Statistical inversion for X-ray tomography with few radiographs II: Application to dental radiology

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

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

Abstract. Diagnostic and operational tasks in dental radiology often require three dimensional information that is difficult or impossible to see in a projection image. A CT-scan provides the dentist with comprehensive three dimensional data. However, often CT-scan is impractical and, instead, only a few projection radiographs with sparsely distributed projection directions are available. Statistical (Bayesian) inversion is well-suited approach for reconstruction from such incomplete data. In statistical inversion, *a priori* information is used to compensate for the incomplete information of the data. The inverse problem is recast in form of statistical inference from the posterior probability distribution that is based on statistical models of the projection data and the *a priori* information of the tissue. In this paper, a statistical model for three dimensional imaging of dentomaxillofacial structures is proposed. Optimization and MCMC-algorithms are implemented for the computation of posterior statistics. Results are given with *in vitro* projection data that was taken with a commercial intraoral X-ray sensor. Examples include limited angle tomography and full-angle tomography with sparse projection data. Reconstructions with traditional tomographic reconstruction methods are given as reference for the assessment of the estimates that are based on the statistical model.

Version 14, Ville 29.11.2002

1. Introduction

The main tool in dental radiology is the X-ray projection image that reveals inner structure of bone and teeth. However, the obvious drawback of a projection (or a panoramic) image is irreversible overlapping of structures. Certain diagnostic and operative tasks often require more precise knowledge of the three dimensional structure of tissue than is available in single projection image. Such tasks include [3, 8, 25]

- Deciding whether two roots have grown together with common root canal or not.
- Detection of alveolar decease, or bone loss between teeth.
- Implant planning.

- Finding out whether certain roots have intimate relationship with the inferior dental canal. This is related to the risk of damaging nerves when removing a tooth.
- Analysis of the form of the condylar process in the temporomandibular joint.

We consider taking a small number of projection images of the tissue from sparsely distributed directions using the dentist's regular X-ray equipment and reconstructing the 3-D structure of tissue from the projections. More precisely, we consider the following two types of sparse projection data:

(A) Sparse limited angle data. In intraoral imaging a few projection radiographs are taken with a small digital sensor in fixed position inside the patient's mouth. Due to geometrical restrictions, the X-ray source positions are limited to a cone with opening angle significantly less than 180°.

(B) Sparse full angle data. In extraoral imaging the region of interest is imaged through the head from a small number of sparsely distributed projection directions.

For both data types, the projection images are often truncated due to small detector size or in order to minimize dose to vital organs. In these cases the image reconstruction has the additional complication of local tomography problem. Both data types, (A) and (B), lead to ill-posed image reconstruction problems (i.e., the solution is sensitive to measurement errors and/or the problem does not have unique solution).

It is well-known that traditional CT algorithms, such as filtered backprojection, are not well-suited for projection data of type (A) or (B) since these data types violate the assumptions of those algorithms. Despite this conflict between the data and the assumptions, traditional methods have been widely used for both data types. For data type (A) a traditional reconstruction method is tuned aperture computed tomography (TACT) method [32, 15, 36], which is basically equivalent to unfiltered backprojection. For data type (B) a popular traditional method is filtered backprojection (FBP) in case of global tomography (i.e., projections are not truncated). For the local tomography data of type (B), a usual method is Λ-tomography which has been developed for local tomography problems with non-sparse full-angle projection data [29, 20, 9, 10].

Statistical inversion (SI) is well-suited approach for 3-D reconstruction with both data types (A) and (B). In statistical inversion, *a priori* knowledge of the tissue is used in the image reconstruction problem in order to compensate for the incomplete information in the sparse projection data. Separate statistical models (probability distributions) are formulated for (a) the acquisition of the projection data and (b) the *a priori* information. Based on these models and the Bayes formula, complete solution of the inverse problem is obtained as the posterior probability distribution. Final images of the target are then obtained as point estimates from the posterior distribution. In contrast to traditional reconstruction methods, the statistical approach gives natural means for the computation of confidence limits for the estimates.

We propose a statistical model for three dimensional dental imaging. In the proposed model, we approximate the three dimensional problem by a stack of two-dimensional problems. In the Bayesian model for each two dimensional problem we use the following prior models:

(i) For each 2D-slice we use a total variation (TV) prior model. Total variation is a feasible prior model for dental structures since they are expected to consist of a few approximately homogeneous regions with sharp, well defined boundaries.

(ii) To take the 3D nature of the problem into account, a L^1-prior is used in the model for the distance of the current slice from the previous one. This is based on the assumption that the cross-section of dental structures does not change much between two consecutive slices.

(iii) Positivity prior, which in short means that X-rays can only attenuate and not intensify inside tissue.

To illustrate the performance of the model, results with *in vitro* sparse projection data are given. For data type (A) we consider two examples. The first example is a model problem with sparse projection data from a tooth phantom. For this test case, the ground truth is given by a full-angle reconstruction. The second example for (A) is reconstruction using truncated intraoral measurements from a realistic head phantom. For both cases, the maximum a posteriori (MAP) estimates are presented as reconstructions of the target. For both cases traditional tomosynthetic (backprojected) reconstructions are shown as reference images for the assessment of the statistical model. For the first test case, we will also give an illustrative example of more complete statistical inference from the posterior distribution using Markov chain Monte Carlo (MCMC) methods.

For data type (B) we consider also two examples. The first example is a model problem using full-angle sparse projection data from the tooth phantom. MAP estimates are represented as images of the target and reconstructions with the widely used filtered backprojection (FBP) method are shown as reference images. The second test case for (B) is full-angle reconstruction from sparse projection data with truncated projections from a jaw phantom. MAP estimates are represented as images of the target. Backprojection and Λ-tomography reconstructions are shown as reference images.

Application of Bayesian inversion to dental radiology appears to be new. Statistical methods has been used for data type (B) in [30], however, only likelihood distribution is used for the reconstruction. With the model introduced in this paper it is possible to further improve [30] with simultaneous reduction in radiation dose.

This paper is organized as follows. In section 2 we discuss the transformation of digital radiographs to tomographic data. We also discuss the experimental imaging geometries used in the examples. In section 3.1 we discuss the statistical model that is used in this paper. The discussion is mainly based on the theory and models that were presented in part I of this paper. We also discuss the computation of the point estimates. A gradient based optimization approach is given for the computation of the MAP estimate, and then computation of other usual statistics that necessitate integration is discussed. Results with the experimental data are given in section 4 and in section 5 we give conclusions.

2. From digital radiographs to tomographic data

The projection radiographs of the targets were acquired using a commercial intraoral X-ray detector Sigma and a dental X-ray source Focus‡. As explained in section 3.2, part I of this article, the input data of tomographic algorithms is a collection of line integrals of the unknown attenuation coefficient function. For each pixel value in each projection image we need to (a) determine the path of the detected X-ray through the pixel/voxel grid and (b) the amount of attenuation of the X-ray through that

‡ Sigma and Focus are registered trademarks of Instrumentarium Corp. Imaging Division

path. We discuss the transformation of the projection data in section 2.1 and the measurement geometries for the experiments in sections 2.2–2.4.

