity of small molecules. This third Dagstuhl Seminar on Computational Metabolomics (following Seminars 15492 and 17491) focused on cheminformatics and machine learning. The seminar was less structured than previous seminars, forming break-out sessions already from Monday afternoon, then collecting participants back into plenary sessions at regular intervals for discussions and further topic exploration. The major topics launched on Monday included cheminformatics, genome mining and autoencoders, which were developed throughout the day. Other topics discussed throughout the week included biosynthesis and gene clusters, confidence and compound identification, spectral versus structural similarity, statistical integration, collision cross section (CCS) and ion mobility separation (IMS), benchmarking data, open feature file format, exposomics, data processing and acquisition. Several evening sessions were also held, including retention time, Bioschemas, MassBank, ethics and philosophy of software development, open biological pathways, mass spec health check, Jupyter notebooks, a mini decoy session and a session on coding tips. The excursion, breaking with previous Christmas Market traditions, was to the Völklingen steelworks. Finally, the entire seminar was wrapped up with a discussion on the future of untargeted metabolomics on Friday – time will tell what the future Computational Metabolomics Seminars will bring. A further seminar in the series may be considered for the end of 2021 or in 2022.

2 Table of Contents

3 Break-Out Group and Plenary Discussions

3.1 Spectral vs. Structural Similarity

Oliver Alka (Universität Tübingen, DE), Adelene Lai (University of Luxembourg, LU), and Justin van der Hooft (Wageningen University, NL)

Spectral similarity underpins many of our analyses, like the use of spectral similarity in library matching and molecular networking. This break-out group tried to reconcile spectral and structural similarity – on a fundamental level, can we equate two molecules structurally if their spectra are considered similar? Feedback collected from the group showed that cosine similarity was the most-used and perhaps well-known measure of spectral similarity because of how easy it is to calculate and wide availability in various vendor software, but that it is an imperfect measure not least because it is hard to test how it works. Further options for measuring spectral similarity discussed include Hybrid (considering fragment and losses), All Mass Differences, and performing both Forward and Reverse comparisons. The impact of different instruments and their respective vendors and options (e.g. ramped collision energy, stepped) on spectra was also discussed, with some suggestions to merge or derive an average spectrum. This could be improved using mass difference in the scoring by creating a hybrid score for example. Some concrete ideas on implementing graph-based extraction of (relevant) mass differences from spectra and using those to calculate a similarity score were also discussed.

Regarding structural similarity, Tanimoto was regarded by many as inadequate, and other methods were discussed, including fingerprint comparison, maximum common edge subgraph, and DICE. On evaluation, chemical classes predicted from spectra were proposed as alternatives to fingerprints.

3.2 Data Processing in Metabolomics

Nikiforos Alygizakis (Environmental Institute – Koš, SK)

Data processing pipelines consist of discrete steps (centroiding, chemical noise removal, peak picking, retention time alignment, grouping of features, componentization of isotopes, adducts and in-source fragments). Even though there is a multitude of software (both open-source and commercial) for each step of the pipeline, there is still space for improvement. Peak picking is an area with great potential for improvement and is a crucial step in metabolomics workflows. It must be highlighted that there are commercial peak pickers (e.g. Genedata Expressionists) that may also be worth implementing as open-source tools and benchmarked against established peak pickers. Little margin for improvement exists for grouping of peaks across samples and retention time alignment. Componentization and especially accurate detection of adducts in MS1 full-scan spectra is a topic that needs further investigation. Adduct formation heavily depends on the mobile phases of chromatography and physicochemical properties of the analytes. This topic has not been addressed and current mass spectral libraries rarely store MS1 spectra. Instrumental developments such as high-resolution (R>500,000) and recording of profile data motivate the need of improved componentization software that can improve annotation in metabolomics workflows. Existing software should be parallelized and new software with sophisticated computational approaches can now be applied, since computer power is readily available. Software developments should take into account the application of strong quality assurance and quality control during metabolomic experiments (e.g. QC charts, spiking of internal standards, standard operational procedures for all parts of the analysis) that needs to be implemented in all analytical laboratories. High-quality data in combination with advanced software tools can significantly improve data processing in metabolomics.

