International Journal of Mathematics
and Computer Science, 1(2006), 63-81

Aristotelian prior boundary conditions

Daniela Calvetti[1], **Jari P. Kaipio**[2], **Erkki Someralo**[3]

[1] Case Western Reserve University, Department of Mathematics
10900 Euclid Ave., Cleveland, OH 44106, USA.
e-mail: dxc57@case.edu

[2] University of Kuopio, Department of Applied Physics
PO Box 1627, FIN-70211 Kuopio, Finland.
e-mail: jari.kaipio@uku.fi

[3] Helsinki University of Technology, Institute of Mathematics
PO Box 1100, FIN-02015 HUT, Finland.
e-mail: erkki.somersalo@hut.fi

(Received June 23, 2005, Accepted October 21, 2005)

Abstract

The selection of boundary conditions when restoring a finite 1D or 2D signal has received a lot of interest in recent times. Dirichlet, periodic, reflective and antireflective boundary conditions may be appropriate selection for some problems, while wholly unsuitable for others. In this paper we show that, in an Aristotelian approach to knowledge, when it is not known a priori which boundary conditions should be chosen, by admitting our lack of information it is possible to let the data itself determine them. We consider 1D and 2D smoothness prior conditions. The application of this Aristotelian approach to boundary conditions to the solution of linear discrete ill-posed problems with Tikhonov regularization or truncated iterative methods is discussed. Computed examples of deblurring and limited angle tomography showing the effectiveness of Aristotelian boundary conditions are presented.

1 Introduction

The selection of suitable boundary conditions when restoring a finite signal from a distorted specimen may play an important role in the success of the outcome. The issue of which boundary conditions are most appropriate has been addressed repeatedly in the image deblurring literature, where Dirichlet, Neumann, peridic, reflective and antireflective boundary conditions have been proposed; see, e.g., [3, 6] and references therein. It is clear from the examples reported in the literature that it is always possible to find problems for which

Research of the first author was supported in part by grant NIH GM-66309-01.
ISSN 1814-0424

one type of boundary conditions work much better than the others. For example, if the signal that we want to recover does not vanish on the boundary, imposing Dirichlet boundary conditions may cause the computed solution to show oscillatory behavior near the boundary. Similar misbehavior of the computed solution can be observed with the other boundary conditions when used inappropriately.

In this paper we consider the case in which we have no definite *a priori* knowledge as to which type of boundary conditions might be most appropriate for our problem. Instead of guessing which boundary conditions might be most appropriate, or selecting the boundary conditions according to how much such selection will simplify the computations, we acknowledge our "tabula rasa" state and let our experience of the signal, i.e., the data itself, determine what is the most likely behavior of the solution at the boundary. We call the boundary conditions determined in this fashion *Aristotelian boundary conditions* because we, like the Greek philosopher, assume that knowledge is a process guided by experiences, built layer after layer starting from a clean slate. [1] In this paper we propose an approach to boundary condition selection that allows a great deal of flexibility in terms of expressing the various degrees of knowledge, or lack thereof, about the behavior of the solution at the boundary, and an algorithm which is computationally efficient also for problems of large dimensions.

We remark that while the basic ideas behind Aristotelian boundary conditions are essentially the same in one, two or three dimensional space, the technical details become more tedious in higher dimensions. Therefore we will first describe the algorithm in one dimension, then explain how it should be modified for application to higher dimensional problems. Furthermore, since a lot of the recent discussions on the choice of boundary conditions has focused on the problem of signal and image deblurring, we will also assume first that we want to recover a signal $\mathbf{x}$ from a blurred, and possibly noisy copy $\mathbf{b}$. After discretization we have the linear system

$$\mathbf{b} = A\mathbf{x} + \mathbf{e}, \qquad \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{b},\ \mathbf{e} \in \mathbb{R}^m, \tag{1}$$

where the matrix $A \in \mathbb{R}^{m\times n}$ is of ill-determined rank, that is, it has singular values of different orders of magnitude close to zero. In order to keep the amplified error components from dominating the computed solution, some form of regularization is typically needed when solving (1). Here we consider only two regularization methods. The first one, Tikhonov regularization, replaces the solution of (1) with the solution of the minimization problem

$$\min\{\|\mathbf{b} - A\mathbf{x}\|^2 + \alpha\|L\mathbf{x}\|^2\}, \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm and L is a regularization operator which defines a seminorm. The minimizer exists and is unique if $\ker(A) \cap \ker(L) =$

[1] Aristotle's position was the opposite of that of his teacher Plato, who regarded knowledge as a trait innate in the human mind.

$\{0\}$. In particular, if $L \in \mathbb{R}^{n\times n}$ is invertible, the problem (2) is well-posed. The second method considered here is truncated iteration. Regularization by truncated iteration consists of applying a few steps of an iterative method to the linear system (1), or to its normal equations, stopping the iteration prior to convergence to avoid the harmful effects of the amplified error components on the computed solution. The use of smoothing preconditioners in connection with truncated iteration may improve the quality of the computed solution. In particular, assuming for simplicity that the matrix A is square and that the matrix L in (2) is invertible, the solution by truncated iteration of the right-preconditioned linear system

$$AL^{-1}\mathbf{w} = \mathbf{b}, \quad L\mathbf{x} = \mathbf{w}, \tag{3}$$

may be computationally more efficient than the solution of (1), see [2].

In this paper we will show how to express our *a priori* knowledge, or lack thereof, about the sought solution and its behavior at the boundary by means of an invertible positive definite matrix L to be used as a regularization operator for Tikhonov or as a right preconditioner in truncated iteration. The matrix L may, for example, express the fact that the solution is smooth along a portion of the boundary, without forcing any particular type of boundary conditions. We will call the matrix L Aristotelian boundary preconditioner when used in connection with iterative methods for linear systems.

The paper is organized as follows. In Section 2 we show how statistical inversion provides the tools both for expressing various levels of uncertainty about the behavior of the solution at the boundary and for having the data itself determine what this behavior should be. In Section 3 we describe how to construct the matrix L in the one-dimensional model case. Section 4 describes its extension to higher dimensions. We also consider smoothness priors related to structural information, and we present computed examples where Aristotelian prior boundary conditions are used in the context of limited angle tomography.

