Systems & Control Letters 92 (2016) 42–45

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Gaussian approximations in high dimensional estimation

Vivek S. Borkar *, Raaz Dwivedi [1], Neeraja Sahasrabudhe *

Department of Electrical Engineering, Indian Institute of Technology, Powai, Mumbai 400076, India

ARTICLE INFO

Article history:
Received 21 April 2015
Accepted 3 March 2016
Available online 29 March 2016

Keywords:
Gaussian approximations
High dimensional estimation
Log-concave distributions
Stochastic sparsity
Random projections

ABSTRACT

Several estimation techniques assume validity of Gaussian approximations for estimation purposes. Interestingly, these ensemble methods have proven to work very well for high-dimensional data even when the distributions involved are not necessarily Gaussian. We attempt to bridge the gap between this oft-used computational assumption and the theoretical understanding of why this works, by employing some recent results on random projections on low dimensional subspaces and concentration inequalities.

1. Introduction

In many geophysical or meteorological signal processing applications, Ensemble Kalman Filter (EnKF) or data assimilation has become a popular methodology. Unlike the extended Kalman filter, it does not linearize the dynamics around a nominal trajectory. Instead, it propagates state-observation dynamics as per the original nonlinear rule, but estimates the next state as though it were conditionally Gaussian, using empirical estimates of covariances based on simulated transitions [1]. The Gaussianity hypothesis remains ad hoc, nevertheless the methodology has been found to be very useful by practitioners. Trying to make sense of this 'unreasonable effectiveness of Gaussianity' (to borrow a phrase from Wigner) is the motivation behind this work. We do not, however, address the dynamic situation handled by EnKF, but consider the simpler problem of estimating a random variable, given another, in a high dimensional set-up and justify the Gaussian approximation thereof.

Traditionally, Gaussian assumption has been justified either by invoking the classical central limit theorem, postulating that the observed randomness is the cumulative effect of a large number of independent small events (e.g., shot noise), or by the maximum entropy principle, which is a 'worst case' analysis. (The two philosophies are not unrelated, as we now know from [2].) What we propose here is a third alternative, also a central limit theorem, but in large dimension asymptotics rather than large sample asymptotics as in the classical case. The key tool is a result regarding approximate Gaussianity of low dimensional marginals of a class of high dimensional distributions due to Klartag and others. The details follow in subsequent sections.

While EnKF remains our original motivation, the applicability and relevance of this work to other domains is not ruled out. More generally, this work is the first step towards providing a rigorous basis for using Gaussian approximations in high dimensional inference, wherever it occurs, subject to the log-concavity and sparsity hypotheses. We use ideas from compressive sensing to claim that given an n-dimensional stochastically sparse random vector, one can recover it from samples or measurements that are fewer than n in number. Compressive sensing essentially deals with the problem of reconstructing a sparse vector from underdetermined measurements. One aims to minimize the l_1-error between the coefficients of the original vector and the reconstructed one. See [3] for details.

The paper is organized as follows. We outline the problem and the notation in the next section. In Section 2, we present our result for the special case of stochastically sparse vectors. As mentioned earlier, this requires some results from the theory of compressive sensing. Section 3 recalls the key result of Klartag and Eldan on low dimensional projection with nearly Gaussian densities and points out its implications in the present context. Throughout, $\|.\|$ denotes the standard Euclidean norm in $\mathbb{R}^n$.

1.1. Outline of the problem

We first show that given a random vector $(X, Y) \in R^{n_1+n_2}$, if X, Y are sparse, $E[Y|X]$ can be approximated by projections

* Corresponding authors.
E-mail addresses: borkar.vs@gmail.com (V.S. Borkar), dwivediraaz@gmail.com (R. Dwivedi), neeraja.sigma@gmail.com (N. Sahasrabudhe).

