Random Walks in Polytopes and Negative Dependence

Yuval Peres[1], Mohit Singh[2], and Nisheeth K. Vishnoi[3]

1 **Microsoft Research, Redmond, USA**
`peres@microsoft.com`
2 **Georgia Institute of Technology, Atlanta, USA**
`mohitsinghr@gmail.com`
3 **Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland**
`nisheeth.vishnoi@gmail.com`

Abstract

We present a Gaussian random walk in a polytope that starts at a point inside and continues until it gets absorbed at a vertex. Our main result is that the probability distribution induced on the vertices by this random walk has strong negative dependence properties for matroid polytopes. Such distributions are highly sought after in randomized algorithms as they imply concentration properties. Our random walk is simple to implement, computationally efficient and can be viewed as an algorithm to round the starting point in an unbiased manner. The proof relies on a simple inductive argument that synthesizes the combinatorial structure of matroid polytopes with the geometric structure of multivariate Gaussian distributions. Our result not only implies a long line of past results in a unified and transparent manner, but also implies new results about constructing negatively associated distributions for *all* matroids.

1998 ACM Subject Classification F.2.1 Numerical Algorithms and Problems

Keywords and phrases Random walks, Matroid, Polytope, Brownian motion, Negative dependence

Digital Object Identifier 10.4230/LIPIcs.ITCS.2017.50

1 Introduction

A basic problem that underlies a large number of approximation algorithms for combinatorial problems is: given a fractional point, how should it be rounded to an integral point? Typically, the fractional point is obtained by solving a linear program and has nice properties which one would like the integeral point to inherit. For instance, the fractional point might belong to the spanning tree polytope of a graph and the goal might be to round the point to a spanning tree whose cost is not much more and, in addition, satisfies some constraints satisfied by the fractional point; e.g., number of edges in the tree across each cut is small. A common approach towards this is to construct randomized rounding algorithms for such a problem which output an unbiased distribution over the underlying set of integral objects (spanning trees for the above example). Since in most interesting cases the set of integral objects is not a box, the distributions have correlations. These correlations end up being quite problematic when it comes to ensure that with high probability the integral point also satisfies the *additional* constraints the fractional point satisfies. The ultimate hope here is that the distribution essentially behaves like a product distribution so that one can apply concentration results such as Chernoff bounds. On the one hand, this has recently led TCS researchers to come up with a wide variety of ingenious rounding algorithms which have resulted in significant

8th Innovations in Theoretical Computer Science Conference (ITCS 2017).
Editor: Christos H. Papadimitrou; Article No. 50; pp. 50:1–50:10
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

progress on fundamental algorithmic problems [11, 2, 6, 7, 12, 19, 1] (see the book [8] for more examples). On the other hand, mathematicians have been investigating what kind of negative dependence properties result in the phenomena of measure concentration [17, 3]. Typically, it is non-trivial to both compute these distributions and prove their negative dependence properties.

In this paper we take a geometric approach and propose a very simple rounding algorithm and prove that it gives an unbiased distribution with *negative association* property for *all* matroids. Negative association is a significant strengthening of the negative correlation property. Roughly a set of random variables $(X_1, \ldots, X_n)$ is said to be negatively associated if for any two monotone functions f and g which act on distinct sets of coordinates S and T respectively, the corresponding random variables $f(X_i)_{i \in S}$ and $g(X_j)_{j \in T}$ are negatively correlated; see Definition 1 and Theorem 4.

A bit more formally, given a matroid over a universe of size n, consider the polytope in $\mathbb{R}^n$ that is the convex hull of all the bases of the matroid. Let θ be a given fractional point that lies inside this polytope and is to be rounded to one of the bases. Our algorithm starts at θ and keeps taking small Gaussian steps centered at the current point – thus maintaining the unbiasedness. If at some point the trajectory hits a constraint on the boundary of the polytope it never leaves it – meaning that the Gaussian in the next step is chosen so as to have no mass outside of this subspace. Thus, eventually, the trajectory gets absorbed at a vertex of the polytope. This rounding algorithm is inspired by the work of [15] on discrepancy and its simplicity is self-evident. At any given time, the algorithm has to keep track of the tight constraints and compute a Gaussian in the space corresponding to the intersection of these constraints.

