

VARIABLE ORDER SMOOTHNESS PRIORS FOR ILL-POSED INVERSE PROBLEMS

DANIELA CALVETTI, ERKKI SOMERSALO, AND RUBEN SPIES

ABSTRACT. In this article we discuss ill-posed inverse problems, with an emphasis on hierarchical variable order regularization. Traditionally, smoothness penalties in Tikhonov regularization assume a fixed degree of regularity of the unknown over the whole domain. Using a Bayesian framework with hierarchical priors, we derive a prior model, formally represented as a convex combination of autoregressive (AR) models, in which the parameter controlling the mixture of the AR models can dynamically change over the domain of the signal. Moreover, the mixture parameter itself is an unknown and is to be estimated using the data. Also, the variance of the innovation processes in the AR model is a free parameter, which leads to conditionally Gaussian priors that have been previously shown to be much more flexible than the traditional Gaussian priors, capable, e.g., to deal with sparsity type prior information. The suggested method, the Weighted Variable Order Autoregressive model (WVO-AR) is tested with a computed example.

1. INTRODUCTION

Consider the one-dimensional inverse problem of the following form: Let $f : I = [0, 1] \rightarrow \mathbb{R}$ be a piecewise continuous function, and $\{G_j\}$, $1 \leq j \leq m$, a family of linear or non-linear functionals acting on those functions. We define a model for noisy indirect observations of f ,

$$(1) \quad b_j = G_j(f) + e_j, \quad 1 \leq j \leq m,$$

where e_j represents a random additive noise. The inverse problem is to estimate the function f on the interval I from the knowledge of $\{b_j\}$.

Examples of problems of this type include deblurring, in which the functional G_j is a convolution with a point spread function g , $G_j(f) = (g * f)(s_j)$, $s_j \in I$, and interpolation with denoising, in which case $G_j(f) = f(s_j)$. Numerical differentiation of a function vanishing at $t = 0$ can also be recast in this form by defining

$$G_j(f) = \int_0^{t_j} f(t) dt.$$

A characteristic feature of these inverse problems is their ill-posedness of various degree, meaning that without proper regularization, the solution may be non-unique, may not exist, or may depend in a discontinuous way on the data, which renders the solutions sensitive to noise and straightforward naïve solutions are doomed to fail.

Received by the editor January 27, 2013 and, in revised form, October 4, 2013.

2010 *Mathematics Subject Classification*. Primary 65F22; Secondary 65C20.

The second author wishes to thank NSF for partial support of this work (DMS 1016183).

©2014 American Mathematical Society

The standard regularization methods of ill-posed inverse problems include the Tikhonov regularization, truncated SVD, and numerous iterative methods; see, e.g., [9, 11, 13] for references. The classical regularization methods include a scalar regularization parameter, such as the Tikhonov parameter or the truncation index, and methods for finding in a given sense optimal values for regularization parameters are well known, including Morozov's discrepancy principle [14], the L-curve method [12], or the generalized cross validation [10]. While performing well in many problems, the compression of the information about regularization into one single scalar may be too restrictive. In the statistical framework, ([6, 13, 15], the Tikhonov regularization can be interpreted as a Maximum A Posteriori (MAP) estimate with Gaussian prior and likelihood, and this interpretation opens rich possibilities to generalize the regularization to be spatially varying, and to deal with prior information not handled well by quadratic regularization functionals ([4, 7, 8]). In particular, the hierarchical models can be used as an alternative for sparsity priors such as the ℓ^1 -prior or total variation prior [1, 16].

We approach the variable smoothness priors through autoregressive (AR) models that are widely used in time series analysis [3]. A commonly encountered problem in AR analysis of time series is to find the coefficients of the autoregressive model, as well as the degree of the model, and a number of standard techniques to do this are known, starting from the classical reference [2]. In this article, we are only borrowing the AR formalism to express a priori local smoothness in the sense of regularization of an inverse problem, and not for prediction. In the latter case, the root condition for the AR model for stability should be checked. In the context of regularization, the stability is not an issue, while the question of determining the model coefficients is tantamount to deciding the degree of the regularizing smoothness penalty based on the signal itself. The methodology developed in this paper starts with the idea of defining an AR model locally as a convex combination of two fixed AR models, and the interpolation parameter is adjusted so as to optimally capture the underlying signal. To assure flexibility, the AR model is allowed to vary from point to point, as the regularity of the underlying signal may be varying.

2. INVERSE PROBLEMS IN THE STATISTICAL FRAMEWORK

To render the problem (1) computationally feasible, we discretize it first, e.g., by representing the unknown function in an appropriate basis, or, as we choose in this article, by collocation, denoting

$$x_j = f(t_j), \quad \text{where } t_j = \frac{j}{n}, \quad j = 0, 1, \dots, n,$$

and by $x \in \mathbb{R}^{n+1}$ we denote the vector with entries x_j .

To regularize the inverse problem, we recast it in the Bayesian framework, defining a multivariate random vector $X \in \mathbb{R}^{n+1}$ with x as its realization, and define a prior probability density that summarizes our a priori beliefs about the unknown, most notably concerning its smoothness or modes of variation.

2.1. Priors and Markov models. A commonly used family of stochastic models leading to numerous useful prior models consists of autoregressive Markov models of the form

$$(2) \quad X_j = \sum_{k=1}^p \alpha_k X_{j-k} + \sqrt{\gamma_j} W_j, \quad W_j \sim \mathcal{N}(0, 1), \quad 0 \leq j \leq n,$$

where the order p and the regression coefficients α_k are given, $\{W_j\}$ is the innovation process of mutually independent standard normal variables, and γ_j is the variance of the j th innovation. For the sake of definiteness, we assume that

$$(3) \quad X_j = 0 \text{ almost certainly for } j < 0.$$

This restriction is natural for time series, but less so when Markov model induced smoothness priors are discussed. We shall return to this condition later.

We will focus, in particular, on smoothness priors in inverse problems corresponding to the following autoregressive models of order 0, 1 and 2,

$$(4) \quad X_j = \sqrt{\gamma_j^{(0)}} W_j,$$

$$(5) \quad X_j = X_{j-1} + \sqrt{\gamma_j^{(1)}} W_j,$$

$$(6) \quad X_j = 2X_{j-1} - X_{j-2} + \sqrt{\gamma_j^{(2)}} W_j,$$

for $j = 0, 1, \dots, n$. To see the connection with smoothness priors, we write the above models in a matrix form, yielding

$$(7) \quad \mathbf{L}_p X = \mathbf{D}_{\gamma^{(p)}}^{1/2} W, \quad p = 0, 1, 2,$$

where $\mathbf{L}_0 = \mathbf{I}_{n+1}$, the unit matrix of size $(n + 1) \times (n + 1)$, and the matrices $\mathbf{L}_p \in \mathbb{R}^{(n+1) \times (n+1)}$, $p = 1, 2$, are lower triangular finite difference matrices,

$$\mathbf{L}_1 = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}, \quad \mathbf{L}_2 = \begin{bmatrix} 1 & & & & \\ -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots \\ & & & 1 & -2 & 1 \end{bmatrix},$$

and, finally, the matrices $\mathbf{D}_{\gamma^{(p)}} \in \mathbb{R}^{(n+1) \times (n+1)}$ are diagonal matrices,

$$\mathbf{D}_{\gamma^{(p)}} = \begin{bmatrix} \gamma_0^{(p)} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \gamma_n^{(p)} \end{bmatrix}.$$

From equation (7), we conclude that the prior distribution for X is

$$\pi_{\text{prior}}(x) \propto \exp \left(-\frac{1}{2} \|\mathbf{D}_{\gamma^{(p)}}^{-1/2} \mathbf{L}_p x\|^2 \right),$$

the symbol “ \propto ” indicating proportionality with an appropriate normalizing constant.