2 Inverse problems, regularization and statistics

The construction of the smoothness prior with Aristotelian boundary conditions is based on the Bayesian interpretation of regularization. Now we briefly recall the basic concepts of the statistical theory of inverse problems. For a comprehensive discussion of this topic, we refer to [4].

Assume that $\mathbf{x} \in \mathbb{R}^n$ represents a quantity we are interested in. Since we are not sure about its value, we model it as a random variable. The randomness expresses our degree of uncertainty, which in turn is encoded in its probability distribution. We adopt the convention that a random variable is denoted by a capital letter, $\mathbf{X}$, and its realizations by lower case letters, $\mathbf{x}$. Further, we

denote the probability density[2] of $\mathbf{X}$ by $\pi(\mathbf{x})$.

Consider the linear inverse problem (1) of estimating $\mathbf{x}$ from the noisy data $\mathbf{b}$. Assume that prior to the measurement of $\mathbf{b}$, we have some information about the possible distribution of the unknown $\mathbf{x}$. This information is encoded in the *prior density* of the corresponding random variable $\mathbf{X}$, denoted by $\pi_{\rm pr}(\mathbf{x})$. Furthermore, assume that the error $\mathbf{e}$ is a realization of a noise process $\mathbf{E}$ whose probability density is $\pi_{\rm noise}(\mathbf{e})$. Then, assuming for the moment that $\mathbf{X} = \mathbf{x}$ is fixed, the equation (1) implies that $\mathbf{b}$ is a realization of a random variable $\mathbf{B}$ whose probability density must be

$$\pi(\mathbf{b} \mid \mathbf{x}) = \pi_{\rm noise}(\mathbf{b} - A\mathbf{x}). \tag{4}$$

The conditional density (4) is known as the *likelihood density*. The joint probability density of the variables $\mathbf{B}$ and $\mathbf{X}$ is then

$$\pi(\mathbf{x}, \mathbf{b}) = \pi(\mathbf{b} \mid \mathbf{x})\pi_{\rm pr}(\mathbf{x}) = \pi_{\rm noise}(\mathbf{b} - A\mathbf{x})\pi_{\rm pr}(\mathbf{x}).$$

Bayes formula states that the *posterior density* of $\mathbf{x}$ is proportional to the joint probability density, that is,

$$\pi(\mathbf{x} \mid \mathbf{b}) \propto \pi_{\rm noise}(\mathbf{b} - A\mathbf{x})\pi_{\rm pr}(\mathbf{x}).$$

A particular but very common case is when the variables $\mathbf{X}$ and $\mathbf{E}$ are independent and Gaussian. Assume, for simplicity, that

$$\mathbf{E} \sim \mathcal{N}(0, \sigma^2 I), \quad \mathbf{X} \sim \mathcal{N}(0, \Gamma),$$

where I is the $m \times m$ identity matrix and $\Gamma \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix. Then Bayes formula implies that the posterior density is, up to a norming constant,

$$\pi(\mathbf{x} \mid \mathbf{b}) \propto \exp\left(-\frac{1}{2}\left\{\frac{1}{\sigma^2}\|\mathbf{b} - A\mathbf{x}\|^2 + \mathbf{x}^{\rm T}\Gamma^{-1}\mathbf{x}\right\}\right). \tag{5}$$

Let $\Gamma^{-1} = \gamma L^{\rm T} L$. Such decomposition exists because Γ is symmetric positive definite and therefore admits a Cholesky factorization. Then the *maximum a posteriori* estimate of $\mathbf{x}$, $\mathbf{x}_{\rm MAP}$, is the maximizer of the posterior density (5) or, equivalently,

$$\mathbf{x}_{\rm MAP} = \operatorname{argmin}\big\{\|\mathbf{b} - A\mathbf{x}\|^2 + \sigma^2\gamma\|L\mathbf{x}\|^2\big\}.$$

Hence the regularization matrix L for Tikhonov regularization corresponds to the covariance matrix of the prior probability density. Thus this matrix reflects our prior information about the unknown. The regularization parameter α in (2) is related to the statistical parameters via $\alpha = \sigma^2\gamma$, giving insight of its statistical interpretation in terms of the noise level and scaling of the prior covariance. The determination of α is usually based on discrepancy principle,

[2]To avoid measure theoretic considerations, all probability densities are assumed to be absolutely continuous with respect to the Lebesgue measure.

while the statistical parameters can be estimated by hierarchical models. For further discussion of these topics, see [4].

For the purposes of this article, the key observation is that the statistical approach gives us means of interpreting the regularization. Indeed, if $L \in \mathbb{R}^{n\times n}$ is an invertible matrix appearing inside the Tikhonov functional (2) or as a preconditioner in (3), the matrix $B = L^{\mathrm{T}}L$ defines, up to a scaling constant, the inverse of a covariance. Exploration of the corresponding prior density e.g. by random draws, gives us a way to check if this matrix corresponds to what we assume to be known *a priori* about the unknown $\mathbf{x}$.

3 One dimensional smoothness prior

We begin this section by assuming that the desired solution is a smooth one-dimensional signal $f(t)$ with support on the unit interval $[0, 1]$ and we denote by $\mathbf{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ its values at the points of a uniformly spaced grid.

Assuming that we believe *a priori* that $f \in C([0, 1]) \cap C^2(]0, 1[)$, we seek to construct a prior density of x that reflects this information. Observe that no assumption whatsoever of the boundary conditions of f is made.

The natural starting point for a smoothness prior is to consider a finite difference approximation of $f''(t)$ at the interior points $t_2, \ldots, t_{n-1}$. We have

$$\begin{bmatrix} f''(t_2) \\ f''(t_3) \\ \vdots \\ f''(t_{n-1}) \end{bmatrix} \approx \frac{1}{h^2} \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{h^2} L\mathbf{x}, \qquad (6)$$

where $h = 1/(n-1)$. This approximation gives rise to a candidate prior density, which we refer to as smoothness *preprior*, defined by

$$\pi_{\mathrm{pre}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\alpha\|L\mathbf{x}\|^2\right) = \exp\left(-\frac{1}{2}\alpha\mathbf{x}^{\mathrm{T}}B\mathbf{x}\right),$$

where

$$B = L^{\mathrm{T}}L \in \mathbb{R}^{n\times n}, \quad L \in \mathbb{R}^{(n-2)\times n}.$$

We observe that since

$$\mathrm{rank}\,(B) = n - 2,$$

π_{pre} is not a proper Gaussian probability density. The rank deficiency of B alone does not prevent from using L as regularizing matrix, although care must be taken if the null spaces of L and A coincide. For certain techniques such as right priorconditioning ([2]), the invertibility of B is essential. Consequently, we want to augment L in such a manner that the resulting prior is a proper density.

