Mathematisches Forschungsinstitut Oberwolfach

Report No. 51/2006

Mini-Workshop: Statistical Methods for Inverse Problems

Organised by
Marc Hoffmann (Marne-la-Vallée)
Markus Reiß (Heidelberg)

November 26th – December 2nd, 2006

ABSTRACT. Inverse problems appear naturally in a broad range of applications. Numerical analysis and statistics have – often independently – developed methods for regularisation and inversion. The aim of this mini-workshop is to bring together these methods and to consider their use in applications, with a focus on mathematical finance.

Introduction by the Organisers

The mini-workshop on "Statistical methods for inverse problems" gathered seventeen people from the area of statistics and numerical analysis. The goal of the workshop was to stimulate discussions around results and methods that are commonly used by different scientific communities. These methods mainly concern regularization of linear and nonlinear inverse problems in presence of deterministic or stochastic noise.

The week was articulated around three main lectures, divided into two talks each. Yuri Golubev (Université of Marseille) started on Monday-Tuesday mornings to give an extended lecture on inverse problems from a statistical perspective. He was quickly followed by Thorsten Hohage (University of Göttingen) who gave lectures on Monday-Tuesday afternoon on inverse problems from a numerical analysis perspective. A third lecture was given on Wednesday-Thursday morning by Rama Cont (École Polytechnique) who addressed the issue of inverse problems in finance, with both deterministic and probabilistic points of view. The main talks were completed by informal contributed talks of approximately one hour each, at the approximate rate of two talks per day except for the mandatory walk in the Schwarzwald, which was unfortunately canceled due to bad weather conditions that concentrated on the only day of rest!

Round tables were planned to follow the lectures, but the friendly and informal atmosphere soon raised constant questions and discussions among the participants during the lectures so that the workshop progressively moved to informal, yet intense scientific discussions: particular focus was given on optimal tuning parameter choice, stochastic or deterministic error modeling and the complexity of numerical schemes. These discussions took place during several extended talks that skipped the initially planned schedule, fortunately tightened by the strict meal hours of the Oberwolfach center.

The excellent atmosphere of the mini-workshop was made possible thanks to the exceptional working conditions at the MFO to which we would like to express our deep gratitude.

Marc Hoffmann and Markus Reiß

Workshop: Mini-Workshop: Statistical Methods for Inverse Problems

Table of Contents

Abstracts

Near optimal approximation of arbitrary signals from highly uncomplete measurements
ALBERT COHEN

The typical paradigm for obtaining a compressed version of a discrete signal represented by a vector $x \in I\!\!R^N$ is to choose an appropriate basis, compute the coefficients of x in this basis, and then retain only the k largest of these with $k < N$. Assuming, without loss of generality, that x already represents the coefficients of the signal in the appropriate basis, and denoting by $S_k \subset \{1, \cdots, N\}$ the set of indices corresponding to these coefficients, the performance that we achieve by such an approximation process, is given in the ℓ_p norm by the *best k-term approximation* error

$$(1) \qquad \sigma_k(x)_{\ell_p} := \|x - x_{S_k}\|_{\ell_p} = \|x_{S_k^c}\|_{\ell_p},$$

where x_S denotes the vector obtained from x by setting to 0 all its component with indices not in S. This approximation process should be considered as *adaptive* since the indices of those coefficients which are retained vary from one signal to another.

The viewed expressed by the theory of *compressed sensing* as developed by Candes, Romberg and Tao [5, 6] and by Donoho [8] is that it is possible to actually compute only a few *non-adaptive* linear measurements and still retain the necessary information about x in order to build a compressed representation. These measurements are represented by a vector

$$(2) \qquad y = \Phi x,$$

of dimension $n < N$ where Φ is the $n \times N$ measurement matrix. The recovery of an approximation x^* of x from these measurements is performed by an operator Δ which we refer to as the decoder. In contrast to Φ, this operator is allowed to be non-linear.

In recent years, considerable progress has been made concerning the possible choices for the measurement matrix Φ and the performances of various decoders Δ for such choices. A natural issue is to compare the performances of compressed sensing and best k-term approximation. In [2], we adressed this issue in the following general terms:

For a given norm $X = \ell_p$ what is the minimal value of n, for which there exists an encoding-decoding pair (Φ, Δ) such that

$$(3) \qquad \|x - \Delta(\Phi x)\|_X \leq C_0 \sigma_k(x)_X,$$

for all $x \in I\!\!R^N$, with C_0 a constant independent of k and N?

We say that a pair (Φ, Δ) which satisfies property (3) is instance optimal of order k. It was shown that the answer to the above question heavily depends on the ℓ_p norm under consideration. Let us illustrate this by two contrasting results from [2]:

(1) In the case $p = 1$, it is possible to build pairs which are instance-optimal of order k with $n \geq ak \log(N/k)$ measurements with a a fixed constant. Moreover the decoder that ensures Δ may be defined as the solution of the minimization problem

$$(4) \qquad \Delta(y) := \mathrm{Argmin}_{\Phi z = y} \|z\|_{\ell_1}.$$

These facts can also be derived from the results of Candes, Romberg and Tao in [6]. Therefore, in order to obtain the same accuracy, the amount n of non-adaptive measurements should only exceed the amount k of adaptive measurements by the small logarithmic factor $a \log(N/k)$.

(2) In the case $p = 2$, if (Φ, Δ) is an instance-optimal pair of order $k = 1$, then the number of measurement n is always larger than aN where a is related to C_0 in (3). Therefore, the needed amount of non-adaptive measurements has to be very large in order to compete with even one single adaptive measurement.

Therefore, instance-optimality does not seem like a viable concept in ℓ_2, which is often the measure of interest in signal processing. A more optimistic result was nevertheless established by Candes, Romberg and Tao in [6]: with $n \geq ak \log(N/k)$, it is possible to build pairs (Φ, Δ) such that for all $x \in \mathbb{R}^N$,

$$(5) \qquad \|x - \Delta(\Phi x)\|_{\ell_2} \leq C_0 \frac{\sigma_k(x)_{\ell_1}}{\sqrt{k}},$$

with the decoder again defined by (4). This implies in particular that k-sparse signals are exactly reconstructed and that signals x such that $\|x\|_{w\ell_p} \leq M$ for some $p < 1$ are reconstructed with accuracy $C_0 M k^{-s}$ with $s = 1/p - 1/2$ which is of the same order as the best estimate available on $\sigma_k(x)_{\ell_2}$ for such signals.

It is also known that instance-optimality in ℓ_2 can be recovered if one accepts a probabilistic statement in which Φ is a random matrix-valued variable. A first result in this direction, obtained by Cormode and Mutukrishnan in [4], shows how to construct a random matrix Φ with $n \sim k(\log N)^{5/2}$ measurements and a decoder Δ such that for any $x \in \mathbb{R}^N$,

$$(6) \qquad \|x - \Delta(\Phi x)\|_{\ell_2} \leq C_0 \sigma_k(x)_{\ell_2}$$

with overwhelming probability (larger than $1 - \varepsilon(n)$ where $\varepsilon(n)$ tends fastly to 0 as $n \to +\infty$).

In [2], we introduced a probabilistic criterion on Φ which ensures instance optimality in the above sense. In addition to the result of [4], this approach allowed us on the one hand to consider fairly general classes of random matrices (such as Gaussian and Bernoulli) and on the other hand to reduce the number of needed measurements down to $n \sim k \log(N/k)$. However, the decoder which was proposed in [2] was defined by minimizing $\|x - \Phi x\|_{\ell_2}$ over all k-sparse vectors, a task which cannot be achieved in any reasonable computational time.