3.3 MS/MS Spectrum Quality and Instrument Control

Corey Broeckling (Colorado State University – Fort Collins, CO, US)

MS/MS Spectrum quality for small molecules has historically depended on spectral similarity to library entries. Computational interpretation tools have opened the possibility to explore spectrum information content in a library independent manner. There is little rigorous description of what constitutes a high quality spectrum for small molecules, particularly in the absence of a library search. In the proteomics field, descriptors for spectrum quality have been suggested and might be adapted to metabolomics. This has yet to be experimentally and statistically determined. In general, it seems that the fragments in the middle between the minimum m/z and the precursor hold the most information, and more fragments are better than few. In addition, using different fragmentation methods, such as CID and HCD, seem not to hold additional information about quality and metabolite identification. Experimental methods offering real-time instrument control could improve the quality by using multiple collision energies or ramps, or refining collision energy on a feature-by-feature basis. The isolation window (MS1), as well as the time of sampling seem to be important. The conclusion is that spectral quality assessment needs more experimental evaluation to find valid descriptors and validate these for different experimental setups

3.4 Exposomics

Xiuxia Du (University of North Carolina – Charlotte, NC, US), Kati Hanhineva (University of Kuopio, FI), and Augustin Scalbert (IARC – Lyon, FR)

The exposome encompasses all environmental exposures including chemical, physical, and biological stressors, as well as lifestyle and social environments, from conception through adulthood (https://hhearprogram.org/). Despite tremendous efforts that have been made by researchers in diverse areas including environmental sciences, metabolomics, nutritional sciences, etc, enormous challenges remain. One of these challenges concerns the tremendous efforts currently required to annotate exposome data. More than half of the session time was spent on discussing the causes of this challenge and potential ways to address it.

The causes include: (1) the huge chemical space that the exposome covers and further biotransformations of the compounds in this space; (2) fragmentation of available resources; (3) onerous efforts required to deposit data in repositories; (4) shortage of reference spectra for assigning spectra to compounds; (5) lack of reference exposome; and (6) shortage of training data to build automated computational tools for annotating the exposome.

This challenge can be addressed from different angles simultaneously. For example, the detected compounds can be prioritized for suspect screening based on metadata that are collected from: (1) specific experiments (e.g. curated in Metabolomics workbench, MetaboLights or GNPS), or (2) literature sources with data eventually curated in existing databases (e.g. HMDB, PubChem, FooDB, Phenol-Explorer, Exposome-Explorer).

Furthermore, additional resources and informatics capabilities would be needed to facilitate exposome annotation. These include: (1) data mining tools to collect information scattered in the literature, mainly in pdf files; (2) training data for priority scoring in annotation (e.g. CRISPR-CAS9, artificial guts, etc); (3) tools for more efficient and rapid annotation and suspect screening in metabolic profiles largely done manually so far; (4) sharing analytical/spectral data from samples and reference compounds to speed up the annotation of the exposome through a community effort; (5) better integration of different types of data from various databases (e.g. links to spectra in PubChem); and (6) resources to support deposition curation and warrant sustainability of databases.

Finally, we discussed how to further address the challenge. We asked Dr. David Wishart to lead an effort to coordinate future research and development activities by researchers. As an actionable item, Dr. Wishart will plan to host a workshop in Edmonton or the Rocky mountain parks (Canmore) in the summer of 2020.

3.5 Mass Spectrometry Coding Standards

Laurent Gatto (University of Louvain, BE) and Ewy Mathé (Ohio State University – Columbus, OH, US)

During this discussion about guidelines on how to share code related to computational mass spectrometry, we decided to remain programming language agnostic, and focus on community-level goals. It was highlighted that for such contributions to be helpful, they need to contain software or code, and at least some testing data and documentation. The extent and "quality" of these elements, especially the latter, should however be regarded as flexible for two main reasons, the first one being that less seasoned contributors shouldn't be barred from disseminating their work due to arbitrarily strict requirements. Second, there is a difference between publishing a method or a (computational) solution to a specific problem and a "finished" software product, and it is important to appreciate the value (novelty or engineering quality, for example) of both of these outputs. Hence the importance for these guidelines to emanate from the community at large to enable/facilitate important goals, and should not become rigid requirements.