When choosing the prior Markov model, it is usually assumed that the model captures the regularity assumptions about the unknown function over the whole interval. However, one may imagine a situation in which one model describes well the prior belief over a subinterval, while another model would be more appropriate over another. Moreover, in general, we may not know a priori how the smoothness of the function f varies over the domain. Therefore we propose a *weighted variable order autoregressive* (WVO-AR) model.

Consider two AR models,

$$(8) \quad \begin{aligned} X_j &= \sum_{k=1}^p \alpha_k X_{j-k} + \sqrt{\gamma_j^{(1)}} W_j^{(1)}, & W_j^{(1)} &\sim \mathcal{N}(0, 1), \\ X_j &= \sum_{k=1}^q \beta_k X_{j-k} + \sqrt{\gamma_j^{(2)}} W_j^{(2)}, & W_j^{(2)} &\sim \mathcal{N}(0, 1), \end{aligned}$$

$j = 0, 1, \dots, n$. For the sake of definiteness, we assume that $p \geq q$. A convex combination of these models results in the model

$$X_j = \sum_{k=1}^p [(1 - \theta_j)\alpha_k + \theta_j\beta_k] X_{j-k} + \sqrt{\gamma_j} W_j, \quad W_j \sim \mathcal{N}(0, 1),$$

where $\theta_j \in [0, 1]$, and the variance of the innovation term is

$$(9) \quad \gamma_j = (1 - \theta_j)^2 \gamma_j^{(1)} + \theta_j^2 \gamma_j^{(2)},$$

and it is interpreted that for $k > q$, $\beta_k = 0$. By varying the parameter θ_j between the extreme values the model varies locally between the Markov models of order p and q . In matrix notation, the model can be written as

$$(10) \quad \mathbf{L}_\theta X = \mathbf{D}_\gamma^{1/2} W,$$

where

$$\mathbf{L}_\theta = (\mathbf{I} - \mathbf{D}_\theta) \mathbf{L}_p + \mathbf{D}_\theta \mathbf{L}_q,$$

the matrix \mathbf{D}_θ being

$$\mathbf{D}_\theta = \begin{bmatrix} \theta_0 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & \theta_n \end{bmatrix}.$$

We write the prior density of X , given the variable weight vector θ and the innovation variance vector γ as

$$(11) \quad \begin{aligned} \pi_{\text{prior}}(x \mid \theta, \gamma) &\propto \det(\mathbf{D}_\gamma^{-1/2} \mathbf{L}_\theta) \exp\left(-\frac{1}{2} \|\mathbf{D}_\gamma^{-1/2} \mathbf{L}_\theta x\|^2\right) \\ &= \exp\left(-\frac{1}{2} \|\mathbf{D}_\gamma^{-1/2} \mathbf{L}_\theta x\|^2 - \frac{1}{2} \sum_{j=0}^n \log \gamma_j\right), \end{aligned}$$

where we used

$$\det(\mathbf{D}_\gamma^{-1/2} \mathbf{L}_\theta) = \det(\mathbf{D}_\gamma^{-1/2}) \det(\mathbf{L}_\theta) = \prod_{j=0}^n \gamma_j^{-1/2},$$

since the matrix \mathbf{L}_θ is lower triangular with unit diagonal. We shall discuss later modifications of the model in which this identity is not valid. We call the model the (p, q) -Markov model.

We remark that the variances of the innovation processes of different order Markov models may be significantly different when the models are used to explain the same data. To compensate this effect, we will later consider also interpolated models where the matrices \mathbf{L}_p are first scaled appropriately.

In the following, we will specify the statistical properties of the model parameters θ_j and γ_j .

2.2. Hierarchical priors. The WVO-AR models yield prior densities depending on parameter vectors θ and γ . In order to convey the idea that the smoothness of f may vary over its domain in an a priori unknown manner, we model the coefficient vector $\theta = [\theta_j]$ and the variance vector of the innovation term, $\gamma = [\gamma_j]$ as multivariate random variables, Θ and Γ . Observe that formula (9) defines a dependency between θ and γ , if $\gamma^{(1)}$ and $\gamma^{(2)}$ are given. However, since these parameter vectors, too, are unknowns, we simply postulate that Θ and Γ are mutually independent, the prior model for the triple (X, Θ, Γ) is given by the density

$$\pi_{\text{prior}}(x, \theta, \gamma) = \pi_{\text{prior}}(x \mid \theta, \gamma) \pi_{\text{hyper},1}(\theta) \pi_{\text{hyper},2}(\gamma),$$

where the densities $\pi_{\text{hyper},1}(\theta)$ and $\pi_{\text{hyper},2}(\gamma)$ constitute the hyperprior models.

The selection of the hyperpriors depends on the particular application. We consider two different choices. In the first one, we express our belief that over most of the domain the function f will have order p smoothness (that is, $\theta_j = 0$ is expected a priori) by writing a hyperprior for the vector Θ of the form

$$(12) \quad \pi_{\text{hyper},1}(\theta) \propto \exp\left(-\frac{1}{2\eta}\|\theta\|^2\right) \chi_{[0,1]^{n+1}}(\theta),$$

which assumes, in particular, that the θ_j are independent identically distributed. Here, $\chi_{[0,1]^{n+1}}$ is the characteristic function of the hypercube in \mathbb{R}^{n+1} . This hyperprior gives preference to solutions with small Euclidean norm.

In another application, we assume that the values θ_j are varying slowly, and we use model

$$(13) \quad \pi_{\text{hyper},1}(\theta) \propto \exp\left(-\frac{1}{2\eta}\|\mathbb{L}_1\theta\|^2\right) \chi_{[0,1]^{n+1}}(\theta),$$

that is, Θ itself satisfies a first order Markov model with non-negativity constraints. This model prefers slowly varying solutions with small first finite differences.

When it comes to choosing a hyperprior for the γ_j , we assume that in general the variable order model describe quite well the function f except at a few places where it may have jumps or corners, which would be difficult for the autoregressive model to explain. Moreover, in order to allow the function f to suddenly change its smoothness, we assume that the variance of the innovation term at each grid point is independent, and that most of the time the variance of the innovation will be close to zero, leaving the option of a few points where it takes on large values. To encode these beliefs, we use the family of *generalized gamma distributions*, $\Gamma_j \sim \text{GenGamma}(r, \beta, \bar{\gamma})$, defined as

$$(14) \quad \begin{aligned} \pi_{\text{hyper},2}(\gamma) &= \pi_{\text{hyper}}(\gamma; r, \beta, \bar{\gamma}) \propto \prod_{j=0}^n \gamma_j^{r\beta-1} \exp\left(-\left(\frac{\gamma_j}{\bar{\gamma}}\right)^r\right) \\ &= \exp\left(-\sum_{j=0}^n \left(\frac{\gamma_j}{\bar{\gamma}}\right)^r + (r\beta-1) \sum_{j=0}^n \log \gamma_j\right). \end{aligned}$$

In particular, we remark that by choosing $r = 1$, we have the *gamma distribution*, $\gamma_j \sim \text{Gamma}(\beta, \bar{\gamma}) = \text{GenGamma}(1, \beta, \bar{\gamma})$,

$$\pi_{\text{hyper},2}(\gamma) \propto \prod_{j=0}^n \gamma_j^{\beta-1} \exp\left(-\frac{\gamma_j}{\bar{\gamma}}\right) = \exp\left(-\sum_{j=0}^n \frac{\gamma_j}{\bar{\gamma}} + (\beta-1) \sum_{j=0}^n \log \gamma_j\right),$$

while with $r = -1$, we obtain $\gamma_j \sim \text{InvGamma}(\beta, \bar{\gamma}) = \text{GenGamma}(-1, \beta, \bar{\gamma})$ which is the *inverse gamma distribution*, and

$$\pi_{\text{hyper},2}(\gamma) \propto \prod_{j=0}^n \gamma_j^{-\beta-1} \exp\left(-\frac{\bar{\gamma}}{\gamma_j}\right) = \exp\left(-\sum_{j=0}^n \frac{\bar{\gamma}}{\gamma_j} - (\beta+1) \sum_{j=0}^n \log \gamma_j\right).$$

Above, we assume that the shape and scale parameters β and $\bar{\gamma}$ are given. The hyperprior family in the context of inverse problems has been discussed in the article [4].

