L01

Proceedings of the 5th International Conference on Inverse Problems in Engineering: Theory and Practice, Cambridge, UK, 11-15th July 2005

STATISTICAL INVERSION WITH GREEN'S PRIORS

S. LASANEN[1] and L. ROININEN[2]
[1] *Department of Mathematical Sciences, University of Oulu, Finland*
e-mail: sari.lasanen@oulu.fi
[2] *Sodankylä Geophysical Observatory, University of Oulu, Finland*
e-mail: lassi.roininen@sgo.fi

Abstract - We consider modeling of the unknown with Green's priors in statistical inverse methods. In statistical inverse theory, the direct theory is often formulated in infinite-dimensional space while the Bayesian inversion is carried out in a finite-dimensional space. This raises the question, what happens when dimensionality of the solution space is increased. We study certain function-valued priors and show that their discretized versions satisfy conditions guaranteeing the convergence of corresponding posterior distributions. With Green's priors one can choose the strength of the corresponding regularization systematically according to the discretization and a priori information of the unknown. Then the solution is essentially independent of the discretization and one gets a clear statistical interpretation for the strength of the regularization. Green's priors are derived from Hilbert space-valued stochastic processes. We consider conditions for the convergence of the solution and point out possible sources of non-convergence. We also demonstrate the power of the method numerically.

1. INTRODUCTION

In statistical inverse theory, the objective is to find out some unknown function, for example velocity or conductivity distribution [2, 6], on the basis of noisy indirect measurements. The problem is formally described with the equation

$$y = AX + \varepsilon, \tag{1}$$

where X is the unknown function, A is a known linear mapping and the noise ε is typically a white noise process. We assume that y is a finite-dimensional vector, i.e. the measurements are spatially fixed. The measurement is then a sample of random vector y. If $A^{-1}(\varepsilon)$ is large (or A^{-1} does not even exist), the solution candidate given by $A^{-1}(y)$ is useless. According to Hadamard [3], such problems are called ill-posed. For bounded error terms ε, linear ill-posed problems may be solved with classical regularization methods [12]. In case of Gaussian noise, the error term ε may have very large values with positive probability. Hence classical regularization methods are not applicable and statistical solution methods are used instead [4, 8, 11]. In statistical inverse methods, we refer to the problem as unstable and within the classical framework as ill-posed.

The main point in Tikhonov's regularization method is applying additional information on X in order to compensate the ill-posedness of the problem. In statistical inverse theory, additional information on X is added in the form of probability distributions. For computational reasons, the unknown function is usually discretized. The problem is then to to find the posterior distribution of a finite-dimensional random vector X_m given a sample of

$$y = A_m X_m + \varepsilon, \tag{2}$$

where X_m is a finite-dimensional random vector and A_m is a finite-dimensional approximation of the linear operator A. We will assume that X_m and ε are statistically independent random vectors and ε has a Gaussian distribution $N(0, C_\varepsilon)$.

According to the statistical inverse theory, the solution of eqn. (2) is the following conditional probability density function

$$D(X_m|y) \propto D_{\mathrm{pr}}(X_m) D(y|X_m), \tag{3}$$

where $D_{\mathrm{pr}}(X_m)$ is the prior density of X_m formed on the basis of some additional information on X_m. By the proportionality sign we mean that the corresponding functions differ only by the normalization constant. By eqn (2), the conditional density $D(y|X_m)$ can be written explicitly and it is proportional to $\exp(-\frac{1}{2}(y - A_m X_m)^T C_\varepsilon^{-1}(y - A_m X_m))$.

The discretization of the unknown function raises the question, how does a change of the dimensionality affect the form of the posterior distribution. In other words, can we discretize the unknown in such a way that the appearance of the posterior distribution is somehow independent of the discretization of the unknown. In this paper, our intention is to formulate a Gaussian prior density

$$D_{\mathrm{pr}}(X_m) \propto \exp\left(-\frac{1}{2}X_m^T C_m^{-1} X_m\right) \tag{4}$$

in such a way that we can guarantee convergence of the posterior mean (maximum a posteriori estimate)

$$\hat{X}_m = C_m A_m^T (A_m C_m A_m^T + C_\varepsilon)^{-1} y \tag{5}$$

and of the posterior covariance

$$K_m = C_m - C_m A_m^T (A_m C_m A_m^T + C_\varepsilon)^{-1} A_m C_m. \tag{6}$$

As prior distributions, we shall choose Green's priors. In short, a Green's prior is a mean zero Gaussian measure with covariance operator C having some Green function as its kernel. That is, the kernel $c(x, y)$ satisfies the equation

$$P(D)c(x, y) = \delta(x - y)$$

in distributional sense. Here the partial differential operator $P(D)$ is a strongly elliptic operator of the form

$$P(D) = \sum_{|\alpha|=|\beta|=m} (-1)^{|\alpha|} a_{\alpha\beta} D^\alpha D^\beta, \tag{7}$$

where $a_{\alpha\beta} = a_{\beta\alpha}$ are constants for all multi-indecis α and β. We shall allow also some fractional order differential operators as a natural generalization of eqn. (7).

