Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Acad. Sci. Paris, Ser. I 340 (2005) 885–888
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Functional Analysis

Reconstruction and subgaussian processes

Shahar Mendelson [a,1], Alain Pajor [b,1], Nicole Tomczak-Jaegermann [c,2]

[a] *Centre for Mathematics and its Applications, The Australian National University, Canberra, ACT 0200, Australia*
[b] *Équipe d'analyse et mathématiques appliquées, université de Marne-la-Vallée, 5, boulevard Descartes, Champs sur Marne, 77454 Marne-la-Vallee cedex 2, France*
[c] *Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1*

Received 3 April 2005; accepted 27 April 2005

Available online 31 May 2005

Presented by Gilles Pisier

Abstract

This Note presents a randomized method to approximate any vector v from some set $T \subset \mathbb{R}^n$. The data one is given is the set T, and k scalar products $(\langle X_i, v\rangle)_{i=1}^k$, where $(X_i)_{i=1}^k$ are i.i.d. isotropic subgaussian random vectors in $\mathbb{R}^n$, and $k \ll n$. We show that with high probability any $y \in T$ for which $(\langle X_i, y\rangle)_{i=1}^k$ is close to the data vector $(\langle X_i, v\rangle)_{i=1}^k$ will be a good approximation of v, and that the degree of approximation is determined by a natural geometric parameter associated with the set T. This extends and improves recent results by Candes and Tao. ***To cite this article: S. Mendelson et al., C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

Résumé

Reconstruction et processus sous-gaussiens. Dans cette Note, on présente une méthode stochastique pour approcher un vecteur v d'une partie $T \subset \mathbb{R}^n$. Les données sont d'une part T et d'autre part k produits scalaires $(\langle X_i, v\rangle)_{i=1}^k$, où $(X_i)_{i=1}^k$ sont des vecteurs aléatoires de $\mathbb{R}^n$, indépendants de type sous-gaussiens, et $k \ll n$. On montre qu'avec une grande probabilité, tout $y \in T$ pour lequel $(\langle X_i, y\rangle)_{i=1}^k$ est proche de $(\langle X_i, v\rangle)_{i=1}^k$ est une bonne approximation de v avec un degré d'erreur déterminé par un paramètre de la géométrie de T. Cette approche permet de généraliser et d'améliorer des résultats d'un récent travail de Candes et Tao. ***Pour citer cet article : S. Mendelson et al., C. R. Acad. Sci. Paris, Ser. I 340 (2005).***

E-mail addresses: shahar.mendelson@anu.edu.au (S. Mendelson), alain.pajor@univ-mlv.fr (A. Pajor), nicole@ellpspace.math.ualberta.ca (N. Tomczak-Jaegermann).

[1] Partially supported by an Australian Research Council grant.
[2] This author holds the Canada Research Chair in Geometric Analysis.

1631-073X/$ – see front matter © 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crma.2005.04.032

The goal of this Note is to investigate the possibility of randomly reconstructing a signal or vector in $\mathbb{R}^n$, taken from a set T, using random projections. To be precise, suppose that one can select a probability measure μ on $\mathbb{R}^n$ and let $X_1, \ldots, X_k$ be independent random vectors distributed according to μ. The reconstruction problem we consider is as follows: the data we are given are the values of the projections of the unknown vector v on X_i – that is, the set of values of scalar products $\langle X_i, v\rangle$, and we study how to use this information (and the fact that $v \in T$) to find some $y \in \mathbb{R}^n$ for which the Euclidean distance $\|y - v\| \leqslant \varepsilon(k)$, where this reconstruction error must hold with high probability. For more information on reconstructions and related coding and decoding we refer to [4].

Questions of a similar flavor have been thoroughly studied in nonparametric statistics and statistical learning theory (see, for example, [3,8] and references therein). Here, we use randomized methods developed in asymptotic geometric analysis to prove a bound on $\varepsilon(k)$ in terms of the geometric structure of the set T.