2.3. Likelihood and posterior densities. The likelihood model is based on the additive noise model (1). We will assume that the additive noise vector $e = [e_j]$ is a realization of a Gaussian random variable E ,

$$E \sim \mathcal{N}(0, \Sigma),$$

and for notational convenience, we denote by S the Cholesky factor of the inverse of the covariance,

$$\Sigma^{-1} = S^T S,$$

yielding a likelihood density

$$\pi_{\text{lh}}(b | x) \propto \exp\left(-\frac{1}{2} \|S(b - G(x))\|^2\right).$$

By Bayes' formula, the posterior density is given by

$$\pi_{\text{post}}(x, \theta, \gamma | b) \propto \pi_{\text{lh}}(b | x) \pi_{\text{prior}}(x | \theta, \gamma) \pi_{\text{hyper},1}(\theta) \pi_{\text{hyper},2}(\gamma).$$

2.4. Approximating the MAP estimate. We outline a simple cyclic updating algorithm for finding an approximate MAP estimate based on the posterior density given in the previous section. The details of each step, and the discussion of the convergence criterion are discussed separately.

Approximate MAP estimator

- (1) Initialize: $x = x^0$, $\theta = \theta^0$, $\gamma = \gamma^0$. Set $k = 0$.
 - (2) Update $(x^k, \theta^k, \gamma^k) \rightarrow (x^{k+1}, \theta^{k+1}, \gamma^{k+1})$,
 - (a) $x^{k+1} = \text{argmax}_x \{\pi(x, \theta^k, \gamma^k | b)\}$,
 - (b) $\theta^{k+1} = \text{argmax}_\theta \{\pi(x^{k+1}, \theta, \gamma^k | b)\}$,
 - (c) $\gamma^{k+1} = \text{argmax}_\gamma \{\pi(x^{k+1}, \theta^{k+1}, \gamma | b)\}$.
 - (3) If convergence criterion is met, stop, else, $k \leftarrow k + 1$ and continue from 2.
-

The maximization at each step in the above algorithm is done by minimizing the negative logarithm of the posterior density, and the particular structure of each problem can be utilized to make the iteration fast.

2.4.1. *Updating x .* To maximize the posterior density with respect to x , it is sufficient to consider the x -dependent factors of the posterior density. We define

$$\begin{aligned} F_x(x, \theta^k, \gamma^k) &= -\log(\pi_{\text{lik}}(b \mid x)\pi_{\text{prior}}(x \mid \theta^k, \gamma^k)) \\ &= \underbrace{\frac{1}{2}\|S(b - G(x))\|^2 + \frac{1}{2}\|D_{\gamma^k}^{-1/2}L_{\theta^k}x\|^2}_{\stackrel{\text{def}}{=} T_k(x)} + \frac{1}{2}\sum_{j=0}^n \log \gamma_j^k. \end{aligned}$$

The minimizer of this expression is the least squares solution that minimizes the functional $T_k(x)$. The problem can be viewed as a Tikhonov regularized solution of the inverse problem. In particular, if G is linear, that is, $G(x) = Gx$ for some matrix G , the update x^k is the least squares solution of the linear system

$$(15) \quad \begin{bmatrix} SG \\ D_{\gamma^k}^{-1/2}L_{\theta^k} \end{bmatrix} x = \begin{bmatrix} Sb \\ 0 \end{bmatrix},$$

and the solution can be approximated, e.g., by using iterative solvers.

2.4.2. *Updating θ .* The updating of θ is similar to the previous step. In this case, the θ -dependent part is

$$\begin{aligned} F_{\theta}(x^{k+1}, \theta, \gamma^k) &= -\log(\pi_{\text{prior}}(x^{k+1} \mid \theta, \gamma^k)\pi_{\text{hyper},1}(\theta)) \\ &= \underbrace{\frac{1}{2}\|D_{\gamma^k}^{-1/2}L_{\theta}x^{k+1}\|^2 + \frac{1}{2\eta}\|C\theta\|^2}_{\stackrel{\text{def}}{=} H_k(\theta)} + \frac{1}{2}\sum_{j=0}^n \log \gamma_j^k, \end{aligned}$$

defined over the hypercube $Q = [0, 1]^{n+1}$, where the matrix C is either the identity or L_1 , corresponding to the hyperpriors (12) or (13), respectively. To minimize H_k , we rewrite the θ -dependency out explicitly. We have

$$\begin{aligned} L_{\theta}x^{k+1} &= (I - D_{\theta})L_p x^{k+1} + D_{\theta}L_q x^{k+1} \\ &= D_{\theta}(L_q - L_p)x^{k+1} + L_p x^{k+1} = Q^{k+1}\theta + v^{k+1}, \end{aligned}$$

where

$$Q^{k+1} = \text{diag}[(L_q - L_p)x^{k+1}], \quad v^{k+1} = L_p x^{k+1}.$$

With these notations, the problem is reduced to the constrained optimization problem,

$$(16) \quad \text{minimize} \quad \left\| \begin{bmatrix} D_{\gamma^k}^{-1/2}Q^{k+1} \\ \eta^{-1/2}C \end{bmatrix} \theta + \begin{bmatrix} D_{\gamma^k}^{-1/2}v^{k+1} \\ 0 \end{bmatrix} \right\|^2 \quad \text{subject to } 0 \leq \theta \leq 1,$$

the inequalities understood componentwise. The minimization, due to the fact that the objective function is quadratic, can be carried out using linear least squares methods with projections to enforce the bound constraints. The minimization problem with bound constraints is computed with the algorithm described in [5].

2.4.3. *Updating γ .* Finally, we define

$$\begin{aligned} F_\gamma(x^{k+1}, \theta^{k+1}, \gamma) &= -\log(\pi_{\text{prior}}(x^{k+1} | \theta^{k+1}, \gamma)\pi_{\text{hyper},2}(\gamma)) \\ &= \frac{1}{2}\|D_\gamma^{-1/2}L_{\theta^{k+1}}x^{k+1}\|^2 - \left(r\beta - \frac{3}{2}\right)\sum_{j=0}^n \log \gamma_j + \sum_{j=0}^n \left(\frac{\gamma_j}{\bar{\gamma}}\right)^r. \end{aligned}$$

We observe that

$$\|D_\gamma^{-1/2}L_{\theta^{k+1}}x^{k+1}\|^2 = \sum_{j=0}^n \frac{(u_j^{k+1})^2}{\gamma_j},$$

where

$$u^{k+1} = L_{\theta^{k+1}}x^{k+1},$$

which makes the minimization of F_k particularly simple for certain values of r . In particular, by differentiating F_γ with respect to γ_j and setting the derivative equal to zero, we have the equation

$$-\frac{1}{2}\frac{(u_j^{k+1})^2}{\gamma_j^2} - \left(r\beta - \frac{3}{2}\right)\frac{1}{\gamma_j} + r\frac{\gamma_j^{r-1}}{\bar{\gamma}^r} = 0, \quad \gamma_j > 0,$$

which in particular cases has an explicit solution: For $r = 1$, corresponding to a gamma hyperprior for Γ , we have

$$(17) \quad \gamma_j = \frac{\bar{\gamma}}{2} \left(\kappa + \sqrt{\kappa^2 + \frac{2(u_j^{k+1})^2}{\bar{\gamma}}} \right), \quad \kappa = \beta - 3/2,$$

while for $r = -1$, which amounts to the inverse gamma hyperprior, it follows that

$$(18) \quad \gamma_j = \frac{(u_j^{k+1})^2 + 2\bar{\gamma}}{2\beta + 3},$$

$0 \leq j \leq n$. For general r , a numerical solution is required.

