Noise Stability Is Computable and Approximately Low-Dimensional*†

Anindya De[1], Elchanan Mossel[2], and Joe Neeman[3]

1 **Northwestern University, Evanston, IL, USA**
de.anindya@gmail.com
2 **Massachusetts Institute of Technology, Cambridge, MA, USA**
elmos@mit.edu
3 **University of Bonn, Bonn, Germany; and**
University of Texas at Austin, Austin, TX, USA
joeneeman@gmail.com

Abstract

Questions of noise stability play an important role in hardness of approximation in computer science as well as in the theory of voting. In many applications, the goal is to find an optimizer of noise stability among all possible partitions of $\mathbb{R}^n$ for $n \geq 1$ to k parts with given Gaussian measures $\mu_1, \ldots, \mu_k$. We call a partition ϵ-optimal, if its noise stability is optimal up to an additive ϵ. In this paper, we give an explicit, computable function $n(\epsilon)$ such that an ϵ-optimal partition exists in $\mathbb{R}^{n(\epsilon)}$. This result has implications for the computability of certain problems in non-interactive simulation, which are addressed in a subsequent work.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Gaussian noise stability, Plurality is stablest, Ornstein Uhlenbeck operator

Digital Object Identifier 10.4230/LIPIcs.CCC.2017.10

1 Introduction

1.1 Isoperimetric Theory and Noise Stability

Isoperimetric problems have been studied in mathematics since antiquity. The solution to the isoperimetric problem in the two dimensional Euclidean plane was known in ancient Greece. The study of isoperimetric problems is central in modern mathematics and theoretical computer science. Some central examples include the study of expanders and mixing of Markov chains.

Our interest in this work is in a central modern isoperimetric problem, i.e, the problem of noise stability. This problem, originally studied by Borell in relation to the study of the heat equation in mathematical physics [3], has emerged as a central problem in theoretical computer science [13], as well as in combinatorics and in voting theory [11] as we elaborate below.

The fundamental question in this area is to find a partition of Gaussian space (with prescribed measures) which maximizes the noise stability of the partition. We equip $\mathbb{R}^n$

* Full version of this paper is available at https://arxiv.org/abs/1701.01483.
† E. M. acknowledges the support of grant N00014-16-1-2227 from Office of Naval Research and of NSF award CCF 1320105. A. D. acknowledges the support of a start-up grant from Northwestern University.

with the standard Gaussian measure γ_n, i.e. the measure with density $(2\pi)^{-n/2}\exp(-|x|^2/2)$, where $|x|$ denotes the Euclidean norm. The Ornstein-Uhlenbeck operator P_t is defined for $t \in [0,\infty)$ and (sufficiently integrable) $f : \mathbb{R}^n \to \mathbb{R}$ by

$$(P_t f)(x) = \int_{y\in\mathbb{R}^n} f(e^{-t}\cdot x + \sqrt{1-e^{-2t}}\cdot y)d\gamma_n(y).$$

The resulting notion of noise stability is defined by $\mathsf{Stab}_t(f) := \mathbf{E}[fP_tf]$, where the expectation is with respect to γ_n. If f denotes the indicator function of a set (call it $\mathcal{A}_f$), then $\mathsf{Stab}_t(f)$ is the probability that two e^{-t}-correlated Gaussian random variables x, y both fall in $\mathcal{A}_f$. In the limit $t \to 0$, noise stability captures the Gaussian surface area of the set $\mathcal{A}_f$ [15]. In particular, we have the following relation for any set:

$$\mathsf{GAS}(\mathcal{A}_f) = \lim_{t\downarrow 0} \frac{\sqrt{2\pi}}{\arccos(e^{-t})}\cdot \mathbf{E}[f\cdot(f-P_tf)],$$

where $\mathsf{GAS}(\mathcal{A}_f)$ denotes the Gaussian surface area of the set $\mathcal{A}_f$.

1.2 Halfspaces are most stable sets

For both noise stability and surface area, halfspaces are known to be the optimal sets [2, 20, 3, 1]. We will state this fact in a slightly convoluted way, in order to more easily generalize it. Let Δ_k denote the probability simplex in $\mathbb{R}^k$ (i.e. the convex hull formed by the standard unit vector $\{\mathbf{e}_1,\ldots,\mathbf{e}_k\}$). Any function with range $[k] := \{1,\ldots,k\}$ naturally embeds in Δ_k by identifying $i \in [k]$ with $\mathbf{e}_i$. Moreover, we extend the Ornstein-Uhlenbeck operator to act on vector valued functions in the obvious way: if $f = (f_1,\ldots,f_k) : \mathbb{R}^n \to \mathbb{R}^k$ then $P_tf = (P_tf_1,\ldots,P_tf_k)$. Finally, say that $f = (f_1, f_2) : \mathbb{R}^n \to \Delta_2$ is a *halfspace* if there exist $a, b \in \mathbb{R}^n$ such that $f_1(x) = 1_{\{\langle x-a,b\rangle\le 0\}}$, where $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product.