We have identified three important end goals that should be highlighted when contributing and disseminating code, namely (1) reproducibility, (2) usability and (3) learnability. Each of these will require code, documentation and test data, albeit to different extents. In some cases, small test data and a README file will suffice to install and reproduce some scripts implementing a novel method. On the other hand, finalised software products will have to provide more in-depth documentation (function-, software-level documentation, how-to's, etc.) and comply with additional (language-specific) software requirements, hence the importance for the contributors to accurately describe the type and scope of the code deliverables they share with the community.

3.6 Cheminformatics for Users

Marcus Ludwig (Friedrich-Schiller-Universität Jena, DE), Steffen Neumann (IPB – Halle, DE), and Egon Willighagen (Maastricht University, NL)

Cheminformatics is the use of computer and informational techniques applied to a range of problems in the field of chemistry [1]. In the context of Computational Metabolomics we represent metabolites as molecular structures, but due to the uncertainty in annotation, we need to be able to represent partially characterised structures. Representation of partial information can be distinguished into two different applications: (1) Listing the occurrence of defined substructures (fingerprint) of the measured molecule or categorizing molecules into classes, and (2) the estimation of the biggest core structure which is supported by the measured data. We concentrated on the estimation of core structures in this discussion. Since the 2017 Dagstuhl Seminar 17491 [2] methods have been developed (e.g. ChemAxon Extended SMILES (CxSMILES), Markush Structures), and examples were now created during an evening session. Discussion topics included how different layers of information provide different pieces of structural evidence, and that CxSMILES provides many solutions, but is limited. For example, for uncertainty of double bond locations in lipid tails, CxSMILES does not have a satisfactory solution. Therefore, the molecular formula and shortlists of specific compounds remain complementary. The need for open source tools to derive a common CxSMILES, depict CxSMILES, and enumerate structures starting with an CxSMILES was established. The Chemistry Development Kit is being explored for this. Another area of cheminformatics is structure generation required to identify metabolites not yet in compound databases. Existing approaches cover a continuum from unconstrained structure generation, to combinatorial decoration of frameworks or backbones and biochemical expansion of structure databases. There are cross-links to the session on autoencoders of chemical structures, which can generate structures, ideally with constraints from prior or experimental knowledge.

References

1 Wikipedia contributors. *Cheminformatics – Wikipedia, The Free Encyclopedia*. https://en.wikipedia.org/w/index.php?title=Cheminformatics&oldid=909899401, Online; accessed 3-February-2020.
2 Dagstuhl seminar 17491 contributors. *Computational Metabolomics: Identification, Interpretation, Imaging*. https://www.dagstuhl.de/17491, Online; accessed 3-February-2020.

3.7 The mzFeature File Format to Bridge Processing and Annotation in Untargeted Metabolomics

Tytus Mak (NIST – Gaithersburg, MD, US), Oliver Alka (Universität Tübingen, DE), Sebastian Böcker (Friedrich-Schiller-Universität Jena, DE), Pieter Dorrestein (University of California – San Diego, CA, US), Markus Fleischauer (Friedrich-Schiller-Universität Jena, DE), Oliver Kohlbacher (Universität Tübingen, DE), Marcus Ludwig (Friedrich-Schiller-Universität Jena, DE), Louis-Felix Nothias-Scaglia (University of California – San Diego, CA, US), and Tomas Pluskal (Whitehead Institute – Cambridge, MA, US)