[1] Raaz Dwivedi is now with Department of EECS, University of California, Berkeley.

http://dx.doi.org/10.1016/j.sysconle.2016.03.001
0167-6911/

on a smaller dimensional subspace. For the final step, where we show that suitable conditional densities can be approximated by Gaussian densities, an additional assumption of log-concavity of conditional density of Y given X is required.

We assume throughout that

(**A1**) $E[\|Y\|^2]^{1/2}$ and $E[\|X\|^2]^{1/2}$ are bounded by some constant $M < \infty$, and,

(**A2**) the regular conditional law $\Psi(\cdot|\cdot)$ of Y given $X = x$ has a Lipschitz version as a map $x \in \mathbb{R}^{n_1} \mapsto \Psi(\cdot|x) \in \mathcal{P}_1(\mathbb{R}^{n_2})$ with Lipschitz constant L, where $\mathcal{P}_1(\mathbb{R}^{n_2})$ is the space of probability measures μ on $\mathbb{R}^{n_2}$ with $\int |x|\mu(dx) < \infty$ under the Wasserstein-1 metric $\rho(\mu', \mu'') := \inf E\left[\|Y' - Y''\|\right]$. Here the infimum is over all pairs of random variables (Y', Y'') with law of Y', resp. Y'', being μ', resp. μ''. We work with this version throughout.

Let $\check{X}, \check{Y}$ denote the orthogonal projection of X, Y on random k_1 and k_2 dimensional subspaces respectively. By suitable choice of basis, we take this to be the first k_1 co-ordinates of X and first k_2 co-ordinates of Y. Let $\hat{X} = \check{X}$ and $\hat{Y} = \sqrt{\frac{n_2}{k_2}}\check{Y}$ be the scaled projection. We denote by $\bar{X}$ and $\bar{Y}$ the vectors obtained by padding $\check{X}$ and $\check{Y}$ by $n_1 - k_1$ and $n_2 - k_2$ zeros respectively.

In Section 2 we show that using results in [4] and under suitable conditions of sparsity of X and Y, one can approximate $E[Y|X]$ by $E[Y^*|X^*]$, where Y^* and X^* are "good" reconstructions of X, Y from only k_1, resp. k_2 observations, where a "good" reconstruction means that Y and Y^* (resp. X and X^*) are close in standard Euclidean norm with high probability. Furthermore, in Section 3, we show that $E[Y^*|X^*]$ and therefore $E[Y|X]$ can be computed approximately using a Gaussian density under a log-concavity assumption on Ψ.

Let $G_{n,l}$ denote the Grassmannian of all l-dimensional subspaces of $\mathbb{R}^n$, and let $\sigma_{n,l}$ stand for the unique rotationally invariant probability measure on $G_{n,l}$ [5].

2. Sparse vectors

Our aim is to estimate $E[Y|X]$ by $E[\widehat{Y}|\widehat{X}]$, thereby reducing the cost of computation. Using the results in [4], we now show that this aim can be achieved for a '*stochastically sparse*' vector. In [4], the authors show that it is possible to reconstruct a sparse vector to high accuracy from a small number of random measurements. Let $|v|_n$ denote the nth largest entry of the vector v, or the nth largest coefficient in a fixed basis. Consider a vector $v \in \mathbb{R}^N$ such that either $|v|_n \le R \cdot n^{-1/p}$ for some $R > 0$ and some $0 < p < 1$, or $\|v\|_{l^1} \le R$ for some $R > 0$ and $p = 1$. Consider a random orthonormal basis $\{\phi_m\}_{m=1}^{\infty}$ of $\mathbb{R}^N$ and let $\theta(g) = [\langle g, \phi_1\rangle, \dots, \langle g, \phi_N\rangle]^T$ for $g \in \mathbb{R}^N$. Suppose we observe only the first K coefficients in this basis. Let F_Ω be the submatrix enumerating those sampled vectors, i.e., the projection operator. One can solve the following optimization problem

$$(\mathbf{P}) \quad \min_{g\in\mathbb{R}^N} \|\theta(g)\|_{l^1} \quad \text{subject to } F_\Omega g = F_\Omega v. \tag{1}$$

