Helsinki University of Technology, Institute of Mathematics, Research Reports
Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja
Espoo 2005 A489

# A COARSE-TO-FINE STRATEGY FOR MAXIMUM A POSTERIORI ESTIMATION IN LIMITED-ANGLE COMPUTERIZED TOMOGRAPHY

**Sampsa Pursiainen**

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology, Institute of Mathematics, Research Reports
Teknillisen korkeakoulun matematiikan laitoksen tutkimusraporttisarja
Espoo 2005 A489

# A COARSE-TO-FINE STRATEGY FOR MAXIMUM A POSTERIORI ESTIMATION IN LIMITED-ANGLE COMPUTERIZED TOMOGRAPHY

**Sampsa Pursiainen**

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics

**Sampsa Pursiainen**: *A coarse-to-fine strategy for maximum a posteriori estimation in limited-angle computerized tomography*; Helsinki University of Technology, Institute of Mathematics, Research Reports A489 (2005).

**Abstract:** *The X-ray tomography problem is to reconstruct a function from its tomographic projections. This paper concerns limited-angle tomography where images are reconstructed from incomplete projection data that are limited in viewing angle and number of radiographs. The limited-angle reconstruction problem is an ill-posed inverse problem. Even if the data are noiseless obtaining an artifact-free reconstruction is problematic without a priori information about the source function. The reconstruction problem is formulated in terms of Bayesian statistics, where all the variables included in the model are defined as random vectors that follow the posterior probability density. In inverse problems, estimation of the properties of the posterior density can be problematic even if the statistical model was simple. This paper introduces a coarse-to-fine strategy, where the maximizing point the mean of a Gaussian posterior density is sought iteratively by projecting the space of possible source images into a subspace of coarser resolution images through wavelet low-pass filtering during the iteration procedure. The maximization problem is written in a preconditioned form in order to obtain better estimates. Numerical results are presented.*

**AMS subject classifications:** 65N21, 65C50, 65F22, 65F10, 65T60

**Keywords:** computerized-tomography, limited-angle tomography, Bayesian statistics, Gaussian densities, diagonalization, wavelet representation, preconditioning, Conjugate Gradient method

**Correspondence**

Sampsa.Pursiainen@hut.fi

ISBN 951-22-7860-X
ISSN 0784-3143

Helsinki University of Technology
Department of Engineering Physics and Mathematics
Institute of Mathematics
P.O. Box 1100, 02015 HUT, Finland
email:math@hut.fi http://www.math.hut.fi/

# 1 Introduction

The *X-ray tomography* problem in two dimensions is to reconstruct a source function $f : \mathbb{R}^2 \to \mathbb{R}$ from its tomographic projections (radiographs) given by the *Radon transform*

$$\mathcal{R}f(t,\theta) = \int_{\mathbb{R}^2} f(x_1,x_2)\delta(x_1\cos\theta + x_2\sin\theta - t)\,dx_1dx_2, \tag{1}$$

where $f \in \mathbf{L}^2(\mathbb{R}^2)$, $\delta$ is the Dirac mass, $\theta \in [-\pi/2, \pi/2)$ and $t \in \mathbb{R}$. It has been shown by Radon in [23] that $f$ can be perfectly recovered if $\mathcal{R}f(t,\theta)$ is known for all $t \in \mathbb{R}$ and $\theta \in [-\pi/2, \pi/2)$, which means that there is noiseless projection data available from all directions.

In real-life applications, the measurement space is finite dimensional and each measured pixel value contains noise. In the discrete Radon transform, each line integral can be approximated as a linear sum of pixel values. The discrete measurement model can be written as a linear system

$$y = Ax + \eta. \tag{2}$$

Through this paper $A$ is a real $m$-by-$n$ *Radon matrix* corresponding to the discrete approximation of the Radon transform, $x \in \mathbb{R}^n$ is a vector corresponding to the image and $\eta \in \mathbb{R}^m$ is an additive noise term. Recovery of an image from discrete tomographic projections is known as Computerized Tomography (CT) [3, 4, 18, 24]. The first commercial CT applications date back to 1970's. At the present CT is an extensively used technology in the field of medical imaging.

The fundamental difficulty in CT is the fact that the Radon transform is a smoothing transform due to which the reconstructions are very sensitive to the measurement noise. In computation of a CT reconstruction, regularization techniques are needed if the measurements are contaminated by noise. The most popular regularization procedures are the *filtered back-projection* (FBP) methods. These are linear Fourier space filtering methods, in which amplification of high frequencies is attenuated to damp out the noisy high frequency components from the reconstruction image. For details on theory of FBP see [21].

This paper concerns *limited-angle tomography* where images are reconstructed from incomplete projection data that are limited in viewing angle and number of radiographs. The problem of reconstructing an object from noisy limited-angle tomographic data arises in many important medical applications [25]. For example, in intraoral dental imaging and mammography radiographs cannot be obtained from all directions. It is also often important to minimize the radiation dose to the patient by keeping the number of radiographs as small as possible which leads to a sparse distribution of projection directions and arises a need for tomographic reconstruction from sparse limited-angle data.

The limited-angle reconstruction problem is an ill-posed inverse problem. Even if the data are noiseless obtaining an artifact-free limited-angle reconstruction is problematic without *a priori* information about the source function. In theory, a perfect limited-angle reconstruction can be obtained from an infinite set of radiographs [26, 21], but a matrix corresponding to a limited-angle Radon transform has a nonzero null-space $\mathbf{N}(A) = \{x \in \mathbb{R}^n \, : \, Ax = 0\}$ about which no information is provided by the measurements. The matrix can also be shown to be *ill-conditioned* [5, 19, 9]; that is, the space of all possible source images $\mathbb{R}^n$ can be decomposed into two orthogonal subspaces

$$\mathbb{R}^n = \mathbf{S}_\varepsilon^- \oplus \mathbf{S}_\varepsilon^+, \tag{3}$$

where $\mathbf{S}_\varepsilon^- = \{x \in \mathbb{R}^n \, : \, ||Ax||/||x|| \leq \varepsilon\}$, $\varepsilon > 0$ is a very small number and $\mathbf{S}_\varepsilon^- \setminus \mathbf{N}(A) \neq \emptyset$. It is very difficult to recover the component of the original source image lying in $\mathbf{S}_\varepsilon^-$ since in that subspace $||Ax||$ is insensitive to variation of $||x||$. One can say that $\mathbf{S}_\varepsilon^-$ forms the 'numerical null-space' of $A$. In further discussion, the decomposition (3) is called a *sensitivity decomposition* of $A$. The term is adopted from [17].

In this work, the limited-angle reconstruction problem is formulated in terms of *Bayesian statistics* [25, 16], where all the unknown variables are defined as random vectors that follow the *posterior probability density*. A priori information about the source function is incorporated into a prior density that together with the likelihood of different measurement outcomes define the posterior. In the Bayesian approach, a reconstruction of the source image is typically an estimate of the maximizing point or the mean of the posterior. These are known as *maximum a posteriori* (MAP) and *conditional mean* (CM) estimates.

In the present paper, the prior, likelihood and posterior densities are all assumed to be *Gaussian*. Due to the property (3), estimation of maximizing point (the mean) of a Gaussian posterior density can be very problematic, if the posterior covariance matrix is ill-conditioned. In order to obtain reasonable estimates based on such a posterior, the reconstruction problem should be represented in a vector basis where the covariance is diagonal or nearly diagonal. This is proposed by Kalifa and Mallat in [14]. This paper discusses a nearly block diagonal representation of the posterior covariance. The concepts of diagonalization and near diagonalization in tomographic reconstruction has been considered by Kalifa et al in the papers [13, 15].

Based on the idea of near block diagonalization of the posterior covariance matrix, this paper introduces a *coarse-to-fine strategy*, where the MAP estimate is computed iteratively by projecting the space of possible source images into a subspace of coarser resolution images through *wavelet* low-pass filtering during the iteration procedure. The maximization problem is written in a *preconditioned* form in order to obtain better estimates. The *Conjugate Gradient* (CG) method is used for maximization. Numerical results are presented.