In [9], Gilbert and Tropp proposed to use a greedy procedure, known as Orthogonal Matching Pursuit (OMP) algorithm, in order to define $\Delta(y)$. The algorithm

can be described as follows : denoting by $(\phi_i)_{i=1,\cdots,N}$ the columns of the matrix Φ, we first define

(7)
$$j_1 := \mathrm{argmax}_{j=0,\cdots,N}\langle y, \phi_j\rangle,$$

and approximate y by its projection $y^1 := z_{j_1}\phi_{j_1}$ with $z_{j_1} := \langle y, \phi_{j_1}\rangle/\|\phi_{j_1}\|^2$. At the step i of the algorithm, we have defined a set of indices $\{j_1,\cdots,j_i\}$ and $y^i = \sum_{l=1}^{i} z_{j_l}\phi_{j_l}$ denotes the orthogonal projection of y onto $\mathrm{Span}\{\phi_{j_1},\cdots,\phi_{j_l}\}$. The new index is defined by

(8)
$$j_{i+1} := \mathrm{argmax}_{j=0,\cdots,N}\langle r_i, \phi_j\rangle,$$

where $r^i := y - y^i$ is the residual. The components $(z_{j_1},\cdots,z_{j_{j_i}})$ define a sparse approximation to x that we denote by x^i and which is supported by $\{j_1,\cdots,j_i\}$:

(9)
$$x_j^i = z_j, \text{ if } j \in \{j_1,\cdots,j_i\}, \ 0 \text{ otherwise.}$$

The following striking result was proved in [9] for fairly general classes of random matrices (such as Gaussian and Bernoulli): under the condition $n \geq ak\log N$, for all k-sparse vector x, the OMP algorithm returns exactly $x^k = x$ after k iterations, with probability greater than $1 - N^{-b}$ where b can be made arbitrarily large by taking a large enough.

Very recently [3], we have proved that the OMP is not only a valid strategy for the recovery of k-sparse vectors, but also for arbitrary N-dimensional vectors, in the sense that these can be recovered up to the accuracy of best k-term approximation. In other words, for such general vectors, the decoder defined by the application of the OMP algorithm on the data y satisfies with high probability the property (6) of instance-optimality in ℓ_2.

It is interesting to remark that for general dictionaries, the OMP algorithm is known to converge slowly: its approximation error $\|y-y^k\|$ can at best be bounded by $k^{-1/2}$ (see [7] and [1] for a general discussion on the rate of convergence). In the present setting, its improved convergence properties are strongly tied to the probabilistic properties of Φ.

More specifically, our main result, which we state below, relies on three properties of the matrix Φ

(P0) : the columns of Φ are statistically independent.

(P1) : for any $x \in \mathbb{R}^N$, we have $\mathrm{Prob}\{|\|\Phi x\|_{\ell_2}^2 - \|x\|_{\ell_2}^2| \geq \delta\|x\|_{\ell_2}^2\} \leq b_1 e^{-c_1 n\delta^2}$, where b_1 and c_1 are absolute constants.

(P2) : for any $z \in \mathbb{R}^n$, we have $\mathrm{Prob}\{|\langle z, \phi_l\rangle| \geq \delta\|z\|_{\ell_2}\} \leq b_2 e^{-c_2 n\delta^2}$, where b_2 and c_2 are absolute constants.

The validity of (P1) and (P2) can be proved for various classes of random matrices such as Gaussian and Bernoulli.

Theorem 1. *There exists a fixed constant $C_0 > 0$ such that if the random matrix satisfies* (P0), (P1) *and* (P2), *then the vector $x^* = x^{2k}$ obtained after $2k$ iterations of the OMP algorithm satisfies*

(10)
$$\|x - x^*\|_{\ell_2} \leq C_0 \sigma_k(x)_{\ell_2},$$

with probability larger than $1 - \varepsilon$ where

$$\varepsilon = (3+4kN)b_1 e^{-c_1 n} + (4k^2 + 2kN)b_2 e^{-c_2 n/k} + 2kN b_2 e^{-c_2 \frac{n}{288k}} + b_1 e^{-n\frac{c_1}{16} + 3k[\log(\frac{4N}{k})]}.$$

As a consequence, for all $b > 0$, there exists $a > 0$ such that (10) holds with probability larger than $1 - N^{-b}$ provided that $n \geq ak \log N$.

REFERENCES

[1] A. Barron, A. Cohen, W. Dahmen and R. DeVore, *Approximation and learning by greedy algorithms*, submitted 2006.
[2] A. Cohen, W. Dahmen and R. DeVore, *Compressed sensing and best k-term approximation*, submitted 2006.
[3] A. Cohen, W. Dahmen and R. DeVore, *Near optimal approximation of arbitrary signals from highly uncomplete measurements*, preprint 2006.
[4] G. Cormode and S. Muthukrishnan, *Towards and algorithmic theory of compressed sensing*, Technical Report 2005-25, DIMACS, 2005.
[5] E. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inf. Theory, **51**, 4203–4215, 2005.
[6] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure and Appl. Math., to appear 2006.
[7] R. DeVore and V. Temlyakov, *Some remarks on greedy algorithms*, Advances in Computational Mathematics **5**, 173-187, 1996.
[8] D. Donoho, Compressed Sensing, EEE Trans. Information Theory, **52**, 1289-1306, 2006.
[9] A. C. Gilbert and J. A. Tropp, *Signal recovery from partial information via Orthogonal Matching Pursuit*, submitted 2005.

Inverse problems in Option pricing
Rama Cont

In these lectures we describe some nonlinear inverse problems which arise in the context of option pricing models in finance and discuss various – deterministic and stochastic – algorithms for solving them. The contents of these lectures are summarized in [7].

Calibration of option pricing models: setting and examples. *Option pricing* problems involve the computation of the expectations $C_i = E^{\mathbb{Q}}[H_i]$ of various functionals $H_i = h_i(X_t, t \in [0, T])$ of a *given* martingale $(X_t)_{t \in [0, T_*]}$, the risk-neutral price process. This is a well-posed problem whose solution typically involves numerical quadrature (Fourier/Laplace transforms), numerical solution of PDEs/ integro-differential equations or Monte Carlo simulation methods. The associated *inverse problem* – called the *model calibration* problem in finance – is to recover the unknown law (martingale measure) $\mathbb{Q}$ from observations $C^* = (C_i^*, i \in I)$ of option prices observed with some noise δ: $|C_i^* - C_i| \leq \delta$. Typically, one observes call option prices, for which $H_i = (X_{T_i} - K_i)^+$ and the unknown probability measure $\mathbb{Q}$ is assumed to have some structure e.g. describe the law of a diffusion process, a Lévy process or more generally, a Markov process solution of a stochastic differential equation. The direct (pricing) problem then corresponds to the solution $F(\theta)$ of a well-posed parabolic PDE, integro-differential equation, or free

boundary problem where the operator involves an (unknown) coefficient θ which belongs to a subset E of $\mathbb{R}^d$, a functional space or a space of measures. The calibration problem $F(\theta) = C^*$ involves the inversion of the nonlinear parameter-to-price map $F : E \to \mathbb{R}^I$ and is typically ill-posed.

In some limiting cases where a continuum of data is assumed to be observed, exact inversion formulas can be obtained: an example is the Dupire formula [16]. But, beyond such special cases, numerical methods are needed to solve the inverse problem. The main issues are the precision of the calibration (output error) and the stability of the solution with respect to the inputs.

Nonlinear least squares and regularization methods. Even in special cases where exact inversion formulas are available, the inverse lacks continuity. Also, due to observational noise, it is not a good idea to fine-tune the model parameters to match the observations exactly. For these reasons one often reformulates the problem in a least-squares sense: the minimization over $\theta \in E$ of the in-sample calibration error $G(\theta) = \|C^* - F(\theta)\|^2$. In all models of interest, the resulting error function(al) $G(.)$ is nonconvex and may exhibit many global/local minima, which renders the computation of such least-squares solution difficult and unstable.

Regularization schemes have been used to overcome this problem. In the case of diffusion models, we have an inverse problem for a parabolc PDE, where the parameter is an unknown functional diffusion coefficient belonging to a suitable Sobolev/Hölder space and Tikhonov regularization methods have been succussfully employed [13, 17]. A Tikhonov regularization method is used in [1] to solve an inverse free-boundary problem related to the calibration of American options. In the case of exponential Lévy models, the unknown parameter is a positive measure –the Lévy measure– describing the jumps of the process: Tikhonov regularization [10] and regularization methods based on relative entropy [12, 11] have been studied for these problems. Belomestny & Reiß [4] propose a spectral regularization method based on a linear reparametrization of the problem. Different regularization methods lead to *different* solutions: Tikhonov regularization approximates the least-squares solution with smallest norm while entropy regularization selects the minimum-entropy least squares solution.

The numerical solution of the resulting optimization problem is usually done by using gradient-based optimization methods, which leads to two issues. One is the efficient computation of the gradient / Fréchet derivative of the regularized functional: this is usually done by solving, at each iteration step, an auxiliary PDE/ boundary value problem [13, 10]. Also, the regularized functional is still not convex so uniqueness of the regularized solution and convergence of gradient-based methods are not obvious. Finally, these methods yield a single set of model parameters calibrated to market data and ignore the non-uniqueness i.e. the uncertainty in the solution, which is an important issue in option pricing [8].

