INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **48** (2003) 1437–1463 PII: S0031-9155(03)57800-9

Statistical inversion for medical x-ray tomography with few radiographs: I. General theory

S Siltanen[1], V Kolehmainen[2], S Järvenpää[4], J P Kaipio[2], P Koistinen[4], M Lassas[4], J Pirttilä[5] and E Somersalo[3]

[1] Instrumentarium Corp. Imaging Division, PO Box 20, FIN-04301 Tuusula, Finland
[2] Department of Applied Physics, University of Kuopio, PO Box 1627, FIN-70211 Kuopio, Finland
[3] Institute of Mathematics, Helsinki University of Technology, PO Box 1100, FIN-02015, Finland
[4] Rolf Nevanlinna Institute, University of Helsinki, PO Box 4, FIN-00014, Finland
[5] Invers Ltd, Tähteläntie 54 A, FIN-99600 Sodankylä, Finland

E-mail: Ville.Kolehmainen@uku.fi

Received 3 January 2003
Published 7 May 2003
Online at stacks.iop.org/PMB/48/1437

Abstract

In x-ray tomography, the structure of a three-dimensional body is reconstructed from a collection of projection images of the body. Medical CT imaging does this using an extensive set of projections from all around the body. However, in many practical imaging situations only a small number of truncated projections are available from a limited angle of view. Three-dimensional imaging using such data is complicated for two reasons: (i) typically, sparse projection data do not contain sufficient information to completely describe the 3D body, and (ii) traditional CT reconstruction algorithms, such as filtered backprojection, do not work well when applied to few irregularly spaced projections. Concerning (i), existing results about the information content of sparse projection data are reviewed and discussed. Concerning (ii), it is shown how Bayesian inversion methods can be used to incorporate *a priori* information into the reconstruction method, leading to improved image quality over traditional methods. Based on the discussion, a low-dose three-dimensional x-ray imaging modality is described.

1. Introduction

Three-dimensional x-ray imaging is based on acquiring several projection images of a body from different directions. If projection images are available from all around a two-dimensional slice of the body, the classical work of Radon (1917) shows that the inner structure of the slice can be determined. This result was reinvented by Cormack and Hounsfield and commercialized

in the 1970s as computerized tomography (CT) imaging technology which is widely used in medicine today (Cormack 1963, 1964, Shepp and Kruskal 1978).

We consider clinical imaging situations where three-dimensional information is helpful but complete CT-type projection data are not available. For instance, in mammography the breast is compressed against a fixed detector and it is possible to move the x-ray source keeping the breast immobilized. However, the detector-beam angle should be relatively small and if the detector plane cannot be rotated, the projections can be obtained only from a relatively narrow aperture. The resulting reconstruction problem[6] is an example of *limited-angle tomography*. Another type of situation occurs in extraoral dental imaging where some teeth are imaged with x-rays passing through other teeth and skull. The region of interest is surrounded by uninteresting tissue which is not attempted to be imaged. This situation is called *local tomography*. Apart from geometric restrictions, keeping the number of radiographs as small as possible to minimize radiation dose to the patient leads to sparse distribution of projection directions. We call the above type of data *sparse projection data* as opposed to traditional CT data.

Three-dimensional medical x-ray imaging using sparse projection data can be viewed as an imaging modality of its own, made feasible by the digital revolution in x-ray imaging technology. It is suited for situations in which the sought-for diagnostic information cannot be retrieved from any single projection image and a CT scan is not feasible due to low resolution, high radiation dose or cost of equipment. By its information content, this kind of imaging is obviously superior to studying a single radiograph. However, it differs significantly from CT imaging since sparse projection data do not contain enough information to completely describe a 3D body. Instead, only certain features of the body can be reliably reconstructed. What these features are depends both on data and available *a priori* information.

It is well known that traditional CT reconstruction algorithms do not produce satisfactory reconstructions when applied to sparse projection data (see Ranggayyan *et al* 1985, Natterer 1986, Hanson 1987 and references therein). In this paper we present and review results suggesting that statistical inversion methods can be successfully used for reconstruction. The statistical inversion approach has the following benefits:

- Any collection of projection data can be used for tomographic reconstruction. In particular, cone beam geometry and truncated projections are not more complicated to work with than parallel beam geometry and full projections.
- Application-dependent *a priori* information on the target can be used in a natural and systematic way to recast the classically ill-posed problem in a well-posed stochastic form. With a well-constructed prior model one can obtain improved image quality over traditional methods.

Part I of this paper is a review paper. It brings together results from physics, mathematics and medical imaging in a way that is not usually considered in the field of CT imaging. The rest of this part is organized as follows. In section 2, we give a review of the mathematical results on the information content of sparse projection data. In section 3, we discuss the theory of statistical inversion, the prior models and the computation of posterior statistics on a rather general level. The likelihood model for the collection of projection data is also discussed. In section 4, we describe a three-dimensional x-ray imaging modality based on sparse projection data and give a review of statistical inversion approaches to x-ray tomography. In section 5,

[6] For historical reasons, we use occasionally the term *reconstruction* to mean any procedure to acquire information on the inner structures from x-ray measurements, although this term is quite imprecise, in particular from the point of view of statistical inference.

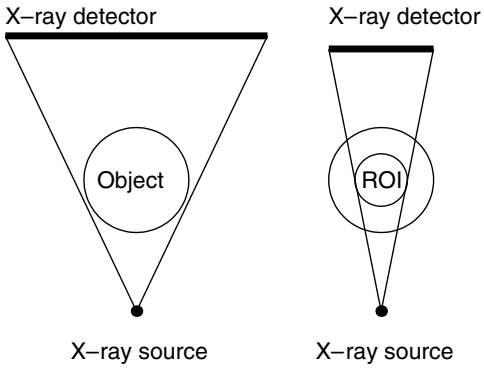

Figure 1. Illustration of cone beam measurement geometry for transmission tomography. Left: global tomography. Right: local tomography. The region of interest is denoted by ROI.

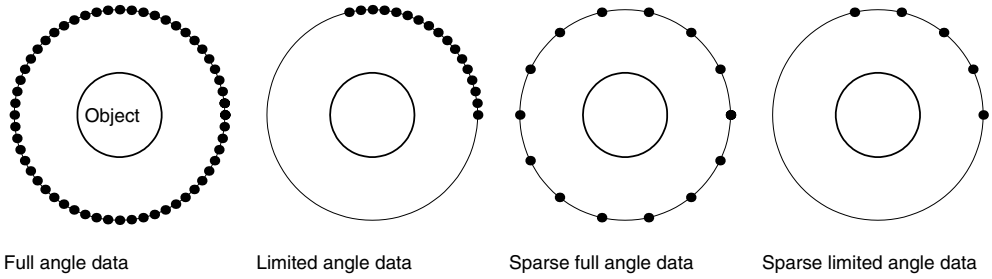

Figure 2. Illustration of four different choices for data collection in transmission tomography. The black dots denote the locations of the x-ray source for the set of projection data. For each such location, the detector is thought to be located opposite to the x-ray source as in figure 1.

we give conclusions. In part II of this paper we apply the general results to practical problems in dental radiology using experimental data.

2. Information content of sparse projection data

Geometrical arrangements of the x-ray source and digital sensor vary according to the diagnostic task and equipment. We illustrate here the types of tomographic data resulting from different imaging situations. For clarity we present two-dimensional examples, but similar situations can be considered in 3D as well.

We separate two cases according to whether the whole object is fully visible in each projection or not, see figure 1. The case on the right in figure 1 is called local tomography.

In traditional CT imaging, projections are taken from all around the object. We sample the angular variable more sparsely in order to lower the radiation dose and due to geometrical limitations, see figure 2. In each of these four cases shown in figure 2, we might additionally have the local tomography situation.

The types of data described above cover a large range of specific imaging tasks. The choice of data collection dictates what kind of features and details we can hope to reconstruct reliably from the data without *a priori* information on the body.

It is well known that reconstruction from the collection of complete projection data is numerically stable (or mildly ill-posed), see e.g. Natterer (1986) and references therein. This

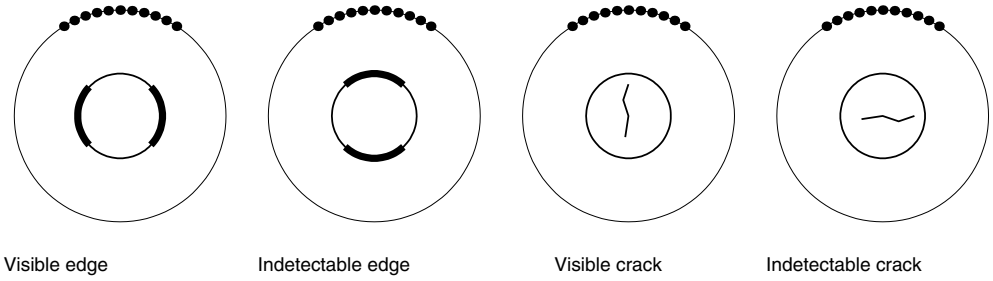

Figure 3. Illustration of discontinuities that can and cannot be recovered based on limited-angle projection data. The black dots denote the locations of the x-ray source in the acquisition of the projection data. For each such location, the detector is thought to be located opposite to the x-ray source as in figure 1.

makes it possible to obtain numerical reconstructions with such algorithms as the filtered backprojection that do not use *a priori* information on the tissue. A careful analysis of collections of x-ray source locations in 3D space giving complete enough projection data for stable recovery has been given by Orlov (1976), Tuy (1986) and Finch (1985). By the term *stable* it is meant here that the reconstruction from such data can be expected to represent the 3D object reliably. This situation falls outside the scope of this paper.

The information content of sparse projection data depends on the type of sparsity. We discuss the effect of limited-angle and local tomography settings and the effect of reducing the number of radiographs.

2.1. Limited-angle tomography

Perfect reconstruction from (an infinite set of) limited-angle tomographic data is possible in principle, as discussed by Smith *et al* (1977) and Natterer (1986). However, the reconstruction problem is extremely ill-posed, or sensitive to measurement noise, as shown by Davison (1983), Louis (1986), Finch (1985) and Tam and Perez-Mendez (1981). Thus, a high quality reconstruction is not possible in practice without *a priori* information on the target.

What features of the target can be reliably reconstructed using only the limited-angle data? A precise answer to this question is given by Quinto (1993). One simple consequence of his results is that a sharp discontinuity, or jump along a curve, is reliably recoverable if and only if some x-ray in some of the projections is tangent to the curve. Otherwise the curve cannot be reconstructed by any algorithm from the projection data alone. We give examples of parts of boundary and cracks that are visible or undetectable in the reconstruction, see figure 3.

In the recent work, Noo *et al* (2002) give conditions for two-dimensional cone beam projections that are needed to reconstruct a region of interest (ROI) within the object. Their result states that a ROI can be recovered accurately if (i) the object is fully visible in each of the projections and (ii) the projection data contain all the possible rays through the ROI.