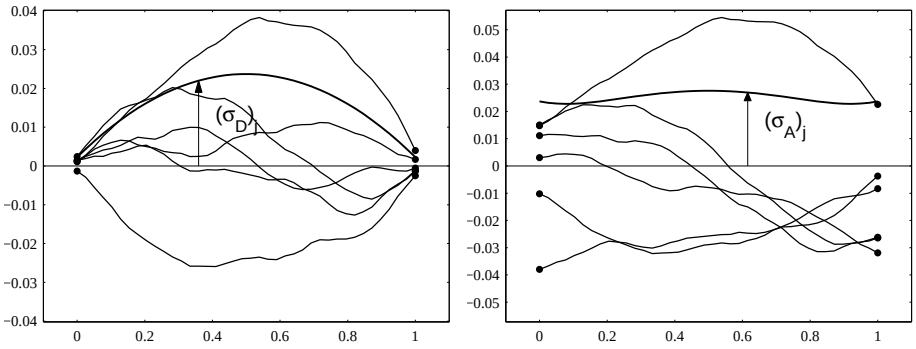

Figure 1: Six random draws from the prior $\pi_{\rm D}$ (left) and $\pi_{\rm A}$ (right) with the parameter value $\alpha = 1$. The draws are generated by solving the system $L\mathbf{x} = \mathbf{w}$, $L = L_{\rm D}$ or $L = L_{\rm A}$, where $\mathbf{w}$ is a white noise Gaussian random draw. The standard deviations of the elements are plotted as a bold curve.

A straightforward way of augmenting L so as to obtain an invertible $n\times n$ matrix is to assume boundary conditions at $t = 0$ and $t = 1$, e.g., Dirichlet, Neumann, Robin or reflecting boundary conditions. Another way is to assume that f admits a C^2-extension outside the interval $[0, 1]$ and the extension vanishes in the extended grid. Such an assumption leads naturally to the extension of L as

$$L = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix} \to \begin{bmatrix} -2 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 \end{bmatrix} = L_{\rm D}$$

The corresponding probability density is

$$\pi_{\rm D}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\alpha\|L_{\rm D}\mathbf{x}\|^2\right),$$

which is a proper density since $L_{\rm D}$, and consequently $L_{\rm D}^{\rm T}L_{\rm D}$, is invertible.

To understand which prior assumptions are implied by this extension of L, we make few random draws from the resulting density. The random draws are facilitated by the observation that if $\mathbf{W}$ is Gaussian white noise, i.e., $\mathbf{W} \sim \mathcal{N}(0, I)$, then $\mathbf{X} = (\sqrt{\alpha}L_{\rm D})^{-1}\mathbf{W}$ is distributed according to $\pi_{\rm D}$. In Figure 1, six independent random draws from this density are shown. The boundary values are near zero, indicating that by augmenting L in this fashion, we have imposed that the values of the solution at the boundary are small. To better understand the relation between the matrix L and the constraint on the solution at the boundary, consider the variances of individual components of $\mathbf{X} \sim \pi_{\rm D}$. We have

$$\big(\sigma_{\rm D}\big)_j^2 = {\rm E}\big\{X_j^2\big\} = \big(\Gamma_{\rm D}\big)_{jj},$$

where the covariance matrix Γ_{D} is

$$\Gamma_{\mathrm{D}} = \left[\alpha L_{\mathrm{D}}^{\mathrm{T}} L_{\mathrm{D}}\right]^{-1}.$$

The standard deviations $\left(\sigma_{\mathrm{D}}\right)_j$ are plotted in Figure 1. It is clear from the plot of σ_{D} that the variance at the endpoints is much smaller than in the interior of the interval. We are now going to relax the constraint at the boundary without affecting the smoothness properties inside the interval.

The extension $L \to L_{\mathrm{D}}$ was based on some implicit assumptions about the boundary values of f. We now analyze the dependence of the interior values on the boundary values. We permute the entries of the vector x so that the interior points come first, followed by the boundary points and we use the notation

$$\mathbf{x} = [x_2, x_3, \ldots, x_{n-1}, x_n, x_1]^{\mathrm{T}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \begin{matrix} n-2 \\ 2 \end{matrix}.$$

Accordingly, we partition the matrix L as

$$L = \left[L_1 \; L_2\right], \quad L_1 \in \mathbb{R}^{(n-2)\times(n-2)}, \quad L_2 \in \mathbb{R}^{(n-2)\times 2}.$$

Observe that the block L_1 is invertible. The partition of L induces a partition of B of the form

$$B = L^{\mathrm{T}} L = \begin{bmatrix} L_1^{\mathrm{T}} L_1 & L_1^{\mathrm{T}} L_2 \\ L_2^{\mathrm{T}} L_1 & L_2^{\mathrm{T}} L_2 \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

Since the matrix L_1 is invertible, so is B_{11}. If the boundary values of f and thus the vector x_2 were known, we could calculate the conditional probability density of x_1 conditioned on this information using the preprior. From now on, we assume that $\alpha = 1$ in the preprior. We have

$$\begin{aligned} \mathbf{x}^{\mathrm{T}} B \mathbf{x} &= \mathbf{x}_1^{\mathrm{T}} B_{11} \mathbf{x}_1 + \mathbf{x}_1^{\mathrm{T}} B_{12} \mathbf{x}_2 + \mathbf{x}_2^{\mathrm{T}} B_{21} \mathbf{x}_1 + \mathbf{x}_2^{\mathrm{T}} B_{22} \mathbf{x}_2 \\ &= \left(\mathbf{x}_1 + B_{11}^{-1} B_{12} \mathbf{x}_2\right)^{\mathrm{T}} B_{11} \left(\mathbf{x}_1 + B_{11}^{-1} B_{12} \mathbf{x}_2\right) \\ &\qquad + \mathbf{x}_2^{\mathrm{T}} \left(B_{22} - B_{21} B_{11}^{-1} B_{12}\right) \mathbf{x}_2. \end{aligned} \tag{7}$$