The solution v^* is such that for $\beta > 0$ sufficiently small

$$\|v - v^*\| \le C_{p,\beta}.R.(K/\log N)^{-r}$$

with probability at least $1 - O(N^{-\rho/\beta})$, where $r = 1/p - 1/2$ and $\rho > 0$ is a universal constant. Also, as noted in [4], the choice of basis is in fact irrelevant. All that is needed is that the vector v be sparse in some fixed basis.

The above optimization problem can be reduced to a linear program by the standard technique of replacing each variable (say) x by $x^+ - x^-$ and defines a map $h : v \in \mathbb{R}^N \mapsto v^* \in \mathbb{R}^N$ whenever the solution v^* is unique. Since the latter holds for a.e. v, h is well defined as a measurable function. From the 1-homogeneity of the objective function and the constraints, it is easy to see that h has linear growth.

To use the above results for random vectors, we define the idea of '*stochastically sparse*' random vectors.

Definition 2.1. Let $Z \in \mathbb{R}^m$ be a random vector. We say that Z is stochastically sparse if for a prescribed $\eta_1 > 0$

$$P\left(\sup_n \frac{|Z|_n}{n^{-1/p}} > R\right) < \eta_1 \tag{2}$$

for some $R > 0$ and $0 < p < 1$.

Let X^* and Y^* denote the solution to the optimization problem (**P**) corresponding to stochastically sparse random vectors X and Y respectively. Then from the above discussion we have that, $Y^* = h(\bar{Y})$ and $X^* = h(\bar{X})$. Define $H(x) = \int y\ \Psi(y|x)dy$ and let $\bar{\Psi}(\cdot|x)$ denote the image of $\Psi(\cdot|x)$ under the projection $\mathbb{R}^{n_1} \mapsto \mathbb{R}^{k_2}$. Now we can prove the following approximation result.

Theorem 2.2. *Let $X \in \mathbb{R}^{n_1}$ and $Y \in \mathbb{R}^{n_2}$ be stochastically sparse. Let η_1, ρ and β be as defined above. Then, given $\epsilon > 0$,*

$$P\left(\left|E[Y|X] - \int h(\bar{y})\bar{\Psi}(\bar{y}|X^*)d\bar{y}\right| > \epsilon\right) \le \frac{4}{\epsilon}\left(\delta_1 + \frac{L\delta_2}{2} + M(2+L)\sqrt{1-q}\right)$$

where, $q = 1 - 2\eta_1 - O(n_2^{-\rho/\beta}) - O(n_1^{-\rho/\beta})$.

Proof. Using the result in [4], we have that on a set B with $P(B) \ge q = 1 - 2\eta_1 - O(n_2^{-\rho/\beta}) - O(n_1^{-\rho/\beta})$,

$$\|Y - Y^*\| \le \delta_1 \quad \text{and} \quad \|X - X^*\| \le \delta_2 \tag{3}$$

where,

$$\delta_1 = C_{p,\beta}.R.(k_2/\log n_2)^{-r} \quad \text{and} \quad \delta_2 = C_{p,\beta}.R.(k_1/\log n_1)^{-r}$$

for $r = 1/p - 1/2$. We have,

$$P\left(\left|E[Y|X] - \int h(\bar{y})\bar{\Psi}(\bar{y}|X^*)d\bar{y}\right| > \epsilon\right) \le P(|E[Y|X] - H(X^*)| > \epsilon/2) + P\left(\left|H(X^*) - \int h(\bar{y})\bar{\Psi}(\bar{y}|X^*)d\bar{y}\right| > \epsilon/2\right).$$

