The Annals of Applied Probability
2015, Vol. 25, No. 2, 465–476
DOI: 10.1214/13-AAP997

A POLYNOMIAL TIME APPROXIMATION SCHEME FOR COMPUTING THE SUPREMUM OF GAUSSIAN PROCESSES[1]

BY RAGHU MEKA

Microsoft Research

We give a polynomial time approximation scheme (PTAS) for computing the supremum of a Gaussian process. That is, given a finite set of vectors $V \subseteq \mathbb{R}^d$, we compute a $(1+\varepsilon)$-factor approximation to $\mathbb{E}_{X \leftarrow \mathcal{N}^d}[\sup_{v \in V} |\langle v, X\rangle|]$ deterministically in time $\text{poly}(d) \cdot |V|^{O_\varepsilon(1)}$. Previously, only a constant factor deterministic polynomial time approximation algorithm was known due to the work of Ding, Lee and Peres [*Ann. of Math.* (*2*) **175** (2012) 1409–1471]. This answers an open question of Lee (2010) and Ding [*Ann. Probab.* **42** (2014) 464–496].

The study of supremum of Gaussian processes is of considerable importance in probability with applications in functional analysis, convex geometry, and in light of the recent breakthrough work of Ding, Lee and Peres [*Ann. of Math.* (*2*) **175** (2012) 1409–1471], to random walks on finite graphs. As such our result could be of use elsewhere. In particular, combining with the work of Ding [*Ann. Probab.* **42** (2014) 464–496], our result yields a PTAS for computing the cover time of bounded-degree graphs. Previously, such algorithms were known only for trees.

Along the way, we also give an explicit oblivious estimator for seminorms in Gaussian space with optimal query complexity. Our algorithm and its analysis are elementary in nature, using two classical *comparison inequalities*, Slepian's lemma and Kanter's lemma.

1. Introduction. The study of supremum of Gaussian processes is a major area of study in probability and functional analysis as epitomized by the celebrated *majorizing measures* theorem of Fernique and Talagrand; see Ledoux and Talagrand (1991), Talagrand (2005) and references therein. There is by now a rich body of work on obtaining tight estimates and characterizations of the supremum of Gaussian processes with several applications in analysis Talagrand (2005), convex geometry Pisier (1999) and more. Recently, in a striking result, Ding, Lee and Peres (2012) used the theory to resolve the *blanket time* conjectures of Winkler and Zuckerman (1996).

Ding, Lee and Peres (2012) used the powerful *Dynkin isomorphism theory* and majorizing measures theory to establish a structural connection between the cover time (and blanket time) of a graph G and the supremum of a Gaussian process

Received July 2013.

[1] A preliminary version of this paper appeared in the conference Foundations of Computer Science, 2012.

MSC2010 subject classifications. Primary 60C05; secondary 68Q87.

Key words and phrases. Gaussian processes, derandomization, cover time, random walks, ε-nets.

associated with the Gaussian Free Field on G. They then use this connection to resolve the Winkler–Zuckerman blanket time conjectures and to obtain the first deterministic polynomial time constant factor approximation algorithm for computing the cover time of graphs. This latter result resolves an old open question of Aldous and Fill (1994).

Besides showing the relevance of the study of Gaussian processes to discrete combinatorial questions, the work of Ding, Lee and Peres gives evidence that studying Gaussian processes could even be an important algorithmic tool; a less investigated aspect in the rich literature on Gaussian processes in probability and functional analysis. Here we address the corresponding computational question directly, which given the importance of Gaussian processes in probability, could be of use elsewhere. In this context, the following question was asked by Lee (2010) and Ding (2014).[2]

QUESTION 1.1. For every $\varepsilon > 0$, is there a deterministic polynomial time algorithm that, given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$, computes a $(1+\varepsilon)$-factor approximation to $\mathbb{E}_{X \leftarrow \mathcal{N}^d}[\sup_i |\langle v_i, X\rangle|]$.[3]

There is a simple randomized algorithm for the problem: sample a few Gaussian vectors and output the median supremum value for the sampled vectors. This, however, requires $O(d \log d/\varepsilon^2)$ random bits. Using Talagrand's majorizing measures theorem, Ding, Lee and Peres give a deterministic polynomial time $O(1)$-factor approximation algorithm for the problem. This approach is inherently limited to not yield a PTAS as the majorizing measures characterization is bound to lose a universal constant factor. Here we give a PTAS for the problem thus resolving the above question.

THEOREM 1.2. *For every $\varepsilon > 0$, there is a deterministic algorithm that, given a set of vectors $v_1, \ldots, v_m \in \mathbb{R}^d$, computes a $(1+\varepsilon)$-factor approximation to $\mathbb{E}_{x \leftarrow \mathcal{N}^d}[\sup_i |\langle v_i, x\rangle|]$ in time $\mathrm{poly}(d) \cdot m^{\tilde{O}(1/\varepsilon^2)}$.*

Our approach uses some classical *comparison inequalities* in convex geometry. To the best of our knowledge these inequalities have not been used before in the context of algorithm design.

