[3] Y. Maday, A. T. Patera, and G. Turinici, Global a priori convergence theory for reduced-basis approximations of single-parameter symmetric coercive elliptic differential equations, C. R. Acad. Sci., Paris, Ser. I, Math., 335,2002, 289–294.

[4] G. Rozza, D.B.P. Huynh, and A.T. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations application to transport and continuum mechanics, Arch. Comput. Method. E., 15, 2008, 229–275.

Anisotropic Representations And Function Spaces In R^n

Shai Dekel

(joint work with Pencho Petrushev and Tal Weissblat)

We investigate representation systems and function spaces over multi-level ellipsoid covers of R^n which may change rapidly from point to point and in depth, from level to level (see previous papers [DHP, DDP]). At this time we are focused on Triebel-Lizorkin spaces and in particular the Hardy spaces. We generalize previous work [B] and classic elements of the Hardy space theory in this setting such as the various maximal function definition, atomic decompositions, the dual BMO spaces, etc.

References

[B]   M. Bownik, Anisotropic Hardy spaces and wavelets, Mem. Amer. Math. Soc.164 (2003), no 781.

[DHP] S. Dekel, Y. Han and P. Petrushev, Anisotropic meshless frames on R^n, Fourier Analysis and Applications 15 (2009), 634-662.

[DDP] W. Dahmen, S. Dekel and P. Petrushev Two-level splits of anisotropic Besov spaces, Constructive Approximation 31 (2010), 149-194.

Sparsity in Learning Theory

Christine De Mol

(joint work with Ernesto De Vito, Sofia Mosci, Lorenzo Rosasco, Magali Traskine and Alessandro Verri)

In contemporary science, we are more and more often faced with the problem of extracting meaningful information or inferring models from a data-rich environment. For example, in bioinformatics, one measures by means of microarrays, for each patient (or experiment) i, a high-dimensional vector of expression levels x_i of p genes. In a supervised learning setting, besides these "input" data, one is given, for each patient i, a response or "output" y_i which can be a real-valued index (survival time or gravity of an illness) or else, in classification problems, a discrete label discriminating between e.g. two different pathological states. A first approach consists in assuming a linear relationship between output and input, i.e. that y_i is just the scalar product of x_i with a p-dimensional vector β. Two distinct problems are of interest: (i) the prediction or "generalization" problem consisting in predicting the response y for new patients to come on the basis of their

gene expression data and (ii) the identification of the vector β defining the model. This latter problem is also referred to as "variable selection" when the vector β is assumed to be sparse, i.e. to contain many zeroes corresponding to irrelevant predictors/variables. To cope with the fact that there are generically many more variables (genes) than patients or experiments, i.e. to address the so-called "large p, small n paradigm", the problem can be reformulated as a multivariate least-squares regression with a regularizing penalty whose aim is to provide the dimension reduction necessary to get stable estimates of β. In so-called "ridge" regression, one uses a quadratic penalty, namely the square of the euclidean norm of the vector β (L_2-norm), whereas in "lasso" regression [Ti], one uses instead a L_1-norm penalty which enforces sparsity and allows for variable selection. In the presence of correlation among the variables, however, the lasso presents some drawbacks which can be overcome by the use of an additional L_2-norm penalty. This allows to select sparse groups of correlated variables without knowing in advance the composition of the groups (for known groups one could use instead the so-called "group lasso" or "joint sparsity" strategies). This "elastic-net" strategy was proposed by Zou and Hastie [ZH] for fixed-design linear regression. In the paper [DDR], we extend this setting to the framework of supervised learning theory, i.e. of random-design nonlinear regression. The regression function f_β is assumed to have a sparse expansion, with coefficients β_γ, on the elements (atoms or features) φ_γ of a possibly infinite dictionary: $f_\beta(x) = \sum_\gamma \beta_\gamma \varphi_\gamma(x)$. For example, one could consider frames of wavelets or of some of their relatives. On the basis of a collection of examples (training set) made of n independent input-output random pairs (X_i, Y_i), $i = 1, \ldots, n$ (the inputs belong to a separable metric space and the outputs to $\mathbb{R}$ or to a separable Hilbert-space), distributed according to an unknown probability distribution, we define the following elastic-net estimator for the regression function

$$\beta_\lambda^n = \mathrm{argmin}_\beta \left[\frac{1}{n} \sum_{i=1}^n |Y_i - f_\beta(X_i)|^2 + \lambda \sum_\gamma (u_\gamma |\beta_\gamma| + \varepsilon v_\gamma |\beta_\gamma|^2) \right]$$