2.1. From detected pixel values to attenuation measurements

The Sigma detector is based on charge coupled device (CCD) technology and it is capable of sensing roughly 2000 gray levels. Size of the active imaging area is 34×26mm and the resolution is 872×664 pixels. After exposure, each pixel contains a positive integer which is directly proportional to the number of X-ray quanta that hit the pixel's area.

A detected pixel value p is transformed to tomographic attenuation measurement P as follows. Let M be the logarithm of the maximum pixel value over all detector pixels. We define the tomographic data as

$$P = M - \log(p). \tag{1}$$

What kind of error is introduced by this transformation? The ideal tomographic data should be the integral of the attenuation coefficient $x(s)$ along the X-ray path L:

$$P' = \int_L x(s)\mathrm{d}s = \log I_0 - \log I_1.$$

The pixel value p is directly proportional to the final intensity: $p = aI_1$. If the detector is partly illuminated by direct radiation, we have $M \approx \log(aI_0) = \log a + \log I_0$. Then

$$P = M - \log(p) \approx \log a + \log I_0 - \log(a) - \log(I_1) = P'.$$

Thus, the above transformation is a feasible choice for problems in which the distance and angle of the X-ray source are fixed with respect to the detector and every projection contains some "air only" observations. Without such observations, $\log(I_0)$ needs to be calibrated from imaging parameters.

2.2. Experimental setup for the tooth phantom model problem

In order to get full-angle reconstructions as a reference for the limited angle reconstructions in the model problem with the tooth phantom, we used the conventional cone beam CT-geometry, which is shown schematically in Figure 1, for our laboratory experiments.

The experiments were carried out as follows. The Sigma CCD-detector and the Focus X-ray source were attached into fixed positions such that the source direction is normal to the detector array. The distance from the focal spot to the detector array was 840mm. The tooth phantom, which was a third mandibular molar removed from a female patient of age 25, was placed on a rotating platform, so that projections from different angles can be obtained. The distance from the center of rotation to the detector was 56mm. Left image in Figure 2 shows one raw 872×664 projection image from the experiments, the middle image shows one row (i.e., raw data for one two-dimensional slice) from the projection image and the right image shows the same row in form of tomographic data. We note that the white triangles in the lower corners of the projection image in Figure 1 do not correspond to detected radiation. They result from the rounded corners of the intraoral detector.

The purpose of the wires that are seen in the lower part of the left image in Figure 2 is to give information about the location and alignment of the rotation axis

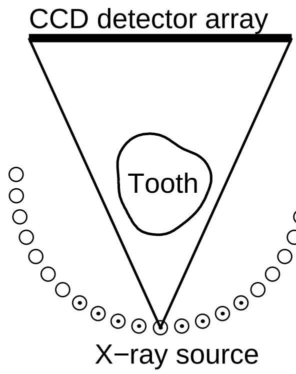

Figure 1. Cone beam imaging geometry for full angle tomography. This geometry was used in the experiments with the tooth phantom. Circles denote the source locations for the full-angle data (23 projections from total view-angle of 187°). The projections that were used in limited angle computations (9 projections from view-angle of 68°) are denoted by black dots within the circles. For clarity, the location and alignment of the detector with respect to the source is depicted only for one source location.

in the projection images. In the sum image of all projections, the wires appear as a sandglass shaped object. The node of this sandglass gives rotation axis for one slice and the inclination angle of the rotation axis can be obtained by computing the normal direction to the path in sum image that is drawn by the upper end of the longer wire. The projection angles were read from a millimeter scale paper that was attached around the rotating platform of the tooth phantom.

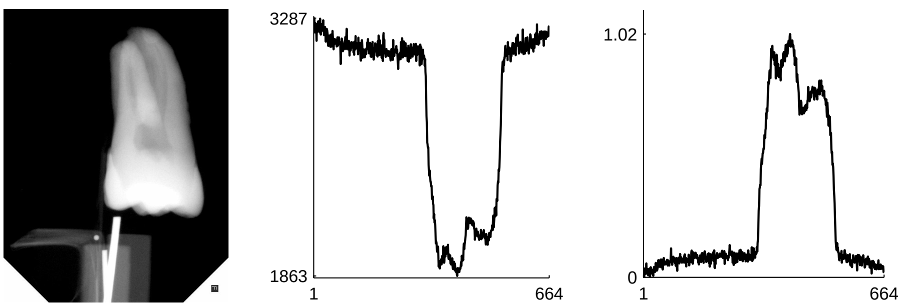

Figure 2. Left: 872×664 projection radiograph from the tooth phantom. Note that the image is shown with inverted colormap (i.e., black correspond to high photon counts). Middle: Pixel values of the 200^{th} row from the raw projection radiograph. Right: Same row in form of tomographic attenuation data.

We note that this experimental geometry corresponds to the case in which the source and detector array move on a horizontal circle. Thus, the projection directions are restricted onto a circular arc. Also, the source to detector distance is relatively long with respect to the physical size of the Sigma detector. These allow us to approximate the 3D reconstruction problem by a stack of two-dimensional problems with reasonably good accuracy. The development of purely 3D methods is left to future studies.

Finally, the transpose of each of the transformed projection images correspond to one block of the tomographic data for the 3D experiment. As an example, the left image in Figure 3 shows one column (i.e., line integral data from all projections for one two-dimensional slice) of this block matrix for the data from the tooth in traditional sinogram form. The data was collected by taking 23 projections with total 187° angle of view. Referring forward to the first model problem, which is the limited angle reconstruction from sparse projection data, the right image in Figure 3 shows the part of sinogram (nine projections with 68° angle of view) that was used in the limited angle reconstructions.

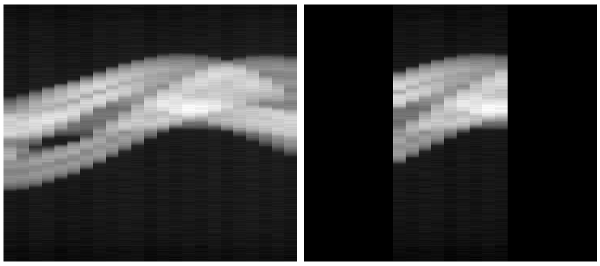

Figure 3. Sinogram for the 200$^{\mathrm{th}}$ slice of the projection data from the tooth phantom. Left: Projections used in this sinogram were collected from 187° angle of view (23 projections with 8.5° projection interval). Right: The 68° part of the sinogram that was used for the limited angle reconstructions.

2.3. Measurement geometry for intraoral imaging

In intraoral dental X-ray imaging, the measurement geometry is such that the detector is in fixed position inside the patient's mouth and the dentist can move the X-ray source, which is mounted on a foldable arm, with respect to the intraoral detector. This geometry is illustrated schematically in Figure 4.

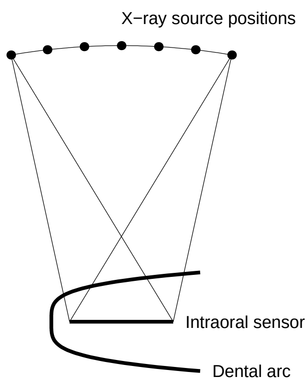

Figure 4. Imaging geometry for intraoral measurements. The detector is in fixed position inside the patient's mouth. This geometry was used in the test case with the head phantom.

