RECOVERY OF FUNCTIONS OF MANY VARIABLES VIA COMPRESSIVE SENSING

Albert Cohen[1], *Ronald A. DeVore*[2], *Simon Foucart*[3], *Holger Rauhut*[4]

[1] Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie
4 Place Jussieu, 75005 Paris, France, cohen@ann.jussieu.fr,
[2] Texas A&M University, College Station, TX 77843-3368, USA, rdevore@math.tamu.edu,
[3] Drexel University, Department of Mathematics, 269 Korman Center,
3141 Chestnut Street, Philadelphia, PA 19104, USA, foucart@math.drexel.edu,
[4] Hausdorff Center for Mathematics & Institute for Numerical Simulation
Endenicher Allee 60, D-53115 Bonn, Germany, rauhut@hcm.uni-bonn.de.

ABSTRACT

Recovery of functions of many variables from sample values usually suffers the curse of dimensionality: The number of required samples scales exponentially with the spatial dimension. In order to avoid this severe bottleneck, one needs to impose further structural properties of the function to be recovered apart from smoothness. Here, we build on ideas from compressive sensing and introduce a function model that involves "sparsity with respect to dimensions" in the Fourier domain. Using recent estimates on the restricted isometry constants of measurement matrices associated to randomly sampled trigonometric systems, we show that the number of required samples scales only logarithmically in the spatial dimension provided the function to be recovered follows the newly introduced high-dimensional function model.

Keywords— Functions in high dimensions, compressive sensing, sparse Fourier expansions, Fourier algebra, restricted isometry property.

1. INTRODUCTION

Many applied problems lead to the problem of recovering a function of a large number of variables. We mention areas like machine learning, mathematical finance, numerical simulation of stochastic PDEs and in quantum mechanics. Typically, one faces the curse of dimensionality, which makes such problems notoriously hard. In this paper we suggest a new function model and an associated reconstruction method that avoids the curse of dimensionality to some extent.

Assume that $f : [0,1]^d \to \mathbb{C}$ is a function of d variables, where d is large. Our goal is to approximate f accurately based on sample values $f(x_1), \ldots, f(x_m)$, with $x_\ell \in [0,1]^d$, $\ell = 1, \ldots, m$.

The following is well-known: Suppose that f is s-times continuously differentiable, $f \in C^s([0,1]^d)$. Then there exists sampling points $t_1, \ldots, t_m$ and a linear reconstruction method, which computes a function $\widetilde{f}$ from the sampling values such that

$$\|f - \widetilde{f}\|_\infty \le C \|f\|_{C^s} m^{-s/d}, \tag{1}$$

where $\|g\|_\infty := \sup_{t \in [0,1]^d} |g(t)|$ as usual. The appearance of d in the exponent s/d is commonly called the curse of dimensionality, as it severely deteriorates the reconstruction error in high dimensions. Expressed differently, we require a number of samples $m \ge c\varepsilon^{-d/s}$ in order to have the reconstruction error below ε, that is, we have an exponential scaling of m in d.

The estimate (1) is sharp in the sense that for each set of m points $x_1, \ldots, x_m$ and each reconstruction map we can find a function in $C^s([0,1]^d)$, such that the reconstruction error cannot be smaller than the right hand side in (1), see [5]. In other words, one needs to impose further assumptions on f aside from smoothness in order to avoid (or at least weaken) the curse of dimension.

2. SPARSITY WITH RESPECT TO DIMENSIONS

Our goal is to introduce a function model which involves more structure than just smoothness, and thereby allows to significantly reduce the number of required samples. We will exploit recent ideas from compressed sensing and introduce a structured sparsity model.

For $k \in \mathbb{Z}^d$, let

$$\phi_k(t) := e^{2\pi i k \cdot t}, \quad t \in [0,1]^d \ .$$

We consider functions which can be well approximated by a sparse expansion of the form

$$g(t) = \sum_{k \in S} x_k \phi_k(t) \ , \tag{2}$$

where $S \subset \mathbb{Z}^d$ is a finite set. Its cardinality is called the sparsity of x or of g, respectively. In sparse approximation or in compressed sensing, one often makes only a constraint on the sparsity; one requires it to be less than s, say. In our context, we

H.R. acknowledges funding through the Hausdorff Center for Mathematics Bonn and the WWTF project SPORTS (MA07-004).

impose further restrictions on S. Indeed, we assume that it is of the form

$$S = [-n_1, n_1] \times [-n_2, n_2] \times \cdots \times [-n_d, n_d] \cap \mathbb{Z}^d , \quad (3)$$

in addition to having cardinality at most s. This means in particular that

$$\prod_{\ell=1}^{d} (2n_i + 1) \le s . \quad (4)$$

This requirement tells us that only a few n_i can be large, and the remaining ones have to be small. Since the vector $n = (n_1, \dots, n_d)$ will not be prescribed, such setup can be interpreted as sparsity with respect to the dimensions: the variables which play the most important role are not known in advance. We introduce $\mathcal{M}_s$ as the collection of all sets $S \in \mathbb{Z}^d$ of the form (3) satisfying (4).