With Green's priors, we consider only constant coefficient case. Taking $a_{\alpha\beta}$ as variables would require specific prior information on the behavior of the unknown. Consider for example Laplacian $\Delta = \sum_{i=1}^n D_i \delta_{ij} D_j$ on a smooth domain with zero boundary condition for $c(x, y)$. Multiplying the kernel $c(x, y)$ with a constant c^2 means multiplying the coefficients δ_{ij} by the factor c^{-2}. In acoustic problems this corresponds to altering the sound speed from one to c. In this case, replacing the coefficients δ_{ij} with functions $a_{ij}(x)$ would roughly mean specifying how fast our unknown changes from place to place.

2. THEORY

In order to get a unified setting for the Bayesian inversion of the infinite-dimensional linear problem in eqn. (1), we assume that the unknown X is a Gaussian distribution-valued random variable [5]. Here the word distribution is used in the sense of Schwartz to represent a continuous linear form on the test function space $\mathcal{D}(U)$ which is the space of all compactly supported smooth functions on an open set $U \subset \mathbf{R}^n$. The benefit is that then X has a characteristic function $\mathcal{L}$ defined by eqns. $\mathcal{L}(\phi) = \mathbf{E}(\exp(iX(\phi)))$, where $\phi \in \mathcal{D}(U)$. According to Minlos theorem [10], $\mathcal{L}$ determines a Gaussian probability measure on the distribution space when $\mathcal{D}'(U)$ is equipped with the Borel sigma algebra with respect to the weak* topology. Therefore, Gaussian generalized random variables have Gaussian probability distributions on $\mathcal{D}'(U)$ and also the Bayesian inversion can be carried out [8]. We take A as a measurable linear mapping, that extends some known linear operator acting on $L^2(U)$. For simplicity, we assume that $A_m = A$.

The class of all distribution-valued random variables is quite large. Obviously, it contains all Gaussian Hilbert space-valued stochastic processes that operate also on $\mathcal{D}(U)$.

Let W be a Gaussian white noise process on $U = (0, 1)$ so that it has the characteristic function $\mathcal{L}(\phi) = \exp(-\frac{1}{2}\|\phi\|^2_{L^2(U)})$ for all $\phi \in \mathcal{D}(U)$. Let T_α be the fractional integral operator

$$T_\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_0^x (x - s)^{\alpha-1} f(s) ds, \; x \in U \tag{8}$$

where $\alpha > \frac{1}{2}$ and $f \in L^2(U)$. When $\alpha = 0$, the operator T_α is the identity. Following [1, 7], we define an extension

$$\hat{T}_\alpha W = \sum_{j=1}^\infty W(\phi_j) T_\alpha \phi_j \tag{9}$$

for any orthonormal basis $\{\phi_j\}$ of $L^2(U)$. This gives a new Hilbert space-valued stochastic process Z. Furthermore, we multiply the process by σ and obtain a new process $X = \sigma Z$.

The covariance operator C_X for the process X is the integral operator $T_\alpha T_\alpha^T$ with kernel $c_\alpha(t,s)$. For example, with white noise process $c_0(t,s) = \delta(t-s)$, and with Brownian motion $c_1(t,s) = \min(t,s)$, and with the integrated Brownian motion $c_2(t,s) = \frac{1}{2}\max(t,s)\min(t,s)^2 - 1/6\min(t,s)^3$. These Hilbert space-valued processes are Green's priors. For integer values of α, the kernel c_α is the Green function for the differential operator $(-1)^\alpha D^{2\alpha}$ with zero boundary value at the origin. For fractional orders of α, the kernel is the Green function for the fractional order differential operator $I_{-\alpha}^T I_{-\alpha}$ defined by analytic continuation with respect to α [9]. They describe the unknown as an element of the corresponding reproducing kernel Hilbert space

$$H_X = \{h \in \mathcal{D}'(U) : \sup\{\langle h, \phi\rangle : \langle C_X\phi, \phi\rangle \le 1, \phi \in \mathcal{D}(U)\} < \infty\}. \tag{10}$$

In this case, the elements of H_X are simply images of $L^2(U)$ functions under the fractional integral operator T_α i.e. $H_X = T_\alpha(L^2(U))$ [1].

Let us define a discretization of X by considering dyadic divisions of the interval U, i.e. U is divided into 2^m consecutive subintervals of equal length h_m. We take

$$X_m = \sigma\hat{T}_\alpha^{(m)}W \tag{11}$$

where $T_\alpha^{(m)}$ is an $O(h_m)$ convergent discretization of the fractional integral operator given by the fractional Euler method [9]. The covariance operator $C_m = T_\alpha^{(m)}(T_\alpha^{(m)})^T$ has the kernel

$$c_\alpha^{(m)}(t,s) = \sigma^2 \sum_{i,j=1}^{2^m} h_m^{2\alpha-1} \sum_{k=1}^{\min(i,j)} (-1)^{i+j} \binom{-\alpha}{j-k}\binom{-\alpha}{i-k} \mathbb{I}_{I_i}(t)\mathbb{I}_{I_j}(s), \tag{12}$$

where $\mathbb{I}_V$ denotes the indicator function of the set V, i.e. $\mathbb{I}_V \equiv 1$ on V and is zero elsewhere. These finite-dimensional operators are naturally identified with matrices. The action of this part of the prior is visualized when we consider the Cholesky factorization of its inverse matrix keeping the connection to the Tikhonov regularization method in mind. For integer values of α, the rows of these matrices are difference operators of order α having $\sqrt{h_m^{1-2\alpha}}$ as multiplier.