Notice that since the last term in (7) does not depend on $\mathbf{x}_1$, it only affects the norming constant in the exponential defining the preprior. The quantity

$$\widehat{B}_{11} = B_{22} - B_{21} B_{11}^{-1} B_{12}$$

is the Schur complement of B_{11}. In this setting, the matrix $\widehat{B}_{11}$ vanishes. In fact, from the relationship between the partition of B and that of L, we have that

$$B_{22} - B_{21} B_{11}^{-1} B_{12} = L_2^{\mathrm{T}} L_2 - L_2^{\mathrm{T}} L_1 \left[L_1^{\mathrm{T}} L_1\right]^{-1} L_1^{\mathrm{T}} L_2 = 0.$$

Consequently, the conditional distribution of $\mathbf{X}_1$ conditioned on $\mathbf{X}_2 = \mathbf{x}_2$ is a proper Gaussian density,

$$\pi_{\mathrm{pre}}(\mathbf{x}_1 \mid \mathbf{x}_2) \sim \mathcal{N}\left(-B_{11}^{-1} B_{12} \mathbf{x}_2 \,,\, B_{11}^{-1}\right).$$

We recall that this density corresponds to the assumption that the Dirichlet boundary values are known, while in reality, they should be treated as random variables. Assume that the boundary values are Gaussian, with marginal distribution

$$\mathbf{x}_2 \sim \mathcal{N}(0, C),$$

where $C \in \mathbb{R}^{2\times 2}$ is a symmetric positive definite matrix whose inverse has a factorization of the form

$$C^{-1} = K^{\mathrm{T}} K.$$

The marginal probability density of $\mathbf{X}_2$ is then

$$\pi_0(\mathbf{x}_2) \propto \exp\left(-\frac{1}{2}\mathbf{x}_2^{\mathrm{T}} C^{-1}\mathbf{x}_2\right) = \exp\left(-\frac{1}{2}\|K\mathbf{x}_2\|^2\right).$$

By using the identity

$$\pi(\mathbf{x}_1, \mathbf{x}_2) = \pi(\mathbf{x}_1|\mathbf{x}_2)\pi(\mathbf{x}_2), \tag{8}$$

we can now define a proper prior probability density π_{pr} as

$$\begin{aligned}
\pi_{\mathrm{pr}}(\mathbf{x}) &= \pi_{\mathrm{pre}}(\mathbf{x}_1 \mid \mathbf{x}_2)\pi_0(\mathbf{x}_2) \\
&\propto \exp\left(-\frac{1}{2}\left[\big(\mathbf{x}_1 + B_{11}^{-1}B_{12}\mathbf{x}_2\big)^{\mathrm{T}} B_{11}\big(\mathbf{x}_1 + B_{11}^{-1}B_{12}\mathbf{x}_2\big) + \|K\mathbf{x}_2\|^2\right]\right) \\
&= \exp\left(-\frac{1}{2}\begin{bmatrix}\mathbf{x}_1^{\mathrm{T}} & \mathbf{x}_2^{\mathrm{T}}\end{bmatrix}\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{21}B_{11}^{-1}B_{12} + C^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix}\right).
\end{aligned}$$

Furthermore, the matrix appearing above allows a factorization in terms of L_1, L_2 and K of the form

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{21}B_{11}^{-1}B_{12} + C^{-1}\end{bmatrix} = \begin{bmatrix} L_1^{\mathrm{T}}L_1 & L_1^{\mathrm{T}}L_2 \\ L_2^{\mathrm{T}}L_1 & L_2^{\mathrm{T}}L_2 + K^{\mathrm{T}}K\end{bmatrix}$$

$$= \begin{bmatrix} L_1 & L_2 \\ 0 & K\end{bmatrix}^{\mathrm{T}}\begin{bmatrix} L_1 & L_2 \\ 0 & K\end{bmatrix} = L_{\mathrm{A}}^{\mathrm{T}} L_{\mathrm{A}},$$

which, in turn, implies that

$$\pi_{\mathrm{pr}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|L_{\mathrm{A}}\mathbf{x}\|^2\right).$$

Note that

$$\det(L_{\mathrm{A}}) = \det(L_1)\det(K) \neq 0.$$

This is equivalent to assuming that we have an observation model of the form

$$\mathbf{X}_1 = A\mathbf{X}_2 + \mathbf{E}, \qquad \mathbf{E} \sim \mathcal{N}\big(0, B_{1,1}^{-1}\big),$$

where

$$A = -B_{11}^{-1}B_{12}$$

is the discretization of an operator that continues the boundary values of a differential equation inside the domain.

If the boundary values are independent Gaussian random variables with equal variance, then

$$C = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad K = \frac{1}{\sigma} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Substituting the expression for L into the matrix L_{A} and repermuting the entries of the vector $\mathbf{x}$ so that the pixels appear in the natural ordering we have that

$$L_{\mathrm{A}} = \begin{bmatrix} 1/\sigma & 0 & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 0 & 1/\sigma & \end{bmatrix}.$$

The selection of the value of the parameter σ can be done so that the variance of the entries of the vector $\mathbf{x}$ is as uniform as possible. This can be achieved by letting

$$\sigma = \max_{1 \le j \le n} \big(\sigma_{\mathrm{D}}\big)_j,$$

where σ_{D} is computed from the density π_{D}. In fact,

$$\big(\sigma_{\mathrm{D}}\big)_j^2 = \mathrm{var}\big(X_j\big) = \big(\big[L_{\mathrm{D}}^{\mathrm{T}} L_{\mathrm{D}}\big]^{-1}\big)_{jj},$$

so denoting by $\mathbf{e}_j$ the unit vector with the jth component equal to one,

$$\big(\sigma_{\mathrm{D}}\big)_j^2 = \mathbf{e}_j^{\mathrm{T}} \big[L_{\mathrm{D}}^{\mathrm{T}} L_{\mathrm{D}}\big]^{-1} \mathbf{e}_j = \|L_{\mathrm{D}}^{-\mathrm{T}} \mathbf{e}_j\|^2.$$

We choose $j = [n/2]$, since the maximum variance occurs at the center of the interval.

Figure 1 shows six random draws from the density π_{pr}, as well as the standard deviation of the components X_j, i.e., the square root of the diagonal of the matrix $\big[L_{\mathrm{A}}^{\mathrm{T}} L_{\mathrm{A}}\big]^{-1}$. The random draws are generated using the same realizations of the white noise as in Figure 1.