Our proof synthesizes polyhedral properties about matroids with well-known properties of Gaussian distribution. Our key structural observation is that if F is a d-dimensional face of a matroid polytope and Σ_F is the covariance matrix of the d-dimensional Gaussian obtained by orthogonally projecting an n-dimensional Gaussian onto F, then all the off-diagonal entries of Σ are non-positive; see Theorem 5. The proof of this relies on an elementary *uncrossing* argument from matroid theory. *Thus, Gaussian distributions on faces of matroid polytopes have the negative correlation property.* This, in turn, immediately leads to an inductive argument that shows that from the beginning to the end of this random walk, the distribution has the negative correlation property. Finally, to go from negative correlation of Gaussian distributions to negative association, we employ a result that implies that for Gaussian distribution negative correlations implies negative association [13]. Roughly, this is a manifestation of the fact that all moments of Gaussian distributions are completely determined by their covariance matrix.

Negative dependence properties for distributions on bases of matroid have been extensively studied. Prior to our work, negatively associated distributions were known only for *balanced* matroids. In particular, [10] shows that the uniform distribution is negatively associated for balanced matroids and gives a Markov chain Monte Carlo method to sample from such a distribution. Unfortunately, the uniform, or more generally, an entropy maximizing distribution on general matroids does not give negative association property. For general matroids, the weaker negative cylindrical property (see Definition 3) was shown for the *pipage rounding* and *randomized swap rounding* [6]. The other two matroids that have been extensively studied are the uniform matroid, where bases are all sets of size k for integer k [20, 5] and the partition matroid [9, 8].

Lovett and Meka [15], who studied this random walk on the polytope obtained by intersecting the hypercube with few *discrepancy* constraints, were interested in the number of integral coordinates of the output vertex. In our setting, we study the walk on integral polytopes, convex hull of bases of matroids, which are typically defined by exponentially many constraints and are interested in negative dependence properties of the output distribution.

While both the algorithm and the proof are simple and the reason for negative dependence is clear, several problems about the distribution remain open. First and foremost, what can we say about non-matroid polytopes? Do non-Gaussian distributions help? While we can understand the random walk locally, a global perspective seems hard. Concretely, can we prove that the distribution obtained by our random walk is the solution to some optimization problem? Finally, do much stronger forms of negative dependence, for example, Strongly Rayleigh property [3] holds for these distributions in specific matroids (see also [4])?

1.1 Preliminaries

We first give the following definition.

▶ **Definition 1** (Negative Association). Let $X_1, \ldots, X_n$ be boolean random variables. Then X_i's are *negatively associated* if for every non-decreasing functions $f : \{0,1\}^n \to \mathbb{R}$ and $g : \{0,1\}^n \to \mathbb{R}$ we have

$$E[f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)] \leq E[f(X_1, \ldots, X_n)]E[g(X_1, \ldots, X_n)] \tag{1}$$

if f and g depend on disjoint sets of coordinates.

While negative association is hard to verify, under very mild assumptions, it implies versions of the central limit theorems (see Yuan et al [21] and Pattersen et al [16]).

A much weaker notion of negative dependence is the negative correlation defined below.

▶ **Definition 2.** Let $X_1, \ldots, X_n$ be real valued random variables. Then X_i's are *negatively correlated* if for each $i \neq j \in [n]$, we have

$$E[X_iX_j] \leq E[X_i]E[X_j]. \tag{2}$$

Another notion of negative dependence is the negative cylindrical property that is stronger than negative correlation but weaker than negative association. This property is also well studied since it is enough to imply tail bounds ala Chernoff bounds.