Statistical approaches: exploring model uncertainty. Viewing the solution of the inverse problem as an *estimator* i.e. a random variable, allows to design

stochastic algorithms which can be used not only to construct solution(s) but also to assess the uncertainty inherent to the solution of the inverse problem.

A first approach is to solve the nonlinear least-squares problem using a stochastic particle algorithm [5] to minimize the in-sample pricing error. This algorithm allows for multiple global minima and which generates a random sample from the set of global minima. Starting from an IID population of candidate solutions drawn from a prior distribution of the set of model parameters, the population of parameters is updated through cycles of independent random moves followed by "selection" according to pricing performance. Using techniques proposed by [15] we exhibit conditions under which such an evolving population converges to a sample of solutions for the inverse problem. The heterogeneity of the obtained sample can then be used to quantify the degree of ill–posedness of the inverse problem and compute the impact of model uncertainty [8]. The algorithm is applied in [5] to the case of a diffusion model, where one aims at retrieving a functional diffusion coefficient from a finite set of option prices: we discuss some implementation details and illustrate the performance on simulated and real data sets of index options.

In the last part of the lecture, we propose a probabilistic construction of a solution to the inverse problem, based on a method studied by Dacunha-Castelle and Gamboa [14]. Starting from a prior distribution on the parameters of a model and a set of observed option prices, we construct the unknown probability measure $\mathbb{Q}$ as a random mixture of (candidate) martingale measures generated from a prior parametric model [9]. The resulting estimator of the unknown pricing measure is a random martingale measure, whose expectation yields an arbitrage-free pricing rule consistent with the observed option prices and whose dispersion properties can be used to quantify model uncertainty. We describe a Monte Carlo algorithm for computing prices under this rule and characterize the limit behavior of the algorithm, which is shown to possess a dual interpretation in terms of minimization of "model risk". This construction only involves a well-posed unconstrained minimization of a convex function, easily performed with gradient-based methods. The resulting algorithm can be seen as a dynamic arbitrage-free version of Avellaneda et al.'s Weighted Monte Carlo algorithm [3], applicable to a wide range of pricing models and products. This approach yields a posterior distribution, instead of a single price, for exotic options and allows to simulate a sample from this posterior distribution. As an example, we compute the posterior distribution for a barrier option given a set of European calls and puts, in a stochastic volatility model with jumps.

References

[1] Achdou, Y. (2005) An inverse problem for a parabolic variational inequality arising in volatility calibration with American options, *SIAM Journal on Control and Optimization*, Volume 43, 1583–1615.
[2] Avellaneda, M. (1998) The minimum-entropy algorithm and related methods for calibrating asset-pricing models, in *Proceedings of the International Congress of Mathematicians*, (Berlin), Vol. III, Documenta Mathematica, 1998, pp. 545–563.

[3] Avellaneda M., R Buff, C Friedman, N Grandchamp, L Kruk, J Newman (2001): Weighted Monte Carlo: A new Technique for Calibrating Asset-Pricing Models, *International journal of theoretical and applied finance*, Vol. 4, 91–119.

[4] Belomestny, D. and Reiß, M. (2006) Spectral calibration of exponential Lévy Models, *Finance and Stochastics*, Vol. 10 No. 4, 449-474.

[5] Ben Hamida, S. & Cont, R. (2004) Recovering volatility from option prices by evolutionary optimization, *Journal of Computational Finance*, Vol. 8, No. 3, 43-76.

[6] Coleman T., Li & Verma, A. (1999): Reconstructing the unknown volatility function, *Journal of Computational Finance*, Vol. 2, Number 3, 1999, 77-102.

[7] Cont, R. (2007) Inverse problems in option pricing, forthcoming in *Statistics and Decisions*.

[8] Cont, R. (2006) Model uncertainty and its impact on the pricing of derivative instruments, *Mathematical Finance*, Vol. 16, No. 3 (July 2006), 519–547.

[9] Cont, R. and Léonard, Ch (2006) A probabilistic approach to inverse problems in option pricing, Working Paper.

[10] Cont, R. & Rouis, M. (2006): Recovering Lévy processes from option prices by Tikhonov regularization, Working Paper.

[11] Cont, R. & Tankov, P. (2004): Nonparametric calibration of jump-diffusion option pricing models, *Journal of Computational Finance*, Vol. 7, No. 3, 1–49.

[12] Cont, R. and Tankov, P. (2005) Recovering Lévy processes from option prices: regularization of an ill-posed inverse problem. *SIAM Journal on Control and Optimization*, Volume 45, Issue 1, pages 1–25.

[13] Crépey, S. (2003) Calibration of the local volatility in a trinomial tree using Tikhonov regularization, *Inverse problems*, 19, 91–127.

[14] Dacunha-Castelle, D. and Gamboa, F. (1990) Maximum d'entropie et problème des moments, *Ann. Inst. H. Poincaré. Probab. Statist.*, 26, 567–596.

[15] Del Moral. P. (2004) *Feynman-Kac formulae*, Berlin: Springer.

[16] Dupire, B. (1994) Pricing with a smile, *RISK*, Vol. 7, January, 18–20.

[17] Engl, H. and Egger, H. (2005) Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates *Inverse problems*, 21, 1027–1045.

Error controlled regularization by spectral projection for a class of severely ill-posed problems

Wolfgang Dahmen

(joint work with Markus Jürgens)

This lecture is concerned with numerical concepts for the regularized solution of severly ill-posed problems represented by the following somewhat simplified setting. Given a Hilbert space X and a linear operator A with domain $\mathcal{D}(A) \subset X$, the solution to the initial value problem

$$\dot{u} + Au = 0, \ u(0) = u_0,$$

is, under suitable assumptions on A, given by $S(u_0)(t) = e^{-tA}u_0$. Although much of what follows can be generalized we confine the discussion for simplicity here to the case that A is a symmetric positive definite operator that is actually H-elliptic for some Hilbert space H that is compactly embedded in X. A guiding example is $X = L_2(\Omega)$, Ω a (sufficiently well behaved) domain in $\mathbb{R}^d$, $H = H_0^1(\Omega)$, the Sobolev space of L_2-functions with first order weak derivatives in L_2 and vanishing trace on the boundary $\partial\Omega$ of Ω, and $Au = \text{div}(a\nabla u)$ where a is a unformly

positive definite matrix on Ω. This corresponds to a heat diffusion process. In order to infer from a current temperature distribution back to the initial value u_0 one has to "invert" the compact operator S which is a severly ill-posed problem, due to the typically exponentially decaying spectrum. The difficulties of contriving good regularization strategies are closely interrelated with the typically heavy computational demands when solving such problems numerically. One of the obstructions to be confronted are the following. One typically discretizes first the forward problem which means to replace the operator A by some finite dimensional approximation A_h. A spectral cut-off regularization (essentially of e^{-tA_h}) is typically prohibitively expensive when dealing with problems in several space dimensions. Tikhonov regularization suffers from a low qualification and requires the solution of increasingly ill-conditioned problems when the regularization parameter decreases. When resorting to iterative methods, which is unavoidable for large scale problems, one has to carry out therefore an increasing number of applications essentially of e^{tA_h} which is not sparse in typical representations.

Therefore, we focus here on a different strategy that exploits the semigroup character of the solution operator $S = e^{-tA}$, see [2, 3, 4]. It is based on regularizing first the infinite dimensional problem by a spectral projection scheme. However, in the present approach we avoid the explicit use of the corresponding eigenfunctions which would be available only in very special cases and whose computation would in general be far too expensive. Instead we employ certain Dunford integral representations of these projections where the choice of curves requires only some estimates for the gaps in the spectrum of A. The numerical scheme consists then of applying such projections within a given accuracy tolerance depending on the (deterministic) noise level. The ability to do so in an efficient way hinges on two pillars. (i) For sufficiently regular curves the trapezoidal rule yields exponential convergence rates for the approximate evaluation of the Dunford integral. (ii) At each quadrature node the resolvent operator has to be applied within a suitable accuracy tolerance which means that one has to solve a boundary value problem for each node which can be done in parallel. Moreover, in contrast to S, these shifted versions of A can be preconditioned efficiently independent of the regularization level. For instance, one can use recent adaptive wavelet methods to solve these problems within certain asymptotically optimal complexity bounds [1]. Thus, regularization and discretization is separated so as to minimize the computational effort needed to realize an ideal regularization strategy on the infinite dimensional level within a suitable tolerance. Using results about the complexity of the underlying wavelet schemes we conclude with a first assessment of the computational cost expressed in terms of the (deterministic) noise level.