Theorem 1 *Let $A : L^2(U) \to \mathbf{R}^n$ be a continuous linear operator. Let $X_m = \sigma\hat{T}_\alpha^{(m)}W$, $X = \sigma\hat{T}_\alpha W$ and ε a Gaussian zero mean random vector on $\mathbf{R}^n$ statistically independent of X. Then the posterior measure for the problem $y = AX_m + \varepsilon$ converges weakly to the posterior measure for the problem $y = \hat{A}X + \varepsilon$.*
Proof — The posterior measures converge weakly on $D'(U)$ if the characteristic functions

$$\int_{D'(U)} \exp(i\langle u, \phi\rangle)d\mu_m(u) = \exp\left(i\langle \hat{X}_m, \phi\rangle - \frac{1}{2}\langle K_m\phi, \phi\rangle\right) \tag{13}$$

converge for every test function ϕ. The posterior mean for the continuous problem is $C_XA'(AC_XA' + C_\varepsilon)^{-1}y$ and the corresponding posterior covariance operator is $C_X - C_XA'(AC_XA' + C_\varepsilon)^{-1}AC_X$ [8]. Similarly as in the resolvent equation,

$$\begin{aligned} & C_mA'(AC_mA' + C_\varepsilon)^{-1} - C_XA'(AC_XA' + C_\varepsilon)^{-1} \\ = \; & (C_m - C)A'(AC_XA' + C_\varepsilon)^{-1} + C_mA'(AC_mA' + C_\varepsilon)^{-1}A(C_X - C_m)A'(AC_XA' + C_\varepsilon)^{-1}. \end{aligned} \tag{14}$$

Note that $C_mA'(AC_mA' + C_\varepsilon)^{-1}A$ are uniformly bounded on L^2. This converges since $(C_m - C_X)A'z$ converges in L^2 for all fixed vectors z.

In the case of the covariance operators, we study convergence of $K_m\phi$ in $L^2(U)$ for every test function ϕ. By the above arguments, these functions converge on $L^2(U)$, even though the convergence can not be seen in the zero order case $\alpha = 0$ from the behavior of the single elements of posterior covariance matrices. This is due to the fact that the kernel contains Dirac's delta function. □

By considering, e.g. approximations of Gamma function, it can be shown that values $c_\alpha^{(m)}(t,s)$ converge to $c_\alpha(t,s)$. This implies that posterior means converge pointwisely for $\alpha > 1/2$. Furthermore, if the kernel of A is smooth, then the posterior means converge pointwisely also for $\alpha = 0$ and the order of convergence is $O(h_m)$.

3. GREEN'S PRIORS IN TWO DIMENSIONS

We suggest the following two dimensional generalization. Let W_1 and W_2 be independent white noise

processes on the rectangular U^2. We generalize operator T_α defined in eqn. (8) by integrating only with respect to one coordinate axis. For $i = 1, 2$, we denote

$$T_{\alpha,i} f(x) = \frac{1}{\Gamma(\alpha)} \int_0^{x_i} (x_i - s)^{\alpha-1} f(x) dx_i, \tag{15}$$

where $\alpha > \frac{1}{2}$ and $f \in L^2(U^2)$. Similarly as before, let $Z_i = \sigma \hat{T}_{\alpha,i} W_i$ for $i = 1, 2$. The covariance operator of Z_i has a kernel

$$\begin{aligned} c_{Z_1}(x, y) &= \delta(x_2 - y_2) c_\alpha(x_1, y_1), \\ c_{Z_2}(x, y) &= \delta(x_1 - y_1) c_\alpha(x_2, y_2). \end{aligned} \tag{16}$$

Taking the distributional derivatives for integer values of α shows that

$$\sigma^{-2}(-1)^\alpha D_i^{2\alpha} c_{Z_i}(x, y) = \delta(x - y) \tag{17}$$

for $i = 1, 2$.

Let μ be a Gaussian measure with mean zero and covariance

$$C = C_{Z_1} - C_{Z_1}(C_{Z_1} + C_{Z_2})^{-1} C_{Z_1} \tag{18}$$

so that it is obtained as a posterior measure for Z_1 given the measurement $0 = Z_1 - Z_2$. Covariance operators are invertible on their reproducing kernel Hilbert spaces, so

$$(C_{Z_1} + C_{Z_2})(C_{Z_1} + C_{Z_2})^{-1} = I$$

on $H_{Z_1+Z_2}$. This leads to the equation $C_{Z_1}(C_{Z_1} + C_{Z_2})^{-1} = I - C_{Z_2}(C_{Z_1} + C_{Z_2})^{-1}$. Inserting this into eqn. (18) gives

$$C = C_{Z_2}(C_{Z_1} + C_{Z_2})^{-1} C_{Z_1} = C_{Z_1}(C_{Z_1} + C_{Z_2})^{-1} C_{Z_2}, \tag{19}$$

where the last equality follows from the fact that the transpose of the covariance operator is the covariance operator itself. Taking now the distributional derivatives and using eqn. (17) gives that

$$\sigma^{-2}(-1)^\alpha (D_1^{2\alpha} + D_2^{2\alpha}) c(x, y) = \delta(x - y). \tag{20}$$

Hence $c(x, y)$ is the Green function for the scaled Laplacian with zero boundary values along the coordinate axis.