While there are open formats for mass spectrometry data (e.g. mzML) and downstream annotation (i.e. mzTab-M), there is currently no existing file interoperable format to bridge the gap between processing and structure annotation tools in non-targeted LC-MS/MS data processing. This proposal aims at designing an intermediate "mzFeature" open file format that would hierarchically store information on the detected spectral features that have been extracted via peak picking/feature finding algorithms (e.g. XCMS, MZmine, OpenMS). Feature objects are storing centroided spectral information (mass traces, associated MS2 spectra, MSn etc.), along with m/z and retention time statistics (i.e. peak apex, peak start/end). These are usually extracted on a file basis and Features of the same file can be grouped as a FeatureMap. The Features can be linked into FeatureGroups across multiple mass spectrometry files, which may consist of an adduct type, isotopologues, and in-source fragments that originate from the same molecule. Ambiguities are accounted for via multiple mappings, such as an MS2 spectrum being assigned to multiple feature objects. The format should accommodate the inclusion of metadata that are specific to various instrument and processing tools.

3.8 Benchmark Data

Ewy Mathé (Ohio State University – Columbus, OH, US)

Benchmarking datasets are needed to test new methods. These data should be well understood and well characterised. One specific area of need is multi-omics datasets. When collecting these data, the proper meta-information (on samples and metabolites) needs to be included. There needs to be a balance between incorporating appropriate meta-information and the difficulty/time required for collecting/inputting that info.

There are multiple complementary efforts for doing this: 1) MANA SODA: community-driven input of data and software; 2) NIH Metabolomics Workbench: well curated datasets collected for benchmarking; 3) a previous Dagstuhl conference had started a similar effort for Proteomics [1]. Incentivizing data generators and software developers to submit their work (e.g. publications, advertisements, recommendations, etc.) is key to the success of these efforts. Also defining use cases, where what people want data for is defined, is important. Benchmarking data could be comparison of existing data or may require the generation of new data. The task of this session will be to define such use cases and best approaches to collecting benchmarking datasets and to make them useful to the larger community.

References

1 Dagstuhl seminar 19351 contributors. *Computational Proteomics*. https://www.dagstuhl.de/19351, Online; accessed 17-April-2020.

3.9 Mining Metabolome and Genome

Ewy Mathé (Ohio State University – Columbus, OH, US)

Many resources are available for supporting the integrated analysis of genomes and metabolomes. However, these resources are largely fragmented and mostly lack interoperability. Computational expertise is most often a requirement to piece together resources for appropriate interpretation of integrated metabolome-genome data. There is thus a need for defining common meta-data/controlled vocabulary, and for automating the process of deriving detailed meta-data for samples and analyte (metabolites, genes).

The task of this group was to define user cases and guidelines on how to use and integrate resources to meet user needs. Guidelines will include defining quantitative metrics to use these databases properly (e.g. FDR confidence, being able to detect discrepancies between different sources), and unit tests for data integration. The steps in integrating resources are modular. Limitations of each module were defined, so that users can then piece together different modules to meet their needs.

3.10 Confidence and Compound Identification

Hunter Moseley (University of Kentucky – Lexington, KY, US)

This session explored how to quantify confidence in compound identification. Three major types of confidence metrics were identified: confidence categories, (continuous) confidence scores, and probabilistic scores. Different types of probabilistic scores were covered, especially probabilistic scores that take into account false discovery. Inaccuracy in estimating low false discovery rates (FDR) in the context of MS/MS-based compound identification was discussed. An alternative or complementary method is to estimate compound identification ambiguity from assignment and dataset specific decoy generation. The supplementation of richer spectral data to improve assignment was mentioned. Using identification confidence and/or ambiguity to limit deposited annotations was discussed. The general consensus was that more assignment annotations with deposition would allow broader data reuse and improve interpretation.

3.11 MassBank: Status Meeting

Steffen Neumann (IPB – Halle, DE)

MassBank was the first open source, open data spectral library. Currently, there are sites in Japan, Europe, and the US (MoNA). In Dagstuhl there was the opportunity to discuss current and future developments among users, developers and related resources. These included the quality assurance in spectral library creation, an upcoming REST interface and opportunities to interchange data with other sites.