The article [4] focuses on specific sets T – the ℓ_1^n unit ball and the unit ball in weak ℓ_p^n space for $0 < p < 1$. Here we show that the reconstruction process holds for an *arbitrary* set $T \subset \mathbb{R}^n$ (of course, the degree of approximation depends on the geometry of T), and using a large class of measures on $\mathbb{R}^n$ that contains the Gaussian measure and the uniform measure on $\{-1, 1\}^n$. In particular, for these measures we obtain the optimal estimates for the sets considered above. The probability we get is exponentially rather than polynomially small. Let us mention that for these sets and for the Gaussian measure exponentially small probability was obtained in [13] (using Candes and Tao's general approach). In fact the ℓ_1^n case follows from [5]. We also demonstrate that the same optimal estimates hold for a much wider family of measures.

Moreover we show that the reconstruction is robust to 'noise', in the sense that if $y \in T$, then to ensure that $\|y - v\|$ is 'small' it suffices to verify that the vector $(\langle X_i, y\rangle)_{i=1}^k$ is close, but not necessarily identical, to the data vector $(\langle X_i, v\rangle)_{i=1}^k$.

It turns out that the reason why such results hold is that under mild assumptions on the probability measure μ and the set T, the random k-dimensional projection operator $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$ satisfies that $\ker \Gamma \cap T$ has a 'small' diameter. Indeed, suppose for simplicity that T is symmetric and convex and that for an unknown $v \in T$ we find $y \in T$ which satisfies that $\langle X_i, y\rangle = \langle X_i, v\rangle$ for $1 \leqslant i \leqslant k$. Then v and y are in an affine section of T with Euclidean diameter less that the one of $T \cap \ker \Gamma$. In other words, the fact that a kernel of random projection Γ generated by μ is a section of small diameter of T suffices for the reconstruction process, and any point in T which is in the same translate of $\ker \Gamma$ as v would be a good approximation. This argument was introduced in [9] for approximating smooth functions.

To formulate the first technical tool, we require the following standard definition.

Definition 1. Let $T \subset \mathbb{R}^n$ and let $g_1, \ldots, g_n$ be independent, standard Gaussian random variables. Denote by $\ell_*(T) = \mathbb{E} \sup_{t\in T} |\sum_{i=1}^n g_i t_i|$, where $t = (t_i)_{i=1}^n \in \mathbb{R}^n$.

We use the following formulation of the main result from [11] (see also [12]). Recall that $T \subset \mathbb{R}^n$ is star-shaped if $t \in T$ and $0 \leqslant \lambda \leqslant 1$ implies $\lambda t \in T$.

Theorem 2. *There are absolute constants c and C for which the following holds. Let $1 \leqslant k \leqslant n$, let $X_1, \ldots, X_k$ be independent $N(0, I_{\mathbb{R}^n})$ Gaussian random vectors in $\mathbb{R}^n$ and let $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$. Then, for every star-shaped set $T \subset \mathbb{R}^n$, with probability larger than $1 - \exp(-ck)$ we have* $\operatorname{diam}(\ker \Gamma \cap T) \leqslant \inf\{\rho > 0;\ \ell_*(T \cap \rho B_2^n) \leqslant C\rho\sqrt{k}\}$.

Remark 1. Although the above statement did not explicitly appear in [11], it follows immediately from the proof there. The parameter $\inf\{\rho > 0;\ \ell_*(T \cap \rho B_2^n) \leqslant C\rho\sqrt{k}\}$ was introduced in [12].

A result similar to Theorem 2 (with the same probabilistic estimate) holds true for random ± 1-vectors – this follows from the result in [1], as observed in [10].

We extend Theorem 2 by showing that the diameter of a 'typical' k-codimensional section of T is small, even for other measures. Recall that the ψ_p norm of a random variable X is defined as $\|X\|_{\psi_p} = \inf\{c > 0$: $\mathbb{E} \exp(|X|^p/c^p) \leqslant 2\}$.