The goal in applying wavelets and preconditioning is to represent the posterior in such a way that the posterior covariance is nearly block diagonal and that the maximization problem can be divided to well-conditioned and ill-conditioned parts. This is closely related to finding an approximative sensitivity decomposition of the Radon matrix. Projective techniques in linear inverse problems have been discussed e.g. in [22, 17]. The first paper by Piana and Bertero introduces a projective approach to a linear inverse image processing problem. In the second article by Liu et al wavelet representation is applied to the heat conduction problem. Preconditioning in linear inverse problems is discussed e.g. in [1, 22, 11]. The first article by Calvetti and Somersalo introduces a concept called priorconditioning and the third paper by Hanke concerns preconditioned Conjugate Gradient method for image restoration problems. Discussion about tomographic reconstruction by wavelets can be found e.g. in [7, 13, 15].

This paper is organized as follows. Section 2 briefly reviews the Bayesian formulation of the reconstruction problem. Section 3 is devoted to Gaussian statistical models. Section 4 discusses the concepts of diagonalization and preconditioning of a posterior covariance matrix. Section 5 briefly reviews the definitions of one- and two-dimensional wavelet bases. In section 6, it is shown numerically by using wavelet low-pass filters, that the condition number of a matrix corresponding to a limited-angle Radon transform can be decreased by projecting the space of possible source images is projected into a subspace of coarser resolution images. In section 7, the proposed coarse-to-fine strategy is described. A few numerical examples are included. Finally, Section 8 collects the conclusions and findings of this paper and discusses the possible directions for the future work.

# 2 The Bayesian model

In the Bayesian formulation of limited-angle tomography the problem to be considered is the following: given the data $y$, find the posterior density $p(x \,|\, y)$. By the Bayes' formula the posterior density can be written in the form

$$p(x \mid y) = \frac{p(x)p(y \mid x)}{p(y)}, \tag{4}$$

where $p(x)$ denotes the prior probability density of the unknown vector $x$ and is assumed to contain all possible information of the target prior to the measurements. The conditional density $p(y \,|\, x)$ denotes the likelihood of measuring $y$ given the vector $x$ and $p(y)$ is the probability density of measuring the data $y$.

By assuming that $x$ and $\eta$ are independent the likelihood density can be written as

$$p(y \,|\, x) = p_{\text{noise}}(y - Ax), \tag{5}$$

that is; the likelihood density can be evaluated if it is known how the measurement noise is distributed. Finally, assume that the measurement data $y$ is given. Then, the posterior density is proportional to the product of the prior and the likelihood densities since $p(y)$ is constant for each given set of measurements $y$. The Bayes' formula states now that the posterior density is

$$p(x \mid y) \propto p(x) p_{\text{noise}}(y - Ax), \tag{6}$$

where $\propto$ means that the distributions are equal up to a constant.

## 2.1 Posterior estimation

In the Bayesian approach, a reconstruction is found as an estimate of some property of the posterior distribution. Typically, the maximum a posteriori or the conditional mean estimate is evaluated. These are defined by

$$x_{MAP} = \arg\max_x p(x \mid y) \quad \text{and} \quad x_{CM} = \int_{\mathbb{R}^n} x\, p(x \mid y)\, dx. \tag{7}$$

Obtaining either of these estimates can be a computationally challenging problem that requires use of advanced optimization and numerical integration algorithms. Difficulties arise whenever the shape of the posterior distribution is such that the algorithms tend to proceed to wrong directions or get stuck in local minima.

In limited-angle CT, these difficulties are caused by the fact that the Radon matrix is ill-conditioned. Under the assumption that the likelihood density (5) is a continuous function of $y$, the values of $p(y \,|\, x)$ and $p(y \,|\, x + z)$ are hardly distinguishable for all $z \in \mathbf{S}_\varepsilon^-$. In a case where the prior is *non-informative*, i.e. the prior density is flat or nearly flat, the posterior $p(x \,|\, y)$ is nearly flat in $\mathbf{S}_\varepsilon^-$ and obtaining appropriate estimates from the posterior can be difficult. In such case, the components of $x_{MAP}$ and $x_{CM}$ lying in $\mathbf{S}_\varepsilon^-$ are also very sensitive to the measurement noise and it is possible that these components do not contain relevant information about the true source image.

# 3 Gaussian densities

Gaussian probability distributions (normal distributions) have an important role in Bayesian computations. Being relatively easy to handle Gaussian densities are attractive in the computational point of view. Due to the central limit theorem they are also often very good approximations for non-Gaussian densities.

In the present paper, an $n$-variate Gaussian density $p(x)$ with mean $\overline{x}$ and covariance matrix $\Gamma$ is defined as

$$p(x) = (2\pi|\Gamma|)^{-n/2} \exp\Big( -\frac{1}{2}(x - \overline{x})^T \Gamma^{-1} (x - \overline{x})\Big), \tag{8}$$

where $\overline{x} \in \mathbb{R}^n$ and $\Gamma$ is real symmetric and strictly positive definite $n$-by-$n$ matrix and $|\Gamma| = \det(\Gamma)$. The pair $\overline{x}$ and $\Gamma$ defines a Gaussian probability distribution $\mathcal{N}(\overline{x}, \Gamma)$.

A distribution of the form $\mathcal{N}(0,\gamma^2 I)$ is called a Gaussian white noise distribution.

### 3.1 The Gaussian measurement model

In X-ray tomography, the measurement noise $\eta$ is often assumed to be Gaussian $\mathcal{N}(\overline{\eta},\Gamma_\eta)$. From (5) one can observe that, in such case, the likelihood is also a Gaussian density given by

$$p(y \mid x) = (2\pi|\Gamma_\eta|)^{-m/2} \exp\Big( -\frac{1}{2}(y-Ax-\overline{\eta})^T \Gamma_\eta^{-1}(y-Ax-\overline{\eta})\Big), \qquad (9)$$

Typically, in numerical simulations the noise term assumes the white noise distribution. In real-life applications, a Gaussian approximation for the distribution of the measurement noise can be determined by careful analysis of the measurement electronics [25].

### 3.2 Gaussian posterior densities

Suppose that the prior density of $x$ is Gaussian with mean $\overline{x}_{\mathrm{pr}}$ and covariance matrix $\Gamma_{\mathrm{pr}}$ and that the likelihood is of the form (9). Then, the posterior density is again Gaussian $p(x \mid y) \propto \exp(-\|b-Fx\|_2^2)$, where

$$b = \begin{bmatrix} L_\eta & 0 \\ 0 & L_{\mathrm{pr}} \end{bmatrix} \begin{bmatrix} y \\ \overline{x}_{\mathrm{pr}} \end{bmatrix} \quad \text{and} \quad F = \begin{bmatrix} L_\eta & 0 \\ 0 & L_{\mathrm{pr}} \end{bmatrix} \begin{bmatrix} A \\ I \end{bmatrix}. \qquad (10)$$

The matrices $L_\eta$ and $L_{\mathrm{pr}}$ are the Cholesky factors of $\Gamma_\eta^{-1}$ and $\Gamma_{\mathrm{pr}}^{-1}$. The mean $\overline{x}_{\mathrm{post}}$ and the covariance matrix $\Gamma_{\mathrm{post}}$ of this distribution are given by

$$\begin{aligned} \overline{x}_{\mathrm{post}} &= (F^TF)^{-1}F^Tb = (A^T\Gamma_\eta^{-1}A + \Gamma_{\mathrm{pr}}^{-1})^{-1}(A\Gamma_\eta^{-1}y + \Gamma_{\mathrm{pr}}^{-1}\overline{x}_{\mathrm{pr}}), && (11) \\ \Gamma_{\mathrm{post}} &= (F^TF)^{-1} = (A^T\Gamma_\eta^{-1}A + \Gamma_{\mathrm{pr}}^{-1})^{-1}. && (12) \end{aligned}$$