[2]We remark that Lee (2010) and Ding (2014) actually ask for an approximation to $\mathbb{E}_{X \leftarrow \mathcal{N}^d}[\sup_i \langle v_i, X\rangle]$. However, this formulation results in a somewhat artificial asymmetry, and for most interesting cases these two are essentially equivalent: if $\mathbb{E}_{X \leftarrow \mathcal{N}^d}[\sup_i \langle v_i, X\rangle] = \omega(\max_i \|v_i\|_2)$, then $\mathbb{E}_{X \leftarrow \mathcal{N}^d}[\sup_I |\langle v_i, X\rangle|] = (1 + o(1))\,\mathbb{E}_{X \leftarrow \mathcal{N}^d}[\sup_i \langle v_i, X\rangle]$. We shall overlook this distinction from now on.

[3]Throughout, $\mathcal{N}$ denotes the univariate Gaussian distribution with mean 0 and variance 1, and for a distribution $\mathcal{D}$, $X \leftarrow \mathcal{D}$ denotes a random variable with distribution $\mathcal{D}$. By a α-factor approximation to a quantity $Z > 0$, we mean a number p such that $p \le Z \le \alpha p$.

We explain our result on estimating semi-norms with respect to Gaussian measures mentioned in the abstract in Section 2.2.

We next discuss some applications of our result to computing cover times of graphs as implied by the works of Ding, Lee and Peres (2012) and Ding (2014).

1.1. *Application to computing cover times of graphs.* The study of random walks on graphs is an important area of research in probability, algorithm design, statistical physics and more. As this is not the main topic of our work, we avoid giving formal definitions and refer the readers to Aldous and Fill (1994), Lovász (1993) for background information.

Given a graph G on n-vertices, the cover time, $\tau_{\mathrm{cov}}(G)$, of G is defined as the expected time a random walk on G takes to visit all the vertices in G when starting from the worst possible vertex in G. Cover time is a fundamental parameter of graphs and is extensively studied. Algorithmically, there is a simple randomized algorithm for approximating the cover time—simulate a few trials of the random walk on G for $\mathrm{poly}(n)$ steps, and output the median cover time. However, without randomness the problem becomes significantly harder. This was one of the motivations of the work of Ding, Lee and Peres (2012) who gave the first deterministic constant factor approximation algorithm for the problem, improving on an earlier work of Kahn et al. (2000) who obtained a deterministic $O((\log\log n)^2)$-factor approximation algorithm. For the special case of trees, Feige and Zeitouni (2009) gave a FPTAS.

Ding, Lee and Peres also conjectured that the cover time of a graph G (satisfying a certain reasonable technical condition) is asymptotically equivalent to the supremum of an explicitly defined Gaussian process, the Gaussian Free Field on G. However, this conjecture though quite interesting on its own, is not enough to give a PTAS for cover time; one still needs a PTAS for computing the supremum of the relevant Gaussian process. Our main result provides this missing piece, thus removing one of the obstacles in their posited strategy to obtain a PTAS for computing the cover time of graphs. Recently, Ding (2014) showed the main conjecture of Ding, Lee and Peres to be true for bounded-degree graphs and trees. Thus, combining his result [see Theorem 1.1 in Ding (2014)] with Theorem 1.2, we get a PTAS for computing cover time on bounded-degree graphs with $\tau_{\mathrm{hit}}(G) = o(\tau_{\mathrm{cov}}(G))$.[4] As mentioned earlier, previously, such algorithms were only known for trees; see Feige and Zeitouni (2009).

2. Outline of algorithm. The high level idea of our PTAS is as follows. Fix the set of vectors $V = \{v_1, \ldots, v_m\} \subseteq \mathbb{R}^d$ and $\varepsilon > 0$. Without loss of generality suppose that $\max_{v \in V} \|v\|_2 = 1$. We first reduce the dimension of V by

[4]The hitting time $\tau_{\mathrm{hit}}(G)$ is defined as the maximum over all pairs of vertices $u, v \in G$ of the expected time for a random walk starting at u to reach v. See the discussion in Ding (2014) for why this is a reasonable condition.

projecting V onto a space of dimension of $O((\log m)/\varepsilon^2)$ á la the classical Johnson–Lindenstrauss lemma (JLL). We then give an algorithm that runs in time polynomial in the number of vectors but exponential in the underlying dimension. Our analysis relies on two comparison inequalities, Fernique–Slepian lemma [Slepian (1962)] for the first step and Kanter's lemma [Kanter (1977)] for the second step. We discuss these modular steps below.

2.1. *Dimension reduction.* We project the set of vectors $V \subseteq \mathbb{R}^d$ to $\mathbb{R}^k$ for $k = O((\log m)/\varepsilon^2)$ to preserve all pairwise (Euclidean) distances within a $(1+\varepsilon)$-factor as in the Johnson–Lindenstrauss lemma (JLL). We then show that the expected supremum of the *projected* Gaussian process is within a $(1+\varepsilon)$ factor of the original value. The intuition is that the supremum of a Gaussian process, though a global property, can be controlled by pairwise correlations between the variables. To quantify this, we use Slepian's lemma, that helps us relate the supremum of two Gaussian processes by comparing pairwise correlations. Finally, observe that using known derandomizations of JLL, the dimension reduction can be done deterministically in time $\text{poly}(d, m, 1/\varepsilon)$; see Engebretsen, Indyk and O'Donnell (2002), Sivakumar (2002).