We now compare the effects of the matrices L_{D} and L_{A} with an example where L_{A} and L_{D} are used as regularization operators for Tikhonov regularization. To show how the matrix L_{A} performs as a Tikhonov penalty, we consider the following deblurring problem. Let

$$b_j = \int_0^1 K(s_j, t) f(t) dt + e_j, \quad 1 \le j \le m,$$

where the kernel K is defined as

$$K(s,t) = c(s) s(1-s) e^{-(t-s)^2/2w^2}, \quad w = 0.03,$$

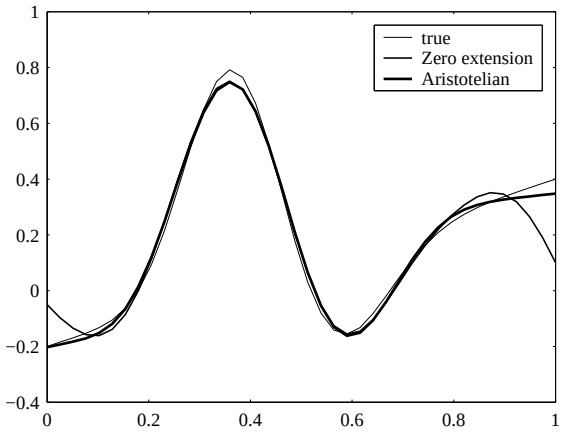

Figure 2: Tikhonov regularized solutions of the deblurring problem using the regularization matrices L_{D} and L_{A}, respectively.

and the scaling function c chosen so that $\max_{t\in[0,1]} K(s,t) = 1$. The sampling points are

$$s_j = (j-1)/(m-1), \quad m = 40.$$

The problem is discretized by approximating

$$\int_0^1 K(s_j,t)f(t)dt \approx \frac{1}{n}\sum_{\ell=1}^{n} K(s_j,t_\ell)x_\ell,$$

where the points t_j and s_j coincide, hence $n = m = 40$. The noise is assumed to be Gaussian white noise,

$$\mathbf{E} \sim \mathcal{N}(0,\gamma^2 I),\ \gamma = 1\% \text{ of max. of the noiseless signal.}$$

To avoid the "inverse crime" (see [4]) of computing the data using the same discretization grid in which the inverse problem is solved, we generate the data using a uniformly spaced grid of 100 discretization points. The inverse problem is solved using the Tikhonov regularization scheme, i.e., by solving

$$\mathbf{x}_\alpha = \mathrm{argmin}\big(\|\mathbf{b} - A\mathbf{x}\|^2 + \alpha\|L\mathbf{x}\|^2\big),$$

where either $L = L_{\mathrm{D}}$ or $L = L_{\mathrm{A}}$. The parameter $\alpha > 0$ in both cases is selected using the Morozov discrepancy principle, i.e., from the condition

$$\|\mathbf{b} - A\mathbf{x}_\alpha\|^2 = \mathrm{E}\big\{\|\mathbf{E}\|^2\big\} = m\gamma^2.$$

The results are shown in Figure 2. The improvements near the boundaries obtained when using L_{A} are clearly visible.

4 Generalizations to higher dimensions

In this section we extend the discussion to two dimensions. The extension is done recursively, and thus can be extended to higher dimensions in a similar manner. We consider two cases, a homogeneous smoothness prior and a structural smoothness prior.

4.1 Homogeneous smoothness prior

We assume that we are seeking a pixelized image supported on the unit square $D = [0,1] \times [0,1]$, and $\mathbf{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ is the value of the solution at appropriately numbered pixels. We classify the pixels into three groups, *interior pixels*, *boundary pixels* and *corner pixels*. To simplify the presentation, assume that the entries of $\mathbf{x}$ are ordered so that the first n_1 entries are the interior pixel values, the following n_2 entries correspond to boundary pixels, and, finally, the last $n_3 = 4$ entries are the corner values. We use the notation

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \begin{matrix} n_1 \\ n_2 \\ n_3 \end{matrix}, \quad n_1 + n_2 + n_3 = n.$$

As in the previous section, we assume second order smoothness prior. Using the five point mask

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{9}$$

we write a second order smoothness prior density,

$$\pi_1(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|L_1\mathbf{x}\|^2\right), \quad L_1 \in \mathbb{R}^{n_1 \times n}.$$

Clearly, π_1 does not define a proper density. As in the one dimensional case, we augment the matrix L_1 by passing through a conditional density which assumes that we know the boundary values. Consider the n_2 pixels $\mathbf{x}_2$ and define the smoothness prior

$$\pi_2(\mathbf{x}_2, \mathbf{x}_3) \propto \exp\left(-\frac{1}{2}\beta^2 \left\| L_2 \begin{bmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \right\|^2\right), \quad L_2 \in \mathbb{R}^{n_2 \times (n_2+n_3)},$$

where the matrix L_2 is assembled as a one-dimensional second order derivative along the edges using the mask

$$\begin{bmatrix} 1 & -2 & 1 \end{bmatrix}.$$

Finally, we define a Gaussian density for the independent identically distributed corner values,

$$\pi_3(\mathbf{x}_3) \propto \exp\left(-\frac{1}{2}\beta^2\|L_3\mathbf{x}\|^2\right), \quad L_3 = \alpha I \in \mathbb{R}^{n_3 \times n_3}.$$

We define the prior density $\pi_{\rm pr}$ as

$$\pi_{\rm pr}(\mathbf{x}) = \pi_1(\mathbf{x}_1 \mid \mathbf{x}_2, \mathbf{x}_2)\pi_2(\mathbf{x}_2 \mid \mathbf{x}_3)\pi_3(\mathbf{x}_3).$$