Definition 3. A probability measure μ on $\mathbb{R}^n$ is called isotropic if for every $y \in \mathbb{R}^n$, $\mathbb{E}|\langle X, y\rangle|^2 = \|y\|^2$, where X is distributed according to μ. A measure μ satisfies a ψ_2 condition with a constant α if for every $y \in \mathbb{R}^n$, $\|\langle X, y\rangle\|_{\psi_2} \leqslant \alpha\|y\|$.

Perhaps the most important example of an isotropic ψ_2 probability measure on $\mathbb{R}^n$ with a bounded constant other than the Gaussian measure is the uniform measure on $\{-1, 1\}^n$. Naturally, if X is distributed according to a general isotropic ψ_2 measure then the coordinates of X need no longer be independent. For example, the normalized Lebesgue measure on an appropriate multiple of the unit ball in ℓ_p^n for $2 \leqslant p \leqslant \infty$ is an isotropic ψ_2 measure with a constant independent of n and p. For more details on such measures see [10]. Our extension of Theorem 2 for isotropic ψ_2 probability measures is based on an approach from [7]. For its formulation we require the notion of the γ_2 functional [14].

Definition 4. For a metric space (T, d) let $\gamma_2(T, d) := \inf \sup_{t\in T} \sum_{s=0}^{\infty} 2^{s/2} d(t, T_s)$, where the infimum is taken with respect to all sequences of subsets $T_s \subset T$ with cardinality $|T_s| \leqslant 2^{2^s}$ and $|T_0| = 1$.

The γ_2 functional plays a central role in the theory of Gaussian processes. By the majorizing measure theorem (see [14] for the most recent survey on the topic) there are absolute constants c_1 and c_2 such that if $\{X_t\colon t \in T\}$ is a Gaussian process, and $d_2^2(s, t) = \mathbb{E}|X_s - X_t|^2$ is its covariance structure, then $c_1\gamma_2(T, d_2) \leqslant \mathbb{E}\sup_{t\in T} X_t \leqslant c_2\gamma_2(T, d_2)$.

The key lemma we require is the following estimate on the empirical L_2 diameter of a random coordinate projection of a set of functions. The proof is based on an argument from [7].

Lemma 5. *There exist $a, c > 0$ for which the following holds. Let $\alpha \geqslant 1$ and let F be a class of functions on a probability space (Ω, μ) such that for every $f, g \in F$, $\|f\|_{\psi_2(\mu)} \leqslant \alpha\|f\|_{L_2(\mu)}$, and $\|f - g\|_{\psi_2(\mu)} \leqslant \alpha\|f - g\|_{L_2(\mu)}$. Let $k \geqslant 1$ and let $X_1, \ldots, X_k$ be independent, distributed according to μ. Then, with probability at least $1 - \exp(-ak)$, we have $\sup_{f\in F}(\frac{1}{k}\sum_{i=1}^k f^2(X_i))^{1/2} \leqslant c\alpha(\sup_{f\in F}(\mathbb{E}f^2)^{1/2} + \gamma_2(F, \|\ \|_{L_2(\mu)})/\sqrt{k})$.*

The proof of the ψ_2 analog of Theorem 2 goes along the same lines as in [11], when Lemma 5 and Bernstein's inequality replace the probabilistic estimates required in the argument.