We define the discretized counterpart for C by replacing the kernel c_{Z_i} with

$$c_{Z_i^{(m)}}(x, y) = \sigma^2 \sum_{j=1}^{2^m} \frac{1}{h_m} \mathbb{I}_{I_j}(x_i') \mathbb{I}_{I_j}(y_i') c_\alpha^{(m)}(x_i, y_i) \tag{21}$$

and denote such discretized covariance operator by C_m. Here x_i' is the vector x without the ith component. As in the previous section, the convergence of $\langle C_m \phi, \psi \rangle$ to $\langle C\phi, \psi \rangle$ for all test functions ϕ and ψ may be reduced to the convergence of $C_{Z_i^{(m)}}$ on fixed functions. The discretization of delta function as well as the the discretization of $T_{\alpha,i}$ converges in norm for any L^2-function. Hence the discretized prior measures $N(0, C_m)$ converge weakly to $N(0, C)$ and also the posterior measures for the statistical inverse problem (2) converge.

When $\alpha = 1$, C_m is the inverse for the discrete Laplacian. The connection is seen from the equation

$$-(\sigma h_m)^{-2}(\Delta_1(h_m) + \Delta_2(h_m)) c_m(x, y) = \sum_{j=1}^{2^{2m}} \frac{1}{h_m^2} \mathbb{I}_{Q_j}(x) \mathbb{I}_{Q_j}(y), \tag{22}$$

where $\Delta_i(h_m)$ is the square of the difference operator in direction i with stepsize h_m and Q_j are squares originating from the dyadic divisions of coordinate axis. Taking $1/h_m \mathbb{I}_{Q_j}$ as basis functions would lead to a matrix equation. Solving the matrix eqn. (22) directly provides another way of producing discretized prior covariance matrices related to a given Green function.

4. NUMERICAL EXAMPLE

Our intention is now to demonstrate the power of the method developed in the previous sections. For

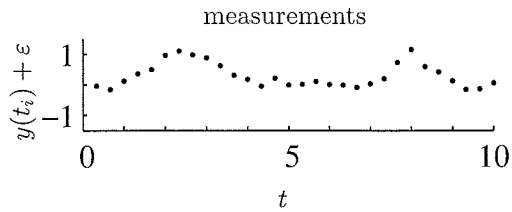

Figure 1: Time-series to be differentiated.

brevity, we shall concentrate here only to the one-dimensional case, even though the theory can be applied to two-dimensional problems. This is due to the complexity of the two-dimensional cases. They need a separate treatment in order to get a unified approach.

In this paper we shall consider one-dimensional time-series problem – differentiation of the following function on the basis of noisy measurements.

$$y(t) = \begin{cases} \exp\left(4 - \frac{25}{t(5-t)}\right) & t \in (0,5) \\ t - 7 & t \in (7,8] \\ -t + 9 & t \in (8,9) \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

The first part of the function is the so-called mollifier function and the second part is a triangle function. The function is parametrized in such a way that the values of its derivative lie approximately between $[-1, 1]$. In this way we can get a clear and systematic interpretation for the maximum a posteriori estimate and for the posterior covariance. With this kind of parametrization we can also avoid unnecessary problems during the computations.

We consider differentiation of the function in the eqn. (23) by transforming differential equation $y'(t) = X(t)$ to the first kind Fredholm integral equation corrupted by white noise. The equation can be written as

$$y(t) = \int H(t - t')X(t')dt' + \varepsilon(t). \tag{24}$$

$H(t - t')$ is the Heaviside step function. The eqn. (24) above is easily discretized and the discretized version can be described with matrix notation as

$$\mathbf{y} = \mathbf{AX} + \varepsilon, \quad \varepsilon \sim N(0, C_\varepsilon). \tag{25}$$

Let us call the vector $\mathbf{y}$ as measurements. For the numerical simulation the measurements shall be spatially sampled with $t_i = 10i/30$, $i = 1, 2, \ldots, 30$. We shall add some Gaussian white noise with distribution $N(0, 0.1^2)$ to measurements. Measurements are visualized in figure 1.

