Journal of Complexity 30 (2014) 290–308

Contents lists available at ScienceDirect

Journal of Complexity

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

Discrepancy based model selection in statistical inverse problems

Shuai Lu [a], Peter Mathé [b,*]

[a] School of Mathematical Sciences, Fudan University, Shanghai 200433, China
[b] Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany

ARTICLE INFO

Article history:
Received 8 April 2013
Accepted 28 January 2014
Available online 7 February 2014

Keywords:
Discrepancy principle
Parameter selection
Regularization
Statistical inverse problem

ABSTRACT

The reconstruction of solutions in statistical inverse problems in Hilbert spaces requires regularization, which is often based on a parametrized family of proposal estimators. The choice of an appropriate parameter in this family is crucial. We propose a modification of the classical discrepancy principle as an adaptive parameter selection. This varying discrepancy principle evaluates the misfit in some weighted norm, and it also has an incorporated emergency stop. These ingredients allow the order optimal reconstruction when the solution owns nice spectral resolution. Theoretical analysis is accompanied with numerical simulations, which highlight the features of the proposed varying discrepancy principle.

1. Introduction

In this study we are concerned with linear operator equations in Hilbert space, given as

$$y^{\sigma} := Tx^{\dagger} + \sigma\xi, \tag{1}$$

where we assume that the operator $T\colon X \to Y$ is a compact injective operator. The element $x^{\dagger}$ denotes the (unknown) exact solution, and the parameter σ denotes the noise level of the noise ξ, specified below as Gaussian white noise. The problem just constitutes a nonparametric statistical inverse problem. We mention the surveying article [4], and the recent [1] for an account on some fundamental issues for such problems.

* Corresponding author.
E-mail addresses: slu@fudan.edu.cn (S. Lu), mathe@wias-berlin.de (P. Mathé).

http://dx.doi.org/10.1016/j.jco.2014.02.002
0885-064X/

The risk of any estimator $\hat{x} = \hat{x}(y^\sigma)$ will be measured in root mean square (RMS) sense, which is given as

$$e(\hat{x}, x^\dagger, \sigma) := \mathbb{E}\left[\|x^\dagger - \hat{x}\|^2\right]^{1/2}.$$

We first motivate this study, and we start with discussing, why to use the discrepancy, which in its plain form is given as $y^\sigma - \hat{y}$, for any estimator $\hat{x}$ of x and with $\hat{y} = T\hat{x}$. The related residual quantity $\|y^\sigma - \hat{y}\|$ is fundamental in parametric statistics, and there is a reason to minimize the residual by some appropriate estimator $\hat{x} = \hat{x}(y^\sigma)$. For inverse problems, i.e., when the operator T does not have a bounded inverse, this is no longer feasible, and this would lead to over-fitting. However, the use of the misfit $y^\sigma - \hat{y}$ for adaptive parameter selection in inverse problems is the basis of the discrepancy principle (DP), which is doubtlessly the most used strategy for parameter selection in the context of bounded deterministic noise, when we assume that the noise obeys $\|\xi\|_Y \leq 1$. So, having just a given set of data y^σ at hand, one might resort to one of the fundamental monographs in inverse problems, as e.g. [5], and they will find that the discrepancy principle is *the* method of choice. However, a sound theory which explains the use of the discrepancy in statistical inverse problems is still lacking. Also, there are several iterative constructions of estimators $\hat{x}_k, k = 1, 2 \ldots,$ for which the discrepancy principle is one of the few provably optimal parameter selection procedures, among them is *conjugate gradient iteration*. In classical regularization theory this was shown in the seminal study [15]. This iterative regularization scheme was actually the major motivation to use the discrepancy in statistical inverse problems in [1]. Therefore, in this study we shall outline, how to use the discrepancy as a means for model selection in inverse problems.

For the setup in (1), the data y^σ will not belong to Y almost surely. Based on the known behavior of the singular value decomposition of the operator T we may find some power μ of the self-adjoint companion $H = TT^*$, such that $H^\mu(y^\sigma - \hat{y}) \in Y$ almost surely. This will be the case exactly if the operator H^μ is a Hilbert–Schmidt operator (has square summable singular values) by Sazonov's Theorem [18]. For simplicity we assume that this is the case for $\mu = 1/2$, and in this case we may use the smoothed misfit $T^*(y^\sigma - \hat{y}) \in X$. One may think of this as follows. We smoothen Eq. (1), and the regularization will be based on the symmetrized equation with operator $A := T^*T$, and smoothed data $z^\sigma := T^*y^\sigma$, by formally letting

$$z^\sigma = T^*y^\sigma = Ax^\dagger + \sigma T^*\xi.$$

The subsequent analysis will start from this symmetrized equation, and we will use z^σ as given data.

As the analysis in [1] revealed, the plain use of this new misfit $z^\sigma - A\hat{x}$ is possible, but it is not appropriate. Instead, some additional weighting should be used, in order to take into account for the statistical nature of the noise. Specifically, the authors in [1] proposed a modified discrepancy principle (MDP) by controlling the weighted discrepancy

$$\|(\lambda I + A)^{-1/2}(z^\sigma - A\hat{x})\| \tag{2}$$

for any proposal solution $\hat{x}$, and for a fixed value λ, together with an emergency stop.

As in other studies we consider model selection from a parametric family $\hat{x}_\alpha, \alpha \in \Delta$, where the parameters (models) range in a grid

$$\Delta := \left\{\alpha_0 > \alpha_1 := q\alpha_0 > \cdots > \alpha_n := q^n\alpha_0 > \cdots > 0\right\}, \tag{3}$$

for a pre-specified value $0 < q < 1$. In its generic form a discrepancy principle prescribes a threshold $\tau = \tau(\sigma)$, and decreases $\alpha \in \Delta$ as long as $\|(\lambda I + A)^{-1/2}(z^\sigma - A\hat{x}_\alpha)\| > \tau(\sigma)$. Under white noise, and for a perfect fit $z^\sigma - A\hat{x}_\alpha = \sigma\zeta$ we have that

$$\mathbb{E}\left[\|\sigma(\lambda I + A)^{-1/2}\zeta\|^2\right] = \sigma^2 \mathrm{Tr}\left[(\lambda I + A)^{-1}A\right] = \sigma^2\mathcal{N}(\lambda),$$

where we use the function $\mathcal{N}(\lambda)$ as the *effective dimension*. In order to have exponential bounds for the deviation from this mean we introduce another auxiliary parameter κ to be specified later, and we shall consider to check whether for a given $\alpha \in \Delta$ we have

$$\|(\lambda I + A)^{-1/2}(z^\sigma - A\hat{x}_\alpha)\| \leq \tau(1 + \kappa)\sigma\sqrt{\mathcal{N}(\lambda)}.$$