▶ **Theorem 1** (Borell). *For any $g : \mathbb{R}^n \to \Delta_2$, a halfspace $f : \mathbb{R}^n \to \Delta_2$ with $\mathbf{E}[f] = \mathbf{E}[g]$ satisfies*

$$\mathbf{E}[\langle f, P_tf\rangle] \ge \mathbf{E}[\langle g, P_tg\rangle].$$

Theorem 1 has many applications in computational complexity. Most famously, it can be combined with an *invariance principle* [16] in order to prove a number of tight hardness-of-approximation results under the unique games conjecture [12].

1.3 More parts?

It is straightforward to extend the notion of noise stability to partitions with many parts. Namely, a partition of $\mathbb{R}^n$ can be described by $f : \mathbb{R}^n \to [k]$. Identifying $[k]$ with $\{\mathbf{e}_1,\ldots,\mathbf{e}_k\}$ as above, we define the noise stability of such a partition by $\mathbf{E}[\langle f, P_tf\rangle]$. One may then ask for an analogue of Theorem 1, but where Δ_2 is replaced by Δ_k for some $k \ge 3$.

Even the three-part version of Theorem 1 turns out to be rather harder than the two-part one. We will try to explain why, by analogy with isoperimetric problems. First of all, in the limit as $t \downarrow 0$, the Euclidean analogue of Theorem 1 for $k = 3$ is known as the "double bubble" problem. The well-known "double bubble conjecture" states that the minimum total surface area of two bodies separating and enclosing two given (Lebesgue) volumes is achieved by two spheres meeting at 120°. After being open for more than a century, this problem was settled rather recently in $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^4$ [9, 10, 19]. For the Gaussian space, [4] showed that for some small but positive constant $c > 0$, the Gaussian surface area of three partitions is

minimized by the "standard simplex partition" as long as the measures of all the three parts is within $1/3 \pm c$; a standard simplex partition is one with three flat boundaries which meet at a single point at 120° angles.

Since halfspaces both maximize noise stability and minimize Gaussian surface area, and since the standard simplex partition is known to minimize multi-part Gaussian surface area in certain cases, it seems natural to guess that the standard simplex partition also maximizes multi-part noise stability. This was explicitly conjectured in [12], in the special case that all of the parts in the partition have equal measure. However, a somewhat surprising (at least to the authors) recent work [7] showed that the standard simplex partition fails to maximize multi-part noise stability *unless* all of the parts have equal measure. On the other hand, there is also some support for the conjecture in the equal-measure case: Heilman [8] showed that the conjecture is true in $\mathbb{R}^n$ for $n \leq n_0(t)$.

1.4 Approximate noise stability of multipartitions?

In light of the uncertainty about optimal partitioning for $k \geq 3$, one can ask a more modest question. Given $k \geq 3$, $t > 0$, and prescribed measures for the k parts, let α_n be the optimal noise stability that can be obtained in $\mathbb{R}^n$ under these constraints. Since the Gaussian measure is a product measure, α_n is clearly non-increasing in n. Since it is bounded below by zero, it has a limit as $n \to \infty$.

Our main result is that there is an explicitly computable $n_0 = n_0(k, t, \epsilon)$ such that $\alpha_{n_0} \geq \alpha_m - \epsilon$ for all $m \in \mathbb{N}$. Although the bound on n_0 that we give is not particularly good, the key point is that it is explicitly computable. As a consequence, up to error ϵ, the noise stable partition is also explicitly computable. We conclude the introduction by an open question:

▶ **Question.** *Does there exists n_0 such that $\alpha_{n_0} = \alpha_n$ for $n > n_0$?*

Our current techniques are not suitable for addressing the question above.

2 Main theorem and overview of proof technique

In order to state the main theorem, we first need to recall the notion of a polynomial threshold function. A function $f : \mathbb{R}^n \to \{0, 1\}$ is said to be a degree-d PTF if there exists a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ of degree d such that $f(x) = 1$ if and only if $p(x) > 0$. We will need a k-ary generalization of this definition. We note that there are several possible ways to generalize the notion of PTFs to k-ary PTFs and our particular choice is dictated by the convenience of using the relevant results from [5].

▶ **Definition 2.** A function $f : \mathbb{R}^n \to [k]$ is said to be a multivariate PTF if there exist polynomials $p_1, \ldots, p_k : \mathbb{R}^n \to \mathbb{R}$ such that

$$f(x) = \begin{cases} j & \text{if } p_j(x) > 0 \text{ and for } i \neq j,\ p_i(x) \leq 0\,, \\ 1 & \text{otherwise}\,. \end{cases}$$

In this case, we denote $f = \mathsf{PTF}(p_1, \ldots, p_k)$. Further, f is said to be a degree-d multivariate PTF if $p_1, \ldots, p_k$ are degree d polynomials.

We now state the main theorem of this paper. We set the convention, that unless explicitly mentioned otherwise, the underlying distribution is γ_n, the standard n-dimensional Gaussian measure. Likewise, given any random variable X over $\mathbb{R}^k$, $\mathbf{E}[X]$ denotes its (vector-valued) expectation and $\mathsf{Var}(X)$ denotes its covariance matrix.