Note that

$$P\left(\left|H(X^*) - \int h(\bar{y})\bar{\Psi}(\bar{y}|X^*)d\bar{y}\right| > \epsilon/2\right) \le P(E[\|Y - Y^*\|I\{B\}|X = x]_{x=X^*} > \epsilon/4) + P(E[\|Y - Y^*\|I\{B^c\}|X = x]_{x=X^*} > \epsilon/4).$$

From stochastic sparsity of Y, we get

$$P(E[\|Y - Y^*\|I\{B\}|X = x]|_{x=X^*} > \epsilon/4) \le \frac{4}{\epsilon}\delta_1 \tag{4}$$

and,

$$P(E[\|Y - Y^*\|I\{B^c\}|X = x] > \epsilon/4) \le \frac{4}{\epsilon}E[\|Y - Y^*\|^2]^{\frac{1}{2}}\sqrt{P(B^c)} \le \frac{8}{\epsilon}M\sqrt{(1-q)}. \tag{5}$$

From Eqs. (4) and (5), it follows that

$$P\left(\left|H(X^*) - \int h(\bar{y})\Psi(\bar{y}|X^*)d\bar{y}\right| > \epsilon/2\right)$$
$$\leq \frac{4}{\epsilon}\delta_1 + \frac{8}{\epsilon}M\sqrt{1-q}. \tag{6}$$

For $P(|E[Y|X] - H(X^*)| > \epsilon/2)$, we use (**A2**) and the stochastic sparsity of X.

$$\begin{aligned} &P(|E[Y|X] - H(X^*)| > \epsilon/2) \\ &\leq P(L\|X - X^*\| > \epsilon/2) \\ &\leq \frac{2L}{\epsilon}E[\|X - X^*\|] \\ &\leq \frac{2L}{\epsilon}E[\|X - X^*\|I\{B\} + \|X - X^*\|I\{B^c\}] \\ &\leq \frac{2L}{\epsilon}\delta_2 + \frac{4L}{\epsilon}M\sqrt{1-q}. \end{aligned} \tag{7}$$

The claim now follows from Eqs. (6) and (7). □

3. Low dimensional projections with Gaussian densities

There are several approaches to study higher dimensional distributions. Just as large deviations and limit theorems exploit the symmetry and/or the independence structure of the random variables, Klartag investigated the classes of densities with certain geometric characteristics. Building upon his earlier work on central limit theorem for convex sets [6,7], Klartag and Eldan [8] proved a pointwise version of the multi-dimensional central limit theorem for convex bodies in 2007. We state some of their important results here. For details see the aforementioned papers or the survey [9] by Klartag.

For a subspace $E \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$ we write $Proj_E(x)$ for the orthogonal projection of x onto E. In [7] Klartag proved a total variation result implying that the density of $Proj_E(X)$ is close to the density of a certain Gaussian random vector in $\mathcal{L}^1$-norm. This later led to the approximation result in pointwise sense (Theorem 1, [8]). Before stating the result of Klartag and Eldan [8], we give a few definitions.

Definition 3.1. A function $f : \mathbb{R}^n \longrightarrow [0, \infty)$ is log-concave if $\log f$ is concave on the support of f.

We consider random vectors in $\mathbb{R}^n$ that are distributed according to a log-concave density. It is important to note at this point that while the Gaussian approximation might work for a larger set of problems, our justification for its effectiveness requires the log-concavity assumption on the underlying density. This restriction is imposed by the nature of the existing works that we build upon and it is not clear to what extent it may be relaxed.