2.2. Local tomography

In local tomography the region of interest is surrounded by tissue that is not reconstructed. This is often the consequence of small detector size forcing truncation of projections, or intentional minimization of radiation dose outside the region of interest. Local tomography was introduced by Smith and Keinert (1985) and Vainberg *et al* (1981). In this problem class the goal is to reconstruct the region of interest using only x-rays passing through it. It turns

out that the actual attenuation function cannot be reconstructed but, instead, another function preserving sharp features can be recovered. This so-called lambda tomography was refined by Faridani *et al* (1992, 1997).

In the above works on local tomography the 3D body is imaged from full view angle. The combination of local and limited-angle tomography is considered by Kuchment *et al* (1995) and Katsevich (1997). The numerical examples in (Kuchment *et al* 1995) are very illuminating. See also the book by Ramm and Katsevich (1996), especially the images on pp 254–257. The results are similar to limited-angle global tomography: certain parts of singularity curves in the region of interest can be stably reconstructed. The parts are exactly those that have some measured x-rays as tangents.

2.3. Few radiographs

Most of the above results on reconstructable features from limited data assume that the data are available from a curve or other continuum of x-ray source positions. In practice datasets are finite, and the number and directions of radiographs have an effect on the information content of the dataset. As noted by Smith *et al* (1977, theorem 4.2), a finite number of projections tell nothing at all about the volume since an almost arbitrary function can be added to the attenuation coefficient without changing the projections. Information content of projections has been studied further by Logan and Shepp (1975), Grünbaum (1982), Hamaker and Solmon (1978), Kazantsev (1991) and Saksman *et al* (1997). The analysis in the above references implies that problems caused by incomplete information content of a finite dataset can be removed or greatly reduced by using *a priori* knowledge on the 3D body to exclude erroneous (oscillatory) features from the reconstruction.

2.4. Conclusion

The discussion in this section suggests that high-quality tomographic reconstruction from sparse projection data is not possible without the use of *a priori* information. Note also that the above theoretical results concerning the reconstructable features in limited-angle and local tomography do not imply that any given practical tomographic algorithm is able to recover those features.

In this paper we discuss how statistical inversion facilitates a systematic and natural way of incorporating *a priori* knowledge in tomographic reconstruction from sparse projection data, leading to improved reconstruction quality over traditional reconstruction methods.

3. Statistical inversion

The following review papers provide more detailed analysis of several issues in statistical inversion approach than is given in this paper: Hanson (1987), Tamminen (1999), Mosegaard and Sambridge (2002), Evans and Stark (2002) and Kaipio *et al* (2000). See also the work of Lehtinen (1985, 1986) and Hämäläinen *et al* (1987).

3.1. Basic definitions

The success in solving ill-posed inverse problems depends heavily on how well one is able to make use of *a priori* information on the target. Particularly useful is information that is complementary to that extracted from the measurement. Such additional prior information is usually available: the practitioner has often a relatively good overall idea of what a typical

target of the measurement should look like. The actual measurement is needed to distinguish additional specific information from the general. From the computational point of view, this prior information may be rather qualitative in nature and work has to be done to translate it in a computationally useful quantitative form. The *statistical inversion* approach is a systematic and flexible way of incorporating in the inversion process extra information of the target of interest.

The main idea in statistical inversion approach is to consider the inverse problem as a problem of Bayesian inference. All variables are redefined to be random variables. The randomness reflects our uncertainty of their actual values and the degree of uncertainty is coded in the probability distributions of these random variables.

To keep the discussion tractable we consider linear measurement models with additive Gaussian errors:

$$m = Ax + \epsilon \tag{1}$$

where the variables $m \in \mathbb{R}^N$, $x \in \mathbb{R}^M$ and $\epsilon \in \mathbb{R}^N$ are vector-valued random variables[7] and A is the deterministic system matrix modelling the measurement. See section 3.2 for an interpretation of tomographic x-ray measurements in the form of (1). When ϵ is Gaussian with zero mean and covariance matrix Γ_{noise}, denoted $\epsilon \sim \mathcal{N}(0, \Gamma_{\text{noise}})$, we have

$$p_{\text{noise}}(\epsilon) \sim \exp\bigl(-\tfrac{1}{2}\epsilon^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}\epsilon\bigr). \tag{2}$$

Let us mention that the statistical formulation does not depend on the Gaussian approximation. For an exposition of how to handle Poisson distributed observation models in inverse problems, see e.g. (Voutilainen *et al* 2000). Further, observe that if the noiseless model would be badly known, we could also model A by a random matrix.

We assume here that the image vector x and the noise are independent. In terms of probability densities, this implies that their joint probability density is of the form

$$p(x, \epsilon) = p_{\text{pr}}(x)\, p_{\text{noise}}(\epsilon). \tag{3}$$

Here, p_{noise} is the probability distribution of the noise that can be approximated, for example, by analysing x-ray images from well-known phantom targets. The probability density p_{pr} is called the *prior density* of the image. It is designed to contain all possible information that we have of the target *prior to the measurement*. It is crucial, in contrast to classical regularization methods, that the choice of the prior distribution should not be based on the data m. The proper design of the prior is an essential part of the statistical inversion procedure. The rule of thumb is that typical image vectors (say, of some existing library) should have high prior probability (density) while atypical or impossible ones should have low or negligible probability. Prior models are discussed in more detail in section 3.3.

Having the joint probability of x and ϵ, we may write the conditional probability of m, given x and ϵ formally as

$$p(m \mid x, \epsilon) = \delta(m - Ax - \epsilon). \tag{4}$$

Here, δ is the Dirac delta, i.e., if x and ϵ were given, m would be completely determined. The joint probability distribution of x and m is then obtained as

$$p(x, m) = \int_{\mathbb{R}^N} p(m \mid x, \epsilon)\, p(x, \epsilon)\, \mathrm{d}\epsilon = p_{\text{pr}}(x)\, p_{\text{noise}}(m - Ax) \tag{5}$$

by straightforward substitution. Finally, the conditional probability distribution of x given the measurement m, or *posterior density* of x is given by the well-known Bayes' formula

$$p(x \mid m) = \frac{p(x, m)}{p(m)} = \frac{p_{\text{pr}}(x)\, p(m \mid x)}{p(m)} \tag{6}$$

[7] For notational convenience, we use the same lower-case notation for both, the random vector and its values.

where $p(m)$ is the marginal density of m and plays the role of a normalization constant. The density $p(m \mid x)$ is called *the likelihood density* and is in this case

$$p(m \mid x) = p_{\text{noise}}(m - Ax). \tag{7}$$

It turns out that the density $p(m)$ is in non-Gaussian posterior cases actually difficult to determine. Fortunately, it also turns out that the most important estimation methods that are based on the posterior distribution do not necessitate the determination of $p(m)$. Thus the posterior density is usually considered in the non-normalized form

$$p(x \mid m) \propto p_{\text{pr}}(x)\, p_{\text{noise}}(m - Ax) \tag{8}$$

that is, the product of the prior and likelihood densities.

In the framework of Bayesian inversion theory, the posterior distribution (6) represents the complete solution of the inverse problem, since it expresses our beliefs about the distribution of x based on all prior information and the measurement. Because the posterior distribution is a probability density in a large-dimensional space, we must have efficient tools to explore it. To produce an image of the target based on the posterior, several alternatives exist. The most common ones are the *maximum a posteriori* estimate (MAP) and *conditional mean* estimate (CM). They are defined by the formulae

$$p(x_{\text{MAP}} \mid m) = \max p(x \mid m) \tag{9}$$

and

$$x_{\text{CM}} = \int_{\mathbb{R}^M} x p(x \mid m)\, \mathrm{d}x. \tag{10}$$

Observe that the MAP estimate is not necessarily unique, while the CM estimate is unique, provided that the integral converges. Finding the MAP estimate is an optimization problem while finding the CM estimate is a problem of integration.

In addition to computing point estimates, the statistical inversion approach strongly suggests the computation of interval and uncertainty estimates as well as the marginal posterior densities of the variables themselves. An example of an interval estimate is the confidence interval, defined for a given $0 < \tau < 1$ as $[a_k, b_k] \subset \mathbb{R}$ where the endpoints a_k and b_k are determined by

$$\int_{a_k}^{b_k} p_k(x_k)\, \mathrm{d}x_k = \tau \qquad p_k(a_k) = p_k(b_k) \tag{11}$$

where p_k is the marginal posterior density

$$p_k(x_k) = \int_{\mathbb{R}^{M-1}} p(x \mid m)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{k-1}\, \mathrm{d}x_{k+1} \cdots \mathrm{d}x_M. \tag{12}$$

Note that the confidence interval is not always well defined.

The most common uncertainty estimate is the posterior covariance

$$\Gamma_{x|m} = \int_{\mathbb{R}^M} (x - x_{\text{CM}})(x - x_{\text{CM}})^{\mathrm{T}} p(x \mid m)\, \mathrm{d}x \tag{13}$$

In the case of non-Gaussian posterior these estimates can usually be obtained only with the aid of sampling methods. Details of how these can be computed are discussed in section 3.6 and a practical illustration of their significance is given in part II of this paper.

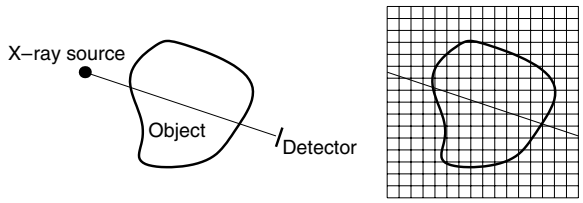

Figure 4. Left: schematic illustration of the pencil beam attenuation model for x-ray transmission tomography. Right: in the discretization the domain Ω under investigation is discretized into a lattice of M pixels Ω_i.

3.2. Likelihood distribution for x-ray imaging

In this section, we discuss in more detail the likelihood model we use for x-ray imaging in this paper.

In x-ray imaging, an almost pointlike x-ray source is placed on one side of an object under imaging. Radiation passes through the object and is detected on the other side, see figure 4. Usually the radiation is detected with x-ray film or a digital sensor that can be thought of as 2D arrays of almost pointlike detectors.

The domain under imaging is modelled by a bounded subset $\Omega \subset \mathbb{R}^3$ (or $\Omega \subset \mathbb{R}^2$ in 2D problems) together with a nonnegative attenuation coefficient $x : \Omega \to [0, \infty)$. The value $x(s)$ gives the relative intensity loss of the x-ray travelling at $s \in \Omega$ within a small distance $\mathrm{d}s$:

$$\frac{\mathrm{d}I}{I} = -x(s)\,\mathrm{d}s.$$

The x-ray has initial intensity I_0 when entering Ω and a smaller intensity I_1 when exiting Ω. We write

$$\int_L x(s)\,\mathrm{d}s = -\int_0^1 \frac{I'(s)\,\mathrm{d}s}{I(s)} = \log I_0 - \log I_1 \tag{14}$$

where I_0 is known by calibration and I_1 from the corresponding point value in a projection image. Thus the measured data are the integral of $x(s)$ along the line L.