Thus, to obtain a PTAS it would be enough to have a deterministic algorithm to approximate the supremum of a Gaussian process in time exponential in the dimension $k = O((\log m)/\varepsilon^2)$. Unfortunately, a naive argument by discretizing the Gaussian measure in $\mathbb{R}^k$ leads to a run-time of at least $k^{O(k)}$; which gives a $m^{O((\log\log m)/\varepsilon^2)}$ algorithm. This question was recently addressed by Dadush and Vempala (2012), who needed a similar sub-routine for their work on computing *M-Ellipsoids* of convex sets and give a deterministic algorithm with a run-time of $(\log k)^{O(k)}$. We resolve this question fully by giving an optimal oblivious estimator for norms in Gaussian space, which when combined with the dimension reduction step gives a PTAS for computing the supremum.

2.2. *Oblivious estimators for semi-norms.* Let $\varphi : \mathbb{R}^k \to \mathbb{R}_+$ be a semi-norm, that is, φ is homogeneous and satisfies triangle inequality. For normalization purposes, we assume that $1 \leq \mathbb{E}_{x \leftarrow \mathcal{N}^k}[\varphi(x)]$ and that the Lipschitz constant of φ is at most $k^{O(1)}$. This is satisfied in most reasonable cases. Note that the supremum function $\varphi_V(x) = \sup_{v \in V} |\langle v, x\rangle|$ satisfies these conditions. Our goal will be to compute a $(1+\varepsilon)$-factor approximation to $\mathbb{E}_{x \leftarrow \mathcal{N}^k}[\varphi(x)]$ in time $2^{O_\varepsilon(k)}$.

THEOREM 2.1. *For every $\varepsilon > 0$, there exists a deterministic algorithm running in time $(1/\varepsilon)^{O(k)}$ and space $\text{poly}(k, 1/\varepsilon)$ that computes a $(1+\varepsilon)$-factor approximation to $\mathbb{E}_{X \leftarrow \mathcal{N}^k}[\varphi(X)]$ using only oracle access to φ.*

Our algorithm has the additional property of being an *oblivious linear estimator*: the set of query points does not depend on φ, and the output is a positive

weighted sum of the evaluations of φ on the query points. Further, the construction is essentially optimal as any such oblivious estimator needs to make at least $(1/\varepsilon)^{\Omega(k)}$ queries; see Section 7. In comparison, the previous best bound of Dadush and Vempala [Dadush and Vempala (2012)] needed $((\log k)/\varepsilon)^{O(k)}$ queries.

A natural first approach to compute $\mathbb{E}_{X \leftarrow \mathcal{N}^k}[\varphi(X)]$, would be to first discretize the one-dimensional Gaussian distribution with a constant granularity $\delta = f(\varepsilon)$ to get a distribution μ and then evaluate the expectation with respect to the product distribution μ^k. We will show that this seemingly naive approach in fact does very well, giving an error bound that does not depend on the dimension k. We do so by using a classical comparison inequality—Kanter's lemma—that allows us to "lift" a simple estimator for the univariate case to the multi-dimensional case.

More concretely, we first construct a symmetric distribution μ on $\mathbb{R}$ that has a simple *piecewise flat graph* and *sandwiches* the one-dimensional Gaussian distribution in the following sense. Let ν be a "shrinking" of μ defined to be the probability density function (p.d.f.) of $(1-\varepsilon)x$ for $x \leftarrow \mu$. We show that if μ has *granularity* about $\varepsilon^{3/2}$, then, for every symmetric interval $I \subseteq \mathbb{R}$, $\mu(I) \le \mathcal{N}(I) \le \nu(I)$.

Kanter's lemma [Kanter (1977)] then says that for p.d.f.'s μ, ν as above that are in addition *unimodal*, the above relation carries over to the product distributions μ^k, ν^k: for every symmetric convex set $K \subseteq \mathbb{R}^k$, $\mu^k(K) \le \mathcal{N}^k(K) \le \nu^k(K)$. This last inequality immediately implies that semi-norms cannot *distinguish* between μ^k and $\mathcal{N}^k$: for any semi-norm φ, $\mathbb{E}_{\mu^k}[\varphi(x)] = (1 \pm \varepsilon)\,\mathbb{E}_{\mathcal{N}^k}[\varphi(x)]$. We then suitably prune the distribution μ^k to have small support and prove Theorem 4.1.

Our main result, Theorem 1.2, follows by first reducing the dimension as in the previous section and applying Theorem 4.1 to the semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}_+$, $\varphi(x) = \sup_i |\langle u_i, x\rangle|$ for the projected vectors $\{u_1, \ldots, u_m\}$.

3. Dimension reduction. The use of JLL type random projections for estimating the supremum comes from the following comparison inequality for Gaussian processes. We call a collection of real-valued random variables $\{X_t\}_{t\in T}$ a Gaussian process if every finite linear combination of the variables has a normal distribution with mean zero. We refer the reader to Corollary 3.14 and the following discussion in Ledoux and Talagrand (1991) for reference.

THEOREM 3.1 (Fernique–Slepian lemma). *Let $\{X_t\}_{t\in T}$ and $\{Y_t\}_{t\in T}$ be two Gaussian processes such that for every $s, t \in T$, $\mathbb{E}[(X_s - X_t)^2] \le \mathbb{E}[(Y_s - Y_t)^2]$. Then, $\mathbb{E}[\sup_t |X_t|] \le \mathbb{E}[\sup_t |Y_t|]$.*

We also need a derandomized version of the Johnson–Lindenstrauss lemma.