Definition 3.2. We say that $f : \mathbb{R}^n \longrightarrow [0, \infty)$ is isotropic if it is the density function of some random variable with zero mean and identity covariance matrix. That is, f is isotropic when

$$\int_{\mathbb{R}^n} f(x)dx = 1, \qquad \int_{\mathbb{R}^n} xf(x)dx = 0 \quad \text{and,}$$
$$\int_{\mathbb{R}^n} \langle x, \theta\rangle^2 f(x)dx = \|\theta\|^2; \quad \forall\, \theta \in \mathbb{R}^n.$$

Any non-negative log concave function with $0 < \int f < \infty$ can be brought to an isotropic position via an affine map. Let (X, Y) be randomly distributed in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ with a log concave density $f > 0$, where $X \in \mathbb{R}^{n_1}$, $Y \in \mathbb{R}^{n_2}$ and n_1, n_2 are large. Then there exists an invertible affine map $A : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \longrightarrow \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ such that

$$(X', Y') := A((X, Y))$$

has an isotropic and log-concave density. Hence without loss of generality, we will assume for now that (X, Y) is a random vector in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ with isotropic log-concave density.

Remark 1. We have $(X, Y) = A^{-1}(X', Y')$. Since we denote $Proj_E$ by Γ, we have (treating Range$(\Gamma) := \mathbb{R}^n$ by zero padding),

$$\Gamma(X, Y) = \Gamma A^{-1}(X', Y')$$
$$\text{Then,} \quad A\Gamma(X, Y) = A\Gamma A^{-1}(X', Y').$$

As $(A\Gamma A^{-1})^2 = A\Gamma^2 A^{-1} = A\Gamma A^{-1}$, we have that $A\Gamma A^{-1}$ is also a projection. Consequently the same results hold for $A\Gamma A^{-1}$ as well. This justifies the assumption.

We state both total-variation and pointwise results from Klartag.

Theorem 3.3 (*[6]*). *There exists a universal constant $c > 0$ for which the following holds: Let $n \geq 3$ be an integer, and let X be a random vector in $\mathbb{R}^n$ with an isotropic, log-concave density. Let $\varepsilon > 0$ and suppose that $1 \leq k \leq c\varepsilon^2 \frac{\log n}{\log\log n}$ is an integer. Then there exists a subset $\mathcal{E} \subset G_{n,k}$ with $\sigma_{n,k}(\mathcal{E}) \geq 1 - e^{-cn^{0.99}}$ such that for any $E \in \mathcal{E}$,*

$$d_{TV}(Proj_E(X), Z_E) \leq \varepsilon$$

where Z_E is a standard Gaussian random vector in the subspace E and d_{TV} denotes the total variation distance between the respective laws.

Theorem 3.4 (*[8]*). *Let X be an isotropic random vector in $\mathbb{R}^n$ with a log-concave density. Let $1 \leq l \leq n^{c_1}$ be an integer. Then there exists a subset $\mathcal{E} \subset G_{n,l}$ with $\sigma_{n,l}(\mathcal{E}) \geq 1 - Ce^{-n^{c_2}}$ such that for any $E \in \mathcal{E}$, the following holds. Denote by f_E the density of the random vector $Proj_E(X)$, then for all $x \in E$ with $\|x\| \leq n^{c_4}$,*

$$\left|\frac{f_E(x)}{\phi_E^l(x)} - 1\right| \leq \frac{C}{n^{c_3}}.$$

Here $C, c_1, c_2, c_3, c_4 > 0$ are universal constants and ϕ_E^l denotes the standard Gaussian density in E.

Remark 2. Klartag observes that c_2 can be made arbitrarily close to 1 at the expense of decreasing the other constants, and that $c_4 \leq 1/4$. See the proof of Theorem 1 and the subsequent remark in [8] for further discussion on optimality of c_1, c_2, c_3, c_4.