The discretization of the attenuation model involves dividing the domain Ω into a lattice of M disjoint 3D voxels Ω_i (or 2D pixels), see figure 4, and measuring the length of the x-ray inside each voxel (or pixel).

Assuming that the attenuation x is constant within each voxel (or pixel) Ω_i, the attenuation map can be approximated in the form

$$x \approx \sum_{i=1}^{M} x_i \chi_i \tag{15}$$

where χ_i is the characteristic function of voxel (or pixel) Ω_i in the lattice. Within the discretization (15), the attenuation map is identified by the coefficient vector $x = (x_1, x_2, \ldots, x_M)^{\mathrm{T}} \in \mathbb{R}^M$. Using the approximation (15), the line integrals through the domain can be approximated by weighted sum of voxel (or pixel) values, that is,

$$\int_{L_j} x(s)\,\mathrm{d}s \approx \sum_{i=1}^{M} x_i |\Omega_i \cap L_j| \tag{16}$$

where the subindex j denotes the measurement index. Arranging the set of projection data into a vector $m = (m_1, m_2, \ldots, m_N)^{\mathrm{T}} \in \mathbb{R}^N$, we get the equation

$$m = Ax \tag{17}$$

where matrix A implements the approximation (16) for the set of projection data. We note that the model (17) can be implemented for any data collection geometry.

In the above model, we neglect scattering phenomena and effects of non-monochromatic radiation, such as beam hardening. See (Sauer *et al* 1994, Antyufeev and Bondarenko 1996) for the former and (Sukovic and Clinthorne 2000) for the latter.

The model (17) is assumed here to represent the noiseless observations. In practice, however, the measurement is corrupted by (at least) two noise types:

- The detector is a photon counter, implying that the attenuated signal at each detector is a count $n_j \in \mathbb{N}$ with expectation $\lambda_j \in \mathbb{N}$, $1 \leqslant j \leqslant N$.
- The photon count of each detector is amplified electronically, causing electronic noise.

The amplification noise of the counting process can reasonably be assumed to be multiplicative. Bearing in mind that the projection data (17) involve a logarithm of the count data, a reasonable model for the electronic noise is additive noise. A feasible model for the count vector n is to assume that each count n_j is independent of the remaining ones and that n_j has Poisson distribution with expectation λ_j. Therefore, we may write

$$\begin{aligned} p(n \mid x) &= \prod_{j=1}^{N} p_{\text{Poisson}}(n_j, \lambda_j) \\ &= \prod_{j=1}^{N} \frac{1}{n_j!} \lambda_j^{n_j} \exp(-\lambda_j), \quad \lambda_j = \lambda_0 I_0 \exp(-(Ax)_j) \end{aligned} \tag{18}$$

where λ_0 is the probability that a photon is absorbed to the detector. However, in x-ray imaging a relatively large number of photons are usually detected at each detector pixel. In such case, the value of the likelihood density (18) can be well approximated with the value of a Gaussian approximation for the attenuation data of form (14) (Bouman and Sauer 1993, Sachs and Sauer 1999). A detailed discussion on this approximation is given in the appendix. An approximation for the distribution of the electronic noise can be determined, for example, by careful analysis of the measurement electronics. However, this falls outside the scope of this paper.

In this study we assume Gaussian distributions for the logarithms of both noise variables. Thus, the overall model for noise becomes the sum of two additive Gaussian random variables. The distribution of this new variable is convolution of two Gaussian distributions, which is again a Gaussian distribution. Therefore, we approximate the statistics of the overall measurement noise by $\epsilon \sim \mathcal{N}(0, \Gamma_{\text{noise}})$, i.e., the measurement noise is assumed to be normally distributed with zero mean and covariance matrix Γ_{noise} that is assumed to be invertible. The likelihood distribution is then

$$p(m \mid x) = \exp\left(-\tfrac{1}{2}(m - Ax)^{\mathrm{T}} \Gamma_{\text{noise}}^{-1} (m - Ax)\right).$$

3.3. Prior models

The most crucial task in statistical inversion is the determination of a feasible prior. In particular, the translation of qualitative prior information into the language of probability densities is often a challenging problem.

The general goal in designing priors is to assign a distribution $p_{\text{pr}}(x)$ with the following property. If E is a collection of *expectable* images and U is a collection of *unexpectable* images, we should have

$$p_{\text{pr}}(x) \gg p_{\text{pr}}(x') \qquad \text{when} \quad x \in E \quad x' \in U.$$

Thus, the prior probability distribution should be concentrated on expectable images and give them a clearly higher prior probability to occur than to those we do not expect to see.

In many cases it turns out that certain expectable features can be formulated in the form of probability densities relatively easily while others may be very tricky if not impossible. A typical example of an easy task is when the image is expected to be smooth. In this case, one can use what is later called a *smoothness prior*. However, if we expect that in addition, the image contains infrequently occurring small anomalies that we would like to locate, the problem becomes more complicated. Indeed, using a smoothness prior we are very likely to fail in detecting the anomaly since its occurrence probability with respect to that prior is extremely low.

In the following we discuss briefly some of the methods for constructing prior models. The emphasis is on finding a qualitative description of these prior models.

In the following, it must be understood that when one talks about an image it is assumed that the pixel values (i.e., values of the x-ray attenuation coefficient) are non-negative. This requirement means that the prior is proportional to the cut-off function,

$$p_{\mathrm{pr}}(x) \propto p_+(x) = \prod_{k=1}^{M} \theta(x_k) \tag{19}$$

where θ is the Heaviside function. This cut-off function is not written out explicitly in the following.

3.3.1. Generic Gaussian priors. The most widely used prior models are the Gaussian *white noise* and *smoothness priors*. Gaussian densities in finite-dimensional spaces are generally of the form

$$p_{\mathrm{pr}}(x) \propto \exp\left(-\tfrac{1}{2}(x - x_*)^{\mathrm{T}} \Gamma_{\mathrm{pr}}^{-1} (x - x_*)\right) \tag{20}$$

where x_* is the mean vector and Γ_{pr} is the covariance matrix. The simplest one is undoubtedly the Gaussian white noise prior,

$$p_{\mathrm{pr}}(x) \propto \exp\left(-\frac{1}{2\sigma^2}\|x - x_*\|_2^2\right) \tag{21}$$

i.e., the covariance is assumed to be diagonal matrix $\Gamma_{\mathrm{pr}} = \sigma^2 I$. This prior is by far the most commonly used *implicit* choice for a prior model in the Tikhonov regularization approach for inverse problems (cf formula (27)). The notion 'white noise' is naturally related to the diagonal covariance structure which means that all image pixels are assumed to be uncorrelated. The standard deviation of each pixel around the assumed mean value x_* is given by σ. No structure of the image is assumed *a priori*.

The smoothness priors in 2D and 3D are typically functions of associated directional derivatives. Perhaps the most commonly used smoothness prior for a 2D pixel image in an equilateral $P \times P = M$ square mesh is given by

$$p_{\mathrm{pr}}(x) \propto \exp\left(-\alpha \sum_{k \in \mathcal{M}} \sum_{\ell \in \mathcal{N}_k} |x_k - x_\ell|^2\right) \tag{22}$$

where α is a scaling parameter, $\mathcal{M}$ is the index set of non-boundary pixels in the lattice (i.e., $\mathcal{M} = \{i \,|\, \partial\Omega_i \cap \partial\Omega = \emptyset\}$) and $\mathcal{N}_k$ is the index set of four nearest pixels for the (non-boundary) pixel k (i.e., $\mathcal{N}_k = \{k-1, k+1, k-P, k+P\}$). The realization of the smoothness prior for higher orders and for non-regular meshes such as arbitrary triangular meshes may turn out to be more tedious. It also depends on the basis in which x is represented, see (Kaipio *et al* 1999) for an example in which x is represented in piecewise linear rather than piecewise constant basis.

A particular class of smoothness priors are *anisotropic priors*. These priors can be viewed as *structural priors* as they reflect structural information on the target. Therefore, we discuss them separately below.

3.3.2. Impulse noise priors. It is easy to generate prior densities based on pixel presentation of the images. It is a less obvious task to describe what sort of qualitative properties they represent as an image. An important class of priors are the *impulse noise priors*. In some applications, we expect to see a low-contrast image with few outstanding pixels as outliers. Such images appear, e.g. in astronomy, where the sky is a nearly black object with bright stars. We mention here three of such priors. These are the L^1-prior

$$p_{\mathrm{pr}}(x) \propto \exp\left(-\alpha \sum_{k=1}^{M} |x_k|\right) = \exp(-\alpha \|x\|_1) \tag{23}$$

the maximum entropy prior

$$p_{\mathrm{pr}}(x) \propto \exp\left(\sum_{k=1}^{M} x_k \log\left(\frac{x_k}{x_0}\right)\right) \tag{24}$$

and the Cauchy distribution prior,

$$p_{\mathrm{pr}}(x) \propto \prod_{k=1}^{M} \frac{1}{1+\lambda x_k^2}. \tag{25}$$

In all these priors, the pixels are uncorrelated. The performance of the maximum entropy prior and the L^1-prior in recovering nearly black objects has been studied in (Donoho *et al* 1992).

3.3.3. Reparametrized priors. In many cases the pixelwise parametrization of the image does not allow easily the coding of the prior information in the form of a prior density. An illustrative example is that we know that the target contains a possibly unknown number of subregions in which the material parameters are constant. We may also know that the boundaries of these subregions are smooth. In this case it is then possible to parametrize the unknown variable for example with the aid of the material parameters inside all subdomains as well as the coefficients of the truncated Fourier series of the boundary curves. For an example in relation to optical tomography, see (Kolehmainen 2001, Kolehmainen *et al* 2000a, 2000b), and in relation to x-ray imaging, see (Hanson *et al* 1997a, 1997b, Mohammad-Djafari and Sauer 1997, Mohammad-Djafari and Soussen 1999).

Also in the case of very fine meshes it may turn out that the covariance matrix becomes almost singular. It is then advisable to consider lower dimensional (subspace) representations of the variable. In addition to making the overall estimation problem smaller dimensional, this may also increase the overall computational stability of the problem, since one often has to work also with the inverse of the covariance matrix.

3.3.4. Sample-based priors. In some cases one has access to a more or less representative *ensemble* of samples/images of the actual variable. One might for example have an ensemble of thoracical topographies of organs based on more extensive measurements (e.g., anatomical atlases), or an ensemble of full-angle reconstructions of teeth. In such cases, it is feasible to assume that the ensemble is distributed according to the prior. The problem is to find a prior that would produce ensembles similar to the one at hand if samples were randomly drawn from the prior. This problem is recognized as a *kernel estimation problem*. As a general reference on kernel estimation methods we give (Thompson and Tapia 1990).