▶ **Definition 3.** Let $X_1, \ldots, X_n$ be real valued random variables. Then X_i's have the *negatively cylindrical property* if for each $S \subseteq [n]$, we have

$$E[\prod_{i \in S} X_i] \leq \prod_{i \in S} E[X_i]. \tag{3}$$

A set system $\mathcal{M} = (U, \mathcal{I})$ is called a matroid if $\mathcal{I} \subseteq 2^U$ satisfies two axioms (i) $A \in \mathcal{I}$ and $B \subseteq A$ implies that $B \in \mathcal{I}$, (ii) $A, B \in \mathcal{I}$ such that $|A| > |B|$ implies that there exists $a \in A \backslash B$ such that $B \cup \{a\} \in \mathcal{I}$. Sets of $\mathcal{I}$ are called *independent* sets and the maximal sets in $\mathcal{I}$ are called *bases* of matroid $\mathcal{M}$. For a matroid $\mathcal{M}$, the corresponding *matroid polytope* is the convex hull of the indicator vectors of all the bases of $\mathcal{M}$.

2 Our Algorithm and Result

2.1 Algorithm

We present a discrete implementation of the random walk algorithm. We first introduce some notation. Let

$$P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq b\}$$

be the inequality description of P. Often we refer to individual inequalities or equalities as $\mathbf{a}_j^\top \mathbf{x} \leq b_j$ or $\mathbf{a}_j^\top \mathbf{x} = b_j$. For any face F of P defined by

$$F = \{\mathbf{x} \in P : A^{=}\mathbf{x} = b^{=}\},$$

let $d(F)$ denote the dimension of F. Let $C(F)$ denote the matrix whose columns form an orthonormal basis of the subspace

$$\{\mathbf{x} \in \mathbb{R}^n : A^{=}\mathbf{x} = 0\}.$$

Observe that the dimension of $C(F)$ is $n \times d(F)$.

Algorithm 1: Algorithm Random Walk

Input: θ, error parameter ε. Let $T = \frac{n^2}{\varepsilon^2}$.
Initialization $\mathbf{x}_0 \leftarrow \theta$. Let $t \leftarrow 0$, $F_0 = P$.
while dimension of $F_t > 0$ or $t \leq T$ **do**
    while there exists j such that $0 < b_j - \mathbf{a}_j^\top \mathbf{x}_t < n\varepsilon$, **do**
        Let $\mathbf{y}$ denote the point in $F_t \cap \{\mathbf{x} \colon \mathbf{a}_j^\top \mathbf{x} = b_j\}$ closest to $\mathbf{x}_t$.
        Let $\mathbf{x}_t \leftarrow \mathbf{y}$, $F_t \leftarrow F_t \cap \{\mathbf{x} \colon \mathbf{a}_j^\top \mathbf{x} = b_j\}$.
    end while
    Let d denote the dimension of F_t and let $\mathbf{g}$ be a normal Gaussian in d dimensions.
    Let $\mathbf{x}_{t+1} = \mathbf{x}_t + \varepsilon \cdot C(F_t)\mathbf{g}$ and let $F_{t+1} \leftarrow F_t$. If $\mathbf{x}_{t+1} \notin P$, then abort.
    Let $t \leftarrow t + 1$.
end while
Return $\mathbf{x}_t$.

If the algorithm ends at time $t^* \leq T$ with the vertex $\mathbf{x}_{t^*}$, we let $\mathbf{x}_t = \mathbf{x}_{t^*}$ for each $t^* \leq t \leq T$.

2.2 Main Result

Our main result is to show that the random walk algorithm described above gives a distribution over vertices of the matroid polytope that is negatively associated.

▶ **Theorem 4.** *Given any matroid $\mathcal{M} = (U, \mathcal{I})$, let P denote the convex hull of bases of $\mathcal{M}$ and let $n = |U|$. Given any $\theta \in P$ and error parameter $\varepsilon > 0$, the random walk algorithm returns a vertex of P before time $T = \frac{n^2}{\varepsilon^2}$ with probability at least $1 - e^{-n}$. Moreover, conditioned on algorithm returning a vertex of P, the output random vertex $\mathbf{x}_T$ satisfies the following properties.*

1. *For each $i \in U$,*

$$\theta(i) - n^2\varepsilon \leq E[\mathbf{x}_T(i)] \leq \theta(i) + n^2\varepsilon.$$

2. *For every 1-Lipschitz non-decreasing functions $f_1 : \mathbb{R}^U \to [0,1]$ and $f_2 : \mathbb{R}^U \to [0,1]$ depending on disjoint set of coordinates, we have*

$$E[f_1(\mathbf{x}_T) f_2(\mathbf{x}_T)] \leq E[f_1(\mathbf{x}_T)]E[f_2(\mathbf{x}_T)] + n\sqrt{n}\varepsilon \tag{4}$$

Moreover, the expected running time of the algorithm is polynomial in n and $\frac{1}{\varepsilon}$.