In Section 2, we showed that for unscaled but stochastically sparse random vectors X and Y, one can construct Y^* and X^* such that $E[Y|X] \approx \int h(\bar{y})\Psi(\bar{y}|X^*)d\bar{y}$. Recall that $X^* = h(\bar{X})$ and $\bar{X}$ is obtained by adding $n_1 - k_1$ zeros to the projection $\check{X}$. Under the conditions of Theorem 3.4, for $1 \leq k_1 \leq n_1^{c_1}$, $\check{X}$ is approximately Gaussian with zero mean and identity covariance matrix of order k_1. If we drop the isotropy condition, we retain Gaussianity but the mean and variance will change. Using Theorems 3.3 and 3.4, we now show that $E[Y^*|\check{X}]$ can be approximated as the expectation of a function of a Gaussian.
We need the following fact:

(**F**) Let B_{C_0} denote a ball of radius C_0 in $\mathbb{R}^{k_1}$ and $\mathcal{N}$ denote the number of balls of radius η_2 required to cover B_{C_0}. Then we have the bound (see [10], Proposition 5, p. 15),

$$\mathcal{N} \leq \left(\frac{4C_0}{\eta_2}\right)^{k_1}.$$

Let $\phi(.|x)$ denote the conditional Gaussian law approximating $\bar{\Psi}(\cdot|x)$. We also assume the following:

(**A3**) For a prescribed $\eta_4 > 0$, we can pick $K_2 > 0$ such that

$$\sup_{\|\check{X}\|<C_0} E[\|h(\bar{Y})\| I\{\|h(\bar{Y})\| > K_2\} | \check{X} = \check{x}] < \eta_4$$

and,

$$\sup_{\|\check{x}\|<C_0} \int \|h(\bar{y})\| I\{\|\bar{y}\| > K_2\} \phi(\bar{y}|\check{x}) dy < \eta_4.$$

Let $(X, Y) \in \mathbb{R}^{n_1+n_2}$ be as above with the projections of X and Y on k_1 and k_2 dimensional euclidean spaces respectively are defined as earlier. Assume in addition that the conditional density $\Psi(\cdot|\cdot)$ is log-concave. Note that this implies in particular that corresponding second moments will be finite [11].

Theorem 3.5. *Let* $1 \leq k_1 \leq n_1^{c_1}$, $1 \leq k_2 \leq c\varepsilon^2 \frac{\log n}{\log \log n}$, *for* c_1 *as in Theorem 3.4 and* c, ε *as in Theorem 3.3. Then, for a prescribed* $\eta_3 > 0$,

$$\left| E[Y^*|\check{X}] - \int h(\bar{y})\phi(\bar{y}|\check{x}) d\bar{y} \right| < (K_1 + K_2 L)\eta_2 + 4\eta_4 + K_2 \epsilon \lambda$$

for some $\lambda, K_1 > 0$, *with probability* $\geq 1 - \eta_3 - \left(\frac{4C_0}{\eta_2}\right)^{k_1} e^{-cn_2^{0.99}}$, *where* C_0, η_2, K_2 *and* η_4 *are as above.*

Proof. Recall that $Y^* = h(\bar{Y})$ and $\check{X}$ is the projection of random vector $X \in \mathbb{R}^{n_1}$ on $\mathbb{R}^{k_1}$. Pick C_0 so that $P(\|\check{X}\| > C_0) < \eta_3$ for a prescribed $\eta_3 > 0$. For some $K_1 > 0$,

$$\left| \int h(\bar{y})\phi(\bar{y}|x) dy - \int h(\bar{y})\phi(\bar{y}|x') d\bar{y} \right| \leq K_1 \|x - x'\|, \quad x, x' \in B_{C_0}.$$

Let $[\check{x}]$ denote the center of the η_2-ball that is closest to $\check{x}$ among all the η_2-balls as above covering B_{C_0}, any tie being resolved arbitrarily. We have,

$$\begin{aligned} &\left| E[h(\bar{Y})|\check{X} = \check{x}] - \int h(\bar{y})\phi(\bar{y}|\check{x}) d\bar{y} \right| \\ &\leq \left| E[h(\bar{Y})|\check{X} = \check{x}] - \int h(\bar{y})\phi(\bar{y}|[\check{x}]) d\bar{y} \right| \\ &\quad + \left| \int h(\bar{y})\phi(\bar{y}|[\check{x}]) d\bar{y} - \int h(\bar{y})\phi(\bar{y}|\check{x}) d\bar{y} \right|. \end{aligned}$$