▶ **Theorem 3.** *Let $f : \mathbb{R}^n \to [k]$ such that $\mathbf{E}[f] = \mu \in \mathbb{R}^k$. Then, given any $t > 0, \epsilon > 0$, there exists an explicitly computable $n_0 = n_0(t, k, \epsilon)$ and $d = d(t, k, \epsilon)$ such that there is a degree-d PTF $g : \mathbb{R}^{n_0} \to [k]$ such that:*

1. $\|\mathbf{E}[f] - \mathbf{E}[g]\|_1 \leq \epsilon$,
2. $\mathbf{E}[\langle g, P_t g\rangle] \geq \mathbf{E}[\langle f, P_t f\rangle] - \epsilon$.

Note that the above theorem automatically implies that a function g satisfying the above properties can be explicitly computed (up to some additional error ϵ). This is because the set of degree-d PTFs on $\mathbb{R}^{n_0}$ admits a finite sized explicitly enumerable ϵ-cover.

2.1 From general partitions to PTF

The first step in the proof of Theorem 3 is to show that given any $f : \mathbb{R}^n \to [k]$, there is a multivariate PTF $g' : \mathbb{R}^n \to [k]$ which meets the two criteria in Theorem 3 and has degree $d = d(t, k, \epsilon)$, for some explicit function $d(t, k, \epsilon)$. In other words, g' satisfies $\|\mathbf{E}[f] - \mathbf{E}[g']\|_1 \leq \epsilon$ and $\mathsf{Stab}_t(g') \geq \mathsf{Stab}_t(f') - \epsilon$. Note that the main difference between the desired conclusion of Theorem 3 and what is accomplished in this step is that the ambient dimension remains n as opposed to a bounded dimension n_0.

Why is this true? The basic intuition is that if f is noise stable then it should have most of its Hermite expansion weight at low degree. Therefore we should be able to replace f with the PTF where the polynomial is the truncated expansion of f. There are a number of challenges in formalizing this intuition:

1. We cannot rule out that a positive fraction of the weight of f is at high degrees (perhaps as large as n);
2. it is not clear that the PTF obtained this way is noise stable; nor that
3. it has the right expected value.

Our analysis proceeds as follows. We would like to construct g' from f by "rounding" $P_t f$ for some small t. The advantage of $P_t f$ over f is that $P_t f$ is guaranteed to have decaying tails. The rounding of $P_t f$ can be performed given some $a \in \mathbb{R}^n$ by considering the function $g_a : \mathbb{R}^n \to [k]$ which takes the value i whenever i is the largest coordinate of $P_t f - a$. It is not hard to prove that it is possible to choose a such that $\mathbb{E}[g_a] = \mathbb{E}[f]$; moreover, one can show that this function g_a has better noise stability than f does. The main obstacle is that the function g_a is not a PTF. Unfortunately the Hermite decay of $P_t f$ does not translate to Hermite decay of g_a. Instead we use smoothed analysis to show that for most a's, g_a has Hermite decay and can therefore be well approximated by PTF. The smoothed analysis argument uses the co-area formula and gradient bounds and draws on ideas from [17, 14].

2.2 From PTF in dimension n to a small PTF of bounded degree polynomials

Given the function $g' : \mathbb{R}^n \to [k]$ of degree $d = d(t, k, \epsilon)$, our next goal is show it is possible to obtain a PTF g on some $n_0 = n_0(t, k, \epsilon)$ variables such that (i) $\|\mathbf{E}[g] - \mathbf{E}[g']\|_1 \leq \epsilon$ and (ii) $\mathrm{Z}|\langle g, P_t g\rangle - \langle g', P_t g'\rangle| \leq \epsilon$. This part builds on and extends the theory and results of [5]. The key notion introduced in [5] is that of an *eigenregular* polynomial. Namely, a polynomial is said to be δ-eigenregular if for the canonical tensor $\mathcal{A}_p$ associated with the polynomial, the ratio of the maximum singular value to its Frobenius norm is at most δ (the tensor notions are explicitly defined later).

The key advantage of this definition is that as shown in [5], when $\delta \to 0$, the distribution of p (under γ_n) converges to a normal. In other words, eigenregular polynomials obey a

central limit theorem. In fact, given k polynomials $p_1, \ldots, p_k$ which are δ-eigenregular, they also obey a multidimensional central limit theorem.

The *regularity lemma* from [5] implies that the polynomials $p_1, \ldots, p_k$ can be jointly expressed as bounded (in terms of t, k and ϵ) size polynomials in eigenregular homogenous polynomials $\{\mathrm{In}(p_{s,q,\ell})\}$ where $1 \le s \le k$, $1 \le q \le d$ and $1 \le \ell \le \mathrm{num}(s,q)$. In other words, we may write $p_s = \mathrm{Out}(p_s)(\{\mathrm{In}(p_{s,q,\ell})\}_{q\in[d],\ell\in[\mathrm{num}(s,q)]})$, where $\mathrm{num}(s) = \sum_{q=1}^{d} \mathrm{num}(s,q)$ is bounded in terms of t, k and ϵ and all the Inner polynomials are δ-eigenregular.

[5] used the statement above to conclude that the joint distribution of $p_1, \ldots, p_k$ can be approximated in a bounded dimension as we can replace each of the inner polynomials by a one dimensional Gaussian. For our application things are more delicate, as we are not only interested in the joint distribution of $p_1, \ldots, p_k$ but also in the noise stability of $p_1, \ldots, p_k$. For this reason it is important for us to maintain the degrees of the inner polynomials (each of which is homogenous) and not replace them with Gaussians.