To prove the theorem, we rely on two crucial observations which also illustrate the crucial role of matroid polytopes. We show that for *every* face F of P, a standard Gaussian projected on F is negatively correlated; see Theorem 5. This result relies crucially on the facial structure of the matroid polytopes and a characterization of every face in terms of tight constraints defining the matroid polytope. Secondly, we use the classical result that for Gaussian random variables, negative correlation implies negative association (see Theorem 12). The above results use an inductive argument to show that the random walk algorithm leads to a distribution that is negatively associated.

▶ Remark. We mention that we describe the discrete version due to algorithmic implications but it naturally leads to error terms in Theorem 4. From a structural point of view, one can construct a Brownian motion that sticks to the face of the polytope as does our random walk and ends at a vertex of the polytope. The resulting distribution over bases of the polytope will satisfy the marginals and the inequality for negative association exactly.

3 Projected Gaussian in Matroids

In this section, we prove that standard Gaussian projected on any face of the matroid polytope is negatively correlated.

▶ **Theorem 5.** *Let $\mathcal{M} = (U, \mathcal{I})$ denote a matroid and P denote the convex hull of all bases of $\mathcal{M}$. Let F be the face of P and $g \in R^U$ be a standard Gaussian random variable. Then for any distinct $i, j \in U$ we have $E[(Cg)_i(Cg)_j] \leq 0$ where C is the projection matrix projecting onto F.*

Proof. Before we prove the general case, we prove the case when $\mathcal{M}$ is a uniform matroid. The convex hull of the bases of this matroid is well understood and has well characterized faces. The general case, whose facial structure is more complicated, will use this case as a building block.

Let k be an integer and $\mathcal{I} = \{S \subseteq U : |S| \leq k\}$. In this case, we have

$$P = \{x \in \mathbb{R}^U : \sum_{i \in U} x_i = k, \;\; 0 \leq x_i \leq 1 \;\; \forall i \in U\}.$$

First consider the face which is P itself. We let n denote the size of $|U|$.

▶ **Lemma 6.** *Let C denote the $n \times (n-1)$ matrix whose columns form the orthonormal basis of subspace $\{x \in \mathbb{R}^U : \sum_{i \in U} x_i = 0\}$. We have $CC^\top = I_n - \frac{1}{n} J_n$ where I_n denotes the identity matrix in n dimensions and J_n denotes the n dimensional matrix with all ones.*

Proof. Let $\hat{C}$ be $n \times n$ formed by adding the column $\frac{1}{\sqrt{n}}(1, \ldots, 1)$ to C. Then all columns of $\hat{C}$ form an orthonormal basis of $\mathbb{R}^n$. Then $\hat{C}\hat{C}^\top = I_n$. But $\hat{C}\hat{C}^\top = CC^\top + \frac{1}{n} J_n$ giving us that $CC^\top = I_n - \frac{1}{n} J_n$. ◀

Since all off-diagonal entries of $I_n - \frac{1}{n} J_n$ are negative, this implies that the standard Gaussian projected on P is negatively correlated. Any other face F is obtained by setting some of the variables to 0 or 1 in P. Thus the covariance matrix of the face will be a block matrix of the form $I_d - \frac{1}{n} J_d$ with the zero matrix. This completes the proof for the case when $\mathcal{M}$ is a uniform matroid.

We now consider the general case. Let $\mathcal{M} = (U, \mathcal{I})$ denote a matroid and P denote the convex hull of all independent sets of $\mathcal{M}$. In the next lemma, we characterize the possible covariance matrices for projections of normal Gaussians on any of the faces of the matroid polytope. It shows that the covariance matrix is of block diagonal form where each block has the same structure as the covariance matrix for the uniform matroid as seen earlier.