3.12 Autoencoders

Jamie Nunez (Pacific Northwest National Lab – Richland, WA, US) and Michael Andrej Stravs (Eawag – Dübendorf, CH)

Methods of generating potentially novel compounds was first covered, which included combinatorics, reactions, rule-based construction, experimental data-driven generation, and autoencoders. Autoencoders were covered in more detail, first describing their general set up and the use of latent space (a compressed version of the data input to the autoencoder which is then interpreted by the decoder). An example of structure-to-structure designs was examined, along with the considerations of its advantages and disadvantages, how training was done, and what latent space truly represents at the end. Other potential designs were then discussed, such as fingerprint-to-structure to generate candidates from experimental data and reactions-to-reactions. It is important to also keep in mind that decoders can often produce invalid output, which has to be checked, showing a need to carefully interpret the real meaning of the output and (non)continuity of latent space.

3.13 Collision Cross Section and Ion Mobility Spectrometry

Tomas Pluskal (Whitehead Institute – Cambridge, MA, US)

Ion mobility spectrometry (IMS) is a technique for separating molecules in a neutral gas phase based on their drift time (time spent in the IMS chamber), which is proportional to the collision cross section (CCS) of the molecule. IMS can be conveniently combined with mass spectrometry for better separation and identification of molecules. Significant progress has been made in predicting CCS values of molecules using deep learning and quantum chemistry calculations. However, IMS presently suffers from relatively poor support in data processing tools and packages. During the session, various hardware approaches for IMS separation were introduced and specific needs for data processing tools were discussed. There was general consensus that IMS has great potential, but the current hardware and software capabilities are limited. In particular, a lack of a good algorithm for 4D (chromatography retention time, IMS drift time, m/z, and intensity) feature detection was identified as a major bottleneck in the field. Development of new visualization tools and CSS distribution databases was also encouraged.

3.14 Jupyter Notebooks for #FAIR Data Science

Stacey N. Reinke (Edith Cowan University – Joondalup, AU)

The Jupyter Notebook is an open-source interactive coding tool that launches in a web browser. It contains cells for descriptive text and live code; outputs of executed code cells (tables, visualisations) are then displayed immediately below the code cell. This framework was developed to enable transparent sharing of code and workflows, therefore promoting FAIR data science in the scientific community. More recently, the launch of the Binder deployment service has allowed researchers to share their Jupyter Notebooks in the cloud with a url link. This session provided a description of Jupyter Notebooks and Binder, as well as their practical utility in workflow sharing and education.

3.15 Statistical Integration

Stacey N. Reinke (Edith Cowan University – Joondalup, AU)

Metabolomics data can be integrated with other types of data, such as other omics or clinical data, to enable a more comprehensive understanding of the biological system. This session aimed to identify and discuss different approaches for data integration of two or more matrices, one being metabolomics data. Three different approaches were identified. Network-driven integration approaches require *a priori* biological knowledge. They can include mathematical models of individual biological processes or pathway mapping. Pathway mapping often suffers from lack of interpretability due to the high level of metabolic interconnection. Dimension reduction integration aims to reduce the metabolic feature space prior to downstream pathway analysis; however, testing has shown lack of robustness with respect to pathway definition. Data-driven integration approaches include methods such as correlation and multivariate analyses. These approaches can enable the identification of novel biology; however, they are limited by lack of usability and interpretability. The outcome of this session included a list of tools for achieving data integration and also an acknowledgement that this is a developing field which needs to be further developed prior to large scale implementation.

3.16 Biosynthesis, Gene Clusters, and Predicting Natural Products from the Genome

Justin van der Hooft (Wageningen University, NL) and Simon Rogers (University of Glasgow, GB)