It gives a good intuition if we give the a priori information given in eqn. (12) as synthetic measurements. Let us consider the case when $\alpha = 1$, i.e. Brownian motion. The prior information can then be given by the following equations

$$\begin{aligned} X_1 &= \tilde{\varepsilon}_0, \quad \tilde{\varepsilon}_0 \sim N(0, \sigma^2 h) \\ X_i - X_{i-1} &= \tilde{\varepsilon}_i, \quad \tilde{\varepsilon}_i \sim N(0, \sigma^2 h), \end{aligned}$$

where h is the discretization step. With matrix notation the a priori information can be written as

$$\mathbf{0} = \mathbf{A}_{\text{pr}}\mathbf{X} + \tilde{\varepsilon}, \quad \tilde{\varepsilon} \sim N(0, \Sigma_{\text{pr}}). \tag{26}$$

In this case, the matrix $\mathbf{A}_{\text{pr}}$ is a lower triangular matrix having values only at the diagonal and at the first band. By combining the two matrix eqns (25) and (26) we obtain a large matrix equation. Solution of the matrix eqn. $\hat{\mathbf{X}}$, i.e. maximum a posteriori estimate, can be computed as a generalized least-squares problem.

$$\hat{\mathbf{X}} = (\mathbf{A}^T C_\varepsilon^{-1} \mathbf{A} + \mathbf{A}_{\text{pr}}^T \Sigma_{\text{pr}}^{-1} \mathbf{A}_{\text{pr}})^{-1} \mathbf{A}^T C_\varepsilon^{-1} \mathbf{y}. \tag{27}$$

For Brownian motion $\alpha = 1$, the covariance matrix Σ_{pr} turns out to be an identity matrix multiplied by the discretization and stochastic factors $\Sigma_{\text{pr}} = \sigma^2 h \mathbf{I}$. With the classical Tikhonov notation in mind we can change the latter eqn. (27) to the form

$$\hat{\mathbf{X}} = \left(\mathbf{A}^T C_\varepsilon^{-1} \mathbf{A} + \frac{1}{\sigma^2 h} \mathbf{A}_{\text{pr}}^T \mathbf{A}_{\text{pr}} \right)^{-1} \mathbf{A}^T C_\varepsilon^{-1} \mathbf{y}. \tag{28}$$

Thus we see that the classical regularization parameter corresponds to $\sigma^2 h$. But with the discretization method developed in the section 2, we get a natural explanation of the regularization parameter. The first term of the parameter defines certain statistical properties of the stochastic process. More precisely, σ is the constant multiplier for the process which means that $\mathbf{E}(X_t X_s) = \sigma^2 \min(t, s)$. The second term takes care of that the solution is essentially independent of the discretization, so that the we have a clear convergence of the posterior distributions.

Similarly, for the integral function of the Brownian motion i.e. $\alpha = 2$ the synthetic measurements are $X_{i+1} - 2X_i + X_{i-1} = \tilde{\varepsilon}_i$, $\tilde{\varepsilon}_i \sim N(0, \sigma^2 h^3)$. The boundary conditions are $X_1 \sim N(0, \sigma^2 h^3)$ and $2X_2 - X_1 \sim N(0, \sigma^2 h^3)$. For the white noise process $\alpha = 0$ one gets $X_i = \tilde{\varepsilon}_i$, $\tilde{\varepsilon}_i \sim N(0, \sigma^2/h)$. Thus one can see that the covariance Σ_{pr} is always an identity matrix multiplied by two factors – the discretization factor and the stochastic factor.

Thus, when α is an integer the regularization matrices turn out to be standard difference matrices. When α is a fractional number, the regularization matrices are fractional difference matrices. They can be written as

$$\sum_{k=0}^{i-1} (-1)^k \begin{pmatrix} \alpha \\ k \end{pmatrix} X_{i-k} = \tilde{\varepsilon}_i, \quad \tilde{\varepsilon}_i \sim N\left(0, \sigma^2 h^{2(\alpha - \frac{1}{2})}\right). \tag{29}$$

We have computed numerical examples with different regularization strengths ($\sigma^2 = 0.01$, 1, 100) and with different orders of regularization ($\alpha = 0$, 1, 1.35, 2). The visualizations of the maximum a posteriori estimates are shown in figures 2 and 3. The number of the unknowns are 10, 20, 50 and 80. We have also computed posterior covariance for the first order prior $\alpha = 1$ and $\sigma^2 = 1$. Visualizations are shown in figure 4.

It can be easily seen that in nearly all the cases – except in the zero order case – the appearance of the estimate seems to be the same when the discretization changes. However, in the zero order case, the situation is a bit more difficult. The kernel function i.e. the Heaviside step function is not a continuous function and thus one cannot see the convergence with pointwise values. Instead, we should consider the convergence for every test function. In the zero order case the estimate consists of "stairs", because the basis functions turn out be of the similar form. There are naturally 30 stairs, because we have got 30 measurements. This is due to the fact that the dimensionality of the solution space cannot be greater than the dimensionality of the measurement space.

It is very difficult to see whether the solution is of the same form for very sparse grids (i.e. 10 unknowns) as for dense grids. Thus one should pose the question, how many unknowns should one have in order to get clear convergence.

It is practically impossible to visualize Gaussian probability distribution with more than two variables. It is common to visualize the distribution by displaying maximum a posteriori estimate with some kind of error boundaries for pointwise values defined by the diagonal values of the posterior covariance. However this approach demands that the matrix is "sufficiently" diagonally dominant. In our case the posterior covariance matrix is not diagonally dominant. Thus one should visualize the whole posterior covariance and maximum a posteriori estimate.