2.4.4. *Stopping criterion.* The stopping criterion for the iterations is based on the relative change of the norm of the unknowns. We define

$$\Delta_x^k = \frac{\|x^k - x^{k-1}\|^2}{\|x^k\|^2}, \quad \Delta_\theta^k = \frac{\|\theta^k - \theta^{k-1}\|^2}{\|\theta^k\|^2}, \quad \Delta_\gamma^k = \frac{\|\gamma^k - \gamma^{k-1}\|^2}{\|\gamma^k\|^2},$$

and set the number of iterations to

$$(19) \quad k_{\max} = \min_k \{ \sqrt{\Delta_x^k + \Delta_\theta^k + \Delta_\gamma^k} < \delta \},$$

where $\delta > 0$ is a given tolerance.

3. BOUNDARY CONDITIONS

The prior models discussed so far are based on the assumption (3) inherited from the time series formulation of the prior, while the values for $j > n$ are not of interest. When applied to inverse problems, however, it is often natural to impose boundary conditions both at $t = 0$ and $t = 1$. For simplicity, we restrict ourselves to homogenous boundary conditions here. Also, for the sake of clarity, consider the case $p = 2$.

Consider the prior model $\pi_{\text{prior}}(x)$, $x = (x_0, x_1, \dots, x_n)$. The model was derived by assuming the condition (3), that is, for $p = 2$,

$$\pi_{\text{prior}}(x) = \pi_{\text{prior}}(x \mid x_{-1} = x_{-2} = 0).$$

We want to modify the density to correspond a symmetric boundary condition,

$$x_{-1} = x_{n+1} = 0,$$

while x_{-2} is arbitrary. To do so, we write the autoregressive models

$$X_j - 2X_{j-1} + X_{j-2} = \sqrt{\gamma_j} W_j = 0, \quad j = 0, 1, \dots, n + 1$$

where, for notational simplicity, we write $\gamma = \gamma^{(2)}$. Assuming that $X_{-1} = X_{n+1} = 0$, while $X_{-2} = a$, where a for the time being is arbitrary, the matrix form of the stochastic model with the $n + 2$ equations above becomes

$$(20) \quad \begin{bmatrix} 1 & & & & & & \\ -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_n \end{bmatrix} + \begin{bmatrix} a \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{\gamma_0} W_0 \\ \vdots \\ \sqrt{\gamma_{n+1}} W_{n+1} \end{bmatrix}.$$

The prior model for the vector $X = (X_0, X_1, \dots, X_n)$ conditioned on zero boundary conditions $X_{-1} = X_{n+1} = 0$, is then obtained by marginalizing the conditional prior with respect to the history $X_{-2} = a$, leading to the probability density

$$\overset{\circ}{\pi}_{\text{prior}}(x) = \int \pi_{\text{prior}}(x \mid x_{-1} = 0, x_{n+1} = 0, x_{-2} = a) da.$$

To find the explicit form, we define the $(n + 1) \times (n + 1)$ matrices

$$\overset{\circ}{L}_2 = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -2 \end{bmatrix}, \quad \overset{\circ}{D}_\gamma = \begin{bmatrix} \gamma_1 & & & \\ & \ddots & & \\ & & & \gamma_{n+1} \end{bmatrix},$$

and write the probability density corresponding to the model (20) as

$$\pi_{\text{prior}}(x \mid x_{-1} = 0, x_{n+1} = 0, x_{-2} = a) \propto \exp \left(-\frac{1}{2\gamma_0} (x_0 + a)^2 - \frac{1}{2} \|\overset{\circ}{D}_\gamma^{-1/2} \overset{\circ}{L}_2 x\|^2 \right),$$

Finally, by integrating with respect to a , we therefore find that

$$\overset{\circ}{\pi}_{\text{prior}}(x) \propto \exp \left(-\frac{1}{2} \|\overset{\circ}{D}_\gamma^{-1/2} \overset{\circ}{L}_2 x\|^2 \right),$$

a formula that is often found in literature as the second order smoothness prior. In a similar fashion, we may derive the smoothness priors with desired left or right boundary conditions for Markov models of any order.

In this work, we are interested in the models of order 0, 1, and 2, and their linear combinations. We will work out in detail the interpolated model between the

second and the first order. The first order matrix we work with is

$$\overset{\circ}{\mathbf{L}}_1 = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & & 1 \\ & & & & -1 \end{bmatrix},$$

and the corresponding (2, 1)-model is

$$(21) \quad \overset{\circ}{\mathbf{L}}_\theta = (1 - \mathbf{D}_\theta) \overset{\circ}{\mathbf{L}}_2 + \mathbf{D}_\theta \overset{\circ}{\mathbf{L}}_1 = \begin{bmatrix} \theta_1 - 2 & 1 & & & \\ 1 - \theta_2 & \theta_2 - 2 & & & 1 \\ & & \ddots & & \\ & & & & \ddots \\ & & & 1 - \theta_{n+1} & \theta_{n+1} - 2 \end{bmatrix}.$$

While the posterior model with this matrix is obtained in a similar fashion as with the non-symmetric model, one significant difference emerges: The matrix is no longer a lower triangular matrix, and consequently, the computation of the determinant is not as straightforward. Moreover, it has to be taken into consideration in the MAP estimation problem. In particular, the objective function in the optimization problem (16) becomes

$$H_k(\theta) = \left\| \begin{bmatrix} \mathbf{D}_{\gamma^k}^{-1/2} \mathbf{Q}^{k+1} \\ \eta^{-1} \mathbf{C} \end{bmatrix} \theta + \begin{bmatrix} \mathbf{D}_{\gamma^k}^{-1/2} v^{k+1} \\ 0 \end{bmatrix} \right\|^2 - 2 \log[\det(\overset{\circ}{\mathbf{L}}_\theta)],$$

with the obvious modifications in the definitions of \mathbf{Q}^{k+1} and v^{k+1} .

3.1. The determinant. In this section, we derive the formulas needed for computing effectively the determinants and their gradients with respect to the model parameter θ for optimization methods. Consider the interpolated matrix $\overset{\circ}{\mathbf{L}}_\theta$ in (21), which has tridiagonal structure,

$$\overset{\circ}{\mathbf{L}}_\theta = \begin{bmatrix} \alpha_1 & 1 & & & \\ \beta_2 & \alpha_2 & 1 & & \\ & \ddots & \ddots & & \\ & & & \beta_{n+1} & \alpha_{n+1} \end{bmatrix},$$

where $\alpha_j = \theta_j - 2$, $j = 1, 2, \dots, n+1$, and $\beta_j = 1 - \theta_j$, $j = 2, 3, \dots, n+1$.

To evaluate the determinant, we use the Gaussian elimination to reduce the matrix to an upper triangular matrix, repeating the elimination step

$$\begin{bmatrix} \alpha_1 & 1 & & & \\ \beta_2 & \alpha_2 & 1 & & \\ & \ddots & \ddots & & \\ & & & \beta_{n+1} & \alpha_{n+1} \end{bmatrix} \rightarrow \begin{bmatrix} \alpha_1 & 1 & & & \\ & \alpha_2 - \beta_2/\alpha_1 & 1 & & \\ & & \ddots & & \\ & & & \beta_{n+1} & \alpha_{n+1} \end{bmatrix},$$

yielding an upper triangular matrix with the diagonal entries d_j , satisfying the recurrence relation

$$(22) \quad d_1 = \alpha_1, \quad d_{j+1} = \alpha_{j+1} - \frac{\beta_{j+1}}{d_j}, \quad 1 \leq j \leq n.$$

Hence, we have

$$(23) \quad \det(\mathring{L}_\theta) = \prod_{j=1}^{n+1} d_j.$$

We have the following result.