In this study, this geometry was used for the limited angle experiments with the realistic head phantom. In the experiments, the Sigma detector was placed in a fixed position inside the mouth of the head phantom such that it was right behind the teeth to be imaged. The X-ray source was mounted on a foldable arm which was used to move the source on a approximately circular arc with distance of ~ 590mm from the detector.

2.4. Measurement geometry for extraoral imaging

In extraoral imaging, the region of interest (ROI) is imaged from different directions through the head. Typical geometry for extraoral imaging is illustrated in Figure 5. The patient is kept in fixed position. The source and the detector array are mounted onto a rotating platform that can be used to move the source and detector array to different projection angles. There are some dedicated devices for such measurements [21, 30]. We note that extraoral imaging (as depicted in Figure 5) leads to local tomography problem.

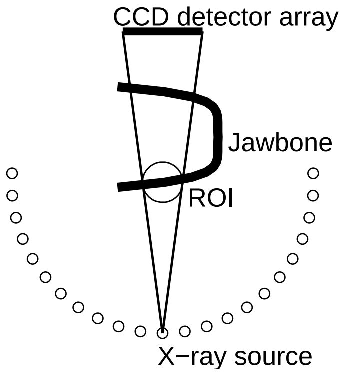

Figure 5. Cone beam imaging geometry for local tomography. This geometry was used in the last test case with the jaw bone phantom. The source positions for the experiments (23 projections from view-angle of 187°) are denoted by circles. The region of interest (ROI) is denoted by thin line. For clarity, the location and alignment of the detector with respect to the source is depicted only for one source location.

Our last test case is extraoral imaging with sparse projection data from a jaw bone phantom. There the experimental setup was implemented similarly to the setup explained in section 2.2: The phantom was placed on the rotating platform and the Sigma detector and X-ray source were mounted into fixed positions. With the exception of different source to detector distance (1292mm) and center of rotation to detector distance (88mm), the geometrical details of the experimental setup were the same as in section 2.2.

3. Statistical inversion in 3D dental imaging

3.1. Statistical model for dental imaging

In this section we discuss the application of the statistical inversion approach to three dimensional dental imaging. As it was discussed previously, we approximate the 3D problem by a stack of $j = 1, 2, \ldots, N_{\mathrm{sli}}$ two-dimensional problems.

Let

$$x^{(j)} = \sum_{i=1}^{M} x_i^{(j)} \chi_i, \tag{2}$$

where χ_i is the characteristic function of pixel Ω_i in the two-dimensional pixel lattice, denote the discrete representation of the j^{th} slice in the stack of two-dimensional slices. In the sequel, we will identify the function (2) by the coefficient vector $x^{(j)} = (x_1^{(j)}, x_2^{(j)}, \ldots, x_M^{(j)})^{\mathrm{T}} \in \mathbb{R}^M$. Further, let

$$m^{(j)} = Ax^{(j)} + \epsilon^{(j)} \tag{3}$$

denote the observation model for the j^{th} 2D-problem. In equation (3), $m^{(j)} = (m_1^{(j)}, m_2^{(j)}, \ldots, m_N^{(j)})^{\mathrm{T}} \in \mathbb{R}^N$ is the vector of tomographic data for j^{th} two dimensional slice and $\epsilon^{(j)} \in \mathbb{R}^N$ denotes the observation noise. It should be noted that with our experimental setup, the model matrix is not the same for each 2D slice due to the possible displacement of the rotation axis between different slices $m^{(j)}$ of the projection data. Also, the size of the data vector $m^{(j)}$ may vary due to the rounded corners of the Sigma detector, see Figure 2. However, due to notational simplicity, we use the notation $A \doteq A^{(j)}$ for the model matrix and $N \doteq N_j$ for the dimension of the projection data vector in the sequel.

The statistical model we use for dental imaging was introduced in sections 3.1–3.2, part I of this paper. As the prior model for the two-dimensional slices $x^{(j)}$ of the dental structures we use the total variation (TV) prior. Within the discretization (2) of the 2D-attenuation coefficient function, the total variation can be written as

$$\mathrm{TV}(x^{(j)}) = \sum_{k=1}^{N_e} l_k |\Delta_k^{\mathrm{T}} x^{(j)}|, \tag{4}$$

where l_k is the length of the edge between the adjacent pixels $\Omega_{i_1^k}$ and $\Omega_{i_2^k}$, $\Delta_k \in \mathbb{R}^M$ is the vector

$$\Delta_k = (0, \ldots, \overset{(i_1^k)}{1}, \; 0, \ldots, 0, \; \overset{(i_2^k)}{-1}, \; 0, \ldots, 0)^{\mathrm{T}},$$

and N_e is the number of edges (s.t. $l_k = |\partial\Omega_{i_1^k} \bigcap \partial\Omega_{i_2^k}| > 0$) connecting two adjacent pixels in the 2D lattice.

The *total variation* prior density for the 2D attenuation coefficient $x^{(j)}$ is defined as

$$p_{\mathrm{TV}}(x^{(j)}) \sim e^{-\alpha \mathrm{TV}(x^{(j)})}, \tag{5}$$

where the total variation is calculated using equation (4). The total variation prior can be considered as a feasible model for dental structures, since it has high probability density for level set type images which consist of a few (almost) constant attenuation levels which are bounded by short, well defined boundary lines. The use of TV prior for the regularization of inverse problems has been discussed for example in [31, 17, 24] and the use of TV constraints for image enhancement in [4, 5, 6].

To take the three dimensional structure of the target into account in the stack of two-dimensional reconstructions, we use a (conditional) L^1-prior between the slices $x^{(j)}$ and $x^{(j-1)}$. Let $\hat{x}^{(j-1)}$ denote an estimate for the slice $x^{(j-1)}$ (with initialisation

$\hat{x}^{(0)} = 0$). Within the discrete framework, this coupling prior density can be written in the form

$$p_{L^1}(x^{(j)}|\hat{x}^{(j-1)}) \sim \exp\left(-\gamma\|x^{(j)} - \hat{x}^{(j-1)}\|_{L^1}\right)$$
$$= \exp\left(-\gamma\sum_{k=1}^{M}|\Omega_k||x_k^{(j)} - \hat{x}_k^{(j-1)}|\right). \quad (6)$$

The L^1-prior is concentrated around images $x^{(j)}$ which are close to $\hat{x}^{(j-1)}$ but may have a few large deviations with small support. For an extensive discussion on this feature, see the article [7]. We chose the model (6) based on the assumption that the cross-section of dental structures does not change much between two consecutive slices.