2.3 A small PTF representation

In the final step of the proof, we maintain $\mathrm{Out}(p_s)$ and show how that polynomials $\{\mathrm{In}(p_{s,q,\ell})\}$ can be replaced by a collection of polynomials $\{\mathrm{In}(r_{s,q,\ell})\}$ in bounded dimensions (in t, k and ϵ) thus completing the proof. The fact that a collection of homogenous polynomials can be replaced by polynomials in bounded dimensions is a tensor analogue of the fact that for any k vectors in $\mathbb{R}^n$, there exist k vectors in $\mathbb{R}^k$ with the same matrix of inner products. Once such polynomials are found, it is not hard to construct eigenregular polynomials from them by averaging the polynomials over independent copies of random variables.

3 Applications

Given the wide applicability of Borell's isoperimetric result to combinatorics and theoretical computer science, we believe that Theorem 3 will also be widely applicable. We will now point out some applications of this theorem. First, by combining Theorem 3 with the invariance principle [16], we derive a weak k-ary analogue of "Majority is Stablest". The analogue of the Ornstein-Uhlenbeck operator is the so-called *Bonami-Beckner operator* defined as follows: for $\rho \in [-1,1]$ and $x \in \{-1,1\}^n$, let $\mathcal{D}_\rho(x)$ be the product distribution over $\{-1,1\}^n$ such that for $y \sim \mathcal{D}_\rho(x)$, for all $1 \le i \le n$, $\mathbf{E}[x_i y_i] = \rho$. Then, for any $f : \{-1,1\}^n \to \mathbb{R}^k$, $T_\rho f(x) = \mathbf{E}_{y\sim D_\rho(x)}[f(y)]$. Likewise, for any $i \in [n]$ and $z \in \{-1,1\}^{n-1}$, let $f_{z,-i} : \{-1,1\} \to \mathbb{R}^k$ denote the function obtained by restricting all but the i^{th} coordinate to z. Then, $\mathsf{Var}(f_{z,-i}) = \mathbf{E}_z[\|f_{z,-i} - \mathbf{E}_z[f_{z,-i}]\|_2^2]$. Define the influence of the i^{th} coordinate on f by

$$\mathsf{Inf}_i(f) = \mathbf{E}_{z\in\{-1,1\}^{n-1}}[\mathsf{Var}(f_{z,-i})].$$

▶ **Theorem 4.** *Given any $k \in \mathbb{N}$ and $\rho \in [0,1], \epsilon > 0$, $\exists n_0 = n_0(k,\epsilon,\rho), C = C(k,\epsilon,\rho)$ (which is explicitly computable) such that the following holds: For any $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k) \in \Delta_k$ and $n \ge n_0$, there is an explicitly computable $g = g_{\boldsymbol{\mu},\rho,\epsilon} : \{-1,1\}^n \to [k]$ such that $\max_i \mathsf{Inf}_i(g) \le C/\sqrt{n}$ and $|\Pr_{x\in\{-1,1\}^{n_0}}[g(x) = i] - \mu_i| \le C/\sqrt{n}$ and for any $f : \{-1,1\}^n \to [k]$ such that $|\Pr_{x\in\{-1,1\}^n}[f(x) = i] - \mu_i| \le \epsilon$ and $\max_i \mathsf{Inf}_i(f) \le \epsilon$,*

$$\mathbf{E}[\langle f, T_\rho f\rangle] \le \mathbf{E}[\langle g, T_\rho g\rangle] + 2k\epsilon.$$

We give a brief sketch of the proof, which is a straightforward consequence of the invariance principle in one direction and the central limit theorem in the other. First, the invariance

principle implies that the discrete noise stability cannot be much larger than the best Gaussian noise stability. It remains, then, to construct a boolean function g whose noise stability is almost the same as the best possible Gaussian noise stability. For this, suppose that $n = n_1 n_2$ (for some n_1 and n_2 that will need to be sufficiently large), and let $h : \mathbb{R}^{n_1} \to [k]$ have almost optimal Gaussian noise stability. Then define the boolean function $g(x) = h(z(x))$, where $z \in \mathbb{R}^{n_1}$ has ith coordinate $n_2^{-1/2} \sum_{j=in_2+1}^{(i+1)n_2} x_j$. When n_2 is sufficiently large, the central limit theorem implies that the boolean noise stability of g is approximately the Gaussian noise stability of h, and so it is almost optimal.

The same result holds for other domains, for example for $f, g : [k]^n \to [k]$. In particular, the case where $(\mu_1, \ldots, \mu_k) = (1/k, \ldots, 1/k)$ implies that in a tied elections between k alternatives, we can find an ϵ-optimally robust noise stable voting rule, where the stability is with respect of each candidate randomizing their vote independently with probability $1 - \rho$.

3.1 Relationship to rounding of SDPs

To the best of our knowledge our results are independent of the the results of Raghavendra and Steurer [18] who showed that for any CSP, there is an a rounding algorithm that is optimal up to ϵ, whose running time is polynomial in the instance size and doubly exponential in $1/\epsilon$. It is natural to suspect that the two results are related, since there are well-known connections between SDP rounding and Gaussian noise stability.