▶ **Lemma 7.** *Let F be a face of P and X be a projection of a normal Gaussian on F. The covariance matrix of X (up to permuting indices of columns and rows) is $\Sigma = CC^\top$ where Σ is a block diagonal matrix with blocks of size $d_1, \ldots, d_r$ such that $\sum_i d_i = n$ and each block of size d_i is exactly $I_{d_i} - \frac{1}{d_i} J_{d_i}$ where I_{d_i} is the identity matrix and J_{d_i} is the all-ones matrix with sizes $d_i \times d_i$.*

Proof. The proof uses the characterization of any face of P by a maximal set of tight constraints which satisfy the chain structure. A collection of sets $A_1, A_2, \ldots, A_k \subseteq U$ form a *chain* if $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k$. The lemma is standard using the *uncrossing* technique (see Chapter 40 [18] or Lemma 5.2.4 [14]).

▶ **Claim 8.** *Given any face F of P, there exists a chain $S'_1 \subseteq S'_2 \subseteq \cdots \subseteq S'_r = U$ where $r = n - d$ and integers $k_j \geq 0$ for each $1 \leq j \leq r$ such that*

$$F = \{x \in P : \sum_{i \in S'_j} x_j = k_j \ \ \forall\, 1 \leq j \leq r\}.$$

Let $S_j = S'_j \setminus S'_{j-1}$ for each $1 \leq j \leq r$ where $S_0 = \emptyset$. Then the following two subspaces are equal

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}(S'_j) = 0 \ \ \forall\, 1 \leq j \leq r\} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}(S_j) = 0 \ \ \forall\, 1 \leq j \leq r\}.$$

Thus columns of C can be chosen to form an orthonormal basis of the subspace

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}(S_j) = 0 \ \ \forall\, 1 \leq j \leq r\}.$$

Since $\{S_j : 1 \leq j \leq r\}$ form a partition of U, we can choose columns of C to be divided into groups of size $|S_j| - 1$ for each $1 \leq j \leq r$ where the j^{th} group is supported on the elements of S_j. Let C_j denote the submatrix with these $|S_j| - 1$ columns. Thus, we can apply Lemma 6 and $C_j^\top C_j$ has the property that the non-zero entries form a block matrix of $I_{d_i} - \frac{1}{d_i} J_{d_i}$ on the elements corresponding to S_j. Since $C'_j s$ have disjoint support, we have $CC^\top = \sum_{i=1}^r C_j C_j^\top$ as required. ◀

This completes the proof of Theorem 5. ◀

4 Negative Dependence Properties

In this section, we show the negative dependence properties of the random walk algorithm in matroid polytopes.

First, we have the following simple claim.

▶ **Lemma 9.** *The algorithm ends in $T = \frac{n^2}{\varepsilon^2}$ iterations with probability at least $1 - e^{-n}$. Conditioned on terminating, if the algorithm terminates with solution x_T, then $x_T = \theta + \sum_{t=1}^T \varepsilon C(F_t) g_t + \sum_{i=1}^n \zeta_i$ where g_t are independent Gaussians and $|\zeta_i| \leq n\varepsilon$ for each i. Here $C(F_t)$ is the projection matrix at time t and is also a random variable.*

Proof. The probability that the algorithm aborts in a single iteration is bounded by the probability of the event that a Gaussian with covariance at most εI is outside the ball of radius εn. Standard concentration result that this probability is bounded by e^{-n^2}. Thus the probability of aborting in the first T iterations is at most $\frac{n^2}{\varepsilon^2} e^{-n^2} \leq \frac{1}{2} e^{-n}$ for large enough n

and $\varepsilon < \frac{1}{n^2}$. Let us condition on the event that the algorithm does not abort. Now we bound the number of iterations. In each iteration, the expected distance squared from θ increases by at least ε^2. This follows since

$$E[\|\mathbf{x}_{t+1} - \theta\|_2^2 | x_t] = \varepsilon^2 E[gC(F_t)C(F_t)^\top g] \geq \varepsilon^2$$

where we use the fact that $C(F_t)^\top g$ is projection of a standard Gaussian in at least dimension one. Taking expectation over x_t, we get the squared distance increases by at least ε. Since the distance is bounded by diameter of P which is $\sqrt{n}$. Thus, with probability at least $1 - \frac{1}{2}e^{-n}$, the algorithm terminates in $n^2\varepsilon^2$ iterations with a vertex of P conditioned on the event that it doesn't terminate. Thus, we get the first claim.