A particularly simple and computationally light special case is when the prior distribution underlying the ensemble can be approximated by a Gaussian distribution. One can then use the ensemble average as the prior mean and a low rank approximation for the inverse of the covariance matrix. In some cases it may also be possible to construct the ensemble artificially, see (Vauhkonen *et al* 1997, 1998) for examples of this approach.

3.3.5. Structural priors. In medical imaging, the overall structure of the target is often well understood based either on anatomical information or on information from other imaging modalities. Different modalities carry information related to different physical parameters such as mass absorbtion coefficient, conductivity and thermal parameters. While there seldom are any simple relationships between these parameters, we can usually expect that certain features are correlated. One such feature is related to detected organ boundaries: it is reasonable to assume that there are jumps with respect to all parameters through the organ boundaries. Hence, if one imaging modality is able to reconstruct an organ boundary, this information could also be used as prior information for other modalities. The procedure how to construct the actual prior density from this information is based on the idea that pixel values of a physical parameter within one type of tissue are strongly correlated while the correlation across the boundaries is low or negligible. In practice, these requirements are met by *anisotropic priors*: different directions close to the boundaries need to be distinguished from each other. These priors are called structural priors. The construction of structural priors is often slightly tedious, starting from the need of powerful image segmentation methods. The good news is that Gaussian structural priors may turn out to be adequate which in turn means that in the case of Gaussian likelihoods most estimates have closed form solutions (Kaipio *et al* 1999).

3.3.6. Anomaly priors. As noted above, one of the main problems in the construction of prior models and densities is that the prior density should be high for such targets that are common or expectable. However, we are often most interested in certain deviations from the usual, cancers being the most obvious examples. This leads to an obvious dichotomy: a good prior distribution for the normal tissue tends to exclude the presence of the anomalous feature and the information could be lost. There is a cure to this problem: it is possible to consider the parameters describing the normal and the anomalous tissues as separate random variables and to assign different prior distributions to them. The statistical inverse problem is then to estimate these two variables separately. Of course, this procedure increases the number of unknowns which may be a computational problem. However, it has the significant virtue that one does not have to discern the cancerous tissue from the normal tissue visually since the estimate for the anomalous part has been computed separately. See (Kaipio and Somersalo 2002) for the treatment of this topic.

3.3.7. Markov random field priors. The smoothness priors discussed above are a special class of Markov random fields (MRF). Let $\mathcal{N} = \{\mathcal{N}_i \mid 1 \leqslant i \leqslant M\}$ denote a neighbourhood system, i.e., $\mathcal{N}_i \subset \{1, 2, \ldots, M\}$ is a list of neighbours of x_i. We require that $i \notin \mathcal{N}_i$ and $i \in \mathcal{N}_j \Leftrightarrow j \in \mathcal{N}_i$. Consider now the conditional probability distribution of a single component x_j. We say that $x = (x_1, \ldots, x_M)^{\mathrm{T}}$ is a MRF with respect to the neighbourhood system $\mathcal{N}$ if it holds that

$$p_{x_j}(x_j \mid x_1, \ldots, x_{j-1}, x_{j+1}, \ldots x_M) = p_{x_j}(x_j \mid x_k, k \in \mathcal{N}_j).$$

In other words, the value of x_j depends on the values of the remaining components only through its neighbours. The probability distributions of MRFs are of a particular form. The

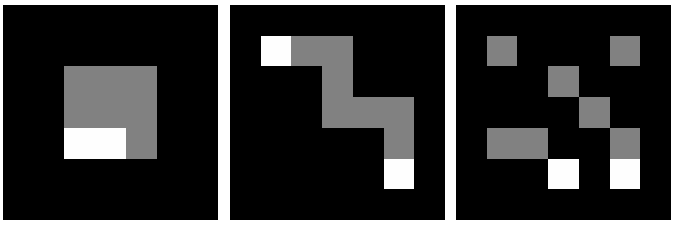

Figure 5. Images having total variations (from left to right) 18, 28 and 40.

well-known Hammersley–Clifford theorem (see e.g. Besag (1974)) states that the probability density of a MRF is necessarily of the form

$$p(x) \propto \exp\left(-\sum_{j=1}^{M} V_j(x)\right)$$

where the functions V_j may depend only on x_j and on those components x_k of x for which $k \in \mathcal{N}_j$. In particular, we have then

$$p_{x_j}(x_j \mid x_1, \ldots, x_{j-1}, x_{j+1}, \ldots x_M) \propto \exp\left(-\sum_{j \in \mathcal{N}_k} V_k(x)\right).$$

A typical example of a MRF prior is the *total variation prior*. Consider again a rectangular pixel image, and define the neighbourhood system in a natural way, i.e., the neighbouring pixels are those that share a common edge. The total variation of the image is defined as

$$\mathrm{TV}(x) = \sum_{i=1}^{M} \sum_{j \in \mathcal{N}_i} l_{ij} |x_i - x_j|$$

where i and j are neighbours with common edge and l_{ij} denotes the length of their common edge (i.e., $l_{ij} = |\partial\Omega_i \cap \partial\Omega_j|$). It turns out that the total variation prior,

$$p_{\mathrm{pr}}(x) \propto \exp(-\alpha \mathrm{TV}(x))$$

is a prior that is concentrated on images that are 'blocky', i.e., the pixels are clustered in blocks with almost equal value and short boundary Dobson and Santosa (1994, 1996). To demonstrate this fact, we have plotted three simple images in figure 5. The values of the pixels are 0 (black), 1 (grey) or 2 (white). The total variations of these images are, from left to right, 18, 28 and 40, respectively. Often, one may also have a relatively good understanding about the distributions of possible values of x within the different blocks in the target. For total variation type MRF priors in such cases, see (Fox and Nicholls 1997, Nicholls and Fox 1998).

The MRF priors are very useful for designing *structural priors*. As an example, consider a medical imaging problem where we know *a priori* (at least approximately) the location of organ boundaries. This information can be based on anatomical data or on data coming from different modalities (CT, MRI, for instance). In medical applications, this information is expressed by saying that we have a *segmented image*, i.e., to each pixel we may assign an index indicating to which tissue type the pixel belongs to. Assume that $T(i) \in \{1, 2, \ldots, K\}$ is the type of the pixel x_i. Now we can define a neighbourhood system $\mathcal{N}$ in such a way that

$$T(j) \neq T(i) \Rightarrow j \notin \mathcal{N}_i$$

i.e., pixels of different type are never neighbours. This means that the pixels are divided into *cliques* according to their type. As a consequence, for a MRF with this type of neighbourhood

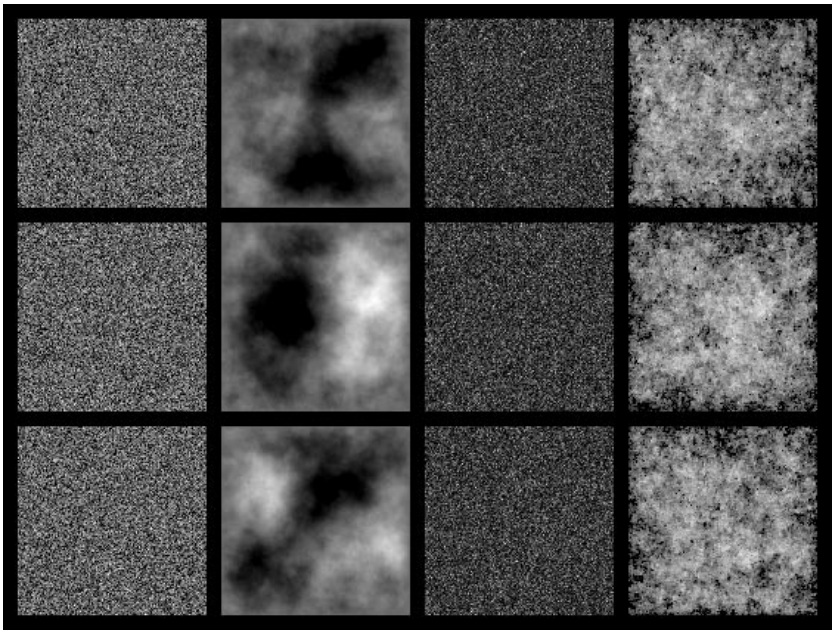

Figure 6. Columns from left to right: three random draws from the Gaussian white noise prior, second-order (isotropic) smoothness prior, L^1-impulse prior and total variation prior, respectively. All priors except the smoothness prior include positivity constraint.

systems, pixels in different cliques are uncorrelated, i.e., the prior allows jumps in the parameter values across the organ boundaries, while within one tissue type, the values may be strongly correlated.

3.4. Visualization of prior models

An enlightening possibility that is very worthwhile but seldom realized, is to draw samples from the prior distribution in order to verify the qualitative nature of the prior. This procedure allows one to get a qualitative understanding of which are actually the features that we assume the target has. The problem in drawing samples from the prior distribution is that the prior is very often not proper (integrable) and the task is impossible as such. However, it is possible to consider certain conditional prior distributions that are proper and can be sampled. With non-Gaussian priors one has to resort to Markov chain Monte Carlo (MCMC) methods for drawing these samples, see section 3.6. In figure 6 random draws from the Gaussian white noise prior, second-order (isotropic) smoothness prior, L^1-impulse prior and total variation prior are shown. For the smoothness and total variation priors the samples are from conditional priors such that the values of the boundary pixels were fixed to zero. The samples that are shown in figure 6 are the three most probable realizations from the set of the samples drawn from each prior.

In addition to drawing samples from the prior, it is also advisable to visualize the prior covariance. In Gaussian prior cases we naturally have to construct the covariance anyway but in non-Gaussian cases the covariance can again be estimated based on MCMC runs. Non-proper priors can be handled as in the case of drawing samples from the priors. The visualization is carried out simply by fixing a pixel, say x_j, and plotting the covariances

$\mathrm{cov}(x_1, x_j), \ldots, \mathrm{cov}(x_M, x_j)$ in the same mesh as the target itself, see (Kaipio *et al* 1999) for details and examples.

We also note that most interesting and feasible priors are nonintegrable or uninformative in the sense that they do not represent proper probability distributions. This means among other things that their covariance matrices may not exist, which again loosely speaking means that the variances of x in certain directions are infinite. More rigidly, let $v \in \mathbb{R}^M$, $\|v\| = 1$. It is possible that the scalar-valued random variable $x_v = (x^{\mathrm{T}} v)$ has infinite variance.