A Gaussian density has only one local maximum, which is also the global maximum, and the maximizing point is simultaneously the mean. Therefore, a Gaussian posterior distribution satisfies

$$x_{\mathrm{MAP}} = x_{\mathrm{CM}} = \arg\min_x \|b - Fx\|^2. \qquad (13)$$

If both the prior and the likelihood follow the Gaussian white noise model, the formulas for $b$ and $F$ are reduced to

$$b = \begin{bmatrix} y \\ 0 \end{bmatrix} \quad \text{and} \quad F = \gamma_\eta^{-1} \begin{bmatrix} A \\ \sqrt{\alpha} I \end{bmatrix} \qquad (14)$$

with $\alpha = \gamma_\eta^2/\gamma_{\mathrm{pr}}^2$. In this particular case, the covariance is $\Gamma_{\mathrm{post}} = \gamma_\eta^2(A^TA+\alpha I)^{-1}$ and the mean coincides with the Tikhonov regularized solution of the equation

$Ax = y$ with regularization parameter $\alpha$ (A.2). Thus, the connection between Tikhonov regularization and the Gaussian statistical model is obvious. Note that if the white noise prior is non-informative in the sense that the prior variance $\gamma_{\text{pr}}^2$ is really large, then $\alpha$ is a very small number and the covariance matrix is nearly singular. If $\alpha$ tends to zero the mean tends to $A^\dagger y$, where $A^\dagger = \lim_{\alpha \to 0^+}(A^T A + \alpha I)^{-1} A^T$ is the Moore-Penrose pseudoinverse, that can be shown to exist for any $m$-by-$n$ matrix [10].

# 4 Diagonalization

Computation of the MAP estimate from a Gaussian posterior density can be problematic if the posterior covariance matrix is ill-conditioned, i.e. if the posterior is extremely flat in some directions, since flatness can cause numerical instability of the estimation procedure. A Gaussian posterior density should, therefore, be represented in a vector basis that distinguishes the flat parts from the other parts. In such a basis the covariance matrix is diagonal or nearly diagonal [14]. If the covariance is not diagonal or nearly diagonal, the posterior can be basically flat in the direction of each basis vector. This is illustrated in Figure 1.

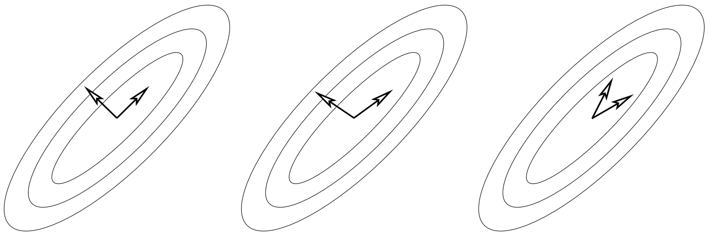

Figure 1: Three different vector bases that provide a diagonal (left), a nearly diagonal (center) and a non-diagonal (right) representations of a Gaussian density.

## 4.1 Diagonal covariance matrices

Suppose that $p(x)$ is a Gaussian density of the form (8). Because the covariance matrix is symmetric and strictly positive definite it has an eigenvalue decomposition of the form [10]

$$\Gamma = V \Lambda V^T, \tag{15}$$

where $\Lambda$ is a strictly positive diagonal matrix containing the eigenvalues of $\Gamma$. The orthogonal eigenvector basis that is formed by the columns of $V$ is said to provide a diagonalization the matrix $\Gamma$. By introducing an orthogonal change of coordinates $z = V^T x$ the original density $p(x)$ is transformed into another Gaussian density with the mean $\overline{z} = V^T \overline{x}$ and the diagonal covariance matrix $\Lambda$.

In the particular case, where the posterior density is given by (14), the columns of $V$ are formed by the eigenvectors of $A^TA$ ( i.e. the right singular vectors of $A$). In the general case (12), the decomposition (15) can be obtained if the eigenvectors of $A^TA$ and the covariances $\Gamma_\eta$ and $\Gamma_{\rm pr}$ are known.

## 4.2 Nearly block diagonal covariance matrices

Instead of a strictly diagonal representation, it is often preferable to represent a Gaussian density in a basis where the covariance matrix is nearly diagonal or nearly block diagonal. This can be due to two reasons. Firstly, the matrix to be decomposed can be so large that it is computationally too expensive to calculate the eigenvalue decomposition. Secondly, if the matrix is ill-conditioned, the eigenvalue decomposition is likely to be contaminated by numerical errors. The usefulness of a nearly diagonal or nearly block diagonal representation can be motivated by using the following result.

Let a joint density of two Gaussian random vectors be of the form

$$p(x_1,x_2) \propto \exp\Big(-\frac{1}{2}\begin{bmatrix} x_1-\overline{x}_1 \\ x_2-\overline{x}_2 \end{bmatrix}^T \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^T & \Gamma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1-\overline{x}_1 \\ x_2-\overline{x}_2 \end{bmatrix}\Big). \tag{16}$$

Then, the conditional density $p(x_1 \,|\, x_2)$ is Gaussian with the conditional mean $\overline{x}_{1.2} = \overline{x}_1 - \Gamma_{12}\Gamma_{22}^{-1}(x_2 - \overline{x}_2)$ and the conditional covariance matrix $\Gamma_{1.2} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{12}^T$. For the proof see [12].

Suppose that the covariance of the joint density (16) is an ill-conditioned matrix but the conditional covariance matrix $\Gamma_{1.2}$ is well-conditioned. Then, estimation of $(\overline{x}_1, \overline{x}_2)$ is extremely difficult but the mean of $p(x_1 \,|\, x_2)$ for a given $x_2$ can be estimated reasonably well. It is easy to see that the conditional mean $\overline{x}_{1.2}$ tends to $\overline{x}_1$ as $\Gamma_{12}$ tends to zero or as $x_2$ tends to $\overline{x}_2$. Therefore, one can estimate $\overline{x}_1$ by estimating the mean of $p(x_1 \,|\, x_2)$ if $x_2$ relatively close to $\overline{x}_2$ and the covariance in (16) is nearly block diagonal.

Consider the Gaussian posterior density defined by (14). An above described partition of the corresponding posterior covariance matrix can be found if an approximation of the sensitivity decomposition (3) is available. This can be shown as follows.

Let the matrices $W_1$ and $W_2$ with orthonormal columns span two orthogonal subspaces of $\mathbb{R}^n$ that approximate the spaces $\mathbf{S}_\varepsilon^+$ and $\mathbf{S}_\varepsilon^-$ respectively. By introducing a change of coordinates $(x_1, x_2) = (W_1x, W_2x)$ the posterior density assumes the form (16) with the covariance matrix $\Gamma_{k\ell} = \gamma_\eta^2 W_\ell^T V(\Lambda^{-1} + \alpha I)V^T W_k$, where $\Lambda$ and $V$ contain the eigenvalues and vectors of $A^TA$. If the subspaces spanned by $W_1$ and $W_2$ provide a reasonable approximation of (3) the covariance is a nearly block diagonal matrix in which $\Gamma_{11}$ is a well-conditioned and $\Gamma_{22}$ is a nearly singular block. If $\Gamma_{12}$ is close enough to a zero matrix, then by the Hoffman-Wielandt inequality (25) the conditional covariance $\Gamma_{1.2}$ is also a well-conditioned matrix and the conditional mean $\overline{x}_{1.2}$ is a reasonable approximation of $\overline{x}_1$.

Note that a similar nearly block diagonal partition of the general posterior covariance matrix (12) into well-conditioned and ill-conditioned blocks can be obtained if an approximative sensitivity decomposition of the radon matrix and the covariances $\Gamma_\eta$ and $\Gamma_{\mathrm{pr}}$ are known.