THEOREM 3.2 [Engebretsen, Indyk and O'Donnell (2002)]. *For every $\varepsilon > 0$, there exists a deterministic $(dm^2(\log m + 1/\varepsilon)^{O(1)})$-time algorithm that given*

vectors $v_1, \dots, v_m \in \mathbb{R}^d$ *computes a linear mapping* $A : \mathbb{R}^d \to \mathbb{R}^k$ *for* $k = O((\log m)/\varepsilon^2)$ *such that for every* $i, j \in [m]$, $\|v_i - v_j\|_2 \le \|A(v_i) - A(v_j)\|_2 \le (1+\varepsilon)\|v_i - v_j\|_2$.

Combining the above two theorems immediately implies the following.

LEMMA 3.3. *For every* $\varepsilon > 0$, *there exists a deterministic* $(dm^2(\log m + 1/\varepsilon)^{O(1)})$*-time algorithm that given vectors* $v_1, \dots, v_m \in \mathbb{R}^d$ *computes a linear mapping* $A : \mathbb{R}^d \to \mathbb{R}^k$ *for* $k = O((\log m)/\varepsilon^2)$ *such that*

$$\begin{aligned} \mathop{\mathbb{E}}_{x \leftarrow \mathcal{N}^d}\Big[\sup_i |\langle v_i, x\rangle|\Big] &\le \mathop{\mathbb{E}}_{y \leftarrow \mathcal{N}^k}\Big[\sup_i |\langle A, (v_i), y\rangle|\Big] \\ &\le (1+\varepsilon) \mathop{\mathbb{E}}_{x \leftarrow \mathcal{N}^d}\Big[\sup_i |\langle v_i, x\rangle|\Big]. \end{aligned} \tag{3.1}$$

PROOF. Let $V = \{v_1, \dots, v_m\} \cup \{-v_1, \dots, -v_m\}$, and let $\{X_v\}_{v \in V}$ be the Gaussian process where the joint distribution is given by $X_v \equiv \langle v, x\rangle$ for $x \leftarrow \mathcal{N}^d$. Then $\mathbb{E}_{x \leftarrow \mathcal{N}^d}[\sup_i |\langle v_i, x\rangle|] = \mathbb{E}[\sup_v X_v]$.

Let $A : \mathbb{R}^d \to \mathbb{R}^k$ be the linear mapping as given by Theorem 3.2 applied to V. Let $\{Y_v\}_{v \in V}$ be the "projected" Gaussian process with joint distribution given by $Y_v \equiv \langle A, (v), y\rangle$ for $y \leftarrow \mathcal{N}^k$. Then $\mathbb{E}_{y \leftarrow \mathcal{N}^k}[\sup_i |\langle v_i, y\rangle|] = \mathbb{E}[\sup_v Y_v]$.

Finally, observe that for any $u, v \in V$,

$$\begin{aligned} \mathbb{E}\big[(X_u - X_v)^2\big] &= \|u - v\|_2^2 \le \big\|A(u) - A(v)\big\|_2^2 \\ &= \mathbb{E}\big[(Y_u - Y_v)^2\big] \le (1+\varepsilon)^2 \mathbb{E}\big[(X_u - X_v)^2\big]. \end{aligned}$$

Combining the above inequality with Lemma 3.1 applied to the pairs of processes $(\{X_v\}_{v \in V}, \{Y_v\}_{v \in V})$ and $(\{Y_v\}_{v \in V}, \{(1+\varepsilon)X_v\}_{v \in V})$ it follows that

$$\mathbb{E}\Big[\sup_v |X_v|\Big] \le \mathbb{E}\Big[\sup_v |Y_v|\Big] \le \mathbb{E}\Big[\sup_v (1+\varepsilon)|X_v|\Big] = (1+\varepsilon)\mathbb{E}\Big[\sup_v |X_v|\Big].$$

The lemma now follows. □

4. Oblivious estimators for semi-norms in Gaussian space. In the previous section we reduced the problem of computing the supremum of a d-dimensional Gaussian process to that of a Gaussian process in $k = O((\log m)/\varepsilon^2)$-dimensions. Thus it suffices to have an algorithm for approximating the supremum of Gaussian processes in time exponential in the dimension. We will give such an algorithm that works more generally for all semi-norms.

Let $\varphi : \mathbb{R}^k \to \mathbb{R}_+$ be a semi-norm. That is, φ satisfies the triangle inequality and is homogeneous. For normalization purposes we assume that $1 \le \mathbb{E}_{\mathcal{N}^k}[\varphi(X)]$ and the Lipschitz constant of φ is at most $k^{O(1)}$.

THEOREM 4.1. *For every $\varepsilon > 0$, there exists a set $S \subseteq \mathbb{R}^k$ with $|S| = (1/\varepsilon)^{O(k)}$ and a function $p : \mathbb{R}^k \to \mathbb{R}_+$ computable in* $\mathrm{poly}(k, 1/\varepsilon)$ *time such that the following holds. For every semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}_+$,*

$$(1-\varepsilon)\Big(\sum_{x\in S} p(x)\varphi(x)\Big) \le \mathop{\mathbb{E}}_{X\leftarrow\mathcal{N}^k}\big[\varphi(X)\big] \le (1+\varepsilon)\Big(\sum_{x\in S} p(x)\varphi(x)\Big).$$

Moreover, successive elements of S can be enumerated in $\mathrm{poly}(k, 1/\varepsilon)$ *time and $O(k\log(1/\varepsilon))$ space.*

Theorem 2.1 follows immediately from the above.