Now consider the case that the algorithm succeeds and returns x_T. At each iteration t, we add $\varepsilon C(F_t)g_t$ to the current vector except when we project on a face. Since the dimension of the face reduces by one, there are at most n projection steps. The error term introduced by the projection is denoted by ζ_i where $|\zeta_i| \leq n\varepsilon$ for each $1 \leq i \leq n$. thus we obtain that

$$\mathbf{x}_T = \theta + \sum_{t=1}^{T} \varepsilon C(F_t) g_t + \sum_{i=1}^{n} \zeta_i.$$ ◀

From now, we condition on the event that the algorithm does not abort and terminates with a vertex at iteration T. Thus all expectations stated from now on are conditioned on this event.

4.1 Negative Cylinder Property

We first show the negative cylinder property as defined in Definition 3. Recall that negative cylinder property is enough to obtain concentration results as given by Chernoff bounds.

▶ **Lemma 10.** *For any subset $R \subseteq U$,*

$$E[\prod_{i\in R} \mathbf{x}_T(i)] \leq \prod_{i\in R} E[\mathbf{x}_T(i)] + 2n^2\varepsilon = \prod_{i\in R} \theta(i) + 2n^2\varepsilon$$

Proof. We will show that $\prod_{i\in R} \mathbf{x}_t(i)$ forms a supermartingale (modulo the error terms). Since $\prod_{i\in R} \mathbf{x}_0(i) = \prod_{i\in R} \theta(i)$, we will have the claim. Let us verify the supermartingale property, by first ignoring the error term introduced due to fixing a constraint. Recall that

$$\mathbf{x}_{t+1} = \mathbf{x}_t(i) + \varepsilon C(F_t) g_t.$$

Let $h_t = C(F_t)g_t$. Consider any time $t+1$.

$$\begin{aligned}
E[\prod_{i\in R} \mathbf{x}_{t+1}(i)|\mathbf{x}_t] &= E[\prod_{i\in R} (\mathbf{x}_t(i) + \varepsilon h_t)\,|\mathbf{x}_t] \\
&= \prod_{i\in R} \mathbf{x}_t(i) + \varepsilon \sum_{j\in R} E[h_t(j)|\mathbf{x}_t] \cdot \left(\prod_{i\in R\setminus\{j\}} \mathbf{x}_t(i)\right) \\
&\quad + \varepsilon^2 \sum_{j,k\in R, j\neq k} E[h_t(j)h_t(k)|\mathbf{x}_t] \cdot \left(\prod_{i\in R\setminus\{j,k\}} \mathbf{x}_t(i)\right) \\
&\quad + \varepsilon^3 \sum_{j,k,l\in R, j\neq k\neq l} E[h_t(j)h_t(k)h_t(l)|\mathbf{x}_t] \cdot \left(\prod_{i\in R\setminus\{j,k,l\}} \mathbf{x}_t(i)\right) + \ldots \\
&\leq \prod_{i\in R} \mathbf{x}_t(i) + 2\varepsilon^3 n^3
\end{aligned}$$

where we use the fact that $E[h_t(j)|\mathbf{x}_t] = 0$ for each $j \in R$ and $\varepsilon^2 \sum_{j,k\in R, j\neq k} E[h_t(j)h_t(k)|\mathbf{x}_t] \leq 0$ for each $j \neq k \in R$ from Theorem 5. The later terms have increasing powers of ε and thus we can bound the error terms using the fact that $\varepsilon \leq \frac{1}{n^2}$.

Applying an inductive argument, and also incorporating the error term introduced due to fixing constraints, we obtain the lemma.

◀

4.2 Negative Association

We now prove the stronger negative dependence property of negative association. This relies on the following classical lemma about Gaussian distribution that says that pairwise negative correlation is enough to obtain negative association. The definition of negative association is extended to real random variables as well.