### 4.3 The concept of preconditioning

Let $S$ and $M$ be two symmetric and strictly positive definite matrices such that $SF = FM$ and let $p$ and $\widetilde{p}$ be two Gaussian densities given by

$$p(x) \propto \exp(-\|b - Fx\|^2) \quad \text{and} \quad \widetilde{p}(x) \propto \exp(-\|S^{-1}(b - Fx)\|^2). \tag{17}$$

Then, it is easy to see that both densities have the same mean $\overline{x} = (F^T F)^{-1} F^T b$ but different covariances. The covariance matrix of $p$ is $\Gamma = (F^T F)^{-1}$ whereas the covariance of $\widetilde{p}$ is given by $\widetilde{\Gamma} = (F^T S^{-2} F)^{-1} = M^2 (F^T F)^{-1}$ and that the same vector basis that diagonalizes $\Gamma$ diagonalizes also $\widetilde{\Gamma}$.

The matrix $S$ is a so-called preconditioning matrix (preconditioner). The goal in preconditioning is to choose the matrix $S$ in such a way that the covariance matrix $\widetilde{\Gamma}$ is as close to diagonal as possible. Yet, one should be able to evaluate $S^{-2}$. If $F$ were a well-conditioned matrix, the optimal choice for $S$ would yield $M^2 = F^T F$ in which case $\widetilde{\Gamma}$ would be an identity matrix. Since the matrix $M$ actually defines a coordinate transformation, that is not necessarily orthogonal, one can deduce that the idea of preconditioning is essentially to represent the posterior in a vector basis where the posterior covariance matrix is nearly diagonal.

Efficiency of a preconditioning matrix can be measured by comparing the condition numbers of the covariance matrices $\Gamma$ and $\widetilde{\Gamma}$. Generally, if the condition number $\kappa(\widetilde{\Gamma}) = \lambda_{\max}/\lambda_{\min}$ is smaller than $\kappa(\Gamma)$, the matrix $\widetilde{\Gamma}$ is likely to be 'closer' to a diagonal matrix than $\Gamma$, since any matrix of the form $V \Lambda V^T$ tends to an identity matrix if $\Lambda$ tends to $I$ and $V$ is kept constant. Again, any orthogonal basis nearly diagonalizes a covariance matrix whose condition number is close enough to one.

## 5 Wavelet bases

A straightforward calculation shows that the Radon transform (1) satisfies the so-called projection-slice formula [6]

$$\int_{\mathbb{R}^2} f(x_1, x_2) e^{-i\lambda(x_1 \sin\theta + x_2 \cos\theta)} \, dx_1 dx_2 = \int_{\mathbb{R}} \mathcal{R}f(t, \theta) e^{-i\lambda t} \, dt. \tag{18}$$

In other words, the Radon transform of $f$ can be obtained by applying the one-dimensional inverse Fourier transform to the two-dimensional Fourier transform of $f$ restricted to radial lines going through the origin. This means that in the Radon transform the two-dimensional Fourier basis is mapped diagonally to the one-dimensional Fourier basis. Unfortunately, the Fourier basis functions have an infinite support in the spatial domain and, therefore, they are not well-suited for representing spatially inhomogeneous data such as images with discontinuities.

For computational purposes, some *multiresolution analyses*, such as wavelet bases [20] can provide a better trade-off between the spatial representation of the data and the representation of the Radon transform than the Fourier basis. In a multiresolution analysis, where the basis functions are relatively well-localized both in frequency and in spatial domain, the Radon transform can be represented in a nearly diagonal form but also the information content of images can be analyzed efficiently.

## 5.1 Orthogonal wavelet representation

A multiresolution analysis is a sequence of subspaces $\{\mathbf{V}_k\}_{k\in\mathbb{Z}}$ of $\mathbf{L}^2(\mathbb{R})$ that is defined by the following conditions

1. $\cdots \subset \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \cdots$.
2. $\bigcap_{k\in\mathbb{Z}} \mathbf{V}_k = \{0\}$ and $\bigcup_{k\in\mathbb{Z}} \mathbf{V}_k$ is dense in $\mathbf{L}^2(\mathbb{R})$.
3. $f \in \mathbf{V}_k$ if and only if $f(2^{-k}\,\cdot) \in \mathbf{V}_0$
4. If $f \in \mathbf{V}_k$, then $f(\cdot - \ell) \in \mathbf{V}_k$ for all $\ell \in \mathbb{Z}$.
5. There exists a scaling function $\varphi \in \mathbf{L}^2(\mathbb{R})$ such that $\{\varphi_{0\ell}(x)\}_{\ell\in\mathbb{Z}}$ forms a basis of $\mathbf{V}_0$.

It follows from these conditions that there exists a so-called scaling vector $\{h_\ell\}_{\ell\in\mathbb{Z}}$ such that the equation $\varphi(x) = \sum_{\ell\in\mathbb{Z}} h_\ell \varphi(2x - \ell)$ holds. The scaling vector is said to generate the multiresolution analysis. If additionally $\varphi$ is such that the family of its translates $\{\varphi(\,\cdot - \ell)\}_{\ell\in\mathbb{Z}}$ is an orthonormal system, the scaling vector generates a family of one-dimensional orthogonal wavelets. The corresponding wavelet basis can be constructed by translating and dilating the scaling function $\varphi$. The wavelet function is defined as $\psi(x) = \sum_\ell (-1)^\ell h_{1-\ell}\varphi(2x - m)$ and the actual wavelet basis functions are obtained from $\varphi$ and $\psi$ as $\varphi_{k\ell}(x) = 2^{k/2}\varphi(2^k x - \ell)$ and $\psi_{k\ell}(x) = 2^{k/2}\psi(2^k x - \ell)$ with $(k, \ell) \in \mathbb{Z}^2$. The set $\{\varphi_{k\ell}(x)\}_{\ell\in\mathbb{Z}}$ forms an orhonormal basis of $\mathbf{V}_k$ and $\{\psi_{k\ell}\}_{\ell\in\mathbb{Z}}$ forms an orthonormal basis of $\mathbf{W}_k = \mathbf{V}_k^\perp \cap \mathbf{V}_{k+1}$. By definition $\bigcup_{k\in\mathbb{Z}} \mathbf{V}_k$ is dense in $\mathbf{L}^2(\mathbb{R})$ and, therefore, any $f \in \mathbf{L}^2(\mathbb{R})$ can be represented in the form $f = \sum_{(k,\ell)\in\mathbb{Z}^2} \langle f, \varphi_{k\ell}\rangle \varphi_{k\ell}$. The scaling and wavelet functions generate two orthonormal bases of the subspace $\mathbf{V}_{k+1} = \mathbf{V}_k \oplus \mathbf{W}_k$ which are $\{\varphi_{k\ell}\}_{(k,\ell)\in\mathbb{Z}^2}$ and $\{\varphi_{k\ell}\}_{(k,\ell)\in\mathbb{Z}^2} \cup \{\psi_{k\ell}\}_{(k,\ell)\in\mathbb{Z}^2}$. This means that for all $f \in \mathbf{L}^2(\mathbb{R})$

$$\sum_{\ell\in\mathbb{Z}} \langle f, \varphi_{k+1,\ell}\rangle \varphi_{k+1,\ell} = \sum_{\ell\in\mathbb{Z}} \langle f, \varphi_{k\ell}\rangle \varphi_{k\ell} + \sum_{\ell\in\mathbb{Z}} \langle f, \psi_{k\ell}\rangle \psi_{k\ell}. \tag{19}$$

The left hand side is a orthogonal projection of $f$ to the space $\mathbf{V}_{k+1}$. The first sum on the right hand side defines a low-pass filter to the the $k$th resolution level and the second sum represents details that are included in $\mathbf{V}_{k+1}$ but not in $\mathbf{V}_k$. The inner products $\langle f, \varphi_{k\ell}\rangle$ and $\langle f, \psi_{k\ell}\rangle$ are called respectively the approximation and the detail coefficients at the $k$th resolution level.