PROOF OF THEOREM 2.1. Follows by enumerating over the set S and computing $\sum_{x\in S} p(x)\varphi(x)$ by querying φ on the points in S. □

We now prove Theorem 4.1. Here and henceforth, let γ denote the p.d.f. of the standard univariate Gaussian distribution. Fix $\varepsilon > 0$, and let $\delta > 0$ be a parameter to be chosen later. Let $\mu \equiv \mu_\delta$ be the p.d.f. which is a piecewise-flat approximator to γ obtained by spreading the mass γ gives to an interval $I = [i\delta, (i+1)\delta)$ evenly over I. Formally, $\mu(z) = \mu(-z)$ and for $z > 0$, $z \in [i\delta, (i+1)\delta)$,

$$\mu(z) = \frac{\gamma([i\delta, (i+1)\delta))}{\delta}. \tag{4.1}$$

Clearly, μ defines a symmetric distribution on $\mathbb{R}$. We will show that for $\delta \ll \varepsilon$ sufficiently small, semi-norms cannot *distinguish* the product distribution μ^k from $\mathcal{N}^k$:

LEMMA 4.2. *Let $\delta = (2\varepsilon)^{3/2}$. Then, for every semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}$,*

$$(1-\varepsilon)\mathop{\mathbb{E}}_{X\leftarrow\mu^k}\big[\varphi(X)\big] \le \mathop{\mathbb{E}}_{Z\leftarrow\mathcal{N}^k}\big[\varphi(Z)\big] \le \mathop{\mathbb{E}}_{X\leftarrow\mu^k}\big[\varphi(X)\big].$$

We first prove Theorem 4.1 assuming the above lemma, whose proof is deferred to the next section.

PROOF OF THEOREM 4.1. Let $\hat{\mu}$ be the symmetric distribution supported on $\delta(\mathbb{Z} + 1/2)$ with p.d.f. defined by

$$\hat{\mu}\big(\delta(i+1/2)\big) = \mu\big([i\delta, (i+1)\delta)\big)$$

for $i \ge 0$. Further, let $X \leftarrow \mu^k$, $\widehat{X} \leftarrow \hat{\mu}^k$, $Z \leftarrow \mathcal{N}^k$.

We claim that $\mathbb{E}[\varphi(\widehat{X})] = (1 \pm \varepsilon)\mathbb{E}[\varphi(Z)]$. Let Y be uniformly distributed on $[-\delta, \delta]^k$ and observe that random variable $X \equiv \widehat{X} + Y$ in law. Therefore,

$$\begin{aligned}
\mathbb{E}\big[\varphi(X)\big] &= \mathbb{E}\big[\varphi(\widehat{X}+Y)\big] = \mathbb{E}\big[\varphi(\widehat{X})\big] \pm \mathbb{E}\big[\varphi(Y)\big] \\
&= \mathbb{E}\big[\varphi(\widehat{X})\big] \pm \delta\mathbb{E}\big[\varphi(Y/\delta)\big] \\
&= \mathbb{E}\big[\varphi(\widehat{X})\big] \pm \delta \mathop{\mathbb{E}}_{Z'\in_u[-1,1]^k}\big[\varphi(Z')\big] \\
&= \mathbb{E}\big[\varphi(\widehat{X})\big] \pm \delta\mathbb{E}\big[\varphi(Z)\big] \qquad \text{(Lemma 5.7)}.
\end{aligned} \tag{4.2}$$

Thus, by Lemma 4.2,

$$\mathbb{E}[\varphi(\widehat{X})] = (1 \pm O(\varepsilon))\mathbb{E}[\varphi(Z)]. \tag{4.3}$$

We next prune $\hat{\mu}^k$ to reduce its support. Define $p : \mathbb{R}^k \to \mathbb{R}_+$ by $p(x) = \hat{\mu}^k(x)$. Clearly, $p(x)$ being a product distribution is computable in $\text{poly}(k, 1/\varepsilon)$ time.

Let $S = (\delta(\mathbb{Z} + 1/2))^k \cap B_2(3\sqrt{k})$, where $B_2(r) \subseteq \mathbb{R}^k$ denotes the Euclidean ball of radius r. As φ has Lipschitz constant bounded by $k^{O(1)}$, a simple calculation shows that throwing away all points in the support of $\widehat{X}$ outside S does not change $\mathbb{E}[\varphi(\widehat{X})]$ much. It is easy to check that for $x \notin S$, $p(x) \le \exp(-\|x\|_2^2/4)/(2\pi)^{k/2}$. Therefore,

$$\begin{aligned}
\mathbb{E}[\varphi(\widehat{X})] &= \sum_x p(x)\varphi(x) = \sum_{x\in S} p(x)\varphi(x) + \sum_{x\notin S} p(x)\varphi(x) \\
&= \sum_{x\in S} p(x)\varphi(x) \pm \sum_{x\notin S} \frac{\exp(-\|x\|_2^2/4)}{(2\pi)^{k/2}} \cdot \left(k^{O(1)}\|x\|_2\right) \qquad (4.4)\\
&= \sum_{x\in S} p(x)\varphi(x) \pm o(1).
\end{aligned}$$

From equation (4.3) and the above equation, we get (recall that $\mathbb{E}[\varphi(Z)] \ge 1$)

$$\mathbb{E}[\varphi(Z)] = (1 \pm O(\varepsilon))\left(\sum_{x\in S} p(x)\varphi(x)\right),$$

which is what we want to show.