However, the usual analysis relating rounding to Gaussian noise stability seems to require that halfspaces maximize noise stability for all possible values of the noise. In our results, this property does not necessarily hold.

In the other direction, it is tempting to try to cast the noise stability problem as an optimization problem on Gaussian graphs and then apply the results of Steurer and Raghavendra to obtain explicit bounds on the dimension where an almost optimal solution can be achieved. It is hard to implement this approach for two reasons: first, we do not know the SDP solution for the Gaussian graph; second, we are interested in the optimal solution and it is not clear what is the relation between the best integral solution and the SDP solution for the Gaussian graph.

While we do not see how to formally relate the two works, connecting the two (if possible) would surely yield important insights.

3.2 Non-interactive correlation distillation

Next, we talk about a basic problem in information theory and communication complexity which was recently considered in the work of Ghazi, Kamath and Sudan [6]. Let there be two non-communicating players Alice and Bob who have access to independent samples from a joint distribution $\mathbf{P} = (\mathbf{X}, \mathbf{Y})$ on the set $\mathcal{A} \times \mathcal{B}$. In other words, Alice (resp. Bob) have access to $(x_1, x_2, \ldots,)$ (resp. $(y_1, y_2, \ldots)$) such that $x_i \in \mathcal{A}$, $y_i \in \mathcal{B}$ and for each $i \in \mathbb{N}$, (x_i, y_i) is distributed according to $\mathbf{P}$, and the random variables $\{(x_i, y_i)\}_{i=1}^n$ are mutually independent. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k) \in \Delta_k$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k) \in \Delta_k$. What is the maximum $\kappa \in [0, 1]$ such that Alice and Bob can non-interactively jointly sample a distribution $\mathbf{Q}$ on $[k] \times [k]$ such that the distribution of the marginals of Alice and Bob are $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ respectively and they sample the same output with probability κ?

To formulate this problem more precisely, we introduce some notation. Let $\mathbf{X}^n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $\mathbf{Y}^n = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, where $(\mathbf{X}_i, \mathbf{Y}_i)$ are independently drawn from $\mathbf{P}$. Now, note that a non-interactive protocol for Alice and Bob is equivalent to a pair (f, g) where $f : \mathcal{A}^n \to [k]$ and $g : \mathcal{B}^n \to [k]$ (for some $n \in \mathbb{N}$). In this terminology, the question

now becomes the following: given $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, do there exist n and $f : \mathcal{A}^n \to [k]$ and $g : \mathcal{B}^n \to [k]$ such that $f(\mathbf{X}^n) \sim \boldsymbol{\mu}$, $g(\mathbf{Y}^n) \sim \boldsymbol{\nu}$, and $\Pr(f(\mathbf{X}^n) = g(\mathbf{Y}^n)) = \kappa$?

Before we state the main result of [6] and our extension, we consider a motivating example. Let $\mathcal{A} = \mathcal{B} = \mathbb{R}$ and let $\mathbf{P} = (\mathbf{X}, \mathbf{Y})$ be two ρ-correlated standard Gaussians. Let $k = 2$ and $\boldsymbol{\mu} = \boldsymbol{\nu}$. Then, Borell's isoperimetric theorem (Theorem 1) states that the maximum achievable κ is given by

$$\kappa = \Pr_{\mathbf{X},\mathbf{Y}}[f(\mathbf{X}) = f(\mathbf{Y})],$$

where $f : \mathbb{R} \to \Delta_2$ is a halfspace with measure $\boldsymbol{\mu}$. Thus, in the above case, $n = 1$ suffices and $f = g$ is the halfspace whose measure is $\boldsymbol{\mu}$. We now state the main result of [6]. For the result below, for probability distribution $\mathbf{P}$, we let $|\mathbf{P}|$ denote the size of some standard encoding of $\mathbf{P}$.

▶ **Theorem 5.** *[Ghazi-Kamath-Sudan] Let $(\mathcal{A} \times \mathcal{B}, \mathbf{P})$ be a probability space, and let $\mathbf{X}^n$ and $\mathbf{Y}^n$ be as above. There is an algorithm running in time $O_{|\mathbf{P}|,\delta}(1)$ such that given $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in Δ_2 and a parameter $\kappa \in [0,1]$, it distinguishes between the following two cases:*

1. *There exist $n \in \mathbb{N}$, $f : \mathcal{A}^n \to \{0,1\}$ and $g : \mathcal{B}^n \to \{0,1\}$ such that $f(\mathbf{X}^n) \sim \boldsymbol{\mu}$, $g(\mathbf{Y}^n) \sim \boldsymbol{\nu}$, and $\Pr(f(\mathbf{X}^n) = g(\mathbf{Y}^n)) \geq \kappa - \delta$. In this case, there is an explicit $n_0 = n_0(|\mathbf{P}|, \delta)$ such that we may choose $n \leq n_0$. Further, the functions f and g are explicitly computable.*
2. *For any $n \in \mathbb{N}$ and $f : \mathcal{A}^n \to \{0,1\}$ and $g : \mathcal{B}^n \to \{0,1\}$, if $g : \mathcal{B}^n \to \{0,1\}$ satisfy $f(\mathbf{X}^n) \sim \boldsymbol{\mu}$ and $g(\mathbf{Y}^n) \sim \boldsymbol{\nu}$ then $\Pr(f(\mathbf{X}^n) = g(\mathbf{Y}^n)) \leq \kappa - 8\delta$.*