3.5. Computation of the MAP estimate

Unless the noise and prior densities are Gaussian, finding the MAP estimate requires iterative optimization methods. If the mapping $x \mapsto p(x \mid m)$ is smooth, Newton–Raphson type methods can be used in mildly large-dimensional cases, while very large-dimensional cases typically require gradient-based methods. Observe that when the prior probability density is of the Gibbs type,

$$p_{\mathrm{pr}}(x) \sim \exp(-\alpha G(x)) \tag{26}$$

the search for the MAP estimate in the case of Gaussian additive noise is tantamount to the minimization problem

$$x_{\mathrm{MAP}} = \arg\min\left(\tfrac{1}{2}(m - Ax)^{\mathrm{T}} \Gamma_{\mathrm{noise}}^{-1}(m - Ax) + \alpha G(x)\right) \tag{27}$$

i.e., a regularized weighted output least squares problem. We see that in the language of traditional inverse problem literature, the function G, called *potential* or *penalty function*, plays the role of Tikhonov regularization functional.

3.6. Computation of the CM estimate

Analytic evaluation of the conditional mean is practically impossible unless the posterior is Gaussian. When the dimensionality of the problem is high, no traditional quadrature rule for computing the integral in (10) is feasible simply because the number of the quadrature points would be unreasonably large. Furthermore, the factor $p(m)$ in (6) should also be computed by integration. The evaluation of this factor is a problem of the same magnitude as the evaluation of the conditional mean itself. Consequently, the integration requires the use of Monte Carlo integration methods.

Markov chain Monte Carlo (MCMC) methods provide a tool for the approximation of high-dimensional posterior integrals. MCMC methods create a set of random samples $\{x^{(t)} \in \mathbb{R}^M\}$ from the density $p(x \mid m)$ and approximate the conditional mean with the sample mean Gilks et al 1996. The key idea in the sample generation is to create a Markov process which has $p(x \mid m)$ as the stationary distribution. Once the simulation of the process has been run long enough, say r iterations, the process will produce a set of samples $\{x^{(t)} \mid t = r + 1, \ldots, S\}$ whose distribution is close to $p(x \mid m)$. The most popular algorithms for the sample generation are the Gibbs sampler and the Metropolis–Hastings algorithm (Gilks *et al* 1996, Hastings 1970, Geman and Geman 1984).

3.6.1. Metropolis–Hastings algorithm. In the Metropolis–Hastings algorithm the states of the Markov chain are generated as follows: given the state $x^{(t)}$ of the chain, a candidate x_{c} for the next state is drawn from the proposal density $q(x_{\mathrm{c}}|x^{(t)})$. Loosely speaking, $q(x_{\mathrm{c}}|x^{(t)})$ is the probability of the move from $x^{(t)}$ to x_{c}. The candidate is not accepted automatically. To understand the acceptance rule, assume first that the proposal density is symmetric, i.e., $q(x|y) = q(y|x)$ for all x, y. It can be interpreted by saying that the probability for moving

from y to x equals the probability of moving from x to y. In this particular case, the acceptance rule is simple: if the proposed state x_c has higher probability than the previous state $x^{(t)}$, the candidate is automatically accepted. However, if it has a lower probability, it is accepted only by a probability that is proportional to the ratio of the probabilities. Hence, the acceptance probability γ of the candidate is simply

$$\gamma = \min\left\{1, \frac{p(x_c \mid m)}{p(x^{(t)} \mid m)}\right\}. \tag{28}$$

If $q(x|y) \neq q(y|x)$, a modification of (28) is needed to compensate the asymmetry:

$$\gamma = \min\left\{1, \frac{p(x_c \mid m)q(x^{(t)} \mid x_c)}{p(x^{(t)} \mid m)q(x_c \mid x^{(t)})}\right\}. \tag{29}$$

If the candidate is accepted, the next state is $x^{(t+1)} = x_c$. Otherwise, $x^{(t+1)} = x^{(t)}$. The distribution of the samples converges asymptotically to $p(x \mid m)$. In practice the acceptance is carried out so that one first draws a sample from the proposal distribution and computes γ. Then a random number from the uniform distribution Uni(0, 1) is drawn and compared with γ.

As it can be seen from equations (28) and (29), the normalization constant $p(m)$ is cancelled out in the computation of the acceptance probability and therefore it is sufficient to know the posterior density up to the normalization constant only. This is a very important feature since the computation of $p(m)$ is a formidable task.

If an arbitrary initial draw $x^{(0)}$ is used, the first subsequent draws do not describe the posterior distribution. Therefore a *burn-in period* $\{x^{(t)}, t = 0, \ldots, r\}$ of the generated Markov Chain is usually discarded in the computation of the conditional mean, and thus, the conditional mean is obtained as

$$\hat{x}_{\mathrm{CM}} \approx \frac{1}{S-r} \sum_{i=r+1}^{S} x^{(i)}. \tag{30}$$

On the other hand, if the MAP estimate has been computed and used as the initial draw, there is no burn-in and the first samples can be used. Correspondingly, an estimate for the posterior covariance can be obtained as

$$\hat{\Gamma}_{\mathrm{CM}} \approx \frac{1}{S-r-1} \sum_{i=r+1}^{S} (x^{(i)} - \hat{x}_{\mathrm{CM}})(x^{(i)} - \hat{x}_{\mathrm{CM}})^{\mathrm{T}}. \tag{31}$$

The convergence of the simulation can be monitored, e.g. by computing confidence intervals for the point estimates by using multiple parallel simulations with different starting points $x^{(0)}$ (30). For more detailed information about the MCMC methods and the convergence properties, see e.g. (Gilks *et al* 1996, Hastings 1970, Gamerman 1997, Smith and Roberts 1993, Tierney 1994, Roberts and Smith 1994). The marginal posterior densities of individual components x_k are obtained simply by using any kernel estimation method for the scalar draws $\left\{x_k^{(t)}, t = r+1, \ldots, S\right\}$.

The key problem in the Metropolis–Hastings method is to find effective proposal distribution. This is especially crucial in the case of large-dimensional problems. If the proposal distribution is not feasible, $\gamma \approx 0$ for almost all draws and very few of the candidates get accepted. On the other hand, the proposal distribution has to be one from which we can perform the draws. In many cases a Gaussian approximation is a feasible choice and at least the one which is tried first. Write the posterior in the form $p(x \mid m) \propto \exp(-g(x \mid m))$. In the case of twice differentiable g, the Gaussian approximation for the posterior density can be

written at point x_* by determining a Taylor's approximation for the exponent term $g(x \mid m)$ of the posterior density

$$g(x \mid m) \approx g(x_* \mid m) + \nabla g(x_* \mid m)(x - x_*) + \tfrac{1}{2}(x - x_*)^{\mathrm{T}} \nabla^2 g(x_* \mid m)(x - x_*). \tag{32}$$

Typically, the MAP estimate is used as the linearization point x_*. Thus, an approximative Gaussian distribution is obtained with covariance

$$\Gamma_{\mathrm{A}} = (\nabla^2 g(x_* \mid m))^{-1}. \tag{33}$$

In many cases the proposal density associated with the exponent (32) is not used as such. Instead, in the proposal density the previous state is used as the expectation while the covariance (33) is retained. A procedure of this kind is often referred to as the random-walk Metropolis algorithm (Gilks *et al* 1996). Thus the proposal density takes the form

$$q(x_{\mathrm{c}} \mid x^{(t)}) = \mathcal{N}(x^{(t)}, \beta\Gamma_{\mathrm{A}}). \tag{34}$$

In equation (34), β is a scaling factor which is used for tuning the mixing properties of the algorithm. The tuning parameter β has to be chosen carefully. If it is too large, almost all candidates x_{c} will be rejected. On the other hand, if β is too small almost all candidates x_{c} will be accepted but the algorithm is again very inefficient due to the fact that the jumps become very small. The draws are very correlated in both extreme cases.

One possible way of choosing the value of β is to monitor the empirical acceptance rate. There is theoretical justification for aiming at acceptance rates between $[0.15, 0.5]$, for more details see (Gilks *et al* 1996) and references therein.

In the case of a Gaussian proposal density the candidates x_{c} can be drawn as follows. Let $\beta\Gamma_{\mathrm{A}} = USU^{\mathrm{T}}$ denote the singular value decomposition of the covariance matrix of the proposal distribution (34). If we now define a multivariate random variable $\varphi \sim \mathcal{N}(0, I)$, we can generate the candidates using the formula

$$x_{\mathrm{c}} = U\sqrt{S}\varphi + x^{(t)}. \tag{35}$$

A visualization of the convergence of Metropolis–Hastings algorithm when $x \in \mathbb{R}^2$ is given in figure 7. Note the relevance of the adjustment of the scaling factor γ. The acceptance rate is in the range $[0.15, 0.5]$ for cases (c) and (d).

3.6.2. Gibbs sampler. Together with the Metropolis–Hastings algorithm and its variants, Gibbs sampler constitutes a popular method of generating samples with a prescribed distribution (Geman and Geman 1984). Rather than using a proposal distribution as in the Metropolis–Hastings scheme, one uses directly the posterior distribution but in a sequential manner. The so-called *single component Gibbs sampler* proceeds as follows:

(i) Fix the initial draw $x^{(0)} = \big(x_1^{(0)}, \ldots, x_M^{(0)}\big)^{\mathrm{T}}$ and set $j = 1$.
(ii) Generate $x^{(j)}$ a single variable at a time:

$$\begin{aligned}
&\text{Draw } x_1^{(j)} \text{ from the density } t \mapsto p\big(t, x_2^{(j-1)}, \ldots, x_M^{(j-1)} \,\big|\, m\big)\\
&\text{draw } x_2^{(j)} \text{ from the density } t \mapsto p\big(x_1^{(j)}, t, x_3^{(j-1)}, \ldots, x_M^{(j-1)} \,\big|\, m\big)\\
&\qquad\vdots\\
&\text{draw } x_M^{(j)} \text{ from the density } t \mapsto p\big(x_1^{(j)}, \ldots, x_{M-1}^{(j)}, t \,\big|\, m\big).
\end{aligned}$$

(iii) Set $j \leftarrow j + 1$ and go to (ii).

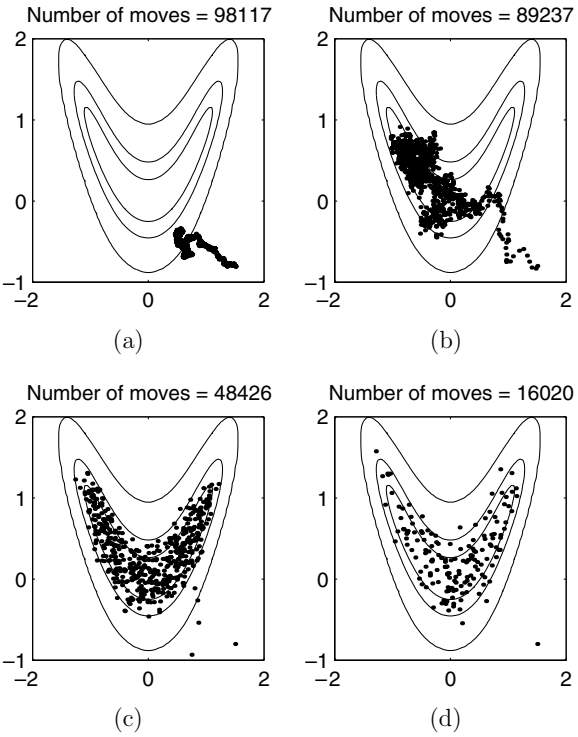

Figure 7. Samples generated by the random walk Metropolis–Hastings algorithm. The boomerang-shaped lines are equipotential lines of the posterior distribution. The initial estimate $x^{(0)}$ is in the lower right-hand corner of the figures. The scaling factors γ were (a) $\gamma = 0.01$, (b) $\gamma = 0.05$, (c) $\gamma = 0.3$ and (d) $\gamma = 1$. 'Number of moves' refers to how many of the 100 000 samples (candidates) have been accepted. Only the accepted samples from the first 1000 candidates are shown.