Taking into account the positivity prior $p_+(x^{(j)})$ for the attenuation coefficient, equation (19) in part I of this paper, and using equations (5) and (6), the overall (conditional) prior density for slice $x^{(j)}$ assumes the form

$$p(x^{(j)}|\hat{x}^{(j-1)}) \sim p_+(x^{(j)})\exp\left(-\alpha\mathrm{TV}(x^{(j)}) - \gamma\|x^{(j)} - \hat{x}^{(j-1)}\|_{L^1}\right). \quad (7)$$

For the observation errors $\epsilon^{(j)}$ we make the assumption that they are Gaussian with zero-mean ($\epsilon^{(j)} \sim \mathcal{N}(0,\Gamma_{\mathrm{noise}})$) and are independent of the attenuation parameters $x^{(j)}$. Using the theory and the likelihood model that were given in sections 3.1-3.2, part I of this paper, the posterior density for the j^{th} two-dimensional problem assumes the form

$$p(x^{(j)}|m^{(j)},\hat{x}^{(j-1)}) \sim p_+(x^{(j)})\exp\left(-F(x^{(j)},\hat{x}^{(j-1)})\right), \quad (8)$$

where

$$F(x^{(j)},\hat{x}^{(j-1)}) = \frac{1}{2}\|m^{(j)} - Ax^{(j)}\|^2_{\Gamma^{-1}_{\mathrm{noise}}} + \alpha\mathrm{TV}(x^{(j)}) + \gamma\|x^{(j)} - \hat{x}^{(j-1)}\|_{L^1}. \quad (9)$$

As it was explained in section 3.1, part I of this paper, the posterior density constitutes the complete solution of the inverse problem in the statistical sense. To summarize and visualise the statistical solution of the inverse problem one needs to compute different statistics from the posterior distribution. Most common choices include the maximum a posteriori (MAP) and conditional mean (CM) estimates, covariance/correlation matrices and marginal densities together with confidence intervals [12, 14, 17]. In the following, we explain the computation of the MAP estimate and then the computation of other, integration based posterior statistics using MCMC methods is briefly discussed.

3.2. Computation of the MAP-estimate

The most usual estimate from the posterior is the maximum a posteriori (MAP) estimate which is defined through the relation

$$p(x^{(j)}_{\mathrm{MAP}}|m^{(j)},\hat{x}^{(j-1)}) = \max(p(x^{(j)}|m^{(j)},\hat{x}^{(j-1)})).$$

As discussed in section 3.5, part I of this paper, the computation of the MAP-estimate from the posterior density in equation (8) amounts to finding the parameter vector that satisfy

$$x^{(j)}_{\mathrm{MAP}} = \arg\min_{x^{(j)}\geq 0} F(x^{(j)},\hat{x}^{(j-1)}),$$

where $F(x^{(j)}, \hat{x}^{(j-1)})$ is as in equation (9). We intend to find the estimate $x_{\mathrm{MAP}}^{(j)}$ by applying gradient based optimization methods. However, here we face two difficulties. First, the total variation and L^1-prior functionals are non–differentiable, due to the presence of the absolute value function. To overcome this problem, we use the smooth approximation

$$|t| \approx h_\beta(t) = \frac{1}{\beta}\log(\cosh(\beta t)), \tag{10}$$

where $\beta > 0$ is a parameter adjusting the accuracy of the approximation. The approximation $h_\beta(t)$ with the value $\beta = 200$ (that is used in this study) and the absolute value function $|t|$ are shown between the interval $t \in [-0.1\ 0.1]$ in Figure 6.

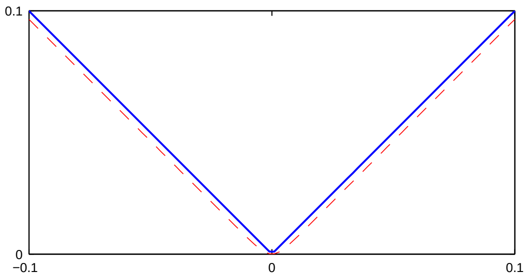

Figure 6. Absolute value function $|t|$ (solid line) and the approximation $h_\beta(t)$ (dashed line), equation (10), of the absolute value function in the interval $t \in [-0.1,\ 0.1]$. In the approximation $h_\beta(t)$, value $\beta = 200$ was used.

Using the approximation (10), the approximate total variation is obtained as

$$\mathrm{TV}_\beta(x^{(j)}) = \sum_{k=1}^{N_e} l_k h_\beta(\Delta_k^{\mathrm{T}} x^{(j)}) \tag{11}$$

and the approximate L^1-norm, which we denote by $L_\beta^1(\cdot)$, is obtained as

$$L_\beta^1(x^{(j)} - \hat{x}^{(j-1)}) = \sum_{k=1}^{M} |\Omega_k| h_\beta(x_k^{(j)} - \hat{x}_k^{(j-1)}). \tag{12}$$

Using these approximations for TV and L^1-functionals, the objective functional (9) is approximated by a differentiable functional of the form

$$F_\beta(x^{(j)}, \hat{x}^{(j-1)}) = \frac{1}{2}\|m^{(j)} - Ax^{(j)}\|^2_{\Gamma^{-1}_{\mathrm{noise}}} + \alpha \mathrm{TV}_\beta(x^{(j)}) + \gamma L_\beta^1(x^{(j)} - \hat{x}^{(j-1)}). \tag{13}$$

Referring to the results about singularities that are reconstructible from limited angle data [26] (see also the review in section 2, part I of this paper), we show in Appendix B that the MAP-estimate with the TV_β-prior does not destroy any of these singularities. In other words, one can expect to see in the MAP-estimate with the TV_β-prior at least the same singularities that are seen in backprojection reconstruction.

The second problem in the computation of the MAP-estimate comes from the positivity constraint which is due to the positivity prior $p_+(x^{(j)})$. To take the positivity prior into account, we use use an exterior point search [11]. In the exterior point search, the original constrained problem is approximated by a sequence of unconstrained problems

$$x_{\mathrm{MAP}}^{(j,t)} = \arg\min \left\{ F_\beta(x^{(j)}, \hat{x}^{(j-1)}) + \Upsilon^{(t)}(x^{(j)}) \right\}, \tag{14}$$

where $\Upsilon^{(t)}(x^{(j)})$ is a penalty functional that is used to penalize the negative components of the solution $x^{(j)}$ and the superindex $x^{(\cdot,t)}$ refers to the t^{th} problem in the sequence of $\{t = 1, \ldots, P\}$ unconstrained problems. Using a suitably chosen sequence of penalty functionals $\{\Upsilon^{(t)}(x^{(j)}), t = 1, \ldots, P\}$, the exterior point method forces the sequence of solutions $\{x^{(j,t)}_{\text{MAP}}, t = 1, \ldots, P\}$ (asymptotically) to the feasible region $x^{(j)} \geq 0$. Then, the solution of the constrained problem is approximated by $x^{(j)}_{\text{MAP}} \approx x^{(j,P)}_{\text{MAP}}$.

In this paper we use a penalty functional $\Upsilon^{(t)}$ of the form

$$\Upsilon^{(t)}(x^{(j)}) = \sum_{k=1}^{M} \phi^{(t)}(x_k^{(j)}), \tag{15}$$

where

$$\phi^{(t)}(x_k^{(j)}) = \begin{cases} \varsigma_t (x_k^{(j)})^2 & , x_k^{(j)} < 0 \\ 0 & , x_k^{(j)} \geq 0 \end{cases}, \tag{16}$$

and $\{\varsigma_t, t = 1, 2, \ldots, P\}$ is a sequence of increasing positive numbers.

In this study the MAP-estimates (14) are computed using the gradient-based Barzilai-Borwein method [1].

3.3. The gradient descent method of Barzilai and Borwein

We briefly describe here the gradient-based method for unconstrained large-scale optimization introduced by Barzilai and Borwein [1]. We chose this method based on the facts that i) the inversion of Hessian matrix is excessively heavy task due to the large dimension (M) of the problem and ii) the computation of the gradient for the objective functional (14) is relatively cheap, enabling relatively fast computation of multiple iterations. Further, gradient based methods are advantageous also in the sense that a more accurate approximation (larger β) for absolute value function can be used. This is due to the fact that we do not need to invert matrices that contain second derivatives of $h_\beta(t)$.