We now reason about the complexity of S. First, by a simple covering argument $|S| < (1/\delta)^{O(k)}$,

$$|S| < \frac{\text{Vol}(B_2(3\sqrt{k}) + [-\delta, \delta]^k)}{\text{Vol}([-\delta, \delta]^k)} = (1/\delta)^{O(k)} = (1/\varepsilon)^{O(k)},$$

where for sets $A, B \subseteq \mathbb{R}^k$, $A + B$ denotes the Minkowski sum, and Vol denotes Lebesgue volume. This size bound almost suffices to prove Theorem 4.1 except for the complexity of enumerating elements from S. Without loss of generality assume that $R = 3\sqrt{n}/\delta$ is an integer. Then, enumerating elements in S is equivalent to enumerating integer points in the n-dimensional ball of radius R. This can be accomplished by going through the set of lattice points in the natural lexicographic order, and takes $\text{poly}(k, 1/\varepsilon)$ time and $O(k \log(1/\varepsilon))$ space per point in S. □

5. Proof of Lemma 4.2. Our starting point is the following definition that helps us *compare* multivariate distributions when we are only interested in volumes of convex sets. We shall follow the notation of Ball (2001).

DEFINITION 5.1. Given two symmetric p.d.f.'s, f, g on $\mathbb{R}^k$, we say that f is less peaked than g ($f \preceq g$) if for every symmetric convex set $K \subseteq \mathbb{R}^k$, $f(K) \le g(K)$.

We also need the following elementary facts. The first follows from the unimodality of the Gaussian density and the second from partial integration.

FACT 5.2. *For any $\delta > 0$ and μ as defined by equation (4.1), μ is less peaked than γ.*

FACT 5.3. *Let f, g be distributions on $\mathbb{R}^k$ with $f \preceq g$. Then for any semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}$, $\mathbb{E}_f[\varphi(x)] \geq \mathbb{E}_g[\varphi(x)]$.*

PROOF. Observe that for any $t > 0$, $\{x : \varphi(x) \leq t\}$ is convex. Let random variables $X \leftarrow f$, $Y \leftarrow g$. Then, by partial integration, $\mathbb{E}[\varphi(X)] = \int_0^\infty \varphi'(t) \times \Pr[\varphi(X) > t]\,dt \geq \int_0^\infty \varphi'(t) \Pr[\varphi(Y) > t]\,dt = \mathbb{E}[\varphi(Y)]$. □

The above statements give us a way to compare the expectations of μ and γ for one-dimensional convex functions. We would now like to do a similar comparison for the product distributions μ^k and γ^k. For this we use Kanter's lemma [Kanter (1977)], which says that the relation $\preceq$ is preserved under tensoring if the individual distributions have the additional property of being *unimodal*.

DEFINITION 5.4. A distribution f on $\mathbb{R}^n$ is unimodal if f can be written as an increasing limit of a sequence of distributions each of which is a finite positively weighted sum of uniform distributions on symmetric convex sets.

THEOREM 5.5 (Kanter's lemma [Kanter (1977)]; cf. Ball (2001)). *Let μ_1, μ_2 be symmetric distributions on $\mathbb{R}^n$ with $\mu_1 \preceq \mu_2$ and let ν be a unimodal distribution on $\mathbb{R}^m$. Then, the product distributions $\mu_1 \times \nu$, $\mu_2 \times \nu$ on $\mathbb{R}^n \times \mathbb{R}^m$ satisfy $\mu_1 \times \nu \preceq \mu_2 \times \nu$.*

We next show that μ "sandwiches" γ in the following sense.

LEMMA 5.6. *Let ν be the p.d.f. of the random variable $y = (1 - \varepsilon)x$ for $x \leftarrow \mu$. Then, for $\delta \leq (2\varepsilon)^{3/2}$, $\mu \preceq \gamma \preceq \nu$.*

PROOF. As mentioned above, $\mu \preceq \gamma$. We next show that $\gamma \preceq \nu$. Intuitively, ν is obtained by spreading the mass that γ puts on an interval $I = [i\delta, (i+1)\delta)$ evenly on the *smaller* interval $(1 - \varepsilon)I$. The net effect of this operation is to push the p.d.f. of μ closer to the origin and for δ sufficiently small the inward push from this "shrinking" wins over the outward push from going to μ.

Fix an interval $I = [-i\delta(1-\varepsilon) - \theta, i\delta(1-\varepsilon) + \theta]$ for $0 \leq \theta < \delta(1-\varepsilon)$. Then

$$\nu(I) = \nu\big([-i\delta(1-\varepsilon), i\delta(1-\varepsilon)]\big) + 2\nu\big([i\delta(1-\varepsilon), i\delta(1-\varepsilon) + \theta]\big) \tag{5.1}$$

$$= \gamma\big([-i\delta, i\delta]\big) + \frac{2\theta \cdot \gamma([i\delta, (i+1)\delta))}{\delta(1-\varepsilon)}. \tag{5.2}$$

We now consider two cases.