The visualizations of the posterior covariance in the case $\alpha = 1$, $\sigma^2 = 1$ are displayed as images in figure 4. They seem to have a clear, even though a slow convergence. If one looks at pointwise values, it is apparent that with sparse grids – 10, 20 unknowns – one cannot see any convergence. For denser grids – 50, 80 unknowns – the convergence can be clearly seen. Thus one should note that the maximum a posteriori estimate shows clear convergence when there are 20 unknowns. For the posterior covariance one needs at least 50 or more unknowns in order to get a proper convergence. A more detailed study of the posterior covariances shall be carried out in later papers.

As a conclusion, one should choose the number of the unknowns large enough and check continuity conditions. One should also take care of that the regularization strength is suitable for the problem by modeling the unknown properly. One should also ask, how many measurements must one have in order to get proper results.

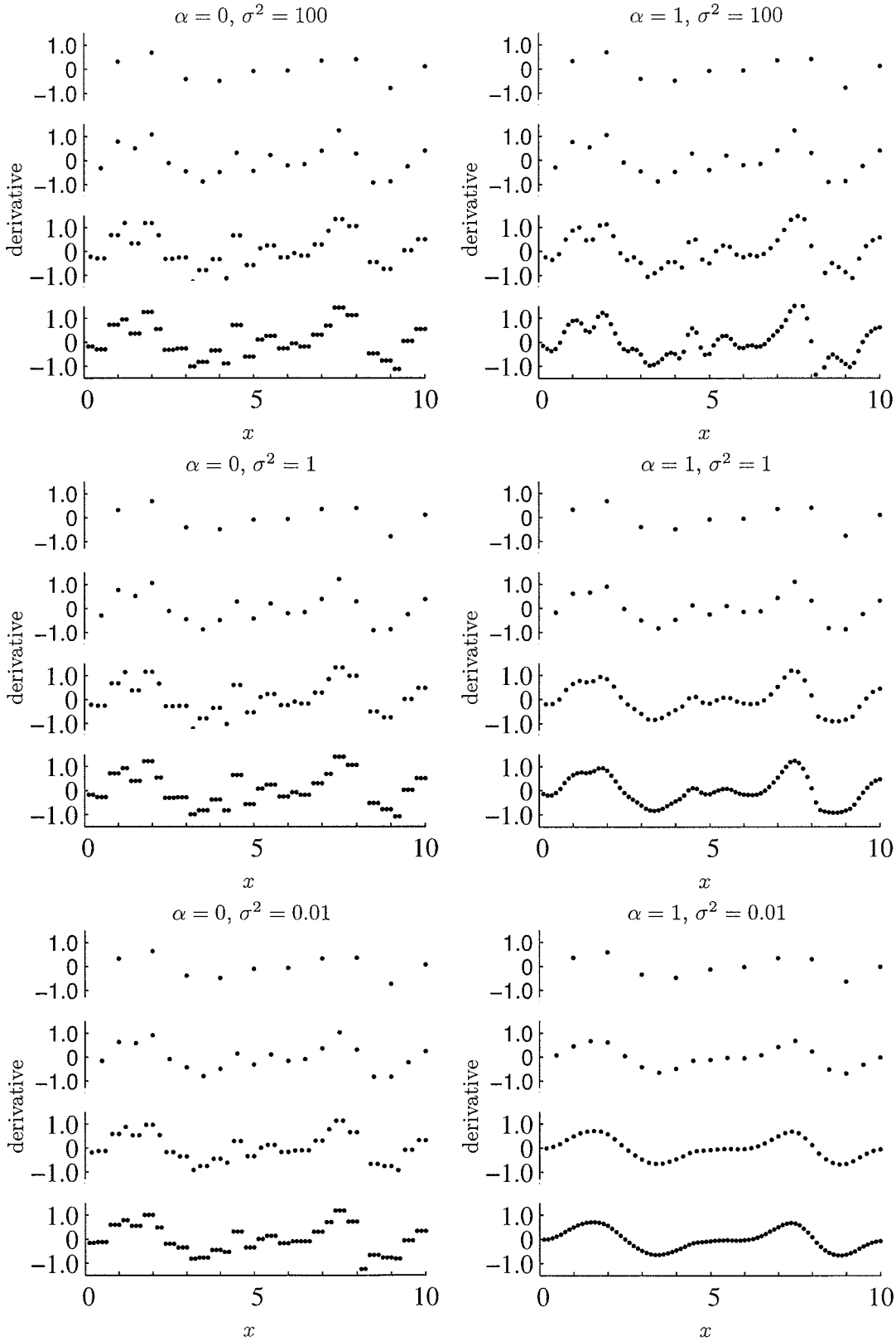

Figure 2: Center-point estimates with different regularization strengths for zero and first order priors.