The determination of the so-called full conditional densities of a single component x_k while the remaining ones are fixed can in some cases be carried out analytically, at least with respect to some variables, but since they are functions of a single variable only, it is relatively straightforward to approximate the associated distribution functions non-parametrically and then employ the well-known golden rule to draw the samples. Compared to the Metropolis–Hastings method, the virtue of the Gibbs sampler is the absence of the problems related to the choice of the proposal distribution as well as questions related to the acceptance rule. The significant drawback is that it becomes easily slow when the number of the components is large as it is in real x-ray tomography problems.

A visualization of componentwise updatings with the density of the previous example is given in figure 8.

4. X-ray tomography from sparse projection data

4.1. Applications to medical imaging

Three-dimensional information is crucial for many radiological tasks. Some examples are

- A dentist wants to know whether the roots of a certain tooth are close to the inferior dental canal (a tunnel for nerves inside the jaw bone). Answer to this question is not necessarily

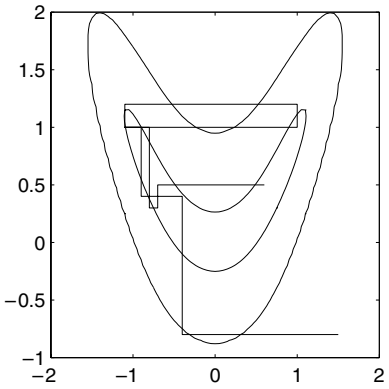

Figure 8. Illustration of the componentwise sampling with the Gibbs sampler.

present in a single intraoral radiograph due to overlapping of structures. In such case, the dentist can use his regular equipment to take a couple of extra images from such angles that the roots and the canal have clear disparity. Tomographic reconstruction from these images is likely to yield the desired information.

- A radiologist suspects breast cancer after examining a screening mammogram. In further study of the finding, seven radiographs are taken of the breast with an opening angle of $\pm 15^\circ$. The reconstructed volume shows three-dimensional distribution of microcalcifications, which helps in deciding whether to perform biopsy or not.
- A surgeon is mounting a screw to bone and wants to check whether the tip of the screw is already close to the surface of the bone. A C-arm unit is used to take an x-ray image. However, the direction of this image has to be carefully chosen to capture the desired information. Instead, several images can be taken from a collection of directions bracketed around the most probably correct direction, making sure that the position of the screw can be reliably seen in the three-dimensional reconstruction.

The above procedures are examples of a novel three-dimensional medical imaging modality. It is intermediate between studying a single radiograph and a CT scan. The radiation dose of the novel modality is equal to the dose of a few radiographs, much lower than the dose of CT. Also, the cost and resources required by the imaging equipment are considerably smaller than those of a CT scanner. Such low dose imaging modalities with applications to dental radiology and mammography have previously been suggested by Webber *et al* (Webber 1998, Webber *et al* 1997, Ramesh *et al* 2002) and McCauley *et al* (2000).

As mentioned in the introduction, three-dimensional reconstruction from sparse projection data is not as straightforward as from the comprehensive dataset provided by CT. In section 4.2 we discuss tuned aperture computed tomography (TACT), a traditional approach using tomosynthesis, or unfiltered backprojection, as reconstruction method. In section 4.3 we discuss how to use the statistical techniques of section 3 to improve the image quality of TACT.

4.2. Tuned aperture computed tomography (TACT)

The TACT method was suggested by Webber (1998). In TACT imaging the directions of the projection images are calibrated from a reference object in the image. The reconstruction

method in TACT is tomosynthesis (unfiltered backprojection). Tomosynthesis was first presented by Ziedses des Plantes (1932) and later promoted by Grant (1972).

The quality of tomosynthetic reconstructions suffer from the blur that is primarily due to the overlapping of structures in the reconstruction due to the poor information content of the limited-angle projection data in the depth direction. Several deblurring methods have been suggested. Ruttimann *et al* (1984) present an iterative enhancement scheme, and Persons (2001) uses Fourier transform methods.

Despite these limitations, TACT imaging has been found useful for many clinical applications, see e.g. (Webber et al 1997, 1996, Webber and Messura 1999, Ramesh *et al* 2002, Lehtimäki *et al* 2003). We note that the proofs of many theorems about the structures that are reconstructable from sparse projection data are based on the analysis of unfiltered backprojection and ideal noise free data. Thus, the analysis in section 2.1 explains the limits for what one can expect to see in TACT reconstructions. So choosing the projection directions is crucial for successful TACT imaging.

4.3. Properties of statistical reconstruction

Statistical inversion can improve the quality of reconstruction over traditional methods (such as TACT) with careful modelling of *a priori* information. It is in principle possible to recover even such features of the target that are not stably reconstructable in the sense of section 2.

Reconstruction algorithms based on statistical inversion are modular because the prior information and the measurement model are separate:

- If different types of tissue are imaged with the same measurement setting (same x-ray device and same projection directions), only the prior distribution needs to be changed in the computation.
- If the measurement setup changes, only the likelihood distribution needs to be changed. This change is accomplished by (i) computing matrix A in equation (17) for the new geometry and (ii) approximating the noise characteristics for the new imaging equipment or imaging parameters.

Statistical inversion can provide more information than just reconstruction. In addition to presenting a single estimate (typically MAP estimate) as final image of the target, statistical inversion gives natural means to obtain more extensive information on the solution by computing different statistics and confidence limits from the posterior. In the case of non-Gaussian posteriors this can be done by sampling the posterior using MCMC methods. Statistical inference from the posterior is conducted using MCMC methods by Hanson *et al* (1997a, 1997b). They verify the prior for the deformable boundaries by samples from the prior in (Hanson *et al* 1997b). In part II of this paper we show marginal distributions and confidence intervals for reconstructed pixel values.

Statistical inversion methods are computationally much more demanding than traditional methods such as TACT or filtered backprojection. Applying statistical inversion to realistic three-dimensional medical imaging leads to large-scale optimization or integration problems. Thus, powerful computers and efficient numerical algorithms are required to reduce reconstruction time to clinically acceptable limits.

In part II of this paper we show how statistical inversion with a feasible prior can improve reconstruction quality over tomosynthesis and other traditional reconstruction methods in dental imaging. We acquire our tomographic data with a commercial medical x-ray device using *in vitro* human phantoms. The examples are computed with realistic resolution.

4.4. Review of statistical inversion approaches to tomography from sparse projection data

In this section we review earlier work on statistical (Bayesian) inversion for reconstruction from sparse projection data.

- Persson *et al* (2001) and Delaney and Bresler (1998) compute MAP estimates with total variation prior using simulated limited-angle projection data. Improved image quality is reported in comparison with filtered backprojection.
- Hanson and his collaborators have worked on Bayesian methods for sparse data transmission tomography for a long time. In the pioneering 1983 work Hanson and Wecksung (1983) applied Bayesian inversion to limited-angle tomography using a Gaussian prior. In later works they used shape parametrization of the problem together with Bayesian inversion methods (Hanson *et al* 1997a, 1997b, Battle *et al* 1997, 1998).
- Shape parametrization and Bayesian approach to limited-angle problem have also been studied by Mohammad-Djafari *et al*, see (Soussen and Mohammad-Djafari 1999, Mohammed-Djafari and Sauer 1997, Mohammad-Djafari and Soussen 1999). See also the discussion about prior models for x-ray imaging in (Mohammad-Djafari and Sauer 1997).
- Sauer *et al* (1994) consider the Bayesian approach to transmission tomography in nondestructive testing. Their data consist of few radiographs collected from full angle, and they compare Markov random field (MRF) priors with different smoothness properties. They also consider binary MRF priors (i.e., possible values of attenuation are known *a priori*). MAP estimates are computed. This work is extended to real measurement data (and modelling of uncertain source locations) in Sachs and Sauer (1999). In (Bouman and Sauer 1996) they present an iterative pixelwise coordinate descent algorithm for the computation of the MAP estimate. See also (Bouman and Sauer 1993). Similar work has been conducted by Hsiao *et al* (2001) with simulated non-sparse data.
- Nygrén *et al* (1997) apply Bayesian approach to ionospheric radiotomography. They use Gaussian priors for reconstructions with simulated and real data. This work is the continuation of the work of Lehtinen (Lehtinen 1985, 1986, Hämäläinen *et al* 1987).
- Anisotropic total variation regularization has been considered by Bleuet *et al* (2002). This approach cannot be interpreted as Bayesian in the strict sense since the regularizing penalty functional depends on the measurement geometry, whereas prior distributions should be independent of the measurements.
- Yu and Fessler (2002) consider structural MRF priors. In their (conditional) MRF prior, the weighting field for the jumps between neighbouring pixels is based on level-set representation of the organ boundaries in the current estimate. Their model contains a hyperprior that prefers relatively short and smooth organ boundaries. A minimization scheme in which the pixel values and organ boundaries are updated in an alternating manner is used for the computation of MAP estimates.
- Bayesian methods have been used widely and successfully in emission tomography modalities (SPECT and PET), see e.g. (Cunningham et al 2998, Chan et al 1997, Aykroyd and Zimeras 1999, Hebert and Leahy 1989, Liang *et al* 1989, Green 1990, Hebert and Gopal 1992, Lalush and Tsui 1992, Gindi *et al* 1993, Mumcuoglu *et al* 1994, Lee *et al* 1995, Bowsher *et al* 1996). Typically, Markov random field (MRF) priors are used as a prior model for the image and MAP estimates are computed as final images from the posterior. In most cases Poisson model is used for the statistics of the emission data. This can give significant improvement over Gaussian noise models when emission data contain very low counts, which is often the case, especially in PET. Many of the articles give detailed discussions on the optimization routines that are used to compute

MAP estimates. The resolution of the reconstructions in SPECT and PET are not expected to be as high as in x-ray tomography. Thus, the numbers of unknowns for these problems remain much lower than in x-ray imaging.

5. Conclusions

In this paper, the application of statistical inversion methods to x-ray tomography with sparse projection data was reviewed and discussed. The main focus was given to the theory of statistical inversion and to the prior models. The discussion included an example on the analysis and visualization of the prior models. We also discussed the likelihood model we use for x-ray tomography and the computation of the point estimates from the posterior. A review of mathematical results on the information content of sparse projection data was also given. Based on the presented theory, a novel statistical model for three-dimensional imaging of dentomaxillofacial structures is proposed in part II of this paper. Results are given with experimental projection data from *in vitro* dental phantoms.