In order to compute an approximation of f of sparsity s, one has to take into account all possible subsets S of the form (3) with the constraint (4). The subset of indices of $\mathbb{Z}^d$ that are contained in one of these sets is a hyperbolic cross,

$$H_s^d := \{k \in \mathbb{Z}^d, \prod_{\ell=1}^{d} (2k_\ell + 1) \le s\} = \bigcup_{S \in \mathcal{M}_s} S.$$

The size of a hyperbolic cross can be estimated as

$$\# H_s^d \le C s \min\{\ln(s)^{d-1}, d^{\ln(s)}\} . \quad (5)$$

3. RECOVERY VIA COMPRESSIVE SENSING

Given sample values $f(t_1), \dots, f(t_m)$ at points $t_1, \dots, t_m \in [0,1]^d$ we propose to compute the Fourier coefficients of the approximation to f via techniques from compressed sensing. Let $A \in \mathbb{C}^{m \times N}$ with $N := \# H_s^d$ be the sampling matrix with entries

$$A_{\ell,k} = \phi_k(t_\ell), \quad \ell = 1, \dots, m, \ k \in H_s^d . \quad (6)$$

A by-now classical reconstruction technique is ℓ_1-minimization,

$$\min_x \|x\|_1 \quad \text{subject to } \|Ax - y\|_2 \le \eta ,$$

where $\|x\|_p = (\sum_{\ell=1}^m |x_\ell|^p)^{1/p}$ for $1 \le p < \infty$ as usual, and η is a suitable parameter reflecting the noise level. Other methods, such as CoSaMP [11], Iterative Hard Thresholding [1, 9] and Hard Thresholding Pursuit [8] are applicable as well.

A basic tool for the analysis of these algorithms is the restricted isometry property (RIP) [2, 3]. The restricted isometry constant δ_s of a matrix A is the smallest number such that

$$(1 - \delta_s)\|x\|_2^2 \le \|Ax\|_2^2 \le (1 + \delta_s)\|x\|_2^2 \quad \text{for all } s\text{-sparse } x .$$

Informally, A is said to possess the RIP if δ_s is small for reasonable large s. If

$$\delta_{\kappa s} \le \delta_*$$

for suitable constants $\kappa \in \mathbb{N}$ and $\delta_* < 1$ then the reconstruction $x^\#$ obtained from applying one of the mentioned algorithms to the noisy measurements $y = Ax + z$ with $\|z\|_2^2 \le \eta$ satisfies

$$\|x - x^\#\|_2 \le C_1 \frac{\sigma_s(x)_1}{\sqrt{s}} + C_2 \eta,$$
$$\|x - x^\#\|_1 \le D_1 \sigma_s(x)_1 + D_2 \sqrt{s} \eta.$$

Here the quantity $\sigma_s(x)_1$ is the best s-term approximation error of x in ℓ_1, that is,

$$\sigma_s(x)_1 := \inf_{z \text{ is } s\text{-sparse}} \|x - z\|_1.$$

The constants κ and δ_* depend on the algorithm. For ℓ_1-minimization, for instance, the best known constants are $\kappa = 2$ and $\delta_* = 0.4652$ [7].

As we would like to apply these error estimates, it is important to clarify whether the measurement matrix (6) associated to our setup satisfies the RIP. Since this question is very difficult to answer in general, we assume that the sampling points $t_1, \dots, t_m$ are chosen independently at random according to the uniform measure on $[0,1]^d$. Set $N := \# H_s^d$. If

$$m \ge C \delta^{-2} s \ln(s)^3 \ln(N), \quad (7)$$

then the restricted isometry constant δ_s of the rescaled matrix $\frac{1}{\sqrt{m}} A$ satisfies $\delta_s \le \delta$ with probability at least $1 - N^{-\gamma \ln^3(s)}$ [10]. Together with the estimate (5), the bound on the samples (7) is satisfied whenever

$$m \ge C' \ln(d) s \ln^4(s) .$$

4. FOURIER ALGEBRAS IN HIGH DIMENSIONS

As the next step we introduce a class of functions in high dimensions which is suitable for our purposes. As we work with trigonometric expansions, it turns out to be natural to measure errors in the Fourier algebra rather than with respect to the supremum norm. For a continuous function $f \in C([0,1]^d)$ we denote by

$$c_k(f) = \int_{[0,1]^d} f(x) e^{-2\pi i k \cdot t} dt, \quad k \in \mathbb{Z}^d,$$

its Fourier coefficients. The Fourier algebra $A_1 = A_1([0,1]^d)$ consists of all functions with summable Fourier coefficients. Its norm is given by

$$\|f\|_{A_1} := \sum_{k \in \mathbb{Z}^d} |c_k(f)| .$$

It is well-known that $\|f\|_\infty \le \|f\|_{A_1}$. Now let $\alpha = (\alpha_1, \dots, \alpha_d)$ be a vector of smoothness indices $\alpha_\ell > 0$. We introduce then the anisotropic smoothness space A_α is the space of functions f with finite norm

$$\|f\|_{A_\alpha} := \sum_{k \in \mathbb{Z}^d} |c_k(f)| (1+|k_1|)_1^{\alpha} \cdot (1+|k_2|)^{\alpha_2} \cdots (1+|k_d|)^{\alpha_d} .$$

This is a kind of anisotropic Hölder space where smoothness is measured in $A_1([0,1]^d)$ instead of $C([0,1]^d)$. A function in A_α is required to be very smooth in directions where α_ℓ is large and it maybe rough in directions where α_ℓ is small.