Before fixing the parameters α and β, we consider the structure of the prior thus obtained. By applying recursively the reasoning of the previous section, we first introduce the partitioning

$$L_1 = \big[L_{11}\ L_{12}\ L_{13}\big], \quad L_{11} \in \mathbb{R}^{n_1\times n_1},\ L_{12} \in \mathbb{R}^{n_1\times n_2},\ L_{13} \in \mathbb{R}^{n_1\times n_3},$$

where $\det\big(L_{11}\big) \neq 0$. We then partition

$$L_2 = \big[L_{22}\ L_{23}\big], \quad L_{22} \in \mathbb{R}^{n_2\times n_2},\ L_{23} \in \mathbb{R}^{n_2\times n_3},$$

where $\det\big(L_{22}\big) \neq 0$. With these notations, the prior density can be written in the form

$$\pi_{\rm pr}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|L_{\rm A}\mathbf{x}\|^2\right), \quad L_{\rm A} = \left[\begin{array}{ccc} L_{11} & L_{12} & L_{13} \\ 0 & \beta L_{22} & \beta L_{23} \\ 0 & 0 & \beta L_3 \end{array}\right] \in \mathbb{R}^{n\times n}, \quad (10)$$

and since

$$\det\big(L_{\rm A}\big) = \det\big(L_{11}\big)\det\big(\beta L_{22}\big)\det\big(\beta L_3\big) \neq 0,$$

it follows that $\pi_{\rm pr}$ is a proper density. Consider now the selection of the parameters α and β. The guiding principle that we adopt here is that the pixel covariances should be as equal as possible in the region of interest. The selection is done recursively from bottom to top. Assume for the time being that β is fixed. Augment first the matrix L_2 to a $(n_2 + n_3) \times (n_2 + n_3)$ matrix by extending the boundary values by zeroes along the extensions of the boundary edges. This corresponds to the matrix $L_{\rm D}$ in the one-dimensional case. We denote this extended matrix by $L_{2,\rm D}$, and we define

$$\pi_{2,\rm D}(\mathbf{x}_2, \mathbf{x}_3) \propto \exp\left(-\frac{1}{2}\beta^2 \left\| L_{2,\rm D}\left[\begin{array}{c} \mathbf{x}_2 \\ \mathbf{x}_3 \end{array}\right]\right\|^2\right), \quad \det\big(L_{2,\rm D}\big) \neq 0.$$

The covariance matrix of this density is

$$\Gamma_{2,\rm D} = \big(\beta^2 L_{2,\rm D}^{\rm T} L_{2,\rm D}\big)^{-1},$$

and the variance of a single pixel value X_j of a vector that is distributed according to $\pi_{2,\rm D}$ is

$$\mathrm{var}\big(X_j\big) = \big(\Gamma_{2,\rm D}\big)_{jj} = \frac{1}{\beta^2}\|L_{2,\rm D}^{-\rm T}\mathbf{e}_j\|^2, \quad (11)$$

where $\mathbf{e}_j \in \mathbb{R}^{n_2+n_3}$ is a unit vector with the jth component equal to one. The maximum variance occurs at the middle of the boundary edge. We assume that j is fixed to yield this maximum variance. We now choose the value of

α in $L_3 = \alpha I$ so that the variance is as uniform as possible over the whole domain. In particular, to allow the variance of the corner pixels to be equal to the maximum variance in the interior, we choose

$$\frac{1}{\alpha^2\beta^2} = \mathrm{var}\big(X_j\big),$$

or

$$\alpha = \frac{1}{\|L_{2,\mathrm{D}}^{-\mathrm{T}}\mathbf{e}_j\|}.$$

Hence, α does not depend on β. Having fixed α, we select β by moving hierarchically to higher dimensional densities. To this end, we augment first the matrix L_1 to a $n \times n$ invertible matrix by extending the image by zeros outside D and using the mask (9). Call this matrix $L_{1,\mathrm{D}}$, and define

$$\pi_{1,\mathrm{D}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\|L_{1,\mathrm{D}}\mathbf{x}\|^2\right).$$

The maximum variance of pixel values distributed according to this distribution is attained in the center of the image. If k is an index to such pixel, the maximum variance is

$$\sigma_{1,\mathrm{D}}^2 = \|L_{1,\mathrm{D}}^{-\mathrm{T}}\mathbf{e}_k\|^2.$$

By imposing that the maximum variance (11) at boundary pixes is equal to the variance above, we arrive at the simple relation

$$\frac{1}{\beta^2}\|L_{2,\mathrm{D}}^{-\mathrm{T}}\mathbf{e}_j\|^2 = \|L_{1,\mathrm{D}}^{-\mathrm{T}}\mathbf{e}_k\|^2, \quad \text{or} \quad \beta = \frac{\|L_{2,\mathrm{D}}^{-\mathrm{T}}\mathbf{e}_j\|}{\|L_{1,\mathrm{D}}^{-\mathrm{T}}\mathbf{e}_k\|}.$$

Notice that the calculation of the maximum variance is numerically light due to the structure of the matrices $L_{j,\mathrm{D}}$. Figure 3 shows the plots of the standard deviation surfaces obtained using the matrices $L_{2,\mathrm{D}}$ and $L_{2,\mathrm{A}}$. We demonstrate the use of the smoothness prior for solving a limited angle tomography problem. The image D represents a cross section of a object with smoothly varying mass absorption. The object is illuminated with parallel beam X-ray source, the projection angles being constrained in a sector of size $\pi/3$. Hence, we have only one third of the full sinogram data available. The measurement configuration is represented schematically in Figure 4. The number of illumination directions is 50 and the number of detectors along a line perpendicular to the rays is also 50. We generate the data by dividing the true image into 80×80 pixels. To avoid the "inverse crime" of solving the inverse problem on the same grid where the data were produced, the inverse problem is solved on a 50×50 grid. After having computed the noiseless data, we add Gaussian white noise with standard deviation 0.1% of the maximum noiseless signal. For details about how to construct the forward map A; see [4]. We solve the problem by using the idea of priorconditioning proposed in [2]. We rewrite the problem in the form (3), where L is the structural prior matrix formed above. The linear

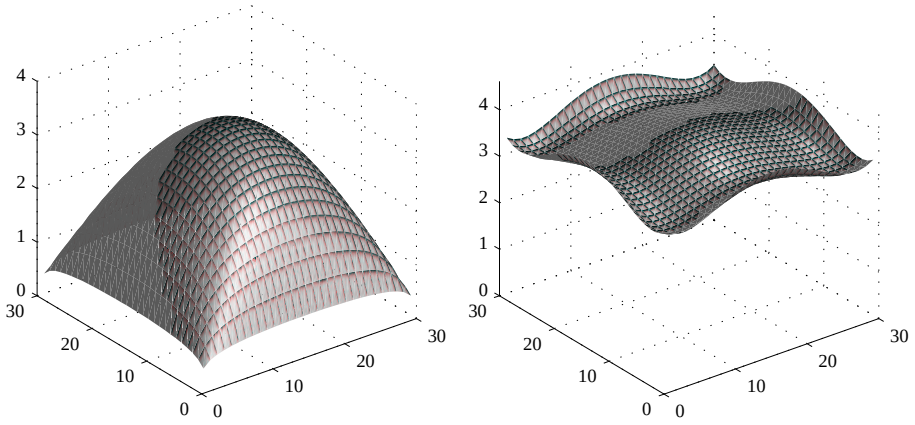

Figure 3: The standard deviations of the pixels according to the priors $\pi_{1,\mathrm{D}}$ and π_{pr}, respectively. In this example, the image size is 30×30.