Case 1: $i \geq (1-\varepsilon)/\varepsilon$ so that $i\delta(1-\varepsilon)+\theta \leq i\delta$. Then, from the above equation,

$$\nu(I) \geq \gamma\big([-i\delta, i\delta]\big) \geq \gamma\big([-i\delta(1-\varepsilon)-\theta, i\delta(1-\varepsilon)+\theta]\big) = \gamma(I).$$

Case 2: $i < (1-\varepsilon)/\varepsilon$. Let $\alpha = (i+1)\delta = \delta/\varepsilon$. Then, as $1-x^2/2 \leq e^{-x^2/2} \leq 1$,

$$\gamma\big((i\delta, i\delta+\theta]\big) \leq \theta\cdot\gamma(0), \gamma\big([i\delta, (i+1)\delta)\big) \geq \delta\cdot\gamma(0)\cdot\big(1-\alpha^2/2\big).$$

Therefore,

$$\begin{aligned}
\nu(I) &= \gamma(I) - 2\gamma\big((i\delta, i\delta(1-\varepsilon)+\theta]\big) + \frac{2\theta\cdot\gamma([i\delta,(i+1)\delta))}{\delta(1-\varepsilon)} \\
&\geq \gamma(I) - 2\gamma\big((i\delta, i\delta+\theta]\big) + \frac{2\theta\cdot\gamma([i\delta,(i+1)\delta))}{\delta(1-\varepsilon)} \\
&\geq \gamma(I) - 2\theta\gamma(0) + \frac{2\theta\cdot\delta\cdot\gamma(0)\cdot(1-\alpha^2/2)}{\delta(1-\varepsilon)} \\
&= \gamma(I) + \frac{2\theta\gamma(0)}{1-\varepsilon}\cdot\big(\varepsilon-\alpha^2/2\big) \geq \gamma(I),
\end{aligned}$$

for $\alpha^2 \leq 2\varepsilon$, that is, if $\delta \leq (2\varepsilon)^{3/2}$. □

Lemma 4.2 follows easily from the above two claims.

PROOF OF LEMMA 4.2. Clearly, μ, ν, γ are unimodal and product of unimodal distributions is unimodal. Thus, from the above lemma and iteratively applying Kanter's lemma we get $\mu^k \preceq \gamma^k \preceq \nu^k$. Therefore, by Fact 5.3, for any semi-norm φ,

$$\mathop{\mathbb{E}}_{\mu^k}\big[\varphi(X)\big] \geq \mathop{\mathbb{E}}_{\gamma^k}\big[\varphi(Y)\big] \geq \mathop{\mathbb{E}}_{\nu^k}\big[\varphi(X)\big] = \mathop{\mathbb{E}}_{\mu^k}\big[\varphi\big((1-\varepsilon)X\big)\big] = (1-\varepsilon)\mathop{\mathbb{E}}_{\mu^k}\big[\varphi(X)\big]. \qquad \square$$

We now prove the auxiliary lemma we used in proof of Theorem 4.1.

LEMMA 5.7. *Let ρ be the uniform distribution on $[-1,1]$. Then, $\gamma \preceq \rho$ and for any semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}$, $\mathbb{E}_{\rho^k}[\varphi(x)] \leq \mathbb{E}_{\gamma^k}[\varphi(x)]$.*

PROOF. It is easy to check that $\gamma \preceq \rho$. Then, by Kanter's lemma $\gamma^k \preceq \rho^k$ and the inequality follows from Fact 5.3. □

6. A PTAS for supremum of Gaussian processes. Our main theorem, Theorem 1.2, follows immediately from Lemma 3.3 and Theorem 2.1 applied to the semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}$, defined by $\varphi(x) = \sup_{i\leq m} |\langle A, (v_i), x\rangle|$.

7. Lower bound for oblivious estimators. We now show that Theorem 4.1 is optimal: any oblivious linear estimator for semi-norms as in the theorem must make at least $(C/\varepsilon)^k$ queries for some constant $C > 0$.

Let $S \subseteq \mathbb{R}^k$ be the set of query points of an oblivious estimator. That is, there exists a function $f : \mathbb{R}_+^S \to \mathbb{R}_+$ such that for any semi-norm $\varphi : \mathbb{R}^k \to \mathbb{R}_+$, $f((\varphi(x) : x \in S)) = (1 \pm \varepsilon)\,\mathbb{E}_{Y \leftarrow \mathcal{N}^k}[\varphi(Y)]$. We will assume that f is monotone in the following sense: $f(x_1, \ldots, x_{|S|}) \le f(y_1, \ldots, y_{|S|})$ if $0 \le x_i \le y_i$ for all i. This is clearly true for any linear estimator (and also for the median estimator). Without loss of generality suppose that $\varepsilon < 1/4$.

The idea is to define a suitable semi-norm based on S: define $\varphi : \mathbb{R}^k \to \mathbb{R}$ by $\varphi(x) = \sup_{u \in S} |\langle u/\|u\|_2, x\rangle|$. It is easy to check that for any $v \in S$, $\|v\|_2 \le \varphi(v)$. Therefore, the output of the oblivious estimator when querying the Euclidean norm is at most the output of the estimator when querying φ. In particular,

$$\begin{aligned}(1-\varepsilon) \mathop{\mathbb{E}}_{Y \leftarrow \mathcal{N}^k}\big[\|Y\|_2\big] &\le f\big(\big(\|x\|_2 : x \in S\big)\big) \le f\big(\big(\varphi(x) : x \in S\big)\big) \\ &\le (1+\varepsilon) \mathop{\mathbb{E}}_{Y \leftarrow \mathcal{N}^k}\big[\varphi(Y)\big].\end{aligned} \tag{7.1}$$

We will argue that the above is possible only if $|S| > (C/\varepsilon)^k$. Let $\mathcal{S}^{k-1}$ denote the unit sphere in $\mathbb{R}^k$. For the remaining argument, we shall view $Y \leftarrow \mathcal{N}^k$ to be drawn as $Y = RX$, where $X \in \mathcal{S}^{k-1}$ is uniformly random on the sphere, and $R \in \mathbb{R}$ is independent of X and has a Chi-squared distribution with k degrees of freedom. Let $S(\varepsilon) = \bigcup_{u \in S}\{y \in \mathcal{S}^{k-1} : |\langle u/\|u\|_2, y\rangle| \ge 1 - 4\varepsilon\}$.