In the sequel, we use the notation $x^{(j,t,\ell)}$ to denote the ℓ^{th} iterate for the t^{th} problem in the sequence of unconstrained problems for the j^{th} slice.

For the problem (14), the Barzilai-Borwein iteration can be written as

$$x^{(j,t,\ell+1)} = x^{(j,t,\ell)} - a_\ell^{-1} d^{(j,t,\ell)}, \tag{17}$$

where the search direction is of the form

$$d^{(j,t,\ell)} = \nabla F_\beta(x^{(j,t,\ell)}) + \nabla \Upsilon^{(t)}(x^{(j,t,\ell)}) \tag{18}$$

and the step-length parameter a_ℓ is computed as

$$a_\ell = \frac{(x^{(j,t,\ell)} - x^{(j,t,\ell-1)})^{\mathrm{T}}(d^{(j,t,\ell)} - d^{(j,t,\ell-1)})}{(x^{(j,t,\ell)} - x^{(j,t,\ell-1)})^{\mathrm{T}}(x^{(j,t,\ell)} - x^{(j,t,\ell-1)})}. \tag{19}$$

The vector $\nabla F_\beta(x^{(j)})$ in equation (18) is of the form

$$\begin{aligned} \nabla F_\beta(x^{(j,t,\ell)}) = & -A^{\mathrm{T}}\Gamma_{\text{noise}}^{-1}(m^{(j)} - Ax^{(j,t,\ell)}) \\ & + \alpha \nabla \mathrm{TV}_\beta(x^{(j,t,\ell)}) + \gamma \nabla L^1_\beta(x^{(j,t,\ell)} - \hat{x}^{(j-1)}), \end{aligned} \tag{20}$$

where the elements of the vector $\nabla \mathrm{TV}_\beta \in \mathbb{R}^M$ are obtained as

$$\Big(\nabla \mathrm{TV}_\beta(x^{(j,t,\ell)})\Big)_m = \sum_{k=1}^{N_e} l_k h'_\beta(\Delta_k x^{(j,t,\ell)}) \Delta_{k,m} \tag{21}$$

and the elements of the vector $\nabla L^1_\beta \in \mathbb{R}^M$ are obtained as

$$\left(\nabla L^1_\beta(x^{(j)})\right)_m = h'_\beta(x^{(j,t,\ell)}_m - \hat{x}^{(j-1)}_m)|\Omega_m|. \qquad (22)$$

In equations (21)-(22) $h^{'}_\beta(\cdot)$ denotes the first derivative of $h_\beta(\cdot)$. The entries of the vector $\nabla\Upsilon^{(t)}(x^{(j,t,\ell)})$ are of the form

$$\left(\nabla\Upsilon^{(t)}(x^{(j,t,\ell)})\right)_m = 2\varsigma_t(x^{(j,t,\ell)}_m)I(x^{(j,t,\ell)}_m < 0) \qquad (23)$$

where $I(x^{(j,t,\ell)}_m < 0)$ denotes the indicator function for the event $x^{(j,t,\ell)}_m < 0$.

3.4. Markov chain Monte Carlo methods

Whereas the computation of the MAP-estimate is an optimization problem, the computation of other usual posterior statistics are problems of integration over a high dimensional parameter space. As discussed in part I of this paper, these tasks necessitate the use of Monte Carlo integration techniques in the case of non-Gaussian posterior density, such as the density function given in equations (8-9).

The basic idea in Monte Carlo integration is to generate a large, representative ensemble $\{x^{(j,\ell)}, \ell = 1, 2, \ldots, S\} \subset \mathbb{R}^M$ of random "sample images" from the posterior density $p(x^{(j)}|m^{(j)}, \hat{x}^{(j-1)})$ and then approximate the integral of function $f(x^{(j)})$ with respect the posterior distribution by the sample mean, that is,

$$\int_{\mathbb{R}^M} f(x^{(j)})p(x^{(j)}|m^{(j)}, \hat{x}^{(j-1)})\mathrm{d}x^{(j)} \approx \frac{1}{S}\sum_{\ell=1}^{S} f(x^{(j,\ell)}). \qquad (24)$$

Often the posterior models for inverse problems, such as the model given by equations (8-9), are such that direct drawing of independent sample images is impossible. In Markov chain Monte Carlo (MCMC) methods the representative ensemble of (dependent) sample images are obtained by generating a realisation of a Markov chain which has its stationary distribution defined by the given posterior density [14, 12]. A more detailed discussion on MCMC methods is given in section 3.6, part I of this paper.

In case of X-ray tomography the large dimension ($M > 10^4$ for a 2D slice) of the parameter space $\mathbb{R}^M$ makes the MCMC-sampling computationally a very demanding task. However, in order to give an illustrative example of more "complete" statistical inference from the posterior distribution, we carry out MCMC analysis for the posterior distribution of one two-dimensional problem (i.e., for one slice $x^{(j)}$) in the first test case, which is the limited angle problem with data from the tooth phantom. The development of efficient MCMC schemes that can be used to carry out inference for 3D reconstruction problems, are left to future studies.

4. Results

4.1. Limited angle tomography from sparse projection data of a tooth

As the first example of data type (A), see section 1, we consider the model problem of limited angle tomography with sparse projection data (9 projections from view-angle of 68°) from a tooth phantom. As it was discussed in section 2.2, the projection images from the tooth phantom were collected using the conventional CT geometry instead of using the fixed detector geometry, which is more typical geometry in clinical dental studies. However, the results for limited angle tomography in these two geometries are qualitatively very similar. We chose to use the conventional CT geometry in order to get full angle reconstructions as the "ground truth" for the limited angle reconstructions in this test problem. The experimental setup is explained in detail in section 2.2.

Results are shown in Figures 7 and 8. The left colum in Figure 7 shows MAP-estimates $x^{(j)}_{\rm MAP}$ for four slices with full-angle data that was collected from view angle of 187° with projection intervals of 8.5° (23 projections). With this set of projection images the size of the data vector $m^{(j)}$ for each two dimensional problem is $N = 15272$. Figure 3 shows one slice $m^{(j)}$ of this data in traditional sinogram form. It should be noted that if the vector $m^{(j)}$ contain values that correspond to the rounded corners of the detector, these values are neglected simply by removing respective rows from $m^{(j)}$ and A.

The domain Ω in the computation of the two dimensional images shown in Figure 7 was a $26 \times 26\text{mm}^2$ square which was divided into a regular $M = 166 \times 166 = 27556$ pixel lattice, leading to pixel size of $\sim 0.16 \times 0.16\text{mm}^2$.

The left column of Figure 8 shows four vertical slices of the (approximate) three dimensional reconstruction with the full-angle data. The three dimensional reconstruction was obtained as a stack of $N_{\rm sli} = 600$ two-dimensional reconstructions. Each of the two-dimensional slices represent 0.045mm thick slice of the three dimensional data, leading to vertical size of 26.1mm for the images in Figure 8. The horizontal size of the images in Figure 8 is the same as in Figure 7, that is, 26mm. Note that the first and fourth slice in Figure 8 are chosen approximately from the front and back surfaces of the tooth.