References

[1] A. Cohen, W. Dahmen, R. DeVore, *Adaptive wavelet methods II - Beyond the elliptic case*, Foundations of Computational Mathematics, **2** (2002), 203–245.
[2] W. Dahmen, M. Jürgens, *Error controlled regularization by projection*, ETNA, **26** (2006).
[3] M. Jürgens, *A Semigroup Approach to the Numerical Solution of Parabolic Differential Equations*, PhD thesis, RWTH Aachen, 2005.

[4] M. Jürgens, *Adaptive application of the operator exponential*, J. Numer. Math., Special Issue on Breaking Complexity: Multiscale Methods for Efficient PDE Solvers, (3)**14** (2006), 217–246.

Spectral methods for estimating change–points in inverse problems
Alexander Goldenshluger

Change-point estimation and detection is one of most important tasks of statistics and as such it retained much attention of statistical and signal processing community. This problem is also well represented in the literature on nonparametric regression estimation. The literature mostly focuses on the models with direct observations. In particular, [2] considered minimax estimation of the change-point in function f observed in Gaussian white noise of level ϵ. A remarkable result in [2] states that the minimax risk over the class of functions having a single change-point and satisfying the Lipschitz condition away from the change-point is ϵ^2, while the minimax rate for the sequential (Markov) estimator is $\epsilon^2 \ln \epsilon^{-1}$. In the problems of sequential estimation of a change-point in the signal, or in its first derivative, precise asymptotic expressions for the minimax risk have been obtained in [1]. It has been proved there that minimax rate of sequential estimation of the change-point in the first derivative is $(\epsilon^2 \ln \epsilon^{-1})^{1/3}$. The problem of change–point estimation from indirect observations is much less studied. We menition only [3] who derived the minimax rates of convergence for change-point estimation in the density convolution model.

In this paper we propose a unified framework to change–point estimation in inverse problems. We show that estimation of the change–point and the jump amplitude can be reduced to the problem of recovering the frequency and the amplitude of a complex harmonic oscillation in the presence of random noise and a deterministic nuisance. To be more precise, consider the following sequence space model

$$y_k = a \exp(2\pi i k\theta) + g_k + \epsilon \sigma_k \xi_k, \quad k \in \mathbb{N},$$

where $g = (g_k) \in C^{\mathbb{N}}$ is an unknown nuisance sequence, $\sigma = (\sigma_k) \in C^{\mathbb{N}}$ is a given sequence, and $\xi = (\xi_k) \in C^{\mathbb{N}}$ is a sequence of independent standard complex-valued normal random variables. We demonstrate that this model includes many problems of change–point estimation in the periodic setups. Then we concentrate on the study of theoretical accuracy limitations in estimating θ and a, and develop corresponding rate-optimal procedures. Our frequency domain estimation technique is closely related to spectral analysis of time series and frequency estimation. We also develop adaptive procedures that do not require prior information on the regularity deterministic nuisance component g.

References

[1] Freidlin, M. I. and Korostelev, A. P. (1995). Image processing for plane domains: change–point problems for the domain's area. *Problems of Information Transmission* **31**, 27–45.

[2] KOROSTELEV, A. P. (1987). Minimax estimation of a discontinuous signal. *Theory Probab. Applic.* **32**, 727-730

[3] NEUMANN, M. H. (1997). Optimal change–point estimation in inverse problems. *Scand. J. Statist.* **24**, 503-521.

On inverse problems from a statistical perspective
YURI GOLUBEV

These lectures deal with a classical problem of recovering an unknown vector $\theta = (\theta(1), \ldots, \theta(n))^\top \in \mathbb{R}^n$ from the noisy data

$$(1) \qquad\qquad\qquad Y = A\theta + \epsilon,$$

where A is a $n \times n$-matrix, $\epsilon \in \mathbb{R}^n$ is a white Gaussian noise with an unknown variance $\sigma = \sqrt{\mathbf{E}\epsilon^2(k)}$. The size n is assumed to be large.

Spectral regularization. The main goal in spectral regularization methods is to suppress high frequency components in the Moore-Penrose inversion

$$\hat{\theta} = (A^\top A)^{-1} A^\top Y.$$

A typical spectral regularization [5] has the form $\hat{\theta}_\alpha = H_\alpha((A^\top A)^{-1})\hat{\theta}$, where $H_\alpha(z)$ is an analytic function such that

$$\lim_{\alpha \to 0} H_\alpha(z) = 1, \quad \lim_{z \to \infty} H_\alpha(z) = 0.$$

As a rule, the choice of $H_\alpha(\cdot)$ is a compromise between numerical complexity and statistical performance of the method. Form a statistical viewpoint, the best H_α is provided by the Pinsker theorem [8]. Popular practical methods include the spectral cut-off, the Tikhonov regularization, the Landweber iterations, ν - method (see [5]). The main advantage of spectral methods is that they represent the operator in the sparsest form. On the other hand, there is no guarantee that these methods provide a sparse representation of the underlying vector.

Regularization with the help of the DWT. This regularization can be viewed as a compromise between sparse representation of the underlying vector and the operator. The discrete wavelet transform (DWT) is a matrix W such that $WW^\top = I$. The principal property of this transform is that Wx and $W^\top x$ may be computed in Cn times. In order to motivate DWT regularization methods, it is assumed that $x = W\theta$ is a sparse vector. To provide a sparse representation of A, the DWT is used twice, namely we make the following linear transformation of $Y = A\theta + \epsilon \rightarrow Z = WAW^\top x + \epsilon'$, where ϵ' is a new white Gaussian noise. Thus the initial problem is reduced to recovering x from Z. Since x is assumed to be sparse, to estimate it, we may use the following soft thresholding estimator

$$\hat{x}_\alpha = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Big\{ \|Z - WAW^\top x\|^2 + \alpha \sum_{k=1}^n |x(k)| \Big\}.$$

The difficulty of this method is related to its numerical complexity. In order to improve this drawback Candès and Tao in [3] proposed to use the so-called Dantzig selectors. This method is given by

$$\hat{x}_\alpha = \operatorname*{argmin}_{x: \|Z - WAW^\top x\|_\infty \leq 1/\alpha} \sum_{i=1}^n |x(i)|,$$

where $\|y\|_\infty = \max_k |y(k)|$.

If the columns of $WAW^\top$ are nearly orthogonal, one may use a very fast optimization method based on a confidence set for x. For given $u \in \mathbb{R}^1$ and integer k denote

$$R(u, k) = \{x \in \mathbb{R}^n : x(k) = u\}$$

Let

$$A_h(k) = \left\{ u \in \mathbb{R}^1 : \inf_{X \in \mathbb{R}^n} \|Y - WAW^\top X\|^2 - \inf_{X \in R(u,k)} \|Y - WAW^\top X\|^2 \geq \log(h) \right\}$$

and

$$\mathcal{A}_h = \left\{ x \in \mathbb{R}^n : x(k) \in A_h(k), k = 1, \ldots, n \right\}$$

Then the soft thresholding may be computed as a solution of the banal optimization problem

$$\hat{x}_h = \operatorname*{argmin}_{x \in \mathcal{A}_h} \sum_{i=1}^n |x(i)|.$$

Empirical risk minimization. When a regularization method is used, the principal issue is related to the data-driven choice of a smoothing parameter. For instance, for the spectral regularization

$$\hat{\theta}_\alpha = H_\alpha[(A^\top A)^{-1}](A^\top A)^{-1}A^\top Y, \quad \text{with } H_0[(AA^\top)^{-1}] = I$$

the issue is to find $\hat{\alpha}(Y)$ that minimizes $\mathbf{E}\|\theta - \hat{\theta}_{\hat{\alpha}(Y)}(Y)\|^2$ uniformly in $\theta \in \mathbb{R}^n$. This problem goes back to [1], and nowadays it is cornerstone in statistics [2]. To construct $\hat{\alpha}(Y)$, one usually relies on the following heuristic arguments. First of all the distance $\|\hat{\theta}_0 - \hat{\theta}_{\hat{\alpha}}\|$ should be small, since if there is no noise, we obviously take $\hat{\theta}_0$. On the other hand, $\|Y - A\hat{\theta}_{\hat{\alpha}}\|$ should not be large, since we want that the estimator fits good the statistical model. Therefore is seems natural to take

$$\hat{\alpha} = \operatorname*{argmin}_\alpha \left\{ ER_{Pen}(Y, \alpha) \right\},$$

where

$$ER_{Pen}(Y, \alpha) = \|\hat{\theta}_0 - \hat{\theta}_\alpha\|^2 + Pen(\alpha)\|Y - A\hat{\theta}_\alpha\|^2 - \sigma^2 \sum_{k=1}^n \lambda_k$$

is called *empirical risk* associated with the penalty function $Pen(\alpha)$. The choice of this function is very important from a statistical viewpoint. Heuristically, we are looking for a 'minimal' function $Pen(\alpha)$ such that uniformly in θ

$$(2) \qquad\qquad \mathbf{E}\|\theta - \hat{\theta}_{\hat{\alpha}}\|^2 \leq \mathbf{E}ER_{Pen}(Y, \hat{\alpha})$$