In other words, the above theorem states that there is an algorithm which, given $\mathbf{P}$, target marginals $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, and correlation κ, can distinguish between two cases: (a) In the first case, it is possible for Alice and Bob to non-interactively simulate a distribution which has the correct marginals and achieves the correlation κ up to δ. In this case, there is an explicit bound on the number of copies of $\mathbf{P}$ required and the algorithm also outputs the functions f, g used for the non-interactive simulation. (b) In the second case, no non-interactive protocol between Alice and Bob can simulate a distribution that has the correct marginals and target correlation to error at most 8δ.

We remark that the requirement for the marginals to match *exactly* is unimportant. Indeed, any protocol where the marginals match approximatly can be "fixed" to have exact marginals, with a small loss in the correlation.

The main restriction of Theorem 5 is that the output of the non-interactive protocol is a pair of bits. Theorem 3 immediately implies the following modification of Theorem 5: we may replace $\{0,1\}$ by $[k]$ wherever it appears, provided that we also assume that $\mathbf{P}$ is a Gaussian measure. In the best of both worlds, we would be able to replace $\{0,1\}$ by $[k]$ without adding the assumption that $\mathbf{P}$ is a Gaussian measure. Using the methods we develop here, this also turns out to be possible; we provide details in a follow-up work.

4 Reduction from arbitrary functions to PTFs

In this section, we will prove that given any k-ary function with a given set of measures for each of the k-partitions, there is a multivariate PTF with nearly the same measures for the induced partitions which is a multivariate PTF and (up to an error ϵ), no less noise stable at a fixed noise rate t. This is the first step ("from general partitions to PTF") of the proof sketch in Section 2.

For technical reasons, we will also require the PTFs to satisfy a property which we refer to as (d, δ)-*balanced* defined below.

▶ **Definition 6.** A degree-d multivariate PTF $f = \mathsf{PTF}(p^{(1)}, \ldots, p^{(k)})$ is said to be (d, δ)-balanced if each $p^{(i)}$ has variance 1 and $|\mathbf{E}[p^{(i)}]| \leq \log^{d/2}(k \cdot d/\delta)$.

We ask the reader to observe that the first condition (namely, $\mathsf{Var}(p^{(i)}) = 1$) can be achieved without loss of generality by simply scaling all the polynomials to have variance 1. This scaling does not change the value of the PTF at any point x. While the condition on expectation is non-trivial, the next proposition says that any multivariate PTF can be assumed to be (d, δ)-balanced while only changing the value of the PTF at δ-fraction of places.

▶ **Lemma 7.** *Let $f : \mathbb{R}^n \to [k]$ defined as $f = \mathsf{PTF}(p^{(1)}, \ldots, p^{(k)})$. Then, there is a (d, δ)-balanced multivariate PTF $g : \mathbb{R}^n \to [k]$ defined as $g = \mathsf{PTF}(q^{(1)}, \ldots, q^{(k)})$ such that $\Pr_x[g(x) \neq f(x)] \leq \delta$. Further, $\Pr_x[x \in \mathsf{Collision}(g)] \leq \Pr_x[x \in \mathsf{Collision}(f)] + \delta$. In fact, the polynomials $\{q^{(i)}\}$ are linear translations of the polynomials $\{p^{(i)}\}$.*

The proof of the lemma is deferred to the full version of the paper. The next theorem is the main result of this section namely, that given any function $f : \mathbb{R}^n \to [k]$, it is possible to obtain a multivariate (d, ϵ) balanced PTF g (for some explicit $d = O_{\epsilon,k}(1)$) such that the resulting PTF g has nearly the same partition sizes and noise stability. Further, $\mathsf{Collision}(g)$ has small probability.

▶ **Theorem 8.** *Let $f : \mathbb{R}^n \to [k]$ satisfy $\mathbb{E}[f] = \boldsymbol{\mu}$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$. Then, for every $\epsilon > 0$, there exists a multivariate PTF $g : \mathbb{R}^n \to [k]$ of degree $d = d(k, \epsilon)$ such that*

- $\|\mathbf{E}[g] - \boldsymbol{\mu}\|_1 \leq \epsilon$,
- $\langle g, P_t g \rangle \geq \langle f, P_t f \rangle - \epsilon$,
- $\Pr[x \in \mathsf{Collision}(g)] \leq \epsilon$, *and*
- *g is (d, ϵ)-balanced.*

We will prove Theorem 8 in two parts. The first step is to show that we can replace f by a function with explicit Hermite decay:

▶ **Lemma 9.** *For every $f : \mathbb{R}^n \to [k]$ and every $\epsilon > 0$, there exists a function $h : \mathbb{R}^n \to [k]$ satisfying the following:*

- $\mathbb{E}[h] = \mathbb{E}[f]$,
- $\langle h, P_t h \rangle \geq \langle f, P_t f \rangle - \epsilon$,
- $\mathsf{W}^{>d}[h] \leq \epsilon$ *for some explicit $d = d(k, \epsilon)$.*