▶ **Definition 11** (Negative Association). Let $X_1, \ldots, X_n$ be real valued random variables. Then X_i's are *negatively associated* if for every non-decreasing functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ we have

$$E[f(X_1, \ldots, X_n)g(X_1, \ldots, X_n)] \leq E[f(X_1, \ldots, X_n)]E[g(X_1, \ldots, X_n)] \tag{5}$$

if f and g depend on disjoint set of coordinates.

Observe that the definitions are consistent and if a set of boolean random variables are negatively associated as per Definition 1 then they are also negatively associated as per Definition 11. This follows since every non-decreasing $f : \{0,1\}^n \to \mathbb{R}$ which depends on $S \subseteq [n]$ can be extended to a non-decreasing function $\hat{f} : \mathbb{R}^n \to \mathbb{R}$ which depends on S. Indeed, for any $x \in [0,1]^n$, let

$$\hat{f}(x) = \sum_{T\subseteq[n]} f(T) \prod_{i\in T} x_i \prod_{i\notin T}(1 - x_i)$$

and for any other x, let

$$\hat{f}(x) = f(x \wedge 1)$$

where $(x \wedge 1)_i = \min\{x_i, 1\}$. A straightforward check shows that $\hat{f}$ and f agree on the boolean hypercube. Moreover, $\hat{f}$ is non-decreasing and only depends on S.

▶ **Theorem 12.** *[13] Let $f_1 : \mathbb{R}^n \to \mathbb{R}$ and $f_2 : \mathbb{R}^n \to \mathbb{R}$ be non-decreasing functions depending on a disjoint set of coordinates. Let $X = (X_1, \ldots, X_n)$ be a Gaussian random variable which is negatively correlated. Then we have*

$$E[f_1(X)f_2(X)] \leq E[f_1(X)]E[f_2(X)].$$

We now prove the following theorem.

▶ **Lemma 13.** *For any non-decreasing functions $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ depending on disjoint set of coordinates that are 1-Lipschitz, we have for each $0 \leq t \leq T$,*

$$E[f_1(\mathbf{x}_t)f_2(\mathbf{x}_t)] \leq E[f_1(\mathbf{x}_t)]E[f_2(\mathbf{x}_t)] + 2n\sqrt{n}\varepsilon$$

Proof. We first ignore the errors introduced due to projection onto tight constraints and prove the inequality without the error term. We let $\mathbf{y}_t$ denote the corresponding fractional points. Thus $\mathbf{y}_t = \theta + \sum_{s=1}^{t-1} C(F_s)g_s$. We prove that

$$E[f_1(\mathbf{y}_t)f_2(\mathbf{y}_t)] \leq E[f_1(\mathbf{y}_t)]E[f_2(\mathbf{y}_t)]$$

by induction on t. For $t = 0$, $\mathbf{y}_0 = \theta$ and both sides are equal. Now, let the statement be true for $t \leq T-1$. We have $\mathbf{y}_{t+1} = \mathbf{y}_t + C_t g_t$ where we denote $C(F_t)$ by C_t. Thus we have

$$E[f_1(\mathbf{y}_{t+1})f_2(\mathbf{y}_{t+1})|\mathbf{y}_t] = E[f_1(\mathbf{y}_t + C_t g_t)f_2(\mathbf{y}_t + C_t g_t)|\mathbf{y}_t]$$

But now observe that $C_t g_t$ is a Gaussian in $\mathbb{R}^n$ which is negatively correlated from Theorem 5. For any $\mathbf{y}_t$, we apply Theorem 12 to the functions $f_1(\mathbf{y}_t + \cdot)$ and $f_2(\mathbf{y}_t + \cdot)$, we obtain

$$\begin{aligned} E[f_1(\mathbf{y}_{t+1})f_2(\mathbf{y}_{t+1})|\mathbf{y}_t] &= E[f_1(\mathbf{y}_t + C_t g_t)f_2(\mathbf{y}_t + C_t g_t)|\mathbf{y}_t] \\ &\leq E[f_1(\mathbf{y}_t + C_t g_t)|\mathbf{y}_t]E[f_2(\mathbf{y}_t + C_t g_t)|\mathbf{y}_t] \\ &= E[f_1(\mathbf{y}_{t+1})|\mathbf{y}_t]E[f_2(\mathbf{y}_{t+1})|\mathbf{y}_t]. \end{aligned}$$