Once such a function has been found, the evaluation of the risk becomes immediate since

$$\mathbf{E} ER_{Pen}(Y, \hat{\alpha}) \leq \inf_\alpha \mathbf{E} ER_{Pen}(Y, \alpha).$$

Unfortunately, the computation of a good penalty is a very delicate problem (see, for instance, [4]). Therefore we focus on the so-called lower penalty approach. Its mitivation is based on the following remark: for any given α there exists θ_α such that the probability distribution of $\hat{\alpha}$ is degenerate. If so, the condition 2 can be rewritten as

$$\mathbf{E}\|\theta - \hat{\theta}_\alpha\|^2 \leq \mathbf{E} ER_{Pen}(Y, \alpha) \text{ for any } \theta \in \mathbb{R}^n.$$

It is not difficult to check with a simple algebra that the above equation results in

$$Pen(\alpha) \geq \underline{Pen}(\alpha) = 2\sum_{k=1}^n \lambda_k h_\alpha(\lambda_k) \bigg/ \sum_{k=1}^n [1 - h_\alpha(\lambda_k)]^2.$$

In what follows it is assumed that smoothers $H_\alpha(\lambda)$ are ordered (see [6]) i.e.:
$\diamondsuit$ for all α, λ, $0 \leq H_\alpha(\lambda) \leq 1$
$\diamondsuit$ $H_{\alpha_1}(\lambda) \geq H_{\alpha_2}(\lambda)$, for all $\alpha_1 \leq \alpha_2$ and all $\lambda > 0$.
Standard examples include the Tikhonov regularization, the spectral cut-off method, the Landweber iterations, the Pinker smoothers.

Theorem 2. *Suppose that there is $\kappa < 1$ such that*

$$\max_k \lambda_k H_\alpha^2(\lambda_k) \leq 2\lambda_1 \Big[\frac{1}{\lambda_1}\sum_{k=1}^n \lambda_k H_\alpha^2(\lambda_k)\Big]^\kappa,$$

Then for some $C_\kappa > 0$, uniformly in $\theta \in \mathbb{R}^d$

$$\mathbf{E}\|\theta - \hat{\theta}_{\hat{\alpha}}\|^2 \leq \underline{R}(\theta) + C_\kappa \lambda_1 \sigma^2 \Big[\frac{\underline{R}(\theta)}{\lambda_1 \sigma^2}\Big]^{(1+\kappa)/2} + C_\kappa \lambda_1 \sigma^2,$$

where $\underline{R}(\theta) = \inf_\alpha \mathbf{E} ER_{\underline{Pen}}(Y, \alpha).$

REFERENCES

[1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle.* Proc. 2nd Intern. Symp. Inf. Theory, Petrov P.N. and Csaki F. eds. Budapest, 267-281.
[2] Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc.* **3** 203–268.
[3] Candès, E. J. and Tao, T. (2006). The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* To appear.
[4] Cavalier, L., Golubev, Yu. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. of Stat.* Vol. 34, No. 4. 1653–1677.
[5] Engl, H. W., Hanke, M. and Neubauer, A. (1996). *Regularization of Inverse Problems.* Kluwer Academic Publishers.
[6] Kneip, A. (1994) Ordered linear smoothers. *Ann. Statist.* **22**, 835–866.
[7] PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission,* **16**, 120–133.
[8] Press, W., Teukolsky, S., Vetterling, W., Flannery, B. (2006) Numerical recipes in C++. Cambridge University Press.

On inverse problems from a numerical analysis perspective

THORSTEN HOHAGE

(joint work with Axel Munk)

This lecture is concerned with the solution of ill-posed operator equations in the presence of deterministic and stochastic noise. More precisely we aim to estimate a vector a in a separable Hilbert space $\mathcal{X}$ given noisy measurements Y of the image of a under a possibily nonlinear operator $F : \mathcal{X} \to \mathcal{Y}$ with values in another Hilbert space $\mathcal{Y}$:

$$(1) \qquad Y = F(a) + \sigma\xi + \delta\zeta$$

Here ξ is a Hilbert space process in Y with a covariance operator $\mathbf{Cov}_\xi$ satisfying $\|\mathbf{Cov}_\xi\| \leq 1$, $\sigma \geq 0$ is the stochastic noise level, $\zeta \in Y$ is the normalized determistic noise ($\|\zeta\| = 1$), and $\delta > 0$ is the deterministic noise level. It is generally assumed that F is one-to-one, but the inverse of F is not continuous.

In the first part of the lecture we consider the case that $F = K$ is linear. We study general spectral regularization methods of the form

$$\widehat{a}_\alpha = g_\alpha(K^*K)K^*Y,$$

see [4, 6]. Here g_α is a family of piecewise continuous functions on the spectrum $\sigma(K^*K)$ parametrized by a regularization parameter $\alpha > 0$ such that $\lim_{\alpha\searrow 0} g_\alpha(t) = 1/t$ for all $t \in \sigma(K^*K) \setminus \{0\}$ and $\sup_{t\in\sigma(K^*K)} |g_\alpha(t)| \leq \frac{C_\mathrm{V}}{\alpha}$. We further assume that there exists a number $\nu_0 > 0$ called *qualification* of the method and constants γ_ν such that

$$\sup_{t\in\sigma(K^*K)} |t^\nu r_\alpha(t)| \leq \gamma_\nu \alpha^\nu \quad \text{for all } \alpha \text{ and } 0 \leq \nu \leq \nu_0.$$

As usual, smoothness of the solution is measured in terms of source-wise representations

$$a = \Lambda(K^*K)w$$

with a continuous, increasing function Λ satisfying $\Lambda(0) = 0$. For $\Lambda(t) = t^\nu$, ν_0 is the largest index for which the bias of the method converges of optimal order. Examples include Tikhonov regularization $g_\alpha(t) = (t + \alpha)^{-1}$, Landweber iteration $g_{1/(n+1)}(t) = \sum_{j=0}^{n-1}(1-t)^j$, spectral cut-off $g_\alpha(t) = \chi_{[\alpha,\infty)}(t)/t$ and ν-methods.

The main result states that under certain conditions on the operator the variances of all these estimators have the same asymptotic behavior. Therefore, the performance of the methods is essentially determined by the bias term, which is known to converge of the same order for all methods up to the qualification of the method. We show that our assumptions on the operator are satisfied for the backwards heat equation, satellite gradiometry, and L^2-boosting in machine learning.

The second part of the lecture is devoted to nonlinear statistical inverse problems. In particular, we study a class of regularized Newton method suggested by

Bakushinskii [2, 1, 3, 5],

$$\widehat{a}_{k+1} := a_0 + g_{\alpha_n}\left(T_k^* T_k\right) T_k^*\left(Y - F(\widehat{a}_k) + T_k(\widehat{a}_k - a_0)\right),$$

where a_0 is an initial guess and $T_k := F'[\widehat{a}_k]$ is the Fréchet derivative of F at $\widehat{a}_k$. We assume that $a - a_0$ satisfies a source condition of the form

$$a - a_0 = (F'[a]^* F'[a])^{\nu} w \qquad \text{for some } \nu \in [\frac{1}{2}, \mu_0] \text{ and } w \in \mathcal{X}.$$

We prove that if the smoothness coefficient ν is known, the expected squared error $\mathbf{E}\|\widehat{\mathbf{a}}_{\mathbf{K}(\sigma,\delta)} - \mathbf{a}\|^{\mathbf{2}}$ with an a-priori stopping rule $K(\sigma, \delta)$ converges of the same order as for linear problems. For unknown smoothness we can still obtain optimal rates of convergence up to a logarithmic factor $\ln(\sigma^{-1})$ using Lepskii's balancing principle [7, 8, 9]. Applications to scattering problems and parallel NMR imaging are discussed.