where λ is a tunable positive regularization parameter, whereas $\varepsilon > 0$ is considered as fixed, and where u_γ and v_γ are two sets of weights (positive and bounded from below) encoding the regularity of the regression function through the assumptions $\sum_\gamma u_\gamma |\beta_\gamma| < \infty$ and $\sum_\gamma v_\gamma |\beta_\gamma|^2 < \infty$. In [DDR], by means of appropriate concentration inequalities, we derive consistency results for this estimator as $n \to \infty$, both for prediction and for variable/feature selection. Our results include finite-sample bounds and an adaptive scheme to select the regularization parameter λ. Thanks to the additional quadratic penalty, stability is guaranteed with respect to the fluctuations arising from random design and no assumptions have to be made to restrict the possible correlations between the features as usually done in the literature dealing with pure lasso strategies.

The paper [DMTV] deals with an application to bioinformatics and proposes a new method, based on linear regression with an elastic-net penalty, to select

relevant groups of correlated genes from microarray data. Using a two-stage approach and an appropriate tuning of the regularization parameters, we are able to demonstrate the good performances of the method on benchmark microarray data sets and to produce expanding gene lists which are almost perfectly nested when increasing the parameter ε. The proposed methodology could also be applied to other practical problems where the goal is to select relevant variables or features in the presence of high correlation among certain groups of these variables.

References

[DDR] C. De Mol, E. De Vito, L. Rosasco, Elastic-net regularization in learning theory, J. Complexity, 25 (2), 2009, 201–230.

[DMTV] C. De Mol, S. Mosci, M. Traskine, A. Verri, A Regularized Method for Selecting Nested Groups of Relevant Genes from Microarray Data, J. Comp. Biol., 16 (5), 2009, 677–690.

[Ti] R. Tibshirani, Regression Selection and Shrinkage via the Lasso, J. R. Statist. Soc. B, 58 (1), 1996, 267–288.

[ZH] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Statist. Soc. B, 67 (2), 2005, 301–320.

Hard Thresholding Pursuit: An Algorithm for Compressive Sensing

Simon Foucart

We introduce a new iterative algorithm to find s-sparse solutions $\mathbf{x} \in \mathbb{C}^N$ of underdetermined linear systems $A\mathbf{z} = \mathbf{y}$, $A \in \mathbb{C}^{m \times N}$, $\mathbf{y} \in \mathbb{C}^m$. The algorithm, called Hard Thresholding Pursuit, is a simple combination of the Iterative Hard Thresholding [BD1, BD2] algorithm and of the Compressive Sampling Matching Pursuit [NT] or Subspace Pursuit [DM] algorithms. It reads:
Start with an s-sparse $\mathbf{x}^0 \in \mathbb{C}^N$, typically $\mathbf{x}^0 = 0$, and iterate the scheme

$$(\text{HTP}_1) \qquad S^{n+1} = \big\{ \text{indices of } s \text{ largest entries of } \mathbf{x}^n + A^*(\mathbf{y} - A\mathbf{x}^n) \big\},$$

$$(\text{HTP}_2) \qquad \mathbf{x}^{n+1} = \operatorname{argmin}\big\{ \|\mathbf{y} - A\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^{n+1} \big\},$$

until the stopping criterion $S^{n+1} = S^n$ is met.
We first notice that the sequence $(\mathbf{x}^n)$ is eventually periodic, so that, assuming convergence of the algorithm, its limit is exactly achieved after a finite number of iterations. Next, we give a short and elegant proof of the following theorem:
Suppose that the $3s$th order restricted isometry constant of the matrix $A \in \mathbb{C}^{m \times N}$ satisfies $\delta_{3s} < 1/\sqrt{3}$. Then, for any s-sparse $\mathbf{x} \in \mathbb{C}^N$, the sequence $(\mathbf{x}^n)$ defined by the Hard Thresholding Pursuit algorithm with $\mathbf{y} = A\mathbf{x}$ converges towards $\mathbf{x}$ at a geometric rate given by

$$\|\mathbf{x}^n - \mathbf{x}\|_2 \le \rho^n \|\mathbf{x}^0 - \mathbf{x}\|_2, \qquad\qquad \rho := \sqrt{\frac{2\delta_{3s}^2}{1 - \delta_{2s}^2}} < 1.$$

We remark that the same result (with a different ρ) holds for fast versions of the algorithm, where the projection step (HTP_2) is replaced by any number of gradient descent iterations. We also remark that the result extends to the case of non-sparse