system $AL^{-1}\mathbf{w} = \mathbf{b}$ is solved approximately by using the truncated conjugate gradient method for the normal equation (CGLS); see e.g. [1] for details. The CGLS is suited well for tomography problems since the matrix $A^{\mathrm{T}}A$ appearing in the normal equation corresponds to the unfiltered backprojection; see [7]. Figure 5 shows the priorconditioned CGLS solutions using the matrix L_{D} and L_{A}, respectively, as priorconditioners. The iterations are truncated after 35 steps. We do not discuss here the optimal truncation criterion, which could be based on the prior statistics; see [2]. The reconstruction results demonstrate that the Aristotelian boundary conditions not only help resolving the boundary values, but also improve the reconstruction in the interior of the domain, since the incorrect boundary conditions inherent to L_{D} are compensated with ringing artifacts.

4.2 Structural smoothness prior

In the previous sections we assumed the solution to be smooth in the interior of the domain. While this assumption is feasible in many cases, there are also examples where the solution displays some clear and known features in its interior. In such cases it is desirable to respect the structure of the problem, while still allowing the data to determine the boundary conditions. We call densities which respect features of the solution *structural smoothness prior densities*. In [5], a procedure for constructing such prior densities is proposed, and the issue is further developed in [4]. Here we show how the Aristotelian boundary conditions work together with structural priors. Consider the two-dimensional prior π_{pr} constructed in the previous section. Assume that we know *a priori* that the image contains a shape whose boundary curve is known, but that the intensity distribution of the image is unknown. Further, assume that we know that the

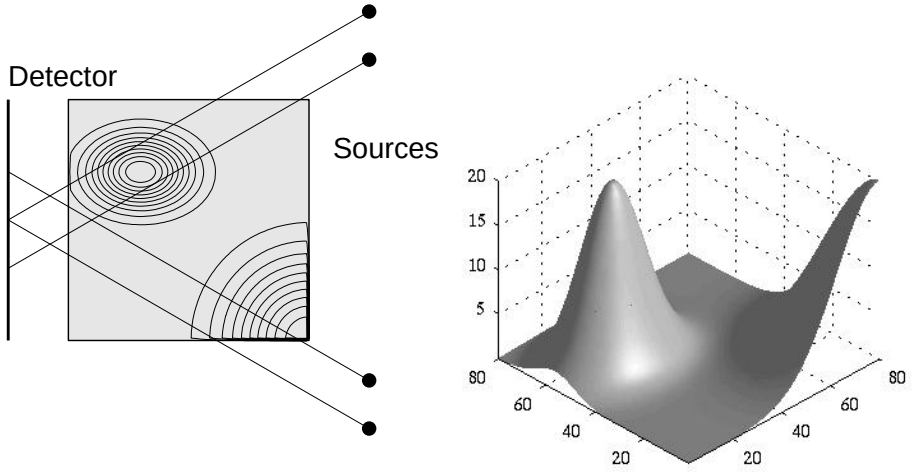

Figure 4: The measurement configuration and the true density. The projection angle ranges over an opening of $\pi/3$ as indicated in the picture.

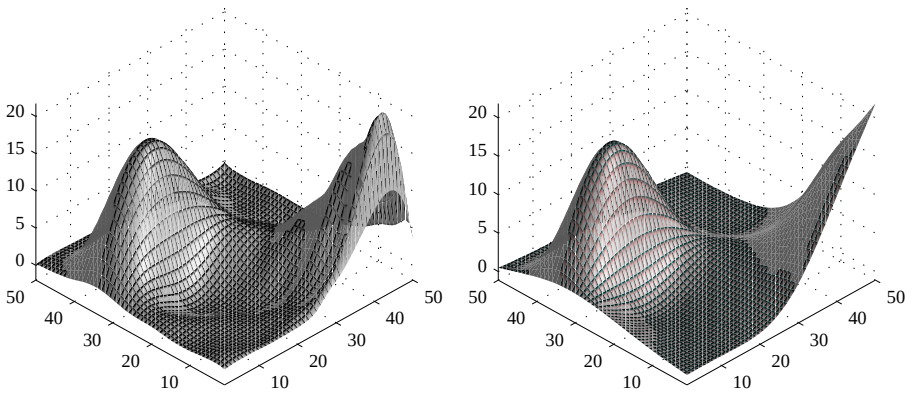

Figure 5: The priorconditioned CGLS solutions with smoothness priors using L_{D} (left) and L_{A} (right) as the priorconditioner. The number of iterations is 35.

Figure 6: The shape $Q \subset D$.

image is smooth inside and outside the boundary curve, and that the values inside and outside are not strongly coupled. Such information can be available e.g. when different imaging modalities are combined in biomedical applications. Let $Q \subset D$ be a set that represents the shape. We divide the pixels in the domain into two cliques, C_1 and C_2, according to whether the centerpoint of the pixel is in Q or not, and indicate this relation by writing $x_j \in C_j$, $j = 1, 2$. An example of such a structure is plotted in Figure 6. Assume now that we have constructed the differencing operator $L_{\rm A}$ and that we want to incorporate the structural information. Consider a pixel with one of its four neighbors belonging to a clique different from its own while the others are in the same one. Assuming that the neighbor in the different clique is on the right, we make a row replacement in $L_{\rm A}$ that corresponds to replacing the mask (9) by

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 0 \\ 1 & -3-\varepsilon & \varepsilon \\ 0 & 1 & 0 \end{bmatrix}.$$

Here, $0 < \varepsilon < 1$ is the coupling constant that controls how strongly the values inside Q and outside are correlated. Similarly, when several neighbors are in the different clique, the unit entries in (9) are replaced by ε and the sum of elements is adjusted to zero. This structural change can be done effectively as follows. Assume that the elements of $\mathbf{x}$ are permuted so that the first n_1 elements are in C_1, the last n_2 in C_2. We partition $L_{\rm A}$ accordingly:

$$L_{\rm A} = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{matrix} n_1 \\ n_2 \end{matrix}.$$

The structural prior is then obtained from the following simple steps.