References

[1] A. B. Bakushinsky. The problem of the convergence of the iteratively regularized Gauss-Newton method. *Comput. Maths. Math. Phys.*, 32:1353–1359, 1992.

[2] A. B. Bakushinsky and M. Y. Kokurin. *Iterative Methods for Approximate Solution of Inverse Problems*. Springer, Dordrecht, 2004.

[3] F. Bauer and T. Hohage. A Lepskij-type stopping rule for regularized Newton methods. *Inverse Problems*, 21:1975–1991, 2005.

[4] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. Technical Report Preprint 2006-02, NAM, University of Göttingen, 2006. http://www.num.math.uni-goettingen.de/bibi/preprintserie/2006.html.

[5] B. Blaschke, A. Neubauer, and O. Scherzer. On convergence rates for the iteratively regularized Gauß-Newton method. *IMA Journal of Numerical Analysis*, 17:421–436, 1997.

[6] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publisher, Dordrecht, Boston, London, 1996.

[7] P. Mathé. The Lepskiǐprinciple revisited. *Inverse Problems*, 22:L11–L15, 2006.

[8] P. Mathé and S. Pereverzev. Geometry of ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19:789–803, 2003.

[9] P. Mathé and S. Pereverzev. Regularization of some linear ill-posed problems with discretized random noisy data. *Math. Comp.*, 2006. to appear.

Reconsidering the Random Coefficient Model

Enno Mammen

(joint work with Stefan Hoderlein, Jussi Klemelä)

In econometrics heterogeneity of individual agents, in particular consumers or firms, is a typical situation. In addition, it is often the case that the individuals are, at least approximately, characterized by a linear relationship between a K-vector of explanatory variables, and a dependent variable. Combining these two notions leads naturally to the random coefficient model (RCM),

$$Y_i = \beta_i^T X_i,$$

where Y_i is an observed continuously distributed random scalar, and X_i denotes an observed random K- vector of individual specific regressors, possibly including an intercept, i.e. $X_{i,1} \equiv 1$. The K-vector of coefficients β_i varies across individuals. It is unobserved and it is modeled as a random variable that captures the unobserved heterogeneity.

Traditionally, this model has been investigated under mean independence, i.e. $\mathbf{E}\left[\beta_i|X_i\right] = \beta$, and homoscedasticity, i.e. $Var\left[\beta_i|X_i\right] = \sigma_\beta^2$. While this allows to identify the average marginal effect and the variance, important features of the joint distribution of marginal effects are left unidentified. These includes the quantiles of the marginals, as well as skewness, curtosis or symmetry of the distribution. Moreover, the question of multimodality, or the related question whether the population consist of a mixture of subpopulations are left unanswered.

It is the aim of this talk to establish that under stronger independence assumptions the joint density of β is identified and can be estimated nonparametricaly.

The structure of our estimator is simple, and very much resembles a standard kernel density estimator. More precisely, the estimator for the joint density of random coefficients at a fixed position, $f_\beta(b)$ is given by

$$\hat{f}_\beta(b) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(S_i^T b - U_i\right)\left(\hat{f}_S(S_i)\right)^{-1}, \qquad b \in \mathbf{R}^K,$$

where U_i and S_i denote suitable transformations of Y_i and X_i, respectively, K_h is an appropriate kernel, and $\hat{f}_S$ denotes an estimator for the density of the transformed regressors. The most prominent difference to standard kernel density estimation are the nonstandard kernel, as well as the normalization by the density of transformed regressors.

Our estimation problem is related to statitistical methods in computer tomography. In fact our estimator is a modification of an approach proposed in Kostelev and Tsybakov (1993) who estimate a Radon transformed regression function. Another approach has been used in Beran, Feuerverger and Hall (1996) and Feuerverger and Hall (2000), who estimate the characteristic function of the response variable and then transform this estimator back. In contrast to this approach, our estimator has a simple direct structure and is obtained by a one-step procedure.

Application of our estimator to consumer demand is discussed using British household data. We analyze budget shares for food in dependence from income and price. The density estimator of β describes the heterogeneity in individual demand.

<div align="center">REFERENCES</div>

[1] R. Beran, A. Feuerverger, P. Hall, *On nonparametric estimation of intercept and slope distributions in random coefficient regression*, Ann. Statist. **24** (1996), 2569–2692.

[2] A. Feuerverger, P. Hall, *Positron emission tomography and random coefficients regression*, Ann. Inst. Statist. **52** (2000), 123–138.

[3] Korostelev, A. P. and Tsybakov, A. B., *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics **82** (1993), Springer, New York.

Discretization of inverse problems
Peter Mathé

When solving (linear) inverse problems in Hilbert space, say

$$y^\delta = Ax + \delta\xi,$$

where $A\colon X \to Y$ is a compact injective operator and ξ is "statistical" noise, the issue of discretization is a serious subject. Often such discretization may be understood as choosing two projections, say Q, describing the way how data are retrieved, and P, describing the way the potential solution shall be represented. In this case we end up with a discretized equation

$$y^\delta = QAPx.$$

Then one can proceed by either taking this equation as regularization (*self-regularization*), see e.g. [6, 2], or to additionally perform some linear regularization, see [7, 4].

In this talk we discuss the following problem: Which properties of the chosen projections P and Q determine the performance of the regularization in either case. This is to be understood in an asymptotical setup when the noise level $\delta \to 0$. We exhibit how different kinds of s-numbers naturally occur and we indicate how to obtain optimal reconstruction rates, if the discretization is chosen appropriately. The analysis is carried out under the assumption that the observations are blurred by white noise ξ. Smoothness of the underlying true solution x is measured in terms of *general source conditions*, a framework which recently became attractive, see [1, 5].

This work extends, summarizes and complements previous work in [1, 4, 3] to statistical setup.

References

[1] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[2] Barbara Kaltenbacher. *Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems*. Inverse Problems, **16** (2000), 1523–1539.

[3] P. Mathé and N. Schöne. *Regularization by projection in variable Hilbert scales*. Weierstrass Institute Berlin, 2006.

[4] Peter Mathé and Sergei V. Pereverzev. *Discretization strategy for linear ill-posed problems in variable Hilbert scales*. Inverse Problems, **19** (2003), 1263–1277.

[5] Peter Mathé and Sergei V. Pereverzev. *Geometry of linear ill-posed problems in variable Hilbert scales*. Inverse Problems, **19** (2003), 789–803.

[6] Frank Natterer. *Regularisierung schlecht gestellter Probleme durch Projektionsverfahren*. Numer. Math., **28** (1977), 329–341.

[7] R. Plato and G. Vainikko. *On the regularization of projection methods for solving ill-posed problems*. Numer. Math., **57** (1990), 63–79.

Jump estimation in statistical inverse problems

AXEL MUNK

(joint work with Leif Boysen, Volkmar Liebscher, Olaf Wittich)

Initially, the use of piecewise constant functions for regression has been proposed by [4], who called the corresponding reconstruction the regressogram. [4] proposed it as a simple explanatory tool. For a given set of jump locations, the regressogram simply averages the data between two successive jumps. A difficult issue, however, is a proper selection of the location of jumps and its convergence analysis.

Nowadays, there are several statistical estimation procedures which use local constant reconstructions. [6] studied the case where the signal is a step function with one jump and showed that in this case the signal can be estimated at the parametric $n^{-1/2}$-rate and that the jump location can be estimated at a rate of n^{-1}. This was generalized by [7] to step functions with a given known upper bound for the number of jumps.

In this work we investigate an l^0 penalized least squares estimator which penalizes the complexity of the reconstruction by the number of intervals where the reconstruction is locally constant, or equivalently by the number of jumps of the reconstruction. The resulting estimator will be called the *Jplse*. Compared to the total variation approach obviously, this method more easily captures extreme plateaus, but is less robust to outliers. This might be of interest in applications where extreme plateaus are informative, like for example in mass spectroscopy.

This talk summarizes the work of the authors on convergence rates of the *Jplse*. In particular we highlight the connection to approximation by step functions as it is well examined in approximation theory, e.g. [5]. In [2] it is shown that given a proper choice of the smoothing parameter γ it is possible to obtain optimal rates for certain classes of approximation spaces under the assumption of subgaussian tails of the error distribution. As special cases the class of piecewise Hölder continuous functions of order $0 < \alpha \leq 1$ and the class of function with bounded total variation are obtained.