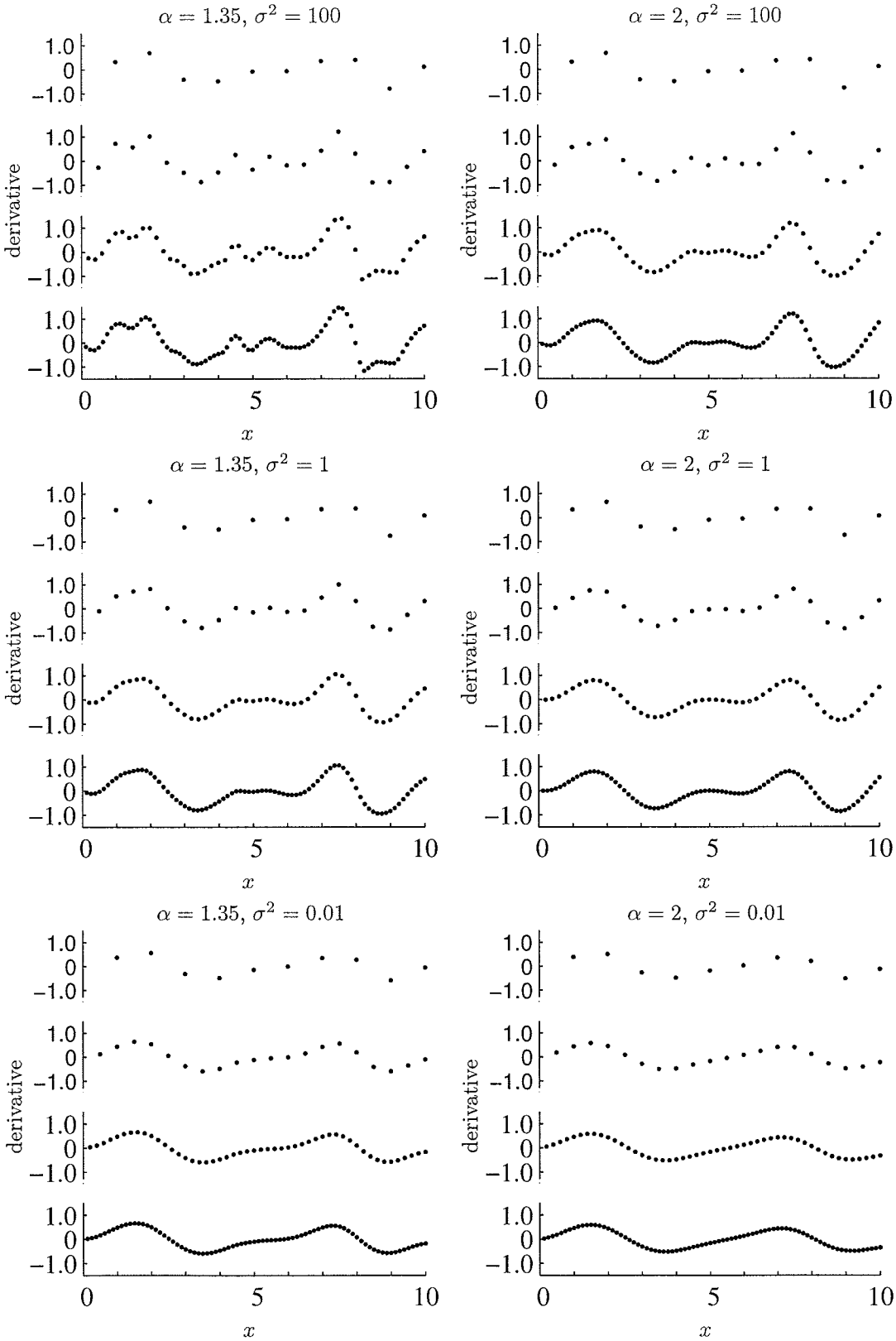

Figure 3: Center-point estimates with different regularization strengths for fractional order 1.35 prior and second order prior.