Theorem 6. *For every $\alpha \geqslant 1$ there exist $c(\alpha), C(\alpha) > 0$ for which the following holds. Let $T \subset \mathbb{R}^n$ be a star-shaped set. Let $\alpha \geqslant 1$, let μ be an isotropic ψ_2 probability measure with constant α. Let $1 \leqslant k \leqslant n$, and let $X_1, \ldots, X_k$ be independent, distributed according to μ. Let $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$. Then, with probability at least $1 - \exp(-c(\alpha)k)$, $\operatorname{diam}(\ker \Gamma \cap T) \leqslant r_k^*$, where $r_k^* = r_k^*(T) := \inf\{\rho > 0\colon \ell_*(T \cap \rho S^{n-1}) \leqslant C(\alpha)\rho\sqrt{k}\}$.*

Applying Theorem 6 one can reconstruct any $v \in T$ using the data $\langle X_i, v\rangle$ for a *general* set T. Indeed, let $\overline{T} = \{\lambda(t - v)\colon t \in T,\ 0 \leqslant \lambda \leqslant 1\}$. Then, if $t \in T$ satisfies $\langle X_i, v\rangle = \langle X_i, t\rangle$ then $t - v \in \ker \Gamma \cap \overline{T}$, and, with high probability, $\|t - v\| \leqslant r_k^*(\overline{T})$. Of course, if T happens to be convex and symmetric then $T - v \subset 2T$ which is star-shaped and thus $\|t - v\| \leqslant r_k^*(2T)$. In a more general case, when T is symmetric and quasi-convex with constant $c \geqslant 1$, i.e., $T + T \subset 2cT$, then it is star-shaped and $t - v \in \ker \Gamma \cap 2cT$. Thus, $\|t - v\| \leqslant r_k^*(2cT)$. This is the case of the unit ball T_p in weak ℓ_p $(0 < p < 1)$ for which $c = 2^{1/p-1}$. (Recall that T_p is the set of all $x = (x_i) \in \mathbb{R}^n$ such that the cardinality $|\{i\colon |x_i| \geqslant s\}| \leqslant s^{-p}$ for all $s > 0$.) Thus, the ability to approximate any point in T using this kind of random sampling depends on the expectation of the supremum of a Gaussian process indexed by the intersection of T and a sphere of a certain radius.

Let us now formulate an 'almost isometric' version of the reconstruction theorem. By almost isometric we mean that in order to obtain a good reconstruction, it suffices that the vector $(\langle X_i, y\rangle)_{i=1}^k$ is close to the data vector $(\langle X_i, v\rangle)_{i=1}^k$. The benefit is that the reconstruction algorithm is more robust. The price paid for this added flexibility is a weaker probability estimate. The proof is based on an argument from [2] and the main result of [7].

Theorem 7. *There exist $c, c' > 0$ for which the following holds. Let $T \subset \mathbb{R}^n$ be star-shaped, set $1 \leqslant k \leqslant n$ and δ such that $\exp(-ck) \leqslant \delta < 1$, and let $0 < \varepsilon < 1$. Put $\alpha > 0$, let μ be an isotropic ψ_2 measure with constant α and set $X_1, \ldots, X_k$ to be independent, distributed according to μ. If $\Gamma = \sum_{i=1}^k \langle X_i, \cdot\rangle e_i$, then with probability at least $1-\delta$, for any $t \in T$ with $\|t\| \geqslant r_k^*(\varepsilon)$ we have $(1-\varepsilon)\|t\|^2 \leqslant \|\Gamma t\|^2/k \leqslant (1+\varepsilon)\|t\|^2$, where $r_k^*(\varepsilon) = r_k^*(\varepsilon, T) := \inf\{\rho > 0\colon\ \ell_*(T \cap \rho S^{n-1}) \leqslant \varepsilon c' \rho \sqrt{k}/(\alpha \log(1/\delta))\}$.*

In particular, with probability at least $1-\delta$, every $t \in T$ satisfies $\|t\|^2 \leqslant \max\{(1-\varepsilon)^{-1}\|\Gamma t\|^2/k, r_k^*(\varepsilon)^2\}$.

Observe that Theorem 7 implies that with probability at least $1-\delta$, the k co-dimensional section $\ker \Gamma \cap T$ has a diameter at most $r_k^*(\varepsilon)$.