Lemma 3.1. *The d_j given by (22) form an increasing sequence bounded below by -2 and above by -1 ,*

$$-2 \leq d_j \leq d_{j+1} \leq -1, \quad j = 1, 2, \dots, n.$$

Proof. First notice that

$$(24) \quad \begin{aligned} d_{j+1} &= \alpha_{j+1} - \frac{\beta_{j+1}}{d_j} = d_j + \left(\alpha_{j+1} - \frac{\beta_{j+1}}{d_j} - d_j \right) \\ &= d_j + \frac{\alpha_{j+1}d_j - \beta_{j+1} - d_j^2}{d_j} \\ &= d_j + \frac{(\theta_{j+1} - 2)d_j - (1 - \theta_{j+1}) - d_j^2}{d_j} \\ &= d_j + \frac{\theta_{j+1}(d_j + 1) - (d_j + 1)^2}{d_j}, \end{aligned}$$

and proceed by induction.

For $j = 1$ we have

$$d_1 = \alpha_1 = \theta_1 - 2,$$

therefore $-2 \leq d_1 \leq -1$. Also, from (24) it follows that

$$d_2 = d_1 + \frac{\theta_2(d_1 + 1) - (d_1 + 1)^2}{d_1}.$$

Since $\theta_2 \geq 0$ and $d_1 + 1 \leq 0$, it follows that

$$\frac{\theta_2(d_1 + 1) - (d_1 + 1)^2}{d_1} \geq 0,$$

and therefore $d_2 \geq d_1$, the equality holding only if $d_1 = -1$. On the other hand,

$$1 + d_2 = 1 + \alpha_2 - \frac{\beta_2}{d_1} = 1 + \theta_2 - 2 - \frac{(1 - \theta_2)}{d_1} = (\theta_2 - 1) \left(1 + \frac{1}{d_1} \right) \leq 0,$$

where the last inequality follows from the fact that $\theta_2 \leq 1$ and $d_1 \leq -1$. Summarizing, we have proved that

$$-2 \leq d_1 \leq d_2 \leq -1.$$

Suppose now that our claim holds for $j - 1$, that is,

$$(25) \quad -2 \leq d_{j-1} \leq d_j \leq -1,$$

and we want to prove that it holds for j . Since (25) implies that

$$\frac{\theta_{j+1}(d_j + 1) - (d_j + 1)^2}{d_j} \geq 0,$$

it therefore follows from (24) that $d_{j+1} \geq d_j$, the equality holding only if $d_j = -1$. Moreover,

$$\begin{aligned} 1 + d_{j+1} &= 1 + \alpha_{j+1} - \frac{\beta_{j+1}}{d_j} = 1 + \theta_{j+1} - 2 - \frac{1 - \theta_{j+1}}{d_j} \\ &= (\theta_{j+1} - 1) \left(1 + \frac{1}{d_j} \right) \leq 0, \end{aligned}$$

where the last inequality follows from the fact that $\theta_{j+1} \leq 1$ and, from (25), $d_j \leq -1$. Thus $d_{j+1} \leq -1$, hence

$$-2 \leq d_j \leq d_{j+1} \leq -1,$$

which completes the proof. \square

It should be pointed out that if $n \rightarrow \infty$, it follows immediately from the last theorem and from the recursive formula (24) that $\lim_{j \rightarrow \infty} d_j = -1$. This is an important result because, since from (23)

$$(26) \quad \log |\det(\mathring{L}_\theta)| = \log \left| \prod_{j=1}^{n+1} d_j \right| = \sum_{j=1}^{n+1} \log |d_j|,$$

then it is possible that if the convergence is fast enough, just a few terms of the sum in (26) will suffice to give a good approximation of the determinant. In this regard, the following result is also important.

Lemma 3.2. *For every $k \in \mathbb{N}$,*

$$|d_k - (-1)| \leq \frac{\prod_{j=1}^k (1 - \theta_j)}{\prod_{j=1}^{k-1} (-\theta_j)} \leq \prod_{j=1}^k (1 - \theta_j).$$

Proof. We write

$$\begin{aligned} 1 + d_{j+1} &= 1 + \alpha_{j+1} - \frac{\beta_{j+1}}{d_j} = 1 + (\theta_{j+1} - 2) - \frac{1 - \theta_{j+1}}{d_j} \\ &= \theta_{j+1} - 1 + \frac{\theta_{j+1} - 1}{d_j} = \frac{(\theta_{j+1} - 1)(d_j + 1)}{d_j}. \end{aligned}$$

Then

$$\begin{aligned} 0 &\leq -(1 + d_{j+1}) = (1 - \theta_{j+1}) \frac{d_j + 1}{d_j} \\ &= \frac{(1 - \theta_{j+1})[-(1 + d_j)]}{-d_j}. \end{aligned}$$

Since $-(1 + d_j) = |d_j - (-1)|$, the above inequality can be written in the form

$$(27) \quad |d_{j+1} - (-1)| \leq (1 - \theta_{j+1}) \frac{|d_j - (-1)|}{-d_j}, \quad \forall j \in \mathbb{N}.$$

It follows from (27) that for all $k \in \mathbb{N}$,

$$|d_k - (-1)| \leq \frac{\prod_{j=1}^k (1 - \theta_j)}{\prod_{j=1}^{k-1} (-d_j)} \leq \prod_{j=1}^k (1 - \theta_j),$$

where the last inequality follows from the fact that $d_j \leq -1$ for all j . \square

It follows from the previous lemma that if $d_k = -1$, or $\theta_k = 1$, then $d_j = -1$ for all $j \geq k$, in which case it follows that

$$\log |\det(\mathring{\mathbf{L}}_\theta)| = \sum_{j=1}^{k-1} \log |d_j|.$$

The determinant depends on the parameter θ , and in order to maximize the posterior density, we need to compute the gradient of the determinant. The necessary formulas for this computation are given in the following lemma.

Lemma 3.3. *The gradient vector $\nabla_\theta \det(\mathring{\mathbf{L}}_\theta) \in \mathbb{R}^{n+1}$ is given by the formula*

$$(28) \quad \nabla_\theta \det(\mathring{\mathbf{L}}_\theta) = \det(\mathring{\mathbf{L}}_\theta) [1/d_1 \quad 1/d_2 \quad \cdots \quad 1/d_{n+1}] \mathbf{M},$$

where the matrix $\mathbf{M} \in \mathbb{R}^{(n+1) \times (n+1)}$ is a lower triangular matrix, and its entries m_{jk} are obtained columnwise by the formulas

$$m_{kk} = 1 + \frac{1}{d_{k-1}} \quad (m_{11} = 1),$$

$$m_{jk} = \frac{\beta_j}{d_{j-1}^2} m_{j-1,k}, \quad j > k.$$

Proof. We define the matrix \mathbf{M} above through

$$m_{jk} = \frac{\partial d_j}{\partial \theta_k}.$$

The matrix \mathbf{M} is lower triangular, because, by the recurrence relation (22), d_k is independent of θ_j for $j > k$.

To prove (28), we observe that

$$\begin{aligned} \frac{\partial}{\partial \theta_k} (d_1 \cdots d_{n+1}) &= \sum_{\ell=k}^{n+1} \frac{\partial d_\ell}{\partial \theta_k} \prod_{j=1, j \neq \ell}^{n+1} d_j = \left(\prod_{j=1}^{n+1} d_j \right) \sum_{\ell=k}^{n+1} \frac{1}{d_\ell} \frac{\partial d_\ell}{\partial \theta_k} \\ &= \left(\prod_{j=1}^{n+1} d_j \right) \sum_{\ell=k}^{n+1} \frac{1}{d_\ell} m_{\ell k}, \end{aligned}$$

which is the claimed identity (28) in component form.