Acknowledgments

This work was supported by the National Technology Agency of Finland (TEKES, contracts 40202/01 and 40288/02), the Saastamoinen Foundation and the Finnish Academy of Science and Letters. We also wish to thank the Finnish IT Center for Science (CSC) for providing computational resources. S Siltanen thanks Richard Webber for inspiring discussions.

Appendix. Gaussian approximation for the likelihood

Let n_i^{meas} denote the value of the detected photon count for the ith observation in the set of projection data. For it we define the value of the attenuation measurement

$$m_i = -\log \frac{n_i^{\text{meas}}}{\lambda_0 I_0} \tag{A.1}$$

modelling the fact that

$$I_i = I_0 \exp(-m_{0,i}) \qquad m_{0,i} = (Ax)_i$$

is the error-less intensity measured on the detector pixel. The photon counts n_i are assumed to be independent and have distributions $n_i \sim \text{Poisson}(\lambda_i)$, $\lambda_i = \lambda_0 I_i$. This models the fact that each photon is absorbed to the detector with probability λ_0.

Next, we would like to approximate the observation errors with a Gaussian distribution. Unfortunately, since the attenuation data is a logarithm of a Poisson distributed quantity this kind of approximation cannot be justified directly. Instead of that, we will consider in detail the likelihood function, that is, the probability mass function of $n = (n_1, \ldots, n_N)^{\mathrm{T}}$ with condition that x is given. At point $u = (u_1, \ldots, u_N)^{\mathrm{T}}$ it is

$$\begin{aligned} p(n = u \mid x) &= \prod_{i=1}^{N} \mathrm{e}^{-\lambda_i} \frac{\lambda_i^{u_i}}{u_i!} \\ &= \prod_{i=1}^{N} \exp[u_i \log(\lambda_0 I_0 \exp(-m_{0,i})) - \lambda_0 I_0 \exp(-m_{0,i}) - \log(u_i!)]. \end{aligned}$$

where $m_{0,i} = (Ax)_i$. Taking logarithms, we get

$$\log p(n = u \mid x) = \sum_{i=1}^{N} [-u_i m_{0,i} - \lambda_0 I_0 \exp(-m_{0,i})] + c(u) \tag{A.2}$$

where the constant $c(u)$ is independent of x. Following Sauer and Bouman (1993), we write

$$\log p(n = u \mid x) = c(u) + \sum_{i=1}^{N} h(m_{0,i}, u_i)$$

where function h is of the form

$$h(t, u_i) = -u_i t - \lambda_0 I_0 \exp(-t).$$

Now, we write Taylor approximations for functions $t \mapsto h(t, u_i)$ at point $\hat{t}_i = \hat{t}_i(u_i)$ where $\partial_t h(\hat{t}_i, u_i) = 0$. This gives

$$\log p(n = u \mid x) \approx c(u) + \sum_{i=1}^{N} h(\hat{t}_i) + \frac{1}{2} \sum_{i=1}^{N} \partial_t^2 h(\hat{t}_i)(m_{0,i} - \hat{t}_i)^2. \tag{A.3}$$

Solving equation $\partial_t h(\hat{t}_i, u_i) = 0$ yields

$$\hat{t}_i(u_i) = -\log \frac{u_i}{\lambda_0 I_0} \qquad \partial_t^2 h(\hat{t}_i, u_i) = -u_i. \tag{A.4}$$

Denoting $\hat{t}(u) = (\hat{t}_1(u_1), \ldots, \hat{t}_N(u_N))^{\mathrm{T}}$ and recalling $m_{0,i} = (Ax)_i$ we obtain

$$p(n = u \mid x) \approx e^{c_1(u)} \exp\left(-\tfrac{1}{2}(\hat{t}(u) - Ax)^{\mathrm{T}} D(u)(\hat{t}(u) - Ax)\right) \tag{A.5}$$

where $c_1(u)$ is a constant and $D(u)$ is a diagonal matrix with elements $D_{ii}(u) = u_i$.

Now, if the values of u_i are the measured photon counts n_i^{meas} and

$$m_i = -\log \frac{n_i^{\mathrm{meas}}}{\lambda_0 I_0}$$

we see that $\hat{t}\left(n_i^{\mathrm{meas}}\right) = m_i$. Thus

$$p(n = n^{\mathrm{meas}} \mid x) \approx \mathrm{e}^{c_1(n^{\mathrm{meas}})} \exp\left(-\tfrac{1}{2}(m - Ax)^{\mathrm{T}} D(n^{\mathrm{meas}})(m - Ax)\right). \tag{A.6}$$

Thus the value of the likelihood function $p(n = n^{\mathrm{meas}} \mid x)$ is approximately the same as some function $c_2(n^{\mathrm{meas}})$, depending only on n^{meas}, multiplied with the likelihood function that we would obtain using an additive Gaussian error model for m_0 with the noise covariance matrix $D^{-1}(n^{\mathrm{meas}})$. When the measured data n^{meas} are given, the function $c_2(n^{\mathrm{meas}})$ can be considered as a normalization constant that does not affect computation of CM or MAP estimates. Summarizing, we have shown that even though Gaussian model may not be the correct model for the logarithm of count data, we obtain right CM and MAP estimates using Gaussian error model in our computations. In practice, the matrix $D(n^{\mathrm{meas}})$ can be approximated by $D(n^{\mathrm{meas}}) \approx \sigma^{-2} I$ whenever the photon counts at each detector pixel are approximately of the same magnitude.

References

Antyufeev V S and Bondarenko A N 1996 X-ray tomography in scattering media *SIAM J. Appl. Math.* **56** 573–87

Aykroyd R G and Zimeras S 1999 Inhomogeneous prior models for image reconstruction *J. Am. Stat. Assoc.* **94** 934–46

Battle X L, Cunningham G S and Hanson K M 1997 3D tomographic reconstruction using geometrical models *Proc. SPIE* **3034** 346–57

Battle X L, Hanson K M and Cunningham G S 1998 Tomographic reconstruction using 3D deformable models *Phys. Med. Biol.* **43** 983–90

Besag J 1974 Spatial interaction and the statistical analysis of lattice systems *J. R. Stat. Soc.* **36** 192–236

Bleuet P, Guillemaud R and Magnin I E 2002 Resolution improvement in linear tomosynthesis with an adapted 3D regularization scheme *Proc. SPIE* **4682** 117–25

Bouman C and Sauer K 1993 A generalized Gaussian image model for edge-preserving MAP estimation *IEEE Trans. Image Process.* **2** 296–310

Bouman C and Sauer K 1996 A unified approach to statistical tomography using coordinate descent optimization *IEEE Trans. Image Process.* **5** 480–92

Bowsher J E, Johnson V E, Turkington T G, Jaszczak R J, Floyd C E Jr and Coleman R E 1996 Bayesian reconstruction and use of anatomical *a priori* information for emission tomography *IEEE Trans. Med. Imaging* **15** 673–86

Chan M T, Herman G T and Levitan E 1997 A Bayesian approach to PET reconstruction using image-modeling Gibbs priors: implementation and comparison *IEEE Trans. Nucl. Sci.* **44** 1347–54

Cormack A M 1963 Representation of a function by its line integrals, with some radiological applications I *J. Appl. Phys.* **34** 2722–7

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

Cunningham G S, Hanson K M and Battle X L 1998 Three-dimensional reconstructions from low-count SPECT data using deformable models *Opt. Express* **2** 227–36

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

Delaney A H and Bresler Y 1998 Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography *IEEE Trans. Image Process.* **7** 204–21

Dobson D C and Santosa F 1994 An image enhancement technique for electrical impedance tomography *Inverse Problems* **10** 317–34

Dobson D C and Santosa F 1996 Recovery of blocky images from noisy and blurred data *SIAM J. Appl. Math.* **56** 1181–98

Donoho D L, Johnstone I M, Hoch J C and Stern A S 1992 Maximum entropy and the near black object *J. R. Stat. Soc.* B **54** 41–81

Evans S N and Stark P B 2002 Inverse problems as statistics *Inverse Problems* **18** R55–97

Faridani A, Finch D V, Ritman E L and Smith K T 1997 Local tomography II *SIAM J. Appl. Math.* **57** 1095–127

Faridani A, Ritman E L and Smith K T 1992 Local tomography *SIAM J. Appl. Math.* **52** 459–84

Faridani A, Ritman E L and Smith K T 1992 Examples of local tomography *SIAM J. Appl. Math.* **52** 1193–8 (a reorganization of the examples which became disorganized while the above article was in press)

Finch D 1985 Cone beam reconstruction with sources on a curve *SIAM J. Appl. Math.* **45** 665–73

Fox C and Nicholls G 1997 Sampling conductivity images via MCMC *The Art and Science of Bayesian Image Analysis: Proc. of the Leeds Annual Statistics Research Workshop (Leeds1–4 July 1997)* ed K V Mardia, C A Gill and R G Aykroyd (Leeds: Leeds University Press) pp 91–100

Gamerman D 1997 *Markov chain Monte Carlo—Stochastic Simulation for Bayesian Inference* (London: Chapman and Hall)

Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41

Gilks W R, Richardson S and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice* (London: Chapman and Hall)

Gindi G, Lee M, Rangarajan A and Zubal I G 1993 Bayesian reconstruction of functional images using anatomical information as priors *IEEE Trans. Med. Imaging* **12** 670–80

Grant D G 1972 Tomosynthesis: a three-dimensional radiographic imaging technique *IEEE Trans. Biomed. Imaging* **19** 20–8

Green P J 1990 Bayesian reconstructions from emission tomography data using a modified EM algorithm *IEEE Trans. Med. Imaging* **9** 84–93

Grünbaum F A 1982 The limited angle reconstruction problem *Proc. Symp. Appl. Math.* **27** 43–61

Hamaker C and Solmon D C 1978 The angles between the null spaces of X rays *J. Math. Anal. Appl.* **62** 1–23

Hämäläinen M, Haario H and Lehtinen M 1987 Inferences about the sources of neuromagnetic fields using Bayesian statistics *Report* TKK-F-A260, TKK offset, Helsinki University of Technology, Low Temperature Laboratory, Helsinki

Hanson K M 1987 Bayesian and related methods in image reconstruction from incomplete data *Image Recovery: Theory and Applications* ed H Stark (Orlando: Academic) pp 79–125

Hanson K M, Cunningham G S and McKee R J 1997a Uncertainty assessment for reconstructions based on deformable geometry *Int. J. Imaging Syst. Technol.* **8** 506–12

Hanson K M, Cunningham G S and McKee R J 1997b Uncertainties in tomographic reconstructions based on deformable models *Proc. SPIE* **3034** 276–86

Hanson K M and Wecksung G W 1983 Bayesian approach to limited-angle reconstruction in computed tomography *J. Opt. Soc. Am.* **73** 1501–9