The MAP-estimates with the full-angle data, left columns of Figures 7-8, are based on the statistical model described in section 3.1. For the covariance matrix $\Gamma_{\rm noise}$ of the observation errors we used the trivial choice $\Gamma_{\rm noise} = \sigma_n^2 I$, i.e., we assumed that the noise is statistically independent Gaussian noise with equal variance in each direction. The noise variance σ_n^2 was estimated from the projection data. This was achieved by taking one approximately homogeneous "air-only" sample (100×50 detector pixels) from one transformed projection image, and then computing estimate for σ_n^2 based on this sample. As the result, we had value $\sigma_n^2 = 0.0004$. In practice, a better estimate for the noise statistics can be obtained from a repeated set of phantom measurements and/or careful analysis of the measurement system. The determination of the accurate noise statistics for the Sigma sensor and the assessment to which extent the reconstruction results improve using a better noise model, are left to future study. The analysis given in section 3.2, part I of this paper suggests, however, that the Gaussian approximation is acceptable.

The prior parameters were chosen by visual inspection from a set of reconstructions with different parameters. With the full-angle data we used $\alpha = 1250$, $\gamma = 1250$ and $\{\varsigma_t, t = 1, 2, \ldots, 5\}$ was a linearly increasing sequence from 12500 to $3.75 \cdot 10^5$. The parameter β in the approximation (10) for the absolute value function was $\beta = 200$. The MAP-estimates were computed using the Barzilai-Borwein method. We computed 6 iteration steps for each problem in the sequence of 5 unconstrained problems and then result from the 5$^{\rm th}$ unconstrained problem was used as an approximation for the MAP-estimate $x^{(j)}_{\rm MAP}$.

The second column in Figures 7-8 show respective slices of an approximate 3D reconstruction with limited angle data that was collected from view angle of 68°

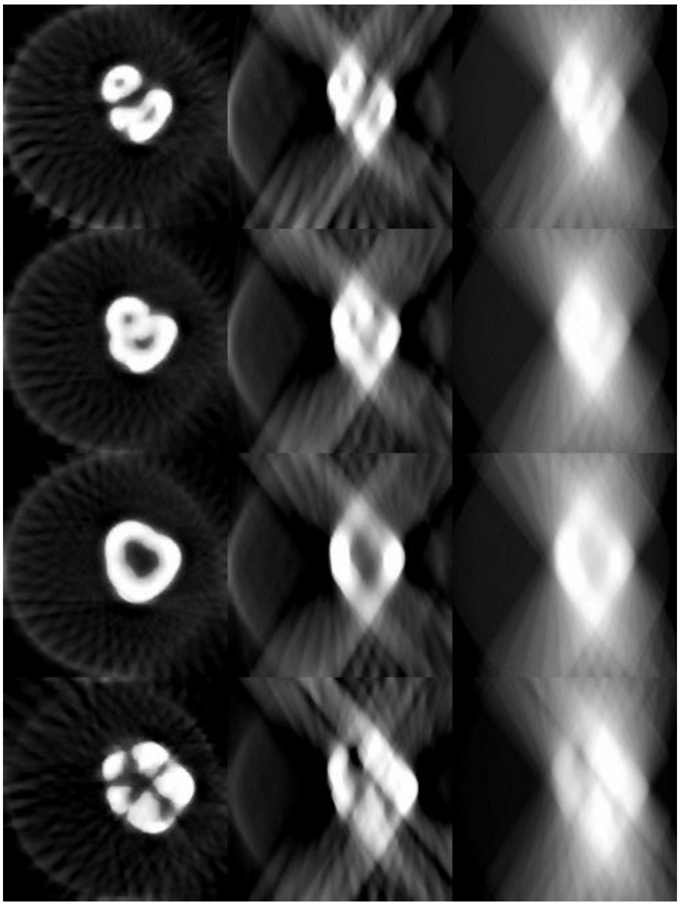

Figure 7. Left column shows horizontal slices (MAP-estimates) from full angle data with the TV_β-prior. The full-angle data consisted of 23 projections from total view-angle of 187° (projection interval 8.5°). The other two columns show MAP-estimates with the TV_β-prior (center column) and tomosynthetic reconstructions (right column) from limited-angle data. In the limited angle case, nine projections from view-angle of 68° were used.

with projection intervals of 8.5° (9 projections). The size of the data vector $m^{(j)}$ for each slice was $N = 5976$. The right image in Figure 3 shows this 68° part of the data, that was used in limited angle computations, for one slice in sinogram form. The prior parameters in the limited angle case were the same as in the full-angle reconstruction. The right columns in Figures 7-8 show the respective slices for a tomosynthetic (backprojection) reconstruction from the same limited angle data with view angle of 68°. For details of the tomosynthesis, see [32, 15, 36].

As can be seen from Figures 7-8, the limited angle MAP-estimates with our statistical model are good in this test problem. Further, the statistical reconstructions are sharper and clearer than the traditional tomosynthetic reconstructions. This is especially evident in the depth direction where the information content of the limited angle projection data is poor. This clear difference in the images gives an illustration for the effect of well chosen prior model in limited angle tomography. Also, note that

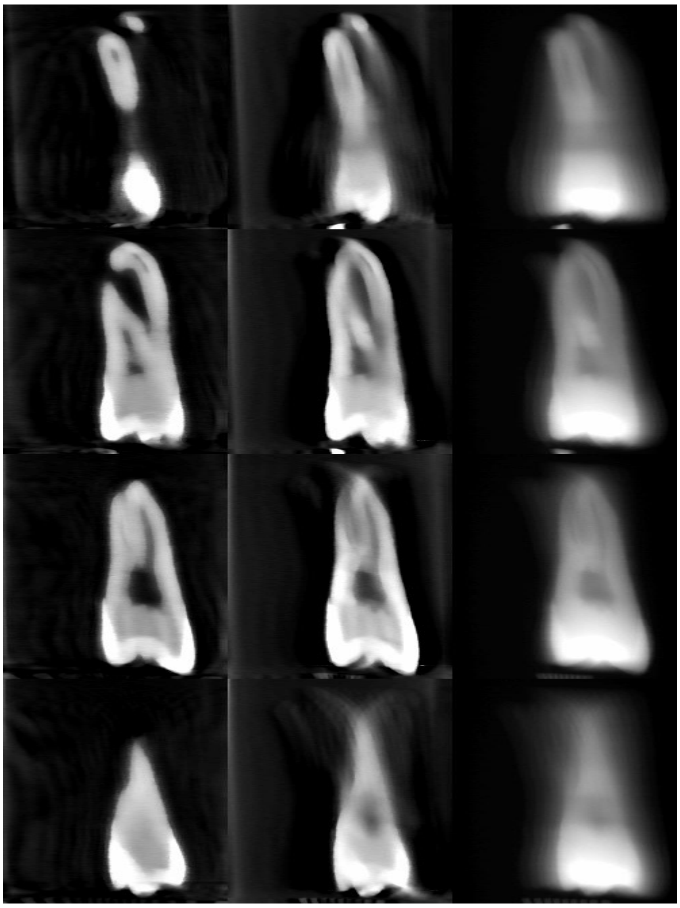

Figure 8. Left column shows four vertical slices of the (approximate) 3D reconstruction from full angle data (23 projections from view-angle of 187°) with the TV_β-prior. The other two columns show 3D reconstruction with the TV_β-prior (center column) and tomosynthetic reconstruction (right column) from limited angle data. In the limited angle case, nine projections from view-angle of 68° were used.

the tomosynthetic reconstructions give an experimental illustration for the analysis about features that are reconstructible based on limited angle projection data in [26]. A brief review of this analysis is given in section 2, part I of this paper. Given Figure 7, see also Figure 3 in part I of this paper.