The second step in the proof of Theorem 8 goes from explicit Hermite decay to an actual PTF:

▶ **Lemma 10.** *Let $h : \mathbb{R}^n \to [k]$ such that $\mathsf{W}^{>d}[h] \leq \epsilon$. Then, there is a PTF $g : \mathbb{R}^n \to [k]$ of degree d such that $\mathbf{E}_x[g(x) \neq h(x)] \leq k^2 \cdot \epsilon$. Further, $\Pr_x[x \in \mathsf{Collision}(h)] \leq k^2 \cdot \epsilon$.*

It is easy to see that Theorem 8 follows in a straightforward way by combining Lemma 9 and Lemma 10. Note that the condition of g being (d, ϵ)-balanced is obtained by simply applying Proposition 7.

While we defer the proof of both Lemma 9 and Lemma 10 to the full version, as the proof of Lemma 9 is more non-standard, we give a brief overview here. The first observation is that bounding $\mathbb{E}[|\nabla h|]$ implies a bound on $\mathsf{W}^{>d}[h]$. This follows a standard spectral argument stated below.

▶ **Lemma 11.** *Let $f : \mathbb{R}^n \to [0, 1]$ be of bounded variation. Then,*

$$\mathsf{W}^{\geq d}[f] \leq O\left(\frac{1}{\sqrt{d}}\right) \cdot \mathbf{E}[|\nabla f|].$$

The second observation is that the function h obtained by thresholding $P_t f$ at a suitable value (chosen, for example, so that $\mathbb{E}[h] = \mathbb{E}[f]$) satisfies $\langle h, P_t h\rangle \geq \langle f, P_t f\rangle$. This is stated below.

▶ **Lemma 12.** *Let $f : \mathbb{R}^n \to \{e_1, \ldots, e_k\}$ and take $t > 0$. If $g \in T_z(f)$ for some $z \in \mathbb{R}^k$ satisfies $\mathbb{E}[g] = \mathbb{E}[f]$ then $\mathsf{Stab}_t(g) \geq \mathsf{Stab}_t(f)$.*

Based on the previous paragraph, it seems like we would like to bound $\mathbb{E}[|\nabla h|]$ where h is obtained by thresholding $P_t f$; let $a \in \mathbb{R}^k$ be the desired threshold value, so that thresholding $P_t f$ at a produces a partition with the right measures. It turns out, unfortunately, that for h defined in this way, $\mathbb{E}[|\nabla h|]$ could be arbitrarily large. A key insight of [14] is that (using the co-area formula and gradient bounds on $P_t f$) the partition produced by thresholding at a random value near a has bounded expected surface area. In particular, although thresholding exactly at a might be a bad idea, there exist many good nearby values at which to threshold. Based on this observation, we construct h in two steps. In the first step, we define h by thresholding $P_t f$, but only on the set of $x \in \mathbb{R}^n$ for which $P_t f(x)$ is not too close to a. By choosing "not too close" in a suitable random way, the observation of [14] implies that this step only contributes a bounded amount to $\mathbb{E}[|\nabla h|]$. Since the first step is almost the same as just thresholding $P_t f$ at a, it is consistent with our desire that $\langle h, P_t h\rangle \geq \langle f, P_t f\rangle - \epsilon$.

In the second step, we partition the remaining part of $\mathbb{R}^n$ by chopping it with halfspaces of the correct size. Since halfspaces have a bounded surface area, this also contributes a bounded amount to $\mathbb{E}[|\nabla h|]$. Crucially, this step does not destroy the value of $\langle h, P_t h\rangle$; fundamentally, this is because $P_t f$ is almost constant on the set we are partitioning. The actual details of the proof of Lemma 10 can be found in the full version. Let us now move to the second major technical ingredient of this paper.

5 Reduction from PTFs to PTFs on a constant number of variables

The second big technical ingredient required in the paper is the following:

▶ **Theorem 13.** *Let $f : \mathbb{R}^n \to [k]$ be a degree-d, (d, ϵ)-balanced PTF with $\mathbb{E}_x[f(x)] = \boldsymbol{\mu}$ where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$. Further, let us assume that $\Pr[x \in \mathsf{Collision}(f)] \leq \epsilon/(40k^2)$. Then, for every $\epsilon > 0$, there exists a degree-d PTF $f_{\mathsf{junta}} : \mathbb{R}^{n_0} \to [k]$ such that:*

- $\|\mathbf{E}_x[f_{\mathsf{junta}}(x)] - \boldsymbol{\mu}\|_1 \leq \epsilon$,
- $\langle f_{\mathsf{junta}}, P_t f_{\mathsf{junta}}\rangle \geq \langle f, P_t f\rangle - \epsilon$.

Further, $n_0 = n_0(d, k, \epsilon)$ is an explicitly defined function.

The above theorem states that given a degree-d multivariate PTF over n variables, there is another multivariate PTF which induces approximately the same partition sizes and has approximately the same noise stability (at any fixed noise rate t) but the new PTF is only over some (explicitly defined) $O_{d,t}(1)$ variables. The main workhorse for proving this thoerem are two structural theorems for low-degree polynomials proven in [5]. Unfortunately, even stating these theorems require significantly cumbersome technical definitions. In particular, we will need to define the relation between polynomials and tensors and then define the notion of an ϵ-*eigenregular* polynomial. We avoid doing that in this version of the paper and refer the reader to the full version of the paper. However, we observe that Theorem 3 follows very easily by combining Theorem 13 and Theorem 8.