Metabolite identification of natural products can be accelerated by linking information gained from genome sequences. This breakout group started with a short historical perspective on using structural information from the genome to inform structural elucidation which started back in 2005 with the first natural product being predicted from the genome of *Streptomyces coelicolor*. The major questions the group addressed were what structural and quantitative information can be predicted from genomes? And how do Biosynthetic Gene Clusters help? A list of resources included the PRISM and antiSMASH ecosystems that sparked the development of tools that link genome and metabolome data. Listed examples are GNP and NRPQuest and RiPPQuest that show relative successful examples for modular structures like peptides and some polyketide classes. The Dorrestein lab developed peptidogenomics and glycogenomics workflows that link the genome and metabolome by predicting amino acid and sugar moieties, respectively, that can be searched for in mass spectrometry data through mass differences and neutral losses. The group then discussed the next steps. Linking genomes directly to structures is a (very) hard problem; linking the genome to spectra is still challenging but can be regarded as "ranking problem". In the genome, gene domains are mainly used to translate the genome into structural information – through (predicted) enzyme activity. In the metabolome/metabolomics, once annotated, structural elements (substructures) can be exploited, for example by using chemical fingerprints. However, spectral patterns on themselves could be used as well to link to specific genetic elements. This could be helpful in prioritizing candidate spectra – gene links found by correlation approaches (based on strain presence/absence). Finally, the group discussed how prioritization of candidate structures could be improved by allowing to select groups of metabolites from one organism or – more widely – from natural products – or even more generic – from molecules that could be found in nature – which could include pesticides. Altogether, accelerating natural product discovery through linking the genome to the metabolome is a promising field!

3.17 Bioschemas

Egon Willighagen (Maastricht University, NL)

Bioschemas (https://bioschemas.org/) is an extension of the schema.org standard used by major search engines like Google and Bing to recognize information or metadata they want to use in their indexing. Bioschemas has an annotation type for the life sciences, like Protein and MolecularEntity, but also types for Tool (e.g. software) and TrainingMaterial (like tutorials). It is supported by the EU ELIXIR community as an interoperability layer and is used in a variety of their projects. In this session we discussed what it is and is not (e.g. it is not an ontology), looked at various annotation types (called "profiles"), and what information one can add to it. We looked at various solutions people have found to use Bioschemas in their project. For example, we looked at how Bioconductor package vignettes can be extended. Additionally, we looked at how ChEMBL uses Bioschemas on their HTML pages. The Bioschemas website has a page with live deployments. The meeting was concluded with hacking on Bioschemas annotation of Bioconductor packages itself, continuing a patch initiated at a computational metabolomics meeting in Wittenberg, DE in April 2019.

3.18 Open Biological Pathways with WikiPathways

Egon Willighagen (Maastricht University, NL)

WikiPathways (https://wikipathways.org/) is a free, online, community-driven, curated knowledgebase of biological pathways. Comparable and complementary to other databases like KEGG and Reactome, WikiPathways has a semantic representation of biological processes, resulting from past and current collaborations with research communities like WormBase, LIPIDMAPS, NetPath, EJP-RD, and many, many more. The WikiPathways Portals reflect this community embedding. The focus has always been on interoperability and semantic meaning which was discussed in the session. The semantic web format and SPARQL application programming interface were also discussed. We walked through a number of further integrations, such as EuropePMC linking to pathways for articles that are cited by that pathway (LabLinks), Wikidata, and named resources that include the WikiPathways, such as RAMP. Finally, we looked at how pathways are drawn with PathVisio, extended with CyTargetLinked in Cytoscape with transcription factors, miRNAs, and drugs (-lead) from DrugBank and ChEMBL. Questions around the underlying GPML format and RDF export were discussed, in addition to the curation process, and how all this is used in systems biological pathway and network enrichment analyses.

3.19 Retention Time

Michael Anton Witting (Helmholtz Zentrum – München, DE)

Retention times represent an interesting orthogonal information for metabolite identification. However, they are less standardized compared to other parameters and represent a property of the metabolite and the employed chromatographic system in comparison to mass, which is a molecular property. In this session we discussed the current state of the art in retention time prediction and how it can be integrated with e.g. analysis of tandem MS data. An approach for prediction of retention orders developed by Juho Rouso and Sebastian Böcker was discussed and what kind of additional data is required to further develop it.