We note that the test case in Figures 7-8 is unrealistic (from the clinical point of view) in the sense that the tooth phantom had no surrounding tissue whereas in practical situations this is always the case. However, the purpose of this example was to test the performance of our statistical model for limited-angle tomography with sparse projections without the added complications coming from the local tomography geometry. We consider a more realistic and complicated case in section 4.3, in which we consider reconstruction from limited angle data that was collected from a head phantom.

4.2. Example of MCMC-analysis for the tooth data

To give an illustrative example of more complete Bayesian inference, we conducted MCMC-analysis for one slice $x^{(j)}$ of the limited angle data that was used in section 4.1. Since the MCMC-analysis was conducted for only one slice, the coupling prior density $p_{L_1}(x^{(j)}|\hat{x}^{(j-1)})$ was not included in the posterior model.

The MAP-estimate $x^{(j)}_{\mathrm{MAP}}$ for the chosen slice with the approximate TV_β-prior is shown in the left image in Figure 9. The MAP estimate was computed using the Barzilai-Borwein method. The parameters α, σ_n^2 of the posterior density, the smoothing parameter β for the approximation of absolute value function and the extrior point search parameter sequence $\{\varsigma_t\}$ were the same that were used in the previous section.

Using the (approximate) MAP-estimate as the initial state in the simulation, we generated an ensemble of 15000 sample images using the Gibbs sampler [13]. For more detailed discussion on Gibbs sampling, see section 3.6, part I of this paper. It should be noted that the Gibbs sampler algorithm samples the original posterior model without any approximations to the TV and positivity priors. Detailed description of a similar algorithm that was applied to electrical impedance tomography problem can be found in [17, 19].

The image on the right in Figure 9 shows the conditional mean estimate $x^{(j)}_{\mathrm{CM}}$ computed as ergodic average based on the simulated Markov chain, see equation (24). The left image in the top row of Figure 10 shows the estimated variances for each pixel, that is, the diagonal entries of the posterior covariance matrix. Notice that the largest uncertainty in the posterior is in the directions corresponding to the pixels located at the boundaries of the tooth. The other plots in Figure 10 show the marginal posterior densities of single pixels marked in the variance image.

Figure 9. Statistical inference from the posterior distribution for one slice of the tooth data. Nine projections from view angle of 68° was used as the data. Left: MAP-estimate with the (smooth) TV_β-prior. Right: CM-estimate with the original TV-prior.

The proper interpretation of the results in Figures 9-10 require care. For example, given a set of 100 realisations of the projection data from the same model, one would be tempted to say that (roughly) in 90 cases the 90% confidence limits would include the true value of the X-ray attenuation coefficient.

However, this interpretation would be incorrect. The key point here is that the posterior distribution reflects our uncertainty based on the i) projection data and ii) the prior information. The pitfall here is that the true attenuation coefficient may have

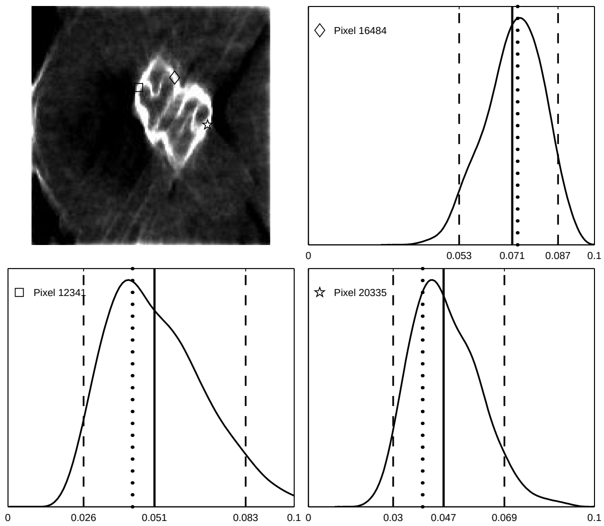

Figure 10. Statistical inference from the posterior distribution for one slice of the tooth data. Nine projections from view angle of 68° was used as the data. Top left: Estimated posterior variance for each pixel (i.e., the diagonal entries of the posterior covariance matrix). Top right and bottom : Marginal densities of single pixels, which are marked in the variance image. Solid line denotes the conditional expectation, dashed lines the 90% confidence limits and the dotted line the initial value (approximate MAP-value found by the Barzilai-Borwein method with TV_β-prior).

small probability with respect to the postulated prior model. The ill–posedness of the problem with sparse tomographic data necessitates that the priors are informative with respect to certain subspaces from which the likelihood (i.e. the projection data) carries only little information. This problem is reflected in the fact that the results are usually sensitive to the selection of the prior. Summarizing, the confidence limits are reliable only in relation to our confidence in the chosen prior. Thus, the choice and construction of the prior model is a crucial step in statistical inversion. See also the discussion about visualization of priors in section 3.4, part I of this paper.

4.3. Limited angle reconstruction of the head phantom

As the second test problem for limited angle tomography for data type (A), we consider reconstruction based on sparse truncated projection data from a head phantom. Using the measurement setup that is explained in section 2.3, seven projection images were taken with approximately equal projection intervals from total view angle of 59.7°. This represents roughly the maximum view angle that can be used in practice. Left image in Figure 11 shows one raw projection image from this data set, the middle image shows one row of detected pixel values from the projection image and the right image shows one slice $m^{(j)}$ of the transformed data in sinogram form. The size of the vector $m^{(j)}$ is $N = 6104$. The projection images were transformed to line integral data with the approximation explained in section 2.1. With the fixed detector geometry used in this example this means that the possible angular dependency of the efficiency of the Sigma detector is neglected.

Projection angles were computed based on the images of the reference ball that was attached in front of teeth with distance of 14mm from the detector array. This metal ball is seen above the middle tooth in left image in Figure 11. The shift of the ball in the projection images was measured, and using this information and known ball-detector distance the angles were obtained by simple trigonometric relations.

Note that in this more realistic example, all the teeth that are imaged are not visible in all projection images. This is also evident from the sinogram which is truncated from the lower side (i.e, it does not go to zero in the lower side). Thus, in addition of being a limited angle case, the problem contains features of local tomography problem.