Proof of Theorem 3

Let us us assume that given measure $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ and noise rate $t > 0$, the most noise stable partition is $f : \mathbb{R}^n \to [k]$. Then, applying Theorem 8 (with error parameter

$\epsilon/(40k^2)$), we obtain a PTF g_1 of degree $d = O_{k,\epsilon}(1)$ such that it is $(d, \epsilon/2)$ balanced and $\Pr[x \in \mathsf{Collision}(g_1)] \leq \epsilon/(40k^2)$. Further, note that $\|\mathbf{E}[g_1] - \mathbf{E}[f]\|_1 \leq \epsilon/2$ and $\langle g_1, P_t g_1 \rangle \geq \langle f, P_t f \rangle - \epsilon/2$.

We next apply Theorem 13 on this function g_1 to obtain f_{junta} which is a degree-d PTF on $n_0 = O_{d,\epsilon,k}(1)$ variables such that $\|\mathbf{E}[g_1] - \mathbf{E}[f_{\mathsf{junta}}]\|_1 \leq \epsilon/2$ and $\langle f_{\mathsf{junta}}, P_t f_{\mathsf{junta}} \rangle \geq \langle g_1, P_t g_1 \rangle - \epsilon/2$.

Combining these two facts, we obtain $\|\mathbf{E}[g_1] - \mathbf{E}[f_{\mathsf{junta}}]\|_1 \leq \epsilon$ and $\langle f_{\mathsf{junta}}, P_t f_{\mathsf{junta}} \rangle \geq \langle f, P_t f \rangle - \epsilon$. Setting $g = f_{\mathsf{junta}}$ concludes the proof.

Acknowledgments. The authors started this research at the workshop on "Information theory and concentration phenomena" at the IMA, Minneapolis in 2015. The authors would like to thank the organizers of this workshop and the IMA for their hospitality.

References

1 S. G. Bobkov. Isoperimetric problem for uniform enlargement. *Studia Math*, 123(1):81–95, 1997.

2 C. Borell. The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30:207–216, 1975.

3 C. Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Z. Wahrsch. Verw. Gebiete*, 70(1):1–13, 1985.

4 J. Corneli, I. Corwin, S. Hurder, V. Sesum, Y. Xu, E. Adams, D. Davis, M. Lee, R. Visocchi, N. Hoffman, and Robert M. Hardt. Double bubbles in gauss space and spheres. *Houston Journal of Mathematics*, 34(1):181–204, 2008.

5 A. De and R. Servedio. Efficient deterministic approximate counting for low-degree polynomial threshold functions. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA*, pages 832–841, 2014. Full version at http://arxiv.org/abs/1311.7178.

6 B. Ghazi, P. Kamath, and M. Sudan. Decidability of non-interactive simulation of joint distributions. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 545–554, 2016.

7 S. Heilman, E. Mossel, and J. Neeman. Standard simplices and pluralities are not the most noise stable. *Israel Journal of Mathematics*, pages 1–21, 2014.

8 Steven Heilman. Euclidean partitions optimizing noise stability. *Electron. J. Probab*, 19(71):1–37, 2014.

9 M. Hutchings, F. Morgan, M. Ritoré, and A. Ros. Proof of the double bubble conjecture. *Res. Announc. Amer. Math. Soc*, 6:45–49, 2000.

10 M. Hutchings, F. Morgan, M. Ritoré, and A. Ros. Proof of the double bubble conjecture. *Ann. of Math.*, 155:459–489, 2002.

11 G. Kalai. A Fourier-theoretic perspective on the Concordet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29(3):412–426, 2002.

12 S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? In *Proc. 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 146–154, 2004.

13 S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM J. Comput.*, 37:319–357, 2007. URL: `http://www.stat.berkeley.edu/~mossel/publications/max_cut_final.pdf`.

14 P. Kothari, A. Nayyeri, R. O'Donnell, and C. Wu. Testing surface area. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1204–1214, 2014.

15 M. Ledoux. Semigroup proofs of the isoperimetric inequality in Euclidean and Gauss space. *Bull. Sci. Math.*, 118:485–510, 1994.

16 E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. *Ann. Math.*, 171(1):295–341, 2010.

17 J. Neeman. Testing surface area with arbitrary accuracy. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 393–397, 2014.

18 Prasad Raghavendra and David Steurer. How to round any csp. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 586–594. IEEE, 2009.

19 Ben W. Reichardt, Cory Heilmann, Yuan Y. Lai, and Anita Spielman. Proof of the double bubble conjecture in r4 and certain higher dimensional cases. *Pacific Journal of Mathematics*, 208(2):347–366, 2003.

20 V. Sudakov and B. Tsirel'son. Extremal properties of half-spaces for spherically invariant measures. *J. Soviet Math.*, 9:9–18, 1978. Translated from Zap. Nauchn. Sem. Leningrad. Otdel. Math. Inst. Steklova. 41 (1974), 14–21.