Hastings W K 1970 Monte Carlo sampling methods using Markov Chains and their applications *Biometrika* **57** 97–109

Hebert T J and Gopal S S 1992 The GEM MAP algorithm with 3-D SPECT system response *IEEE Trans. Med. Imaging* **11** 81–90

Hebert T and Leahy R 1989 A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors *IEEE Trans. Med. Imaging* **8** 194–202

Hsiao I-T, Rangarajan A and Gindi G 2001 Bayesian image reconstruction for transmission tomography using mixture model priors and deterministic annealing algorithms *Proc. SPIE* **4322** 899–908

Kaipio J, Kolehmainen V, Somersalo E and Vauhkonen M 2000 Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography *Inverse Problems* **16** 1487–522

Kaipio J P, Kolehmainen V, Vauhkonen M and Somersalo E 1999 Inverse problems with structural prior information *Inverse Problems* **15** 713–29

Kaipio J P and Somersalo E 2002 Estimating anomalies from indirect observations *J. Comput. Phys.* **181** 398–406

Katsevich A 1997 Local tomography for the limited-angle problem *J. Math. Anal. Appl.* **213** 160–82

Kazantsev I G 1991 Information content of projections *Inverse Problems* **7** 887–98

Kolehmainen V 2001 Novel approaches to image reconstruction in diffusion tomography *PhD Thesis*, Kuopio University Publications C. Natural and Enviromental Sciences 125 (http://venda.uku.fi/vkolehma)

Kolehmainen V, Arridge S R, Vauhkonen M and Kaipio J P 2000a Simultaneous reconstruction of internal tissue region boundaries and coefficients in optical diffusion tomography *Phys. Med. Biol.* **15** 1375–91

Kolehmainen V, Vauhkonen M, Kaipio J P and Arridge S R 2000b Recovery of piecewise constant coefficients in optical diffusion tomography *Opti. Epress* **7** 468–480

Kuchment P, Lancaster K and Mogilevskaya L 1995 On local tomography *Inverse Problems* **11** 571–89

Lalush D S and Tsui B M W 1992 Simulation evaluation of Gibbs prior distributions for use in maximum *a posteriori* SPECT reconstructions *IEEE Trans. Med. Imaging* **11** 267–75

Lee S-J, Rangarajan A and Gindi G 1995 Bayesian image reconstruction in SPECT using higher order mechanical models as priors *IEEE Trans. Med. Imaging* **14** 669–80

Lehtimäki M, Pamilo M, Raulisto L, Roiha M, Kalke M, Siltanen S and Ihamäki T 2003 Diagnostic clinical benefit of digital spot and digital 3D mammography on analysis of screening findings *SPIE Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display (San Diego, 15–20 Feb. 2003)* ed R L Galloway, at press

Lehtinen M 1985 On statistical inversion theory *Theory and Applications of Inverse Problems Proc. 4th National Meeting of Physics in Industry (Helsinki, 4–6 Sept. 1985)* ed H Haario (Essex: Longman Scientific & Technical)

Lehtinen M 1986 Statistical theory of incoherent scatter measurements EISCAT *Technical Note* 86/45 (*PhD Thesis*, University of Helsinki)

Liang Z, Jaszczak R and Greer K 1989 On Bayesian image reconstruction from projections: uniform and nonuniform *a priori* source information *IEEE Trans. Med. Imaging* **8** 227–35

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

Louis A K 1986 Incomplete data problems in X-ray computerized tomography: I. Singular value decomposition of the limited angle transform *Numer. Math.* **48** 251–62

McCauley T G, Stewart A, Stanton M, Wu T and Phillips W 2000 Three-dimensional breast image reconstruction from a limited number of views *Proc. SPIE* **3977** 116–27

Mohammad-Djafari A and Sauer K 1997 Shape reconstruction in x-ray tomography from a small number of projections using deformable models *17th Int. Workshop on Maximum Entropy and Bayesian Methods (MaxEnt97) (Boise, Idaho, USA, 4–8 Aug. 1997)*

Mohammad-Djafari A and Soussen C 1999 Reconstruction of compact homogeneous 3D objects from their projections *Discrete Tomography—Foundations, Algorithms and Applications* (Basel: Birkhäuser) ch 14 pp 317–42

Mosegaard K and Sambridge M 2002 Monte Carlo analysis of inverse problems *Inverse Problems* **18** R29–54

Mumcuoglu E Ü, Leahy R, Cherry S R and Zhou Z 1994 Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images *IEEE Trans. Med. Imaging* **13** 687–701

Natterer F 1986 *The Mathematics of Computerized Tomography* (Chichester/Stuttgart: Wiley/Teubner)

Nicholls G K and Fox C 1998 Prior modelling and posterior sampling in impedance imaging *Proc. SPIE* **3459** 116–27

Noo F, Defrise M, Clackdoyle R and Kudo H 2002 Image reconstruction from fan beam projections on less than a short scan *Phys. Med. Biol.* **47** 2525–46

Nygrén T, Markkanen M, Lehtinen M, Tereshchenko E D and Khudukon B Z 1997 Stochastic inversion in ionospheric radiotomography *Radio Sci.* **32** 2359–72

Orlov S S 1976 Theory of three-dimensional reconstruction: 1. Conditions of a complete set of projections *Sov. Phys. Crystallogr.* **20** 312

Persons T 2001 Three-dimensional tomosynthetic image restoration for brachytherapy source localization *Med. Phys.* **28** 1812

Persson M, Bone D and Elmqvist H 2001 Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography *Phys. Med. Biol.* **46** 853–66

Quinto E T 1993 Singularities of the X-ray transform and limited data tomography in $\mathbb{R}^2$ and $\mathbb{R}^3$ *SIAM J. Math. Anal.* **24** 1215–25

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

Ramesh A, Ludlow J B, Webber R L, Tyndall D A and Paquette D 2002 Evaluation of tuned-aperture computed tomography in the detection of simulated periodontal defects *Oral and Maxillofac. Radiol.* **93** 341–9

Ramm A G and Katsevich A I 1996 *The Radon Transform and Local Tomography* (Boca Raton, Fl: CRC Press)

Ranggayyan R M, Dhawan A T and Gordon R 1985 Algorithms for limited-view computed tomography: an annotated bibliography and a challenge *Appl. Opt.* **24** 4000–12

Roberts G O and Smith A F M 1994 Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms *Stoch. Process. Appl.* **49** 207–16

Ruttimann U E, Groenhuis R A J and Webber R L 1984 Restoration of digital multiplane tomosynthesis by a constrained iteration method *IEEE Trans. Med. Imaging* **3** 141–8

Sachs J Jr and Sauer K 1999 3D reconstruction from sparse radiographic data *Discrete Tomography—Foundations, Algorithms and Applications* (Basel: Birkhäuser) ch 16 pp 363–83

Saksman E, Nygrén T and Markkanen M 1997 Ionospheric structures invisible in satellite radiotomography *Radio Sci.* **32** 605–16

Sauer K and Bouman C 1993 A local update strategy for iterative reconstruction from projections *IEEE Trans. Signal Process.* **41** 534–48

Sauer K, James S Jr and Klifa K 1994 Bayesian estimation of 3-D objects from few radiographs *IEEE Trans. Nucl. Sci.* **41** 1780–90

Shepp L A and Kruskal J B 1978 Computerized tomography: the new medical X-ray technology *Am. Math. Mon.* **85** 420–39

Smith K T and Keinert F 1985 Mathematical foundations of computed tomography *Appl. Opt.* **24** 3950–7

Smith A F M and Roberts G O 1993 Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods *J. R. Statist. Soc.* B **55** 3–23

Smith K T, Solmon D C and Wagner S L 1977 Practical and mathematical aspects of the problem of reconstructing objects from radiographs *Bull. AMS* **83** 1227–70

Soussen C and Mohammad-Djafari A 1999 Multiresolution approach to the estimation of the shape of a 3D compact object from its radiographic data *Mathematical modeling, Bayesian Inference and Inverse Problems SPIE 99 (Denver, CO)* vol 3816, pp 150–60

Sukovic P and Clinthorne N H 2000 Penalized weighted least squares image reconstruction for dual energy X-ray transmission tomography *IEEE Trans. Med. Imaging* **19** 1075–81

Tam K C and Perez-Mendezs V 1981 Limits to image reconstruction from restricted angular input *IEEE Trans. Nucl. Sci.* **28** 179–83

Tamminen J 1999 MCMC methods for inverse problems *Geophysical Publications* vol 48 (Helsinki: Finnish Meteorological Institute)

Thompson J R and Tapia R A 1990 *Nonparametric Function Estimation, Modeling, and Simulation* (Philadelphia, PA: SIAM)

Tierney L 1994 Markov chains for exploring posterior distributions, with discussion *The Ann. Stat.* **22** 1701–62

Tuy H 1983 An inversion formula for cone-beam reconstruction *SIAM J. Appl. Math.* **43** 546–52

Vainberg E I, Kazak I A and Kurczaev V P 1981 Reconstruction of the internal three-dimensional structure of objects based on real-time integral projections *Sov. J. Nondestruct. Test.* **17** 415–23

Vauhkonen M, Kaipio J P, Somersalo E and Karjalainen P A 1997 Electrical impedance tomography with basis constraints *Inverse Problems* **13** 523–30

Vauhkonen M, Vadász D, Karjalainen P A, Somersalo E and Kaipio J P 1998 Tikhonov regularization and prior information in electrical impedance tomography *IEEE Trans. Med. Imaging* **17** 285–93

Voutilainen A, Kolehmainen V and Kaipio J P 2000 Statistical inversion of aerosol size distribution data *J. Aerosol Sci.* **31** S767–8

Webber R L 1998 Method and system for creating three-dimensional images using tomosynthetic computed tomography *US Patent Application* 09/034, 922, filed on March 5, 1998

Webber R L, Horton R A, Tyndall D A and Ludlow J B 1997 Tuned aperture computed tomography (TACT) Theory and application for three-dimensional dento-alveolar imaging *Dentomaxillofac. Radiol.* **26** 53–62

Webber R L, Horton R A, Underhill T E, Ludlow J B and Tyndall D A 1996 Comparison of film, direct digital, and tuned aperture computed tomography images to identify the location of crestal defects around endosseous titanium implants *Oral Surg. Oral Med. Oral Pathol.* **81** 480–90

Webber R L and Messura J K 1999 An *in vivo* comparison of diagnostic information obtained from tuned-aperture computed tomography and conventional dental radiographic imaging modalities *Oral Surg. Oral Med. Oral Pathol.* **88** 239–47

Yu D F and Fessler J A 2002 Edge-preserving tomographic reconstruction with nonlocal regularization *IEEE Trans. Med. Imaging* **21** 159–73

Ziedses des Plantes 1932 Eine neue Method zur Differenzierung der Röntgenographie *Acta Radiol.* **13** 182–92