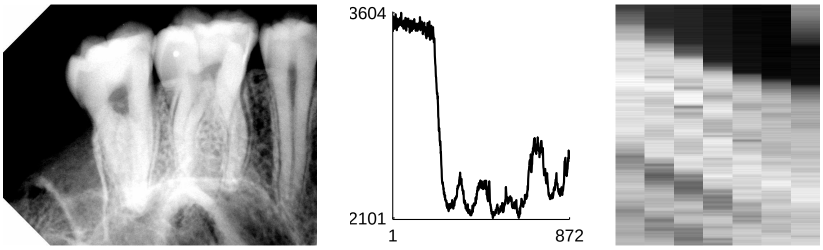

Figure 11. Left: Raw 664×872 projection image from the head phantom (black correspond to high counts). Middle: Detected pixel values for the 300^{th} row of the projection image. Right: Sinogram for the 300^{th} horizontal slice. The data consisted of 7 projections that were collected from 59.7° angle of view. Note that the sinogram does not go to zero from the lower side. This reflects the fact that the problem has features of local tomography.

The results for the head phantom case are shown in Figure 12. The left column shows four vertical slices of a tomosynthetic reconstruction and the right column shows respective slices of an approximate 3D reconstruction that was obtained as stack of $N_{\text{sli}} = 664$ two-dimensional MAP estimates with the TV_β-prior. In the computation of the 2D slices, the width of the two dimensional rectangular domain $\Omega \subset \mathbb{R}^2$ was 61.25mm and the depth was 25mm, respectively. The domain Ω was divided into a $M = 400 \times 160 = 64000$ pixel grid, leading to pixel size of $0.153 \times 0.156\text{mm}^2$ (width$\times$depth). The MAP-estimates were computed using the methods described in section 3. The posterior parameters were the same as in section 4.1.

As can be seen, the statistical approach with the TV_β-prior yields good reconstructions in this more realistic and difficult test case. The effect of well chosen prior is also seen clearly in Figure 12: Statistical inversion can capture relatively accurately the three dimensional structure of the teeth despite the poor depth information content of the limited angle projection data. Further, slices from the statistical reconstruction are sharper than tomosynthetic slices.

4.4. Full-angle tomography with sparse projection data

As the first test problem for data type (B) we consider full-angle global tomography problem with sparse projection data. The data was collected with same imaging geometry that is explained in section 2.2 and from the same tooth phantom that was used in section 4.1.

The results for this example are shown in Figure 13. Each image is a reconstruction of the same two-dimensional slice. The domain Ω and the number of pixels in the images are the same as in Figure 7, that is, the domain is $26 \times 26\text{mm}^2$ and the number

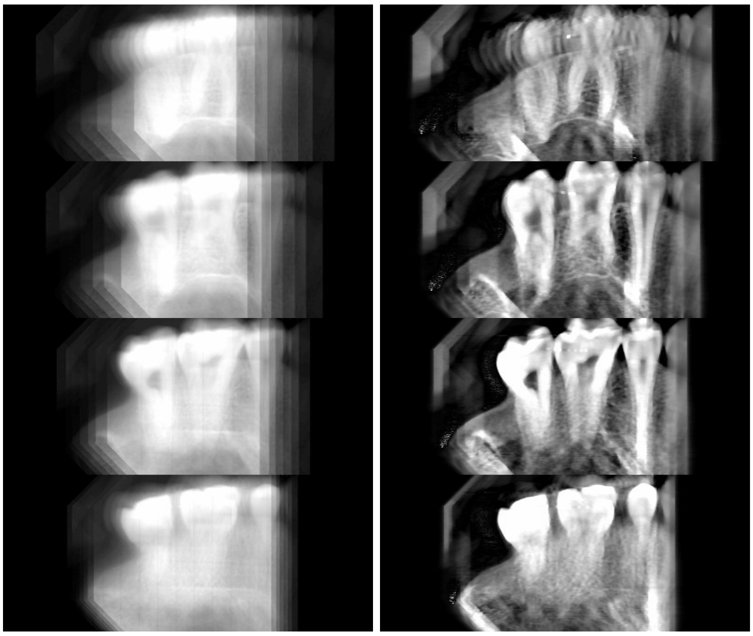

Figure 12. The left column shows four vertical slices of a tomosynthetic, or backprojected, reconstruction from limited angle data (seven projections from 59.7° view-angle). The right column shows respective slices of an approximate 3D reconstruction with the TV_β-prior from the same data.

of pixels is 166×166 with pixel size $0.16 \times 0.16\text{mm}^2$. The data that was used for the reconstructions in the first column of Figure 13 was collected from a view-angle of 204° with regularly spaced projection intervals of 8.5° (total 25 projections, size of data vector $m^{(j)}$ is $N = 16600$). In the second column, the total view-angle was the same but only 13 regularly spaced projections were used, leading to a projection interval of 17° ($N = 8632$). In the third column number of projections was 7 with projection interval of 34° ($N = 4648$) and in the fourth column the number of projections was 5 with projection interval of 51° ($N = 3320$).

The top row shows reconstructions with the filtered back projection (FBP) algorithm and the bottom row MAP estimates with the same statistical model that was explained in section 3 and used in Figures 7-8. The noise covariance and prior parameters for the statistical method were the same that were used in Figures 7-8.

For the FBP-reconstructions, the 664 element projection data vectors for each slice were averaged into bins of four data points, leading to data vector of 166 elements in each projection. With this operation, the pixel size for the FBP reconstruction becomes the same that was used in the statistical approach. Further, this operation improves the signal to noise ratio in the data for FBP, leading to less noisy reconstructions with the cost of reduced resolution. To de-emphasize the effects of (high-frequency) observation noise, Hanning-window was applied to the filtering of the projections in the frequency domain. Nearest neighbour interpolation was used in the backprojection process. For details on theory and implementations of FBP-methods, see e.g. [18, 23] and references therein.

As can be seen from Figure 13, the statistical approach provides good

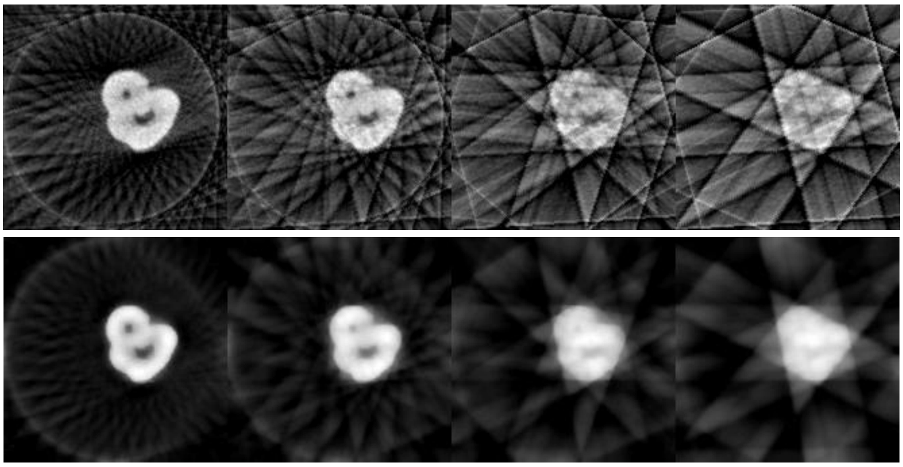

Figure 13. Full-angle (204°) reconstructions using different amount of projections. Columns from left to right: 25 projections (projection interval 8.5°), 13 projections (17°), 7 projections (34°) and 5 projections (51°). Top row: Filtered back projection. Bottom row: MAP-estimates with the TV_β-prior.

reconstructions of the tooth. The MAP-estimate in the second column with projection interval 17° is almost as good the first one with