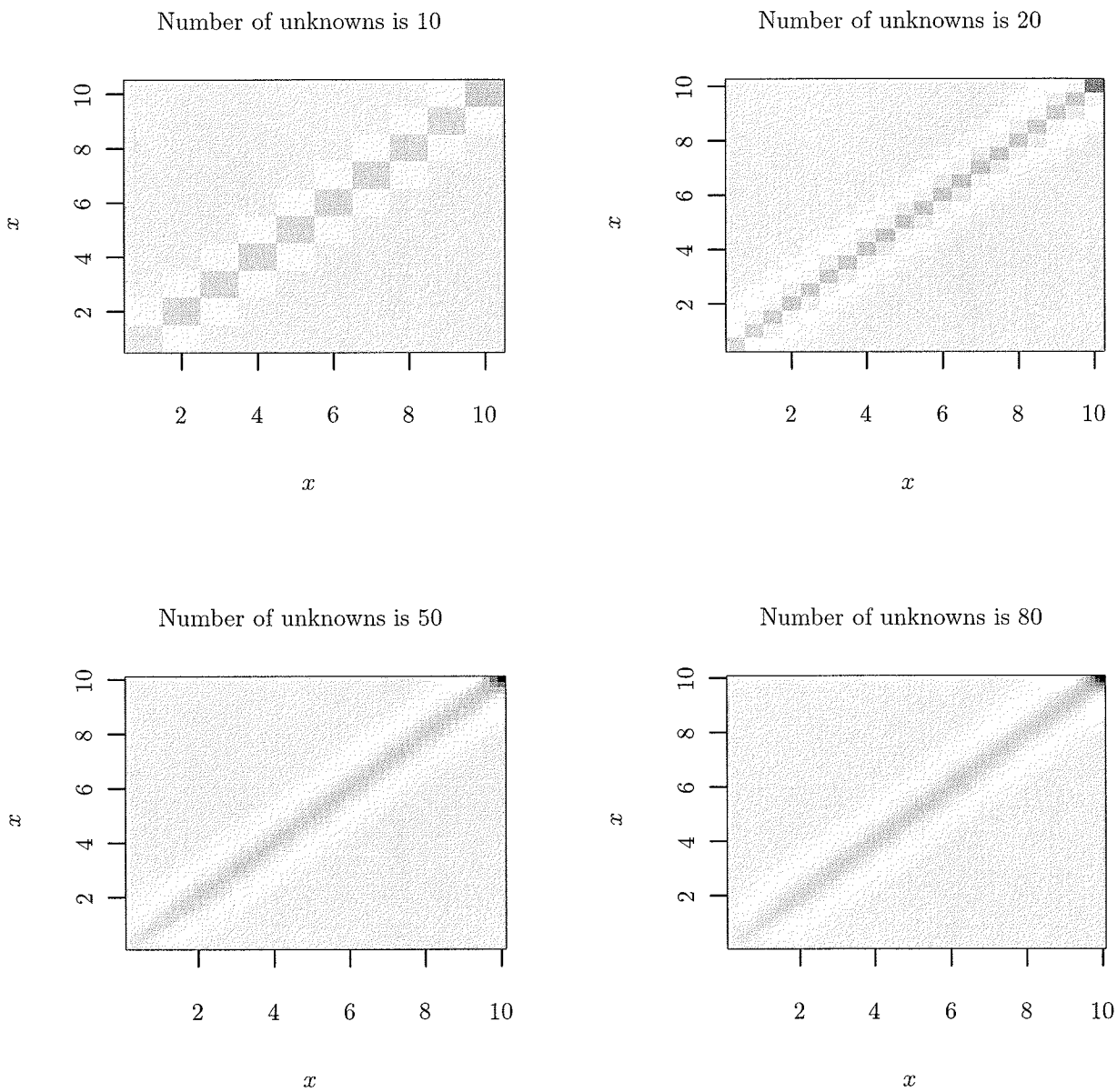

Figure 4: Posterior covariances for the first order prior, i.e. Brownian motion with regularization strength $\sigma^2 = 1$.

5. CONCLUSIONS

We have established a discretization method in statistical inverse theory in such a way that the posterior distribution is essentially independent of the discretization. We have shown convergence of the posterior distributions and demonstrated the power of the method numerically in the one-dimensional case.

Compared with the classical regularization, the regularization parameter has a clear statistical and systematical interpretation. Moreover, the prior information can be viewed as characteristics of the used Hilbert-space valued stochastic process. This allows for every applicant a systematic way of modeling the unknown.

Acknowledgement

The authors want to thank M. Markkanen for helpful discussions. This work was partially supported by Academy of Finland.

REFERENCES

1. V.I. Bogachev, *Gaussian Measures*, AMS, Providence RI, 1998.
2. H. Haario, M. Laine, M. Lehtinen, E. Saksman and J. Tamminen, Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *J. Roy. Stat. Soc.: Series B (Statistical Methodology)* (2004), **66**, 591-607.
3. J. Hadamard, *Le Probléme de Cauchy et les Equations aux Dérivées Partielles Linéaires Hyperboliques*, Herman et Cie, Paris, 1932.
4. J.N. Franklin, Well-posed stochastic extension of ill-posed linear problems. *J. Math. Anal. Appl.* (1970), **31**, 682-715.
5. I.M. Gelfand and N.Y. Vilenkin,*Generalized Functions. Vol. 4: Applications of Harmonic Analysis*, Academic Press, New York, 1968.
6. J.P. Kaipio, V. Kolehmainen, E. Somersalo and M. Vauhkonen, Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems* (2000), **16**, 1487–1522.
7. S. Lasanen, Discretizations of generalized random variables with applications to inverse problems (Dissertation). *Ann. Acad. Sci. Fenn. Math. Diss.* (2002), No. **130**.
8. M.S. Lehtinen, L Päivärinta and E. Somersalo, Linear inverse problems for generalized random variables. *Inverse Problems* (1989), **5**, 599-612.
9. C. Lubich, Discretized fractional calculus. *SIAM J. Math. Anal.* (1986), **17**, 704-719.
10. L. Schwartz, *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*, Oxford University Press, London, 1973.
11. A. Tarantola and B. Valette, Inverse problems = quest for information. *J. Geophys.* (1982), Vol. **50**, 159-170.
12. A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, John Wiley & Sons, Washington DC, 1977.