Here,

$$\left| \int h(\bar{y})\phi(\bar{y}|[\check{x}]) d\bar{y} - \int h(\bar{y})\phi(\bar{y}|\check{x}) d\bar{y} \right| \leq K_1 \eta_2$$

on $\{\check{x} : \|\check{x}\| < C_0\}$, which has probability $\geq 1 - \eta_3$. Note that $\phi(.|\check{x})$ is not isotropic but has uniformly bounded mean and variance for $\|x\| \leq C_0$. From Remark 1, we know that it can be brought to an isotropic position via an affine map. Then using Theorem 3.3 (with a factor of constant $\lambda > 0$ to take into account the non-isotropic conditional law) along with assumption (**A2**), (**A3**), we get

$$\left| E[h(\bar{Y})|\check{X} = \check{x}] - \int h(\bar{y})\phi(\bar{y}|[\check{x}]) d\bar{y} \right| < (K_1 + K_2 L)\eta_2 + 4\eta_4 + K_2 \epsilon \lambda$$

with probability $\geq 1 - \eta_3 - \left(\frac{4C_0}{\eta_2}\right)^{k_1} e^{-cn_2^{0.99}}$. $\square$

4. Concluding remarks

We have argued through a series of approximation steps that conditional expectations in high dimensions may be approximated by Gaussian integrals. From the nature of these results one expects this effect to kick in for very high dimensions. In the context of our original motivation, viz., to give some intuition why EnKF and its variants work, this makes sense because most applications of EnKF such as geophysics, meteorology, oceanography, etc., deal with *infinite dimensional* systems that are being approximated by finite dimensional caricatures for computational purposes. Nevertheless the present results are a far cry from a rigorous theory for such dynamical scenarios. Our hope is to provide an initial framework to start thinking about it.

Acknowledgments

Research of this author was supported in part by a J. C. Bose Fellowship and a grant for '*Distributed computation for optimization over large networks and high dimensional data analysis*' from the Department of Science and Technology, Government of India. This author thanks Prof. Mokshay Madiman, Uni. of Delaware, for introducing him to Klartag's work.

References

[1] S. Gillijns, O.B. Mendoza, J. Chandrasekar, B.L.R. De Moor, D.S. Bernstein, A. Ridley, What is the ensemble Kalman filter and how well does it work? in: Proc. Amer. Control Conf., 2006, pp. 4448–4453.
[2] S. Artstein, K.M. Ball, F. Barthe, A. Naor, Solution of Shannon's problem on the monotonicity of entropy, J. Amer. Math. Soc. 17 (2004) 975–982.
[3] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Birkhäuser, 2013.
[4] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inform. Theory 52 (2006) 12.
[5] D. Sullivan, For $n > 3$ there is only one finitely additive rotationally invariant measure on the n-sphere defined on all Lebesgue measurable subsets, Amer. Math. Soc. Bull. New Ser. 4 (1981) 121–123.
[6] B. Klartag, A central limit theorem for convex sets, Invent. Math. 168 (2007) 91–131.
[7] B. Klartag, Power-law estimates for the central limit theorem for convex sets, J. Funct. Anal. 245 (2007) 284–310.
[8] B. Klartag, R. Eldan, Pointwise estimates for marginals of convex bodies, J. Funct. Anal. 254 (2008) 2275–2293.
[9] B. Klartag, High-dimensional distributions with convexity properties, in: European Congress of Mathematics, Eur. Math. Soc., Zürich, 2010.
[10] F. Cucker, S. Smale, On the mathematical foundations of learning, Amer. Math. Soc. Bull. New Ser. 39 (1) (2002) 1–49.
[11] R. Adamczak, R. Latala, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Tail estimates for norms of sums of log-concave random vectors, Proc. Lond. Math. Soc. 108 (Part 3) (2014) 600–637.