Now taking expectations w.r.t. $\mathbf{y}_t$, we obtain that

$$E[f_1(\mathbf{y}_t)f_2(\mathbf{y}_t)] \leq E[f_1(\mathbf{y}_t)]E[f_2(\mathbf{y}_t)].$$

Now bound the distance between $\mathbf{x}_t$ and $\mathbf{y}_t$. We have

$$\|x_t - \mathbf{y}_t\|_2 = \|\sum_{i=1}^{t-1} \zeta_i\|_2 \leq \sqrt{n}\max_i \|\zeta_i\|_2 \leq n\sqrt{n}\varepsilon.$$

Since f_1 and f_2 are 1-Lipschitz, we have the result of the lemma. ◀

Acknowledgements. We thank Robin Pemantle for many discussions and comments.

References

1 Nima Anari, Shayan Oveis Gharan, and Alireza Rezaei. Monte carlo markov chain algorithms for sampling strongly rayleigh distributions and determinantal point processes. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, volume 49 of *JMLR Workshop and Conference Proceedings*, pages 103–115. JMLR.org, 2016. URL: `http://jmlr.org/proceedings/papers/v49/anari16.html`.

2 Arash Asadpour, Michel X. Goemans, Aleksander Madry, Shayan Oveis Gharan, and Amin Saberi. An o(log n/ log log n)-approximation algorithm for the asymmetric traveling salesman problem. In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 379–389. SIAM, 2010. `doi:10.1137/1.9781611973075.32`.

3 Julius Borcea, Petter Brändén, and Thomas Liggett. Negative dependence and the geometry of polynomials. *Journal of the American Mathematical Society*, 22(2):521–567, 2009.

4 Petter Brändén and Rafael S González D'León. On the half-plane property and the tutte group of a matroid. *Journal of Combinatorial Theory, Series B*, 100(5):485–492, 2010.

5 Petter Brändén and Johan Jonasson. Negative dependence in sampling. *Scandinavian Journal of Statistics*, 39(4):830–838, 2012.

6 Chandra Chekuri, Jan Vondrak, and Rico Zenklusen. Dependent randomized rounding via exchange properties of combinatorial structures. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 575–584. IEEE, 2010.

7 Chandra Chekuri, Jan Vondrák, and Rico Zenklusen. Multi-budgeted matchings and matroid intersection via dependent rounding. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1080–1097. SIAM, 2011.

8 Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

9 Devdatt P Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *BRICS Report Series*, 3(25), 1996.

10 Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 26–38. ACM, 1992.

11 Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, and Aravind Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM (JACM)*, 53(3):324–360, 2006.

12 Shayan Oveis Gharan, Amin Saberi, and Mohit Singh. A randomized rounding approach to the traveling salesman problem. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 550–559. IEEE, 2011.

13 Kumar Joag-Dev and Frank Proschan. Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295, 1983.

14 Lap Chi Lau, Ramamoorthi Ravi, and Mohit Singh. *Iterative methods in combinatorial optimization*, volume 46. Cambridge University Press, 2011.

15 Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. *SIAM J. Comput.*, 44(5):1573–1582, 2015. `doi:10.1137/130929400`.

16 Ronald F Patterson, Wendy D Smith, Robert L Taylor, and Abolghassem Bozorgnia. Limit theorems for negatively dependent random variables. *Nonlinear Analysis: Theory, Methods & Applications*, 47(2):1283–1295, 2001.

17 Robin Pemantle. Towards a theory of negative dependence. *Journal of Mathematical Physics*, 41(3):1371–1390, 2000.

18 Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2002.

19 Mohit Singh and Nisheeth K. Vishnoi. Entropy, optimization and counting. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 50–59. ACM, 2014. `doi:10.1145/2591796.2591803`.

20 Aravind Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 588–597. IEEE, 2001.

21 Ming Yuan, Chun Su, and Taizhong Hu. A central limit theorem for random fields of negatively associated processes. *Journal of Theoretical Probability*, 16(2):309–323, 2003.