To prove the formulas for the components of the matrix, we use the recurrence relation (22), implying that

$$m_{kk} = \frac{\partial}{\partial \theta_k} \left(\alpha_k - \frac{\beta_k}{d_{k-1}} \right) = 1 + \frac{1}{d_{k-1}},$$

and for $j > k$,

$$m_{jk} = \frac{\partial}{\partial \theta_k} \left(\alpha_j - \frac{\beta_j}{d_{j-1}} \right) = \frac{\beta_j}{d_{j-1}^2} \frac{\partial d_{j-1}}{\partial \theta_k} = \frac{\beta_j}{d_{j-1}^2} m_{j-1,k},$$

by the definition of the matrix \mathbf{M} . □

We point out that the gradient of the logarithm of the determinant is found by the formula

$$\nabla_{\theta} \log \det(\overset{\circ}{\mathbf{L}}_{\theta}) = \frac{\nabla_{\theta} \det(\overset{\circ}{\mathbf{L}}_{\theta})}{\det(\overset{\circ}{\mathbf{L}}_{\theta})} = [1/d_1 \quad 1/d_2 \quad \cdots \quad 1/d_{n+1}] \mathbf{M}.$$

In the following section, further details of the algorithm are given in connection with a computed example.

4. COMPUTED EXAMPLE

The computed test involves a signal that changes its smoothness characteristics. Consider a deconvolution problem with noisy data,

$$b_j = \int_0^1 K(s_j - t) f(t) dt + e_j,$$

where the convolution kernel is the Airy function appearing in optical applications,

$$K(t) = \left(\frac{J_1(\kappa t)}{\kappa t} \right)^2,$$

where J_1 is the Bessel function of first kind of order 1, and k is a parameter controlling the width of the function. Here we use $\kappa = 50$. We discretize the convolution model to obtain a Toeplitz matrix $\mathbf{G} \in \mathbb{R}^{(n+1) \times (n+1)}$,

$$\mathbf{G}_{ij} = K(s_i - t_j) = K(h(i - j)), \quad h = 1/n, \quad 0 \leq i, j \leq n,$$

where $n = 200$. The input signal f generated as

$$f(t) = \frac{1}{1 + e^{(t-t_0)/\tau}}, \quad 0 \leq t \leq 1/2,$$

where $t_0 = 1/2$ and $\tau = 0.08$, and

$$f(t) = A \sum_{j=1}^{20} \sin(2\pi\nu_j(t - \eta_j)), \quad 1/2 < t \leq 1,$$

where $A = 0.8$, and the frequencies ν_j and phase shifts η_j are randomly generated,

$$\nu_j = 20(1 + \xi_j), \quad \xi_j, \eta_j \sim \text{Uniform}([0, 1]).$$

The additive noise components e_j are drawn independently from zero mean normal distribution with standard deviation $\sigma = 0.02$. Figure 1 shows one realization of the input signal and the computed noisy data.

Observe that because of the narrowness of the convolution kernel, this problem is close to a denoising problem; however, the noise level in this example is quite significant.

For this signal, one could expect that a smoothness prior gives optimal results over the interval $[0, 1/2]$ while over $[1/2, 1]$, the rough nature of the signal could be better captured by the L_0 -prior. Therefore, we consider here the $(2, 0)$ -WVO-AR model.

As a choice of the hyperpriors, we use the Gamma hyperprior for γ , that is, $r = 1$ in (14), while for θ , we use the L_1 -hyperprior (13).

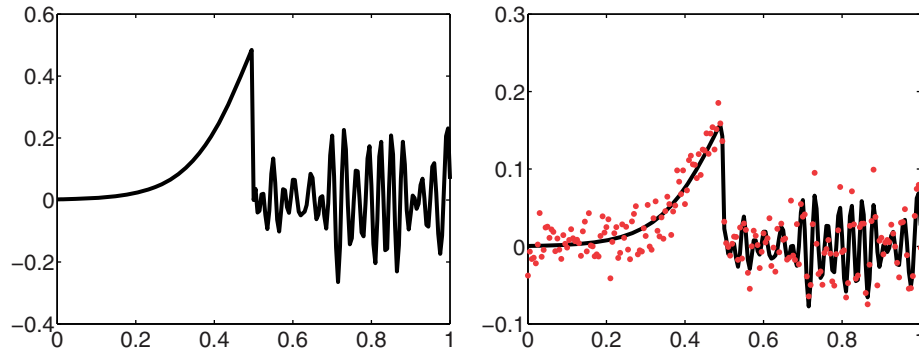


FIGURE 1. The input signal of the computed example (left panel) and the corresponding convolved noisy data (right). (*Color available online.*)

The algorithm as described in the previous section, runs into a scaling problem due to the difficulty in comparing the variations in the signal measured by the L_0 -prior, and the variations in its curvature measured by the L_2 -prior. As the hyperparameter γ controls the trade-off between the prior and the likelihood, an imbalance ensues: To avoid resorting to large values of γ with low hyperprior probability, the algorithm chooses values of θ which keep the prior probability of the unknown relatively low without increasing γ , even at the cost of fidelity. Thus, effectively, the smoothness prior is strongly favored, even when the signal is non-smooth.

As a remedy, we suggest a scaling of the matrices L_j using the following strategy. Assume first that the underlying signal in the deconvolution problem is a smooth, slowly varying signal, and an estimate of the noise level is available. The standard non-statistical method for solving the inverse problem is to use the Tikhonov regularization, defining the solution as

$$x_\mu = \operatorname{argmin}\{\|Gx - b\|^2 + \mu^2 \|L_2x\|^2\},$$

and to seek the regularization parameter μ from Morozov's discrepancy principle,

$$\|Gx_\mu - b\| = \tau \times \{\text{estimated norm of the noise vector}\},$$

where $\tau \geq 1$ is a safeguard factor. If the noise vector is zero mean normally distributed with variance σ^2 , a rather standard estimate for the noise vector norm would be $\sqrt{n+1}\sigma$, square root of its expected variance. We denote the value for μ thus defined by μ_* . On the other hand, the Tikhonov regularized solution corresponds to the MAP estimate for x with the L_2 -smoothness prior with constant prior variance. Indeed, considering the least squares problem (15) with $\gamma_j = \bar{\gamma} = \text{constant}$ and with the second order smoothness matrix, we find that the regularization parameter, the noise variance and the prior variance are related by the formula

$$\frac{\bar{\gamma}}{\sigma^2} = \mu^2.$$

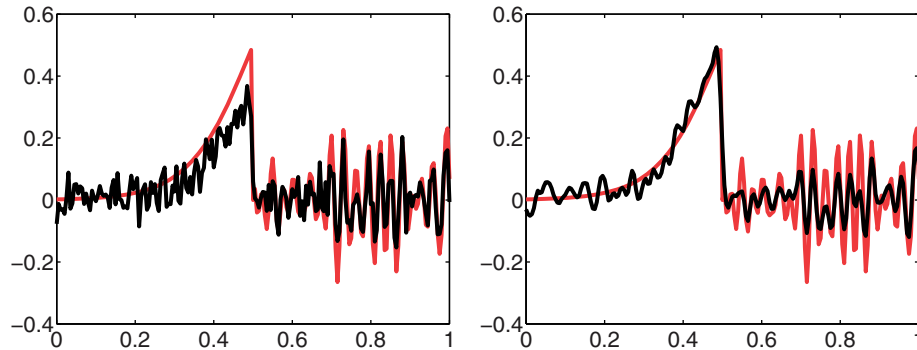


FIGURE 2. The reconstruction using either the L_0 - (left) or L_2 -penalty (right) in the Tikhonov regularization and Morozov's discrepancy principle for selecting the regularization level. The red curve is the true signal used for data generation, the black one is estimated from the data. It is evident that both methods perform poorly, since a single regularization parameter has to compromise between a small value at the rough end and a larger value at the smooth end of the signal. (*Color available online.*)

In order to scale the prior model (7) in such a way that the regularization parameter value found by the Morozov discrepancy principle corresponds to an innovation process with unit variance, we use the scaling

$$L_2 \rightarrow \frac{1}{\sqrt{\bar{\gamma}}} L_2, \quad \bar{\gamma} = \sigma^2 \mu_*^2.$$

The value μ_* is different from problem to problem, and even from one realization of the noise to another. We therefore run the Tikhonov regularization scheme with different smooth inputs and several realizations to find an order of magnitude of $\bar{\gamma}$. We then repeat the scaling with L_0 , using different smooth and non-smooth input functions.