Now, by a standard volume argument, for any $y \in \mathcal{S}^{k-1}$, $\Pr_X[|\langle X, y\rangle| \ge 1 - 4\varepsilon] < (O(\varepsilon))^k$. Thus, by a union bound, $p = \Pr_X[X \in S(\varepsilon)] < |S| \cdot (O(\varepsilon))^k$. Further, for any $y \in \mathcal{S}^{k-1} \setminus S(\varepsilon)$, $\varphi(y) < 1 - 4\varepsilon$. Therefore,

$$\begin{aligned}\mathop{\mathbb{E}}_X\big[\varphi(X)\big] &= \Pr\big[X \notin S(\varepsilon)\big] \cdot \mathbb{E}\big[\varphi(X) | X \notin S(\varepsilon)\big] \\ &\quad + \Pr\big[X \in S(\varepsilon)\big] \cdot \mathbb{E}\big[\varphi(X) | X \in S(\varepsilon)\big] \\ &\le (1-p)(1-4\varepsilon) + p.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\big[\varphi(Y)\big] &= \mathbb{E}\big[\varphi(RX)\big] = \mathbb{E}[R] \cdot \mathbb{E}\big[\varphi(X)\big] \\ &\le \mathbb{E}\big[\|Y\|_2\big] \cdot \big((1-p)(1-4\varepsilon) + p\big).\end{aligned} \tag{7.2}$$

Combining equations (7.1) and (7.2), we get

$$1 - \varepsilon \le (1+\varepsilon) \cdot \big((1-p)(1-4\varepsilon) + p\big) < 1 - 3\varepsilon + 2p.$$

As $p < |S| \cdot (O(\varepsilon))^k$, the above leads to a contradiction unless $|S| > (C/\varepsilon)^k$ for some constant $C > 0$.

REFERENCES

ALDOUS, D. and FILL, J. (1994). Reversible Markov chains and random walks on graphs. Available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

BALL, K. (2001). *Convex Geometry and Functional Analysis*. *Handbook of the Geometry of Banach Spaces* **1**. North-Holland, Amsterdam. MR1863692

DADUSH, D. and VEMPALA, S. (2012). Deterministic construction of an approximate M-ellipsoid and its applications to derandomizing lattice algorithms. In *Proceedings of the Twenty-Third Annual ACM–SIAM Symposium on Discrete Algorithms* 1445–1456. SIAM, Philadelphia. MR3205304

DING, J. (2014). Asymptotics of cover times via Gaussian free fields: Bounded-degree graphs and general trees. *Ann. Probab.* **42** 464–496. MR3178464

DING, J., LEE, J. R. and PERES, Y. (2012). Cover times, blanket times, and majorizing measures. *Ann. of Math.* (2) **175** 1409–1471. MR2912708

ENGEBRETSEN, L., INDYK, P. and O'DONNELL, R. (2002). Derandomized dimensionality reduction with applications. In *Proceedings of the Thirteenth Annual ACM–SIAM Symposium on Discrete Algorithms* 705–712. SIAM, Philadelphia, PA.

FEIGE, U. and ZEITOUNI, O. (2009). Deterministic approximation for the cover time of trees. Preprint. Available at arXiv:0909.2005.

KAHN, J., KIM, J. H., LOVÁSZ, L. and VU, V. H. (2000). The cover time, the blanket time, and the Matthews bound. In 41*st Annual Symposium on Foundations of Computer Science* (*Redondo Beach*, *CA*, 2000) 467–475. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1931843

KANTER, M. (1977). Unimodality and dominance for symmetric random vectors. *Trans. Amer. Math. Soc.* **229** 65–85. MR0445580

LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. Springer, Berlin. MR1102015

LEE, J. (2010). Open question: Cover times and the Gaussian free field. Available at https://tcsmath.wordpress.com/2010/12/09/open-question-cover-times-and-the-gaussian-free-field/.

LOVÁSZ, L. (1993). Random walks on graphs: A survey. *Combinatorics*, Paul Erdős Is Eighty, Vol. 2 (Keszthely, 1993). Bolyai Soc. Math. Stud. 2353–2397. János Bolyai Math. Soc., Budapest. MR1395866

PISIER, G. (1999). *The Volume of Convex Bodies and Banach Space Geometry*. Cambridge Univ. Press, Cambridge.

SIVAKUMAR, D. (2002). Algorithmic derandomization via complexity theory. In *IEEE Conference on Computational Complexity* 619–626. ACM, New York.

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183

TALAGRAND, M. (2005). *The Generic Chaining*: *Upper and Lower Bounds of Stochastic Processes*. Springer, Berlin. MR2133757

WINKLER, P. and ZUCKERMAN, D. (1996). Multiple cover time. *Random Structures Algorithms* **9** 403–411. MR1605407

MICROSOFT RESEARCH
1288 PEAR AVENUE
MOUNTAIN VIEW, CALIFORNIA 94043
USA
E-MAIL: meka@microsoft.com