As already announced we do not wish to prescribe directions in which f is allowed to be rough, but we would only like to restrict the number of "rough directions". We model such functions by a union of anisotropic Fourier algebras. For a parameter $r > 0$ we set

$$\mathcal{A}_r := \bigcup_{\sum_\ell \alpha_\ell^{-1} \le r^{-1}} A_\alpha \ .$$

An example of a function f in this class would be a function that is constant in most directions, and only depends on $k \ll N$ variables, $f(x) = g(x_{i_1}, \dots, x_{i_k})$, where $g \in A_\alpha([0,1]^k)$, $\alpha = kr(1,1,\dots,1)$. This is the case, for instance, if $\frac{\partial^\beta g}{\partial x_j^\beta}$, $j = 1, \dots, k$, exists for all $\beta = 1, \dots, n$ with $n > kr + 1$. The crucial point is that the variables $x_{i_1}, \dots, x_{i_k}$ on which f depends are not prescribed in advance. This example is similar to the setup in [6]; but the set $\mathcal{A}_r$ allows more general functions, so that our setup seems to be more flexible than [6].

Functions in $\mathcal{A}_r$ can be well-approximated by sparse expansions of the form (2), where $S \in \mathcal{M}_s$, that is, S satisfies (3) and (4). In order to measure the approximation error, we introduce

$$\widetilde{\sigma}_s(f) := \widetilde{\sigma}_{\mathcal{M}_s}(f)_{A_1} := \inf_{S \in \mathcal{M}_s, x:\operatorname{supp} x \subset S} \|f - \sum_{k \in S} x_k \phi_k\|_{A_1}$$
$$= \inf_{S \in \mathcal{M}_s} \sum_{k \in \mathbb{Z}^d \setminus S} |c_k(f)| \ .$$

If f is contained in $\mathcal{A}_r$ then one can show that

$$\widetilde{\sigma}_s(f) \le C \min_{\alpha: \sum \alpha_\ell^{-1} \le r^{-1}} \|f\|_{A_\alpha} s^{-r} \ .$$

Our goal is to achieve the same rate of convergence, having only sampling values of f at our disposal. Of course, we use all the preparations above for attacking this task. In particular, we exploit the restricted isometry property and error estimates from compressed sensing. Working out all the details leads to the following result.

Theorem 1. *Let $r > 0$ and $s > 0$. Assume that m is such that*

$$m \ge C \left(1 + \frac{2}{r}\right) \ln(d) s \ln^4(s)$$

Then there eixsts a set of m sampling points $t_1, \dots, t_m \in [0,1]^d$ such that for every $f \in \mathcal{A}_r$ an approximation $\widetilde{f}$ can be reconstructed from the samples $f(t_1), \dots, f(t_m)$ with $\operatorname{supp}(c_k(\widetilde{f})) \subset H_s^d$. The approximation error satisfies

$$\|f - \widetilde{f}\|_{A_1} \le C' \min_{\alpha: \sum \alpha_\ell^{-1} \le r^{-1}} \|f\|_{A_\alpha} \ s^{-r} \ .$$

The sampling points in the theorem are taken at random (which allows to use corresponding estimates for the RIP outlined above). Since $\|\cdot\|_\infty \le \|\cdot\|_{A_1}$ we obtain an estimate in the supremum norm as well.

The crucial point in the above result is, of course, that the number of required sampling points only scales logarithmically in d opposed to the exponential scaling in (1). Therefore, the curse of dimensionality can be avoided (at least to some extend) using ideas from compressed sensing.

Detailed proofs will appear in the paper [4].

5. REFERENCES

[1] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.

[2] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[3] E.J. Candès, J.K.. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[4] A. Cohen, R.A. DeVore, S. Foucart, and H. Rauhut. Revory of functions in high dimensions via compressive sensing. in preparation.

[5] R. DeVore, R. Howard, and C. Micchelli. Optimal nonlinear approximation. *Manuscripta Math.*, 63:469–478, 1989.

[6] R. DeVore, G. Petrova, and P. Wojtaszczyk. Approximation of functions of few variables in high dimensions. *Constr. Approx.*, 33(1):125–143, 2011.

[7] S. Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Appl. Comput. Harmon. Anal.*, 29(1):97–103, 2010.

[8] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *preprint*, 2010.

[9] S. Foucart. Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants. In *Proceedings of the 13th International Conference on Approximation Theory*, 2010.

[10] H. Rauhut. Compressive Sensing and Structured Random Matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.

[11] J.A. Tropp and D. Needell. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2008.