### 5.2 Two-dimensional orthogonal wavelets

Let $\varphi$ and $\psi$ be the one-dimensional scaling and wavelet functions of some orthogonal wavelet family. The corresponding two-dimensional scaling and wavelet functions are $\phi(x_1,x_2) = \varphi(x_1)\varphi(x_2)$, $\psi^1(x_1,x_2) = \psi(x_1)\varphi(x_2)$, $\psi^2(x_1,x_2) = \varphi(x_1)\psi(x_2)$ and $\psi^3(x_1,x_2) = \psi(x_1)\psi(x_2)$. The corresponding wavelet basis functions are obtained from these through translation and dilation: $\phi_{k\ell j}(x_1,x_2) = 2^k\phi(2^k x_1-\ell, 2^k x_2-j)$ and $\psi^i_{k\ell j}(x) = 2^k\psi^i(2^k x_1-\ell, 2^k x_2-j)$ with $(k,\ell,j) \in \mathbb{Z}^3$ and $i = 1,2,3$. The sum $\sum_{(\ell,j)\in\mathbb{Z}^2}\langle f, \phi_{k\ell j}\rangle\phi_{k\ell j}$ defines a low-pass filter to the $k$th resolution level. The approximation coefficients at the $k$th level are formed by the set $\{\langle f, \phi_{k\ell j}\rangle\}_{(\ell,j)\in\mathbb{Z}^2}$ and the corresponding detail coefficients $\{\langle f, \psi^i_{k\ell j}\rangle\}_{(\ell,j,i)\in\mathbb{Z}^2\times\{1,2,3\}}$ can be subdivided to horizontal, vertical and diagonal coefficients.

## 6 Numerical sensitivity analysis

Because larger structures provide the image 'context', it is natural to analyze first the image details at a coarse resolution level and then gradually increase the resolution [20]. In this this section, it is shown numerically by using the wavelet representation that limited-angle Radon projections are, in general, more sensitive to a coarse- than fine-scale fluctuations, which suggests that the coarse-to-fine strategy can be applied in limited-angle reconstruction.

Let $\varphi$ be a scaling function that generates a family of two-dimensional orthogonal wavelets and let $\{\mathbf{V}_k\}_{k\in\mathbb{Z}}$ be the related multiresolution analysis. Assume that the source image to be reconstructed lies in the $n$-dimensional space $\mathbf{V}_0$ (i.e. the finest resolution level is indexed by zero) and denote by $P_k$ a matrix that defines an orthogonal projection between the coordinates of $f \in \mathbf{V}_0$ and the coordinates of the low-pass filtration $\sum_{(\ell,j)\in\mathbb{Z}^2}\langle f, \varphi_{-k,\ell j}\rangle\varphi_{-k,\ell j} \in \mathbf{V}_{-k}$. The ranges of the matrices $P_k$ and $I - P_k$ decompose the space of possible source images $\mathbb{R}^n$ into two orthogonal parts.

By Section 4.2, a near block diagonal representation of the posterior covariance matrix can be obtained if an approximation of the sensitivity decomposition (3) is available. Whether the ranges of matrices $P_k$ and $I - P_k$ provide approximations of the spaces $\mathbf{S}^+_\varepsilon$ and $\mathbf{S}^-_\varepsilon$ respectively, is studied below by analyzing the eigenvalues of the matrices $P_k A^T A P_k$, $(I - P_k)A^T A(I - P_k)$ and $(I - P_k)A^T A P_k$ (see A.1).

## 6.1 Results and discussion

Figure 2 shows how $\lambda_{\min}(P_k A^T A P_k)$, $\lambda_{\max}((I - P_k) A^T A (I - P_k))$ and $\lambda_{\max}((I - P_k) A^T A P_k)$ behave as a function of $k = 1, \ldots, 5$. Also illustrated in the figure is $\lambda_{\mathrm{rank}(P_k)}(A^T A)$, which shows what $\lambda_{\min}(P_k A^T A P_k)$ would be if $P_k$ would be spanned by the first $\mathrm{rank}(P_k)$ right singular vectors of $A$. Here, the original image resolution is $64 \times 64$ pixels, the number of Radon projections is 15, the angle between the projection beam and the vertical image axis obtains values $0°, \pm 3°, \pm 6°, \ldots, \pm 18°$ and $\pm 21°$ and the matrix $P_k$ is constructed by using the Meyer wavelets which are bandlimited in the Fourier domain.

Figure 2 indicates that $\lambda_{\min}(P_k A^T A P_k)$ grows monotonically as $k$ increases. In other words, for each $k = 0, \ldots, 4$ the Radon projections are more sensitive to details at level $k + 1$ than details at level $k$. One can see also from the figure that the gap between $\lambda_{\min}(P_k A^T A P_k)$ and $\lambda_{\mathrm{rank}(P_k)}(A^T A)$ becomes narrower as $k$ increases, which means that the range of $P_k$ approximates better the most sensitive $\mathrm{rank}(P_k)$ dimensional subspace at coarse than at fine scales.

Because the projection beams are basically vertical the measurements are much more sensitive to horizontal than to vertical details in the source image at each resolution level. This can be observed from the projection-slice formula (18) according which vertical projections do not give any information about horizontal spectrum of the reconstructed object. The measurements are very sensitive to even the finest scale horizontal fluctuations and, therefore, $\lambda_{\max}((I - P_k) A^T A (I - P_k))$ does not depend very much on $k$ as can be seen from Figure 2.

For each $k$, the maximal eigenvalue of $(I - P_k) A^T A P_k$ is large compared to the other computed eigenvalues. Therefore, one cannot say that the ranges of $A P_k$ and $A(I - P_k)$ would be close to orthogonal.

The fact that the maximal eigenvalues of the matrices $(I - P_k) A^T A (I - P_k)$ and $(I - P_k) A^T A P_k$ are clearly above zero for each $k$ indicates that the spaces spanned by the columns of $P_k$ and $(I - P_k)$ do not provide an optimal sensitivity decomposition of $A$. However, these spaces can still provide an excellent approximative sensitivity decomposition in practice.

# 7 The coarse-to-fine strategy

In this section, it is demonstrated how the coarse-to-fine strategy discussed in the previous section can be applied to finding the maximizing point $x_{\mathrm{MAP}}$ (the mean) of a Gaussian posterior density. A simple model is adopted, where both the prior and the likelihood densities follow the Gaussian white noise model. With this choice, the posterior density assumes the form $p(x \mid y) \propto \exp(-||b - Fx||^2)$ where $F$ and $b$ are given by (14) and the related maximum a posteriori estimation problem is given by (13). It is assumed that the prior variance $\gamma_{\mathrm{pr}}^2$ is so large that a direct solution of the problem, e.g. through Cholesky factorization [10], will be

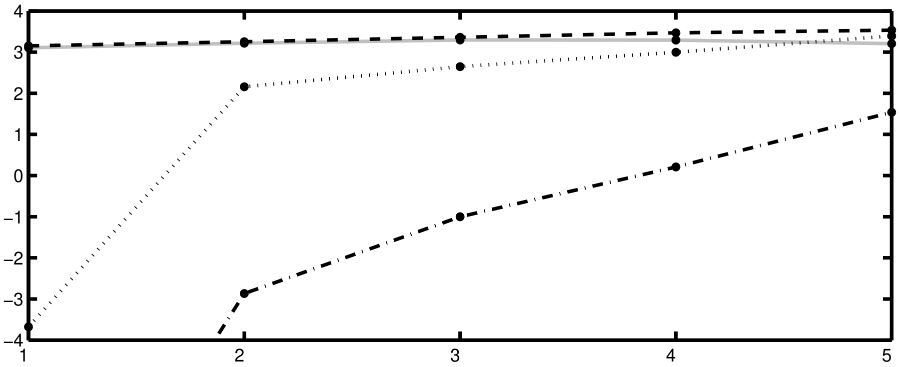


Figure 2: The base 10 logarithm of $\lambda_{\min}(P_k A^T A P_k)$ (dash-dotted line), $\lambda_{\max}((I - P_k)A^T A(I - P_k))$ (dashed line), $\lambda_{\max}((I - P_k)A^T A P_k)$ (solid line) and $\lambda_{\mathrm{rank}(P_k)}(A)$ (dotted line) as a function of $k$.

contaminated by numerical errors.