3.20 Conclusion: The Future of Computational Metabolomics

Sebastian Böcker (Friedrich-Schiller-Universität Jena, DE)

In this plenary discussion, we tried to identify upcoming research questions in computational metabolomics, but also identify new possibilities that computational methods will provide for metabolomics in general. Computational methods for, say, small molecule annotation have evolved greatly in recent years, as demonstrated by CASMI contests [1]; how can we continue with method development in this speed, and how do we best utilize the developed methods? One particular topic of discussion was how to attract experts from machine learning to work on metabolomics problems. Here, it is of utmost importance to lower the barrier to enter the field for scientists from machine learning; e.g., to formalize problem(s) and to describe them in terms that machine learning scientists can understand (graph theory, optimization, etc). We will try to use the Kaggle platform (https://www.kaggle.com/) to attract ML experts; we identified some topics such as anomaly detection in clinical environments (i.e. high cholesterol) and retention time/order prediction as topics where this may be possible. Another topic of discussion was the disruptive changes of MS and computational technology: Where do we expect them to be, and what impact will these changes have? Discussed topics included prediction accuracy, quantum computing, the use of GPUs, and substantial increase in annotation rates. A third topic of discussion was metabolic modelling and stable isotope labelling experiments: these can lead to improved biological insight, with or without stable isotope labeling. We discussed the potential to use existing and new datasets that link metabolomics, transcript, genomics, or proteomics to improve interpretability; the importance of FAIR data was mentioned in this context. Finally, improved annotation can produce better metabolic/system modelling and allow us to generate new biological hypotheses.

References

1 Schymanski, Emma L., et al. Critical Assessment of Small Molecule Identification 2016: automated methods. Journal of Cheminformatics, 9.1 (2017): 22.

Participants

- Oliver Alka
Universität Tübingen, DE
- Nikiforos Alygizakis
Environmental Institute – Koš, SK
- Sebastian Böcker
Universität Jena, DE
- Evan Bolton
National Institutes of Health – Bethesda, US
- Corey Broeckling
Colorado State University – Fort Collins, US
- Celine Brouard
INRA – Toulouse, FR
- Andrea Brunner
KWR Water Research Institute – Nieuwegein, NL
- Jacques Corbeil
University Laval – Québec, CA
- Alexis Delabriere
ETH Zürich, CH
- Pieter Dorrestein
University of California – San Diego, US
- Xiuxia Du
University of North Carolina – Charlotte, US
- Timothy Ebbels
Imperial College London, GB
- Markus Fleischauer
Universität Jena, DE
- Laurent Gatto
University of Louvain, BE
- Kati Hanhineva
University of Kuopio, FI
- Rick Helmus
University of Amsterdam, NL
- Lukas Käll
KTH Royal Institute of Technology – Solna, SE
- Oliver Kohlbacher
Universität Tübingen, DE
- Adelene Lai Shuen Lyn
University of Luxembourg, LU
- Jan Lisec
BAM – Berlin, DE
- Marcus Ludwig
Universität Jena, DE
- Tytus Mak
NIST – Gaithersburg, US
- Hiroshi Mamitsuka
Kyoto University, JP
- Ewy Mathé
Ohio State University – Columbus, US
- Hunter Moseley
University of Kentucky – Lexington, US
- Steffen Neumann
IPB – Halle, DE
- Louis-Felix Nothias-Scaglia
University of California – San Diego, US
- Jamie Nunez
Pacific Northwest National Lab. – Richland, US
- Tomas Pluskal
Whitehead Institute – Cambridge, US
- Stacey N. Reinke
Edith Cowan University – Joondalup, AU
- Simon Rogers
University of Glasgow, GB
- Juho Rousu
Aalto University, FI
- Augustin Scalbert
IARC – Lyon, FR
- Tobias Schulze
UFZ – Leipzig, DE
- Emma Schymanski
University of Luxembourg, LU
- Christoph Steinbeck
Universität Jena, DE
- Michael Andrej Stravs
Eawag – Dübendorf, CH
- Justin van der Hooft
Wageningen University, NL
- Philip Wenig
Lablicate – Hamburg, DE
- Egon Willighagen
Maastricht University, NL
- David Wishart
University of Alberta – Edmonton, CA
- Michael Anton Witting
Helmholtz Zentrum – München, DE
- Oscar Yanes
Rovira i Virgili University – Reus, ES
- Nicola Zamboni
ETH Zürich, CH