The parameters $r_k^*(T)$ and $r_k^*(\varepsilon, T)$ can be estimated in many cases. For example, using [6] one can estimate them for T the unit ball T_p of weak ℓ_p^n for $0 < p \leqslant 1$, or for ℓ_p^n, for $0 < p < \infty$, thus recovering and extending results from [4,13]. We will only state here an almost isometric reconstruction estimate for weak ℓ_p^n.

Corollary 8. *There is $c > 0$ such that the following holds. Let $0 < p < 1$. Let $1 \leqslant k \leqslant n$ and δ such that $\exp(-ck) \leqslant \delta \leqslant 1/4$, and let $\theta > 0$. Let μ be an isotropic ψ_2 probability measure on $\mathbb{R}^n$ with constant α, and let $X_1, \ldots, X_k$ be independent, distributed according to μ. With probability at least $1-\delta$, for any $v, y \in T_p$ for which $(\sum_{i=1}^k \langle X_i, v-y\rangle^2/k)^{1/2} \leqslant \theta$, we have $\|y-v\| \leqslant c_1(\alpha, \delta, p)(\theta + (\log(k/n)/k)^{1/p-1/2})$, where $c_1(\alpha, \delta, p)$ depends on α, δ and p.*

Proofs and related results will be presented elsewhere.

Acknowledgements

This work was done when authors A.P. and N.T.-J. were visiting the ANU, Canberra; they wish to thank the Centre for Mathematics and its Applications for its hospitality.

References

[1] S. Artstein, The change in the diameter of a convex body under a random sign-projection, in: V.D. Milman, G. Schechtman (Eds.), GAFA Seminar Notes 2002–2003, in: Lecture Notes in Math., vol. 1850, 2004, pp. 31–39.

[2] P.L. Bartlett, S. Mendelson, Empirical minimization, Probab. Theory Related Fields, in press.

[3] S. Boucheron, O. Bousquet, G. Lugosi, Theory of classification: a survey of recent advances, available from http://www.econ.upf.es/lugosi/esaimsurvey.pdf.

[4] E. Candes, T. Tao, Near optimal recovery from random projections: universal encoding strategies, Preprint.

[5] A.U. Garnaev, E.D. Gluskin, The widths of a Euclidean ball, Dokl. Acad. Nauk SSSR 277 (1984) 1048–1052 (in Russian); English translation: Soviet Math. Dokl. 30 (1984) 200–204.

[6] Y. Gordon, A. Litvak, S. Mendelson, A. Pajor, private communication.

[7] B. Klartag, S. Mendelson, Empirical processes and random projections, J. Funct. Anal., in press.

[8] S. Mendelson, in: V.D. Milman, G. Schechtman (Eds.), GAFA Seminar Notes 2002–2003, in: Lecture Notes in Math., vol. 1850, 2004, pp. 193–235.

[9] S. Mendelson, N. Tishby, Statistical Sufficiency for classes in empirical L_2 spaces, in: N. Cesa-Bianchi, S.A. Goldman (Eds.), Proceedings of the 13th Annual Conference on Computational Learning Theory COLT00, 2000, pp. 81–89.

[10] V.D. Milman, A. Pajor, Regularization of star bodies by random hyperplane cut off, Studia Math. 159 (2003) 247–261.

[11] A. Pajor, N. Tomczak-Jaegermann, Subspaces of small codimension of finite-dimensional Banach spaces, Proc. Amer. Math. Soc. 97 (4) (1986) 637–642.

[12] A. Pajor, N. Tomczak-Jaegermann, Nombres de Gelfand et sections euclidiennes de grande dimension, in: Séminaire d'Analyse Fonctionelle 1984/1985, in: Publ. Math. Univ. Paris VII, vol. 26, Univ. Paris VII, Paris, 1986, pp. 37–47 (in French) (Gelfand numbers and high-dimensional Euclidean sections).

[13] M. Rudelson, R. Vershynin, Geometric approach to error correcting codes and reconstruction of signals, Preprint.

[14] M. Talagrand, The Generic Chaining, Springer-Verlag, in press.