Further in [1] consistency of regressograms for arbitrary L^2 functions is shown under more general assumptions on the error. If the true function is *càdlàg*, we additionally show consistency in the Skorokhod topology. This a substantially stronger statement than the L_2 convergence and yields consistency of the whole *graph* of the estimator.

Finally, we adress the problem of estimating a jump function in the context of an inverse regression equation $Y_i = Kf(x_i) + \epsilon_i$, $i = 1, \cdots, n$, where K is a known (linear) integral operator and $f : [0, 1] \longrightarrow \mathbb{R}$ is the unknown function to be estimated. The x_i are (regular, possibly random) design points. It turns out that here a $\sqrt{n}$-rate of convergence is generic and minimax, provided the kernel of K is bounded and continuous. In fact, the jump locations together with the jump sizes are asymptotically multivariate normal. To this end we require an identifiabilty condition on the kernel K, which turns out to be crucial for recovering jump

functions in nosiy inverse problems. Let

$$\Delta(x, a, b) := \begin{cases} (b - a)^{-1} \displaystyle\int_a^b K(x, y) dy & : & b \neq a \\ K(x, a) & : & b = a \end{cases} \quad .$$

Assume, that the functions

$$\Delta(x, 0, \kappa_1), \, \Delta(x, \kappa_2, \kappa_3), \, \dots, \, \Delta(x, \kappa_r, 1)$$

are linearly independent in $L_2([0, 1])$ for every $r \in \mathbb{N}$ and every choice of $0 < \kappa_1 \leq \kappa_2 < \dots \leq \kappa_r < 1$, where only two subsequent κ_i are allowed to be equal. Then it is possible to show that $K : f \in S_k([0, 1)) \to L_2([0, 1])$ is one-to-one, where $S_k([0, 1))$ denotes the linear space of jump functions with a finite but arbitrary number of jumps $k \in \mathbb{N}$. Using arguments from total positivity and the theory of radial basis functions it can be shown that this assumption holds, either if K is extended totally positive or, if K is a deconvolution kernel, positive definite, $K \in C(\mathbb{R}) \cap L^1(\mathbb{R})$, and it holds that

$$\sup_{x \in \mathbb{R}} \frac{|\hat{K}(x)|^{-1}}{1 + |x|^s} < \infty, \qquad \text{for some } s \in \mathbb{N},$$

where $\hat{K}$ denotes the Fourier transform of K. Losely speaking, the latter condition rules out functions which are too smooth. Asymptotic normality can be used to construct confidence bands for jump functions or for a piecewise linear regression function in multiphase regression.

Motivated by a problem from material science, we extend this to estimation problems with certain nonlinear operators. More specifically we show similar results for the class of generalized Hammerstein equations of the type

$$K_\varphi(f)(\cdot) \quad = \quad \int K(x, \cdot) \varphi \circ f(x) dx,$$

for some φ injective, C^1. It is an open and challenging problem how general nonlinear problems can be treated.

Finally, we would like to stress that our analysis for jump spaces is completely different to the situation where the underlying function space is of some smoothing type, such as a Sobolev space, where the spectral behaviour of K determines the asymptotics [3]. We show that for jump spaces the localisation behaviour of the kernel determines the rate of convergence, rather than the spectral behaviour. In this sense a bounded integral kernel is most difficult. We obtain, e.g. for singular kernels with decay of the singularity of the order $|x|^{-\alpha}$, $\alpha \in [1/2, 1)$ the minimax rate (which is attained by the *Jplse*) as $n^{-1/\min(2, (3 - 2\alpha))}$.

REFERENCES

[1] L. Boysen, V. Liebscher, A. Munk, O. Wittich, *Scale space consistency of piecewise constant least squares estimators - another look at the regressogram*, IMS Lecture Notes (2006), submitted.
[2] L. Boysen, L., A. Kempe, V. Liebscher, A. Munk, O. Wittich, *Consistencies and rates of convergence of jump-penalized least squares estimators*, Ann. Statist, in revision.

[3] A. Goldenshluger, A.Tsybakov, A. Zeevi, *Optimal change-point estimation from indirect observations*, Ann. Statist. **34** (2006), 350-372.

[4] J. W. Tukey *Curves as parameters, and touch estimation*, In: Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I. Univ. California Press, Berkeley, Calif. (1961), 681–694.

[5] R. A. DeVore, G. G. Lorentz, *Constructive Approximation*, In: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **303**, Springer-Verlag, Berlin (1993).

[6] D. V. Hinkley, *Inference about the change-point in a sequence of random variables*, Biometrika **57** (1970), 1–17.

[7] Y.-C. Yao, S. T. Au, *Least-squares estimation of a step function*, Sankhyā Ser. A **51** (1989), 370–381.

Adaptive regularization algorithms in learning theory

Sergei V. Pereverzev

(joint work with Ernesto De Vito, Lorenzo Rosasco)

We investigate the problem of an adaptive parameter choice for regularization of learning algorithms. In the theory of ill-posed problems there is a long history of choosing regularization parameters in optimal way without a priori knowledge of a smoothness of the element of interest (see, e.g., [1], [2]). But known parameter choice rules cannot be applied directly in Learning Theory. The point is that these rules are based on the estimation of the stability of regularization algorithms measured in the norm of the space where unknown element of interest should be recovered. But in the context of Learning Theory this norm is determined by an unknown probability measure, and is not accessible (see, e.g., [3]).

In the talk we present a new parameter choice strategy consisting in adaptive regularization performed simultaneously in a Hypothesis space and in a space equipped with an empirical norm. Both these spaces are accessible and known parameter choice rules such as a balancing principle (see, e.g., [4])can be used there. Then a parameter for the regularization in the inaccessible space is chosen as the minimal among the parameters selected for above mentioned accessible spaces. We prove that under rather mild assumptions such strategy guarantees an optimal order of the risk estimated in [5].

References

[1] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of inverse problems*, Kluwer (1996).

[2] S. Pereverzev, E. Schock, *On the adaptive selection of the parameter in regularization of ill-posed problems*, SIAM J. Numer. Analysis **43** (2005), 2060 – 2076.

[3] F. Cucker, S. Smale, *Best choices for regularization parameters in learning theory: on the bias-variance problem*, Found. Comput. Math. **2** (2002), 413–428.

[4] P. Mathe, S.V. Pereverzev, *Geometry of linear ill-posed problems in variable Hilbert scales*, Inverse Problems **19** (2003), 789–803.

[5] F. Bauer, S. Pereverzev, L. Rosasco, *On regularization algorithms in Learning Theory*, Journal of Complexity (2006).

Estimation in inverse problems and second-generation wavelets
Dominique Picard
(joint work with Gérard Kerkyacharian, Pencho Petrushev, Thomas Willer)

We consider the following linear inverse problem of recovering a function f, when we receive a blurred and noisy version :

$$Y_\varepsilon = Af + \varepsilon \dot{W}.$$

Our purpose here is to emphasize the fact that in such a problem, there generally exists a basis which is fully adapted to the problem, where for instance the inversion remains very stable : this is the Singular Value Decomposition basis. On the other hand, the SVD basis might not be appropriate for the accurate description of the solution with a small number of parameters.

Also in many practical situations, the signal provides inhomogeneous regularity, and its local features are especially interesting to recover. In such cases, other bases (in particular localised bases such as wavelet bases) may be much more appropriate to give a good representation of the object at hand. Our approach here will be to produce estimation procedures trying to keep the advantages of localisation without loosing the stability and computability of SVD decompositions. We will detail two cases. In the first one (which is the case for instance of the deconvolution example) we show that a fairly simple algorithm (WAVE-VD) using an appropriate thresholding technique performed on a standard wavelet system, enables us to estimate the object with rates which are almost optimal up to logarithm factors for any L_p loss function, and on the whole range of Besov spaces.

In the second case (which is the case of the Wicksell example where the SVD bases lies in the range of Jacobi polynomials), we prove that quite a similar algorithm (NEED-VD) can be performed provided replacing the standard wavelet system by a second generation wavelet-type basis : the Needlets.

We use here the construction (essentially following the work of Petrushev and co-authors) of a localised frame linked with a prescribed basis (here Jacobi polynomials) using a Littlewood Paley decomposition combined with a cubature formula.