In the coarse-to-fine strategy $x_{\mathrm{MAP}}$ is estimated first at a coarse level and then gradually refined. This is done by estimating

$$x_k^* = \arg\min_x ||S^{-1}(b - FP_k x)||^2 \tag{20}$$

where $S$ is a preconditioning matrix and $P_k$ is a projection matrix corresponding to a Meyer wavelet low-pass filter similarly as in the previous section. During the minimization process the index $k$ is gradually decreased. For each $k$ the problem is solved numerically by applying the Conjugate Gradient method [10] to the normal equation

$$(P_k F^T S^{-2} F P_k) x_k^* = S^{-1} b. \tag{21}$$

Each time when $k$ is decreased by one the current iterate is used as an initial guess for $x_{k-1}^*$. Note that this is not the classical Preconditioned Conjugate Gradient (PCG) method, in which preconditioning is applied to speed up the convergence of the iteration. Here, preconditioning is applied to obtain better reconstructions.

Finding $x_k^*$ is equivalent to finding the mean of the conditional density $\widetilde{p}(x_1 \mid x_2 = 0, y)$, where $\widetilde{p}$ is obtained from the posterior density $p$ as shown in (17) and $(x_1, x_2) = (P_k x, (I - P_k)x)$. By the discussion in Section 4.2, the mean of $\widetilde{p}(x_1 \mid x_2, y)$ is a reasonable estimate of $\overline{x}_1 = P_k \overline{x}_{\mathrm{post}}$, if the covariance $\widetilde{\Gamma}$ of $\widetilde{p}$ is nearly diagonal and $||x_2 - \overline{x}_2||$ is close enough to zero. In this case, one can expect that $x_k^* \approx P_k \overline{x}_{\mathrm{post}}$ if $\widetilde{\Gamma}$ is close enough to identity and $||(I - P_k)\overline{x}_{\mathrm{post}}||$ is small enough.

The fact that $||(I - P_k)\overline{x}_{\mathrm{post}}||$ should be small means that the image corresponding to $\overline{x}_{\mathrm{post}}$ should not be dominated by fine scale fluctuations. Therefore, the coarse-to-fine strategy (20) is experimented by using phantoms that contain mainly coarse-scale features. Again, since the measurement noise is likely to induce a high-frequency component to $\overline{x}_{\text{post}}$ the noise variance $\gamma_\eta^2$ is assumed to be small. The preconditioning matrix that is used in computations is given by

$$S^2 = \begin{bmatrix} AA^T + \alpha I & 0 \\ 0 & I \end{bmatrix}. \tag{22}$$

It is easy to see that with this choice the preconditioned covariance matrix becomes $\widetilde{\Gamma} = \gamma_\eta^2 (A^T(AA^T + \alpha I)^{-1} A + \alpha I)^{-1}$. By using the singular value decomposition one can verify that there exists a positive definite matrix $M$ such that $SF = FM$ and that the preconditioned covariance matrix has a smaller condition number than the original posterior covariance matrix. Hence, by Section 4.3 the equation (22) defines a feasible preconditioner. In [22] an analogous preconditioning scheme is called *Tikhonov preconditioning.*

## 7.1 Results and discussion

Figure 3 illustrates the three $64 \times 64$ binary phantoms used in the computations. Figures 4 and 5 both contain three different reconstructions of each phantom that were computed by

1. using the filtered back projection algorithm provided by the MATLAB iradon function

2. directly solving $x_{\text{MAP}}$ through Cholesky factorization of the posterior covariance matrix

3. finding $x_{\text{MAP}}$ iteratively through the coarse-to-fine strategy.

In the coarse-to-fine strategy five CG iterations were performed at the coarsest level $k = 3$ and after that one iteration at each level from $k = 2$ to $k = 0$. Each reconstruction was computed by choosing $\alpha = \gamma_\eta^2/\gamma_{\text{pr}}^2 = 10^{-5}$ and $||\eta||/||y|| \approx 10^{-5}$. In Figure 4 the number of Radon projections used in the reconstructions is 15. The angle between the projection beam and the vertical image axis obtains the values $0°, \pm 3°, \pm 6°, \ldots, \pm 18°$ and $\pm 21°$. In Figure 5, the number of projections is 22 and the directions are $\pm 1°, \pm 3°, \pm 5°, \ldots, \pm 19°$ and $\pm 21°$.

As observed in Section 6.1, the measurements are much more sensitive to vertical than to horizontal fluctuations, because the projection beams are basically vertical. Therefore, it is difficult to recover any horizontal details from the data. From Figures 4 and 5 one can see that the coarse-scale horizontal details of the phantoms are best recovered by applying the coarse-to-fine strategy. However, also these reconstructions contain artifacts. In FBP reconstructions horizontal details are completely lost.

By comparing Figures 4 and 5 one can observe that the sparseness of the projection beams affects substantially to the quality of the MAP estimates.

Figure 3: The three phantoms that were used in computations.

Preconditioning enhances the quality of coarse-to-fine reconstructions considerably. Without preconditioning minimization of (20) yields very similar images than direct solution of the posterior mean by using the Cholesky factorization.

Due to the relatively large prior variance used in the computations, increasing the measurement noise level from that used easily causes high-frequency artifacts to the MAP estimates. Removing these artifacts without changing the prior density would necessitate decreasing the prior variance. This would again rapidly decreases the distinguishability of the horizontal fluctuations in the reconstruction images.

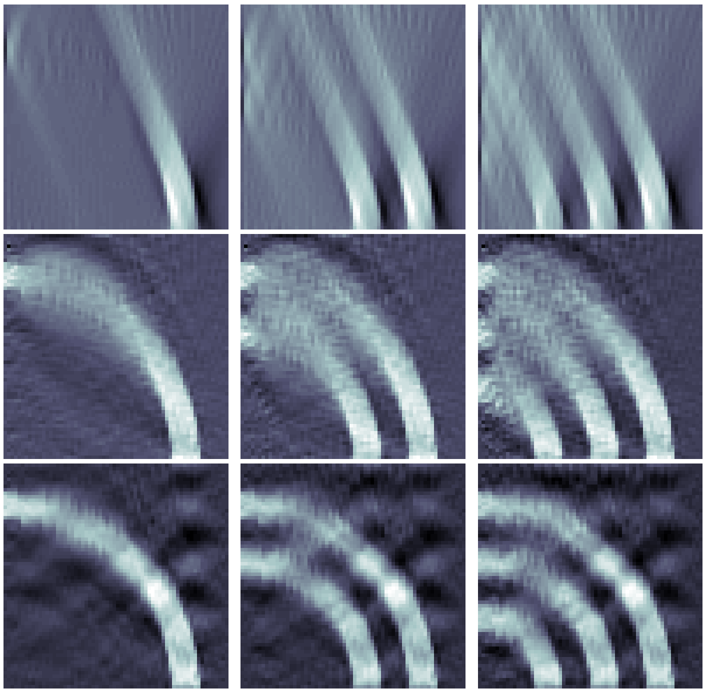

Figure 4: Reconstructions of the phantoms obtained by applying three different methods: FBP reconstruction (top row), direct solution of $x_{\mathrm{MAP}}$ (center row) and the coarse-to-fine strategy (bottom row).