To have a reference point, we first solve the inverse problem using the Tikhonov regularized solution with Morozov's discrepancy principle with either the L_0 - or the L_2 -penalty alone. The results are shown in Figure 2. It is obvious from these reconstructions that since the underlying signal does not quite correspond to either one of the assumptions underlying these penalties, the reconstructions suffer from the compromise between the smooth and rough part of the signal.

We then run the alternating MAP approximation. The updating of x requires a least squares solution in each step, while the updating of γ is done using formula (17). The updating of θ is performed by a simple steepest descent algorithm with projected gradient in the orthogonal complement of the active set. In Figures 3–5, we show four snapshots of the progression of the algorithm, following the evolution of the estimates of the three variables. The initial value for θ is $\theta_j = 1/2$, $j = 1, \dots, n+1$. Due to the smoothness prior with assumed boundary value at the left ($\theta_{-1} = 0$), the value at the left end settles quickly to small values, while the right end, due to the roughness of the signal, starts to increase towards the maximum value $\theta_j = 1$. As soon as this value is reached, the corresponding pixel x_j becomes independent of its neighbors, and the pointwise estimate of the signal

becomes significantly good (Figure 3). At the points where $\theta_j < 1$, the prior correlation with the neighboring values keep the estimated function rather rigid and thus unable to estimate the fast oscillations well. Figure 6 plots the relative change $r_k = \sqrt{\Delta_x^k + \Delta_\theta^k + \Delta_\gamma^k}$ appearing in the stopping criterion (19). The algorithm requires a relatively high number of iterations (597 for this example) to converge, but seems to work very consistently, and the results over numerous restartings are stable.

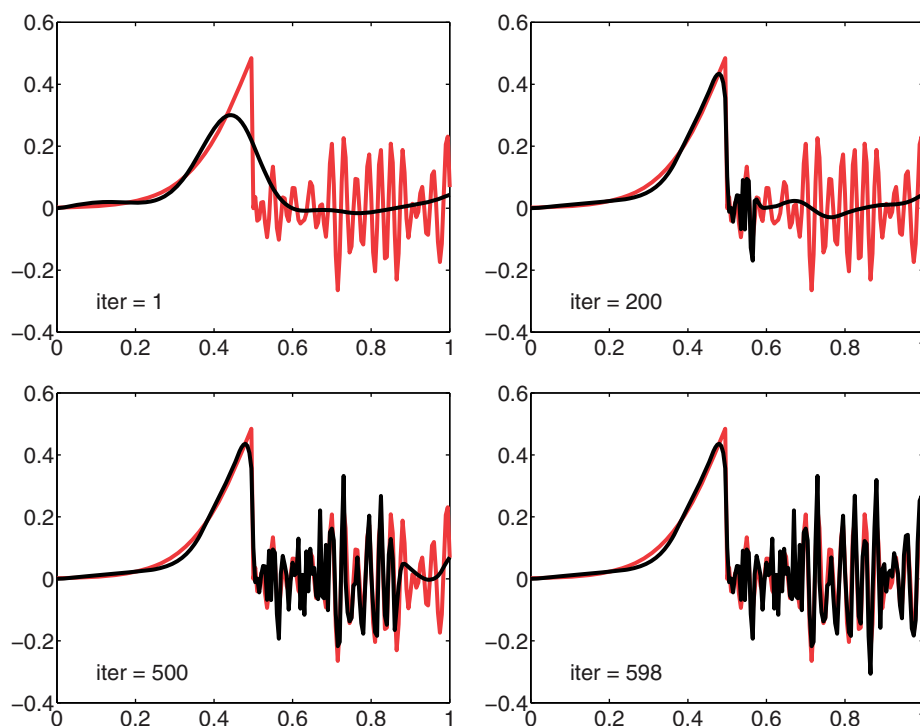


FIGURE 3. Four snapshots of the progression of the updating of the estimate of the signal. The true signal is indicated by the red curve, the current estimate is plotted in black. The number of iterations in indicated in the figure. (*Color available online.*)

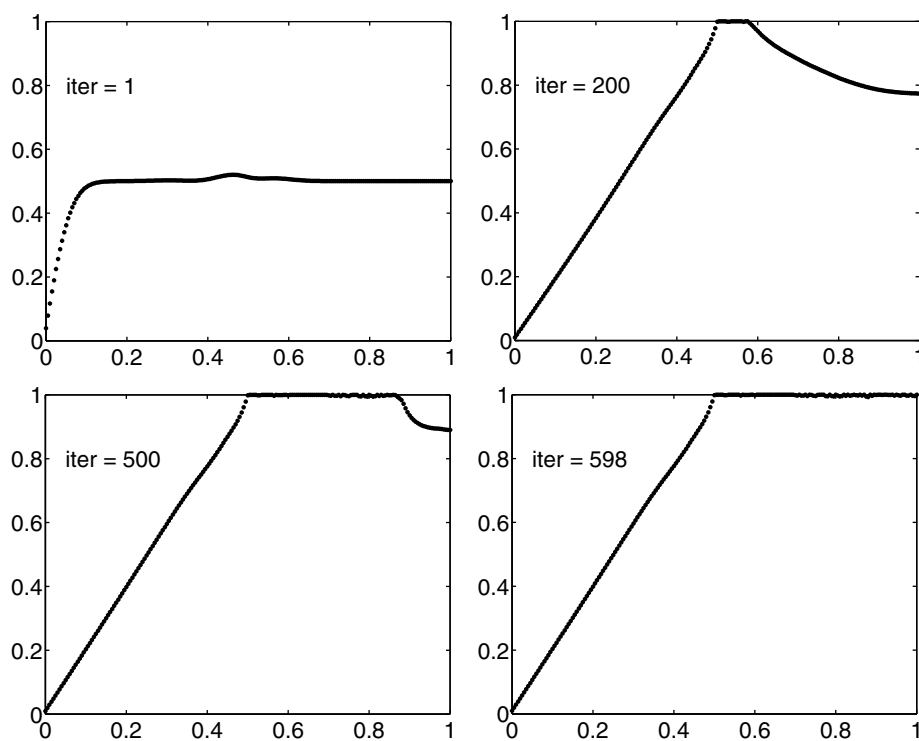


FIGURE 4. Snapshots of the estimate of the θ vector. Observe that the boundary condition implied by the Markov model keeps the value θ_0 close to zero. This must be thought of as part of the prior information: If no reason to assume that the signal starts as a smooth signal exists, the boundary condition should be removed.