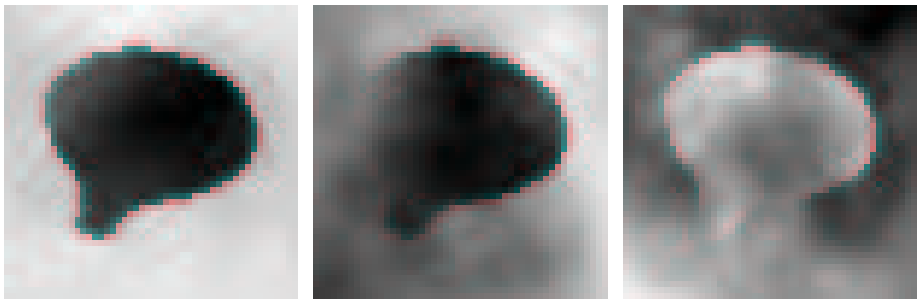

Figure 7: The shape $Q \subset D$ corresponding to the structural information and three random draws from the structural prior with Aristotelian boundary condition.

1. Calculate the vectors

$$\mathbf{d}_1 = \text{column sum of } L_{12} \in \mathbb{R}^{n_1}, \quad \mathbf{d}_2 = \text{column sum of } L_{21} \in \mathbb{R}^{n_2},$$

and replace

$$L_{\mathrm{A}} \to L_{\mathrm{A}} + (1-\varepsilon)\mathrm{diag}[\mathbf{d}_1; \mathbf{d}_2].$$

2. Replace the off-diagonal blocks

$$L_{12} \to \varepsilon L_{12}, \quad L_{21} \to \varepsilon L_{21}.$$

In Figure 7 we display three random draws of the prior density corresponding to the structural matrix thus defined. The coupling parameter here is $\varepsilon = 0.01$, and the image size is 50×50. The set Q is clearly visible in these draws. However, the structural prior does not force jumps across the boundary. Let us mention that the structural prior constructed above can be seen as an approximation of the diffusion operator $\nabla \cdot D\nabla$, where the diffusion matrix $D = D(x) \in \mathbb{R}^{2\times 2}$ is such that the diffusion across the boundary of Q is weak. Such an interpretation brings the method very close to ideas of using anisotropic nonlinear diffusion for image processing; see [8]. We demonstrate the use of the structural prior with a limited angle tomography example analogous to the one described in the previous subsection. The true mass absorption density is shown in Figure 8. In this example, D could be a cross section of a pillar containing a cavity (e.g. water pipe, a casing of electric wires etc.) of known shape, while the density distribution is unkown and of interest. Figure 9 shows the outcome based on truncated CGLS with the structural smoothness preconditioner, either with or without the Aristotelian boundary conditions. The number of CGLS iterations in this example is 20. A remarkable feature of using the structural prior is the lack of the shadow artifacts that are typical for limited angle tomography reconstructions when no prior information is available or used. In fact, the discontinuities along the structure boundaries are well resolved even in the projection directions, and the smooth object in the corner is also clearly localized.

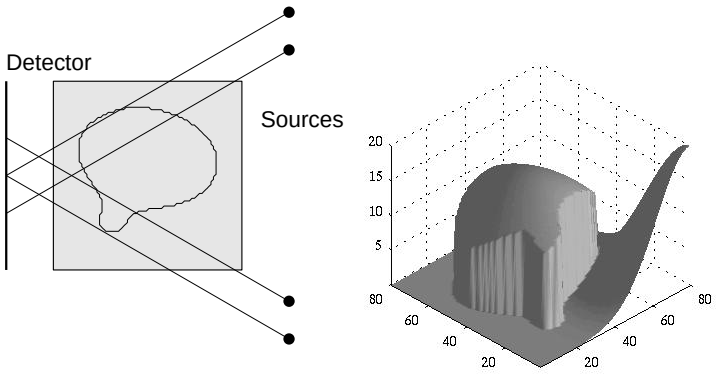

Figure 8: The measurement configuration and the true structural density.

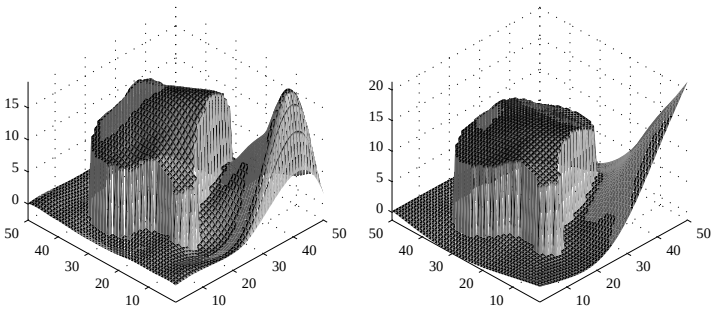

Figure 9: The priorconditioned CGLS solutions with structural smoothness priors based on a structural modification of L_{D} (left) and L_{A} (right). The number of iterations is 20.

References

[1] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.

[2] D. Calvetti and E. Somersalo: Priorconditioners for linear systems. Inverse Problems **21** (2005) 1397–1418.

[3] M. K. Ng, R. H. Chan and W-C. Tang: A fast algorithm for deblurring models with Neumann boundary conditions, SIAM J. Sci. Comput. **21** (1999), 851–866.

[4] J. P. Kaipio and E. Somersalo: *Statistical and Computational Inverse Problems*, Springer-Verlag, New York, 2004.

[5] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen and E. Somersalo: Inverse problems with structural prior information. Inverse Problems **15** (1999) 713–729.

[6] S. Serra-Capizzano: A note on antireflective boundary conditions and fast deblurring models, SIAM J. Sci. Comput. **25** (2004) 1307–1325.

[7] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä and E. Somersalo: Statistical inversion for medical X-ray tomography with few radiographs I: General theory. Phys. Med. Biol. **48** (2003) 1437–1463.

[8] J. Weickert: *Anisotropic Diffusion in Image Processing.* Teubner-Verlag 1998.