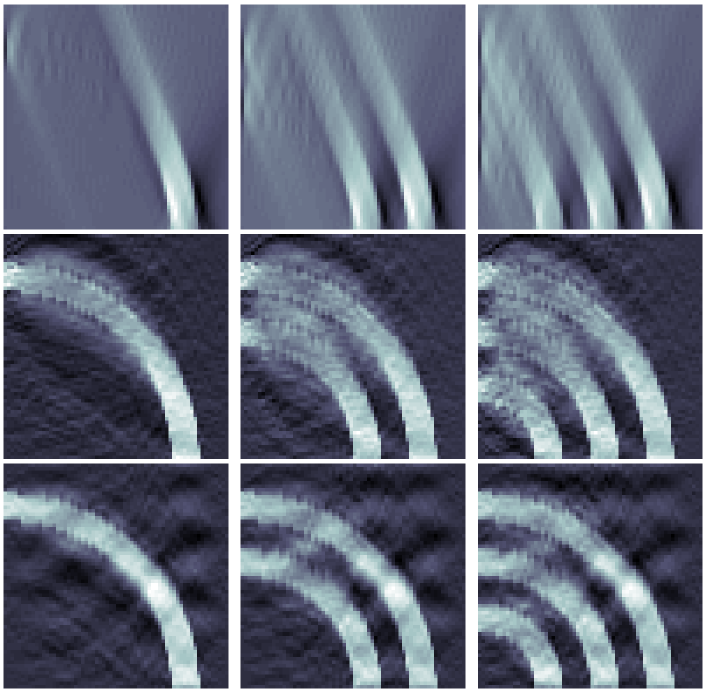

Figure 5: Reconstructions of the phantoms obtained by applying three different approaches: FBP reconstruction (top row), direct solution of $x_{\mathrm{MAP}}$ (center row) and the coarse-to-fine strategy (bottom row).

# 8 Conclusions and future work

In this paper, it was found numerically by using wavelet low-pass filters, that the condition number of a matrix corresponding to a limited-angle Radon transform can be decreased by projecting the space of possible source images into a subspace of coarser resolution images. The numerical results concerning the coarse-to-fine strategy show that wavelet filters can also be successfully applied to approximatively decompose the space of possible source functions into sensitive and insensitive parts. It was also observed, that the results obtained by using such an approximative decomposition in the MAP estimation process can be substantially enhanced by preconditioning of the covariance matrix. When preconditioning was applied the reconstructions obtained through the coarse-to-fine strategy were superior to the reconstructions obtained from direct solution of the posterior mean.

One possibility for the future work would be to design more sophisticated multiresolution filters that decompose the space of sought functions to detectable and undetectable fluctuations. In such work, multiresolution bases that analyze spectral data in more than just horizontal, vertical and diagonal directions, such as curvelet [2] and contourlet [8] bases, might turn out to be useful. Another topic for future studies could be to develop a regularizing prior by using which the proposed coarse-to-fine strategy would not be as sensitive to the measurement noise as they are in the present numerical implementation.

# A Appendix : The singular value decomposition and its applications

A general real $m \times n$ matrix has a singular value decomposition (SVD) [10]

$$A = U\Sigma V^T, \tag{23}$$

where $U = (u_1, \ldots, u_m)$ is a real $m$-by-$m$ matrix and $V = (v_1, \ldots, v_n)$ is a real $n$-by-$n$ matrix both of which have orthogonal columns so that $U^T U = V^T V = I$. The columns $u_i$ and $v_i$ of $U$ and $V$ are called the *left* and *right singular vectors*, respectively. $\Sigma$ is a real $m$-by-$n$ diagonal matrix whose diagonal elements $\sigma_1, \ldots, \sigma_{\min(m,n)}$ are the *singular values* of $A$ and ordered so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(m,n)} \geq 0$. Additionally, it is defined that $\sigma_k = 0$ for $\min(m,n) < k \leq n$. The singular values can be shown to satisfy [10]

$$\sigma_k(A) = \max_{W \in \mathbf{U}_{n,k}} \min_{0 \neq z \in \mathbb{R}^k} \frac{||AWz||}{||z||} = \min_{W \in \mathbf{U}_{n,n-k+1}} \max_{0 \neq z \in \mathbb{R}^k} \frac{||AWz||}{||z||}, \tag{24}$$

where $\mathbf{U}_{n,k}$ denotes a space of $n$-by-$k$ matrices with orthonormal columns. This is known as the Courant-Fischer min-max theorem. The proof is a simple application of SVD.

Some important matrix norms can be expressed in terms of singular values. From the Courant-Fischer theroem it follows that the Euclidean norm $||A|| = \max_{||x||_2=1} ||Ax||$ is equal to the largest singular value; that is, $||A|| = \sigma_1(A)$. The Frobenius norm is defined as $||A||_F^2 = \sum_{k=1}^m \sum_{\ell=1}^n |A_{k\ell}|^2$ and satisfies $||A||_F^2 = \sum_{k=1}^n \sigma_k(A)^2$. The Hoffman-Wielandt inequality [10] shows that the singular values depend continuously on $A$:

$$\sum_{k=1}^n |\sigma_k(A+E) - \sigma_k(A)|^2 \leq \|E\|_F^2. \tag{25}$$

The singular value decomposition is an important tool in solving ill-conditioned linear systems. If $A$ is an ill-conditioned matrix, the value of $||Ax||/||x||$ varies tremendously depending on $x$. From (24) one sees that this is equivalent to an extremely widespread distribution of singular values in the sense that the condition number $\kappa(A) = \sigma_{\max}/\sigma_{\min}$, i.e. the ratio between the largest and the smallest singular value, is extremely large.

If the matrix is symmetric the singular values and vectors are similarly the eigenvalues and vectors. Eigenvectors of the matrix $A^T A$ coincide with the right singular vectors of $A$ and the equation between the eigenvalues and the singular values is $\lambda_i(A^T A) = \sigma_i^2(A)$ for $i = 1, \ldots, n$.

## A.1 The sensitivity decomposition

The sensitivity decomposition (3) introduced in [17] divides the space of sought functions into two parts where $||Ax||$ is respectively sensitive and insensitive to variation of $||x||$. The subspaces $\mathbf{S}_\varepsilon^-$ and $\mathbf{S}_\varepsilon^+$ can be obtained from SVD. The subspace $\mathbf{S}_\varepsilon^+$ is spanned by the right singular vectors $v_1, \ldots, v_k$ and $\mathbf{S}_\varepsilon^-$ is spanned by the vectors $v_{k+1}, \ldots, v_n$, if $k$ is such that $\sigma_k > \varepsilon \geq \sigma_{k+1}$. Supposing that $\varepsilon$ has a reasonable value, the sensitivity decomposition subdivides $\mathbb{R}^n$ into sensitive and insensitive parts. The subspaces $\mathbf{S}_\varepsilon^+$ and $\mathbf{S}_\varepsilon^+$ are orthogonal due to the orthogonality of the singular vectors. The subspaces $\{Ax : x \in \mathbf{S}_\varepsilon^-\}$ and $\{Ax : x \in \mathbf{S}_\varepsilon^+\}$ are also orthogonal since $Av_i = \sigma_i u_i$.

Consider now an approximative sensitivity decomposition into two orthogonal subspaces that are spanned by the matrices $W_1$ and $W_2$ with orthogonal columns. Due to the min-max property (24), sensitivity of $A$ in these subspaces can be analyzed by examining the distribution of the singular values of $AW_1$ and $AW_2$. A decomposition into sensitive and insensitive parts should satisfy $\sigma_{\min}(AW_1) \geq \sigma_{\max}(AW_2)$. However, by (24) the inequality $\sigma_{\min}(AW_1) > \sigma_{\max}(AW_2)$ can be satisfied only if the columns of $W_1$ are spanned by the first $k$ right singular vectors. Therefore, as an approximative sensitivity decomposition this is, in general, the better the smaller is the gap between $\sigma_{\min}(AW_1)$ and $\sigma_{\max}(AW_2)$ (or the gap between $\lambda_{\min}(W_1^T A^T AW_1)$ and $\lambda_{\max}(W_2^T A^T AW_2)$). The quality of the decomposition depends also on how close to orthogonal are the ranges of the matrices $AW_1$ and $AW_2$, i.e. how large is $\lambda_{\max}(W_1^T A^T AW_2)$. The singular value decomposition is optimal in the sense that by choosing $W_1 = (v_1, \ldots, v_k)$ and $W_2 = (v_{k+1}, \ldots, v_n)$ one has $\lambda_{\max}(W_1^T A^T AW_2) = 0$ and the gap between $\sigma_{\min}(AW_1)$ and $\sigma_{\max}(AW_2)$ is minimized.