This estimation methods yields minimax rates of convergence, also for L_p norms and various Besov-type constraints. This minimax rates show the special role played here by the indices (α, β) of the Jacobi polynomials :

$$E\|\hat{f} - f\|_p^p \le C[\log(1/\varepsilon)]^\delta [\varepsilon \sqrt{\log(1/\varepsilon)}]^{\mu p},$$

$$\mu = \min\{\mu(s),\ \mu(s,\alpha),\ \mu(s,\beta)\} \quad \text{with}$$

$$\mu(s) = \frac{s}{s + \nu + \frac{1}{2}}, \ \mu(s,\alpha) = \frac{s - 2(1+\alpha)(\frac{1}{\pi} - \frac{1}{p})}{s + \nu + 2(1+\alpha)(\frac{1}{2} - \frac{1}{\pi})}.$$

References

[1] Petrushev, Pencho and Xu, Yuan, *Localized polynomial frames on the interval with Jacobi weights*, J. Fourier Anal. Appl. **11** (2005), 557–575.

[2] Kerkyacharian, G., Petrushev, P., Picard, D. and Willer, T., *Needvd: second generation wavelets for estimation in inverse problems*, http://hal.archives-ouvertes.fr/hal-00080849/en/.

Local model selection in inverse problem

VLADIMIR SPOKOINY

(joint work with C. Vial)

This paper presents a new method for estimation in linear inverse problem. We introduce a new procedure called *local model selection* (LMS) which aims to select in an automatic way one estimate out of a class of the given estimates ordered by their variability. We also propose a new approach towards selecting the parameters of the procedure by providing the prescribed behavior of the resulting estimate in the simple parametric situation. We establish a number of important theoretical results concerning the optimality of the aggregated estimate. In particular, our "oracle" results claims that its risk is up to some logarithmic multiplier equal to the smallest risk for the given family of estimates. The performance of the procedure is illustrated by application to the classification problem. A numerical study demonstrates its nice performance in simulated and real life examples.

REFERENCES

[1] Cavalier L. (2001). On the problem of local adaptive estimation in tomography. *Bernoulli* **7**, 63–78.

[2] Cavalier L., Golubev, G.K., Picard, D., Tsybakov, A. (2002). Oracle inequalities for inverse problems. *Annals of Statistics*, **30**, n.3, 843-874.

[3] Cavalier L., Hengartner N.W. (2005). Adaptive estimation for inverse problems with noisy operators. *Inverse Problems* **21**, 1345–1361.

[4] Cavalier L., Tsybakov, A. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Mathematical Methods of Statistics*, **10**, 247-282.

[5] Cavalier L., Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, **123**, 323-354.

[6] Goldenshluger A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5**, 907-925.

[7] Goldenshluger A., Pereverzev S. (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli* **9**(5), 783–807.

[8] Goldenshluger A., Pereverzev S. (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory and Related Fields* **118**, 169-186.

[9] Golubev G.K. (2004). The Method of Risk Envelope in Estimation of Linear Functionals. *Problems of Information Transmission*, **40**, No. 1, 2004, pp. 53–65. Translated from *Problemy Peredachi Informatsii*, No. 1, 2004, pp. 58–72.

[10] Lepskii, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35**, No. 3, 454–466. Translated from *Teor. Veroyatnost. i Primenen.* 35 (1990), no. 3, 459–470.

[11] Mathé P., Pereverzev S. V. (2001) Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods, *SIAM J. Numer. Anal.* **38**, No. 6, pp. 1999–2021.

[12] Tsybakov A.B.(2000) Adaptive estimation for inverse problems: a logarithmic effect in L_2. *C.R. Acad. Sci. Paris*, ser. 1, t.330, 835-840.

Retrieving exponential Lévy models from option prices using relative entropy

PETER TANKOV

(joint work with Rama Cont)

We consider a problem of calibrating an exponential Lévy model to a set of market option prices. In this model, stock price is represented as $S_t = S_0 e^{rt+X_t}$, where r is the interest rate and (X_t) is a Lévy process on probability space $(\Omega, \mathcal{F}, Q)$ such that e^{X_t} is a Q-martingale. Call option prices can be evaluated as discounted expectations of terminal payoffs:

$$C^Q(T, K) = e^{-rT} E^Q[(S_0 e^{rT+X_T} - K)^+].$$

The probability measure Q is parametrized by the characteristic triplet (σ, ν, γ) of (X_t), where σ is the volatility of the continuous Gaussian component, ν is the Lévy measure and γ denotes the drift coefficient. The martingale condition allows to compute γ from σ and ν.

The calibration problem consists in finding the law $Q(\sigma, \nu)$ such that the model option prices $C^{Q(\sigma,\nu)}(T_i, K_i)$ coincide with the market prices $C^M(T_i, K_i)$ for a set of strikes and maturities $(T_i, K_i)_{i \in I}$. The problem of reproducing the observed option prices exactly is ill-posed: it does not always admit a solution and when it does, the solution is not stable with respect to perturbations in the data. We therefore formulate a regularized version of the calibration problem: find a risk-neutral Lévy process $Q(\sigma, \nu)$, minimizing the following *calibration functional*:

$$J(Q) = \mathcal{E}(Q) + \alpha I(Q|P),$$

over the set of all risk-neutral Lévy processes, where

$$\mathcal{E}(Q) := \sum_{i \in I} w_i (C^Q(T_i, K_i) - C^M(T_i, K_i))^2$$

is the pricing error (w_i are weights assigned to the observations) and

$$I(Q|P) := E^P \left[\frac{dQ}{dP} \log \frac{dQ}{dP} \right]$$

is the relative entropy (Kullback-Leibler distance — see [6]) of $Q(\sigma, \nu)$ with respect to a reference Lévy process $P(\sigma^P, \nu^P, \gamma^P)$. This reference process (prior) allows to introduce additional information into the calibration problem, in order to gain stability. It can be estimated from historical data or, more generally, chosen based on our views about the model underlying the market option prices.

For the relative entropy $I(Q|P)$ to be finite, necessarily $\sigma = \sigma^P$ and $\nu \ll \nu^P$. Therefore, the calibration problem can be expressed as that of finding the Lévy measure $\nu \ll \nu^P$, minimizing

$$(1) \qquad\qquad J(\nu) = \mathcal{E}(\nu) + \alpha I(\nu).$$

After deriving the explicit expression of the relative entropy in terms of ν, ν^P and σ^P, we show that

- The regularized calibration problem always admits a solution (though the solution need not be unique).
- The solutions are continuous with respect to input data and the prior Lévy measure.

Moreover, when the noise level in the data tends to zero, we give the conditions under which the solutions of the regularized problem converge to the solutions of the initial calibration problem.

To solve the regularized calibration problem numerically, we discretize the prior measure on a finite grid of points: $\nu^P = \sum_{i=1}^{N} a_i \delta_{x_i}$. The set of measures ν such that $\nu \ll \nu^P$ is then finite-dimensional, and the calibration functional (1) can be minimized using a gradient descent method. Finally, the continuity of solutions of the calibration problem with respect to the prior Lévy measure allows to conclude that if $\{\nu_n^P\}$ is a sequence of discrete priors, converging weakly to a continuous measure ν^P, the sequence of solutions will also converge weakly to the solution of the calibration problem with prior measure ν^P.

Details of the above results can be found in [5] and the numerical solution of the discretized problem is discussed in [4]. See [3] for background on Lévy processes and exponential Lévy models and [1, 2] for other applications of relative entropy in model calibration.

References

[1] M. Avellaneda, R. Buff, C. Friedman, N. Grandchamp, L. Kruk, and J. Newman, *Weighted Monte Carlo: a new technique for calibrating asset-pricing models*, Int. J. Theor. Appl. Finance, **4** (2001), pp. 91–119.
[2] M. Avellaneda, C. Friedman, R. Holmes, and D. Samperi, *Calibrating volatility surfaces via relative entropy minimization*, Appl. Math. Finance, **4** (1997), pp. 37–64.
[3] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall / CRC Press, 2004.
[4] R. Cont and P. Tankov, *Non-parametric calibration of jump-diffusion option pricing models*, J. Comput. Finance, **7** (2004), pp. 1–49.
[5] R. Cont and P. Tankov, *Retrieving Lévy processes from option prices: Regularization of an ill-posed inverse problem*, SIAM J. Control Optim., **45** (2006), pp. 1–25.
[6] I. Csiszar, *I-divergence geometry of probability distributions and minimization problems*, Ann. Proba., **3** (1975), pp. 146–158.

Reporter: Markus Reiß