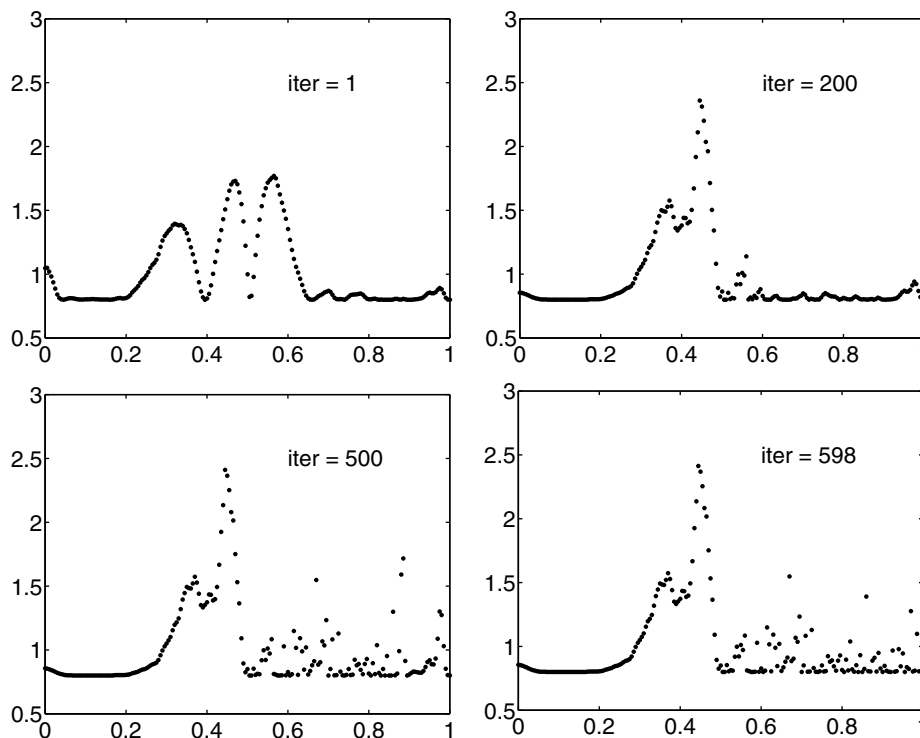


FIGURE 5. Four snapshots of the update of the innovation variance γ . Observe a maximum close to where the signal switches from smooth to rough at $t = 1/2$, demonstrating a significant jump at that point.

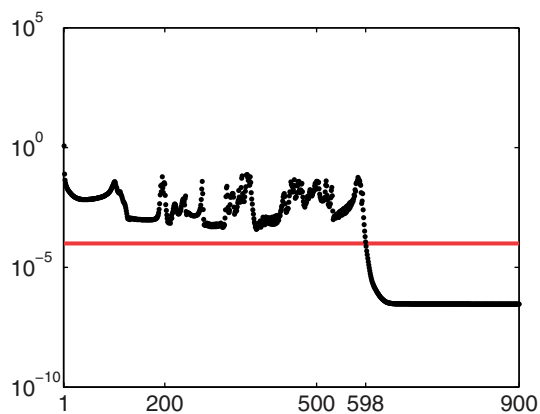


FIGURE 6. The relative change of the update as the iteration proceeds. The horizontal line indicates the stopping tolerance $\delta = 1 \times 10^{-4}$. At $k_{\max} = 598$, the stopping criterion is satisfied, but for demonstrating the stability, the iteration is run until $k = 900$. (Color available online.)

5. DISCUSSION

The article discusses the regularization of ill-posed inverse problems approaching the problem from the Bayesian statistical viewpoint, in which the regularization corresponds to prior information about the unknown. In particular, smoothness penalties correspond to autoregressive prior models. In this context, the degree of the AR model corresponds to the a priori smoothness of the unknown, which may itself be poorly known, but can be addressed by using hierarchical Bayesian models. The computed example demonstrates that if the forward model is not too strongly smoothing, the data may contain enough information about the smoothness properties of the underlying signal for judiciously adjusting the order of the AR prior. The example suggests that the method could be applicable for imaging problems (denoising, deblurring) when parts of the image are characterized by texture of different degree of smoothness. For wider convolution kernels, it is to be expected that resolving the degree of smoothness of the prior model based on data becomes increasingly difficult. Extensions from one dimension to two dimensions pose further challenges, such as the anisotropic nature of the underlying textures.

REFERENCES

- [1] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems **10** (1994), no. 6, 1217–1229. MR1306801 (95i:65092)
- [2] Hirotugu Akaike, *Fitting autoregressive models for prediction*, Ann. Inst. Statist. Math. **21** (1969), 243–247. MR0246476 (39 #7780)
- [3] Peter J. Brockwell and Richard A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics, Springer-Verlag, New York, 1987. MR868859 (88k:62001)
- [4] Daniela Calvetti, Harri Hakula, Sampsa Pursiainen, and Erkki Somersalo, *Conditionally Gaussian hypermodels for cerebral source localization*, SIAM J. Imaging Sci. **2** (2009), no. 3, 879–909, DOI 10.1137/080723995. MR2540174
- [5] D. Calvetti, G. Landi, L. Reichel, and F. Sgallari, *Non-negativity and iterative methods for ill-posed problems*, Inverse Problems **20** (2004), no. 6, 1747–1758, DOI 10.1088/0266-5611/20/6/003. MR2107234 (2005j:65044)
- [6] Daniela Calvetti and Erkki Somersalo, *Introduction to Bayesian Scientific Computing*, Surveys and Tutorials in the Applied Mathematical Sciences, vol. 2, Springer, New York, 2007. Ten lectures on subjective computing. MR2351679 (2008g:62002)
- [7] Daniela Calvetti and Erkki Somersalo, *A Gaussian hypermodel to recover blocky objects*, Inverse Problems **23** (2007), no. 2, 733–754, DOI 10.1088/0266-5611/23/2/016. MR2309673 (2008a:62007)
- [8] Daniela Calvetti and Erkki Somersalo, *Hypermodels in the Bayesian imaging framework*, Inverse Problems **24** (2008), no. 3, 034013, 20, DOI 10.1088/0266-5611/24/3/034013. MR2421950 (2009g:94016)
- [9] Heinz W. Engl, Martin Hanke, and Andreas Neubauer, *Regularization of Inverse Problems*, Mathematics and its Applications, vol. 375, Kluwer Academic Publishers Group, Dordrecht, 1996. MR1408680 (97k:65145)
- [10] Gene H. Golub, Michael Heath, and Grace Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics **21** (1979), no. 2, 215–223, DOI 10.2307/1268518. MR533250 (81e:62079)
- [11] Per Christian Hansen, *Rank-deficient and Discrete Ill-posed Problems*, SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Numerical aspects of linear inversion. MR1486577 (99a:65037)
- [12] Per Christian Hansen and Dianne Prost O’Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput. **14** (1993), no. 6, 1487–1503, DOI 10.1137/0914086. MR1241596 (94h:65045)
- [13] Jari Kaipio and Erkki Somersalo, *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences, vol. 160, Springer-Verlag, New York, 2005. MR2102218 (2005g:65001)

- [14] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl. **7** (1966), 414–417. MR0208819 (34 #8628)
- [15] Albert Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. MR2130010 (2007b:62011)
- [16] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput. **17** (1996), no. 1, 227–238, DOI 10.1137/0917016. Special issue on iterative methods in numerical linear algebra (Breckenridge, CO, 1994). MR1375276

CASE WESTERN RESERVE UNIVERSITY, DEPARTMENT OF MATHEMATICS, APPLIED MATHEMATICS AND STATISTICS, 10900 EUCLID AVENUE, CLEVELAND OHIO 44106

E-mail address: `daniela.calvetti@case.edu`

CASE WESTERN RESERVE UNIVERSITY, DEPARTMENT OF MATHEMATICS, APPLIED MATHEMATICS AND STATISTICS, 10900 EUCLID AVENUE, CLEVELAND OHIO 44106

E-mail address: `erkki.somersalo@case.edu`

INSTITUTO DE MATEMÁTICA APLICADA DEL LITORAL, IMAL, CONICET-UNL, GÜEMES 3450, S3000GLN, SANTA FE, ARGENTINA

E-mail address: `rspies@santafe-conicet.gov.ar`