## A.2 The Tikhonov regularized solution of a linear system

It is easy to see that $x_{\text{ls}}$ minimizes the least-squares norm $||y - Ax||^2$ if and only if the so-called normal equation $A^T Ax_{\text{ls}} = A^T b$ is satisfied. If $A^T A$ is a full-rank matrix the solution of the normal equation is unique and given by $x_{\text{ls}} = \sum_{k=1}^n (y^T v_k/\sigma_k^2) v_k$. If $A$ is of full-rank but an ill-conditioned matrix the singular values tend to zero rapidly as $k$ increases. Then, $1/\sigma_k^2$ is very large for large $k$, which causes instability of the solution.

Tikhonov regularized solution $x_\alpha$ is the minimizer of the regularized least-squares problem $\arg\min_x ||Ax - y||^2 + \alpha||x||_2^2$ and satisfies the equation $(A^T A + \alpha I)x_\alpha = A^T y$. The matrix on the left hand side is of full rank for all $A \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^m$ and $\alpha > 0$. The parameter $\alpha$ controls the level of regularization. By using SVD this can be written as

$$x_\alpha = (A^T A + \alpha I)^{-1} A^T y = \sum_{k=1}^n \frac{\sigma_k}{\sigma_k^2 + \alpha}(y^T u_k) v_k. \tag{26}$$

Assuming that the classical least squares solution exists $x_\alpha$ is close to $x_{\mathrm{ls}}$ in the sense that $x_\alpha^T v_k \approx x_{\mathrm{ls}}^T v_k$ for $\sigma_k \gg \alpha$, since $\sigma_k/(\sigma_k^2+\alpha) \approx 1/\sigma_k$ for small $k$. Instability is, however, not a problem, if $\alpha$ is chosen properly, because $\sigma_k/(\sigma_k^2 + \alpha) \to 0$ when $\sigma_k \to 0$ for all $\alpha > 0$.

# References


[1] Calvetti D and Somersalo 2005 E Priorconditioners for linear systems *Inverse Problems* **21** 1397–1418

[2] Candés E J and Donoho D L 2000 Curvelets - a surprisingly effective non-adaptive representation for objects with edges. In *Curves and Surfaces* ed C Rabut, A Cohen and L L Schumaker (Nashville, TN : Vanderbilt University Press) pp 105–120

[3] Cormack A M 1963 Representation of a function by its line integrals, with some radiological applications I *J. Appl. Phys.* **34** 2722–27

[4] Cormack A M 1964 Representation of a function by its line integrals, with some radiological applications II *J. Appl. Phys.* **35** 195–207 [13]

[5] Davison M E 1983 The ill-conditioned nature of the limited-angle tomography problem *SIAM J. Appl. Math.* **43** 428–48

[6] Deans S 1983 *The Radon transform and some of its applications* (John Wiley and Sons)

[7] Delanay A H and Bresler Y 1995 Multiresolution tomographic reconstruction using wavelets *IEEE Transactions on Image Processing* **4** 799–813

[8] Do M N and Vetterli M 2003 Contourlets *Beyond Wavelets* ed G V Welland (Academic Press)

[9] Finch D 1985 Cone beam reconstruction with sources on a curve. *SIAM J. Appl. Math.* **45**(4) 665–673

[10] Golub G and van Loan C 1989 *Matrix Computations* (Baltimore : The John Hopkins University Press)

[11] Hanke M 2000 Iterative regularization techniques in image restoration *Surveys on Solution Methods for Inverse Problems* ed D Colton et al (Wien : Springer-Verlag) pp 35–52

[12] Kaipio J P and Somersalo E 2004 *Statistical and Computational Methods for Inverse Problems* (Berlin: Springer)

[13] Kalifa J, Jin Y, Laine A F and Esser P D 2000 Tomographic Reconstruction with Non-linear Diagonal Estimators *Proceedings of European Association of Nuclear Medicine* (Paris, France, September 2000)

[14] Kalifa J and Mallat S 2003 Thresholding estimators for linear inverse problems and deconvolutions *Ann. Statist.* **31**(1) 58–109

[15] Kalifa J, Laine A and Esser P D 2003 Regularization in Tomographic Reconstruction Using Thresholding Estimators *IEEE Transactions on Medical Imaging* **22**(3) 351–359

[16] Kolehmainen V, Siltanen S, Järvenpää S, Kaipio J P, Koistinen P, Lassas M, Pirttilä J and Somersalo E 2003 Statistical inversion for medical x-ray tomography with few radiographs: II. Application to dental radiology *Physics in Medicine and Biology* **48** 1465–1490

[17] Liu J, Guerrier B and Benard C 1995 A sensitivity decomposition for the regularized solution of inverse heat conduction problems by wavelets *Inverse Problems* **11** 1177–1187

[18] Logan B F and Shepp L A 1975 Optimal reconstruction of a function from its projections. *Duke Math. J.* **42** 645–59

[19] Louis A K 1986 Incomplete Data Problems in X-Ray Computerized Tomography I. Singular Value Decomposition of the Limited Angle Transform. *Numer. Math.* **48** 251–262

[20] Mallat S G 1989 A Theory for Multiresolution Signal Decomposition: The Wavelet Representation *IEEE Transactions on Pattern Analysis and Machine Intelligence* **7** 674–693

[21] Natterer F 1986 *The Mathematics of Computerized Tomography* (New York : Wiley)

[22] Piana M and Bertero M 1997 Projected Landweber method and preconditioning 1997 *Inverse Problems* **13** 441–463

[23] Radon J 1917 Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte Über die Verhandlungen der Sächsischen Akademien der Wissenschaften, Leipzig. Mathematisch-physische Klasse* **69** 262–7

[24] Shepp L A, Kruskal J B 1978 Computerized tomography: the new medical X-ray technology. *Americal Mathematical Monthly* **85**(6) 420–439

[25] Siltanen S, Kolehmainen V, Järvenpää S, Kaipio J P, Koistinen P, Lassas M, Pirttilä J and Somersalo E 2003 Statistical inversion for medical x-ray tomography with few radiographs: I. General theory *Physics in Medicine and Biology* **48** 1437–1463

[26] Smith K T, Solmon D C and Wagner S L 1977 Practical and mathematical aspects of the problem of reconstructing objects from radiographs. *Bulletin of the AMS* **83**(6) 1227–1270

(continued from the back cover)

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The list of reports is continued inside. Electronical versions of the reports are available at *http://www.math.hut.fi/reports/* .

A486 Hanna Pikkarainen
A Mathematical Model for Electrical Impedance Process Tomography
April 2005

A485 Sampsa Pursiainen
Bayesian approach to detection of anomalies in electrical impedance tomography
April 2005

A484 Visa Latvala , Niko Marola , Mikko Pere
Harnack's inequality for a nonlinear eigenvalue problem on metric spaces
March 2005

A482 Mikko Lyly , Jarkko Niiranen , Rolf Stenberg
A refined error analysis of MITC plate elements
April 2005

A481 Dario Gasbarra , Tommi Sottinen , Esko Valkeila
Gaussia Bridges
December 2004

ISBN 951-22-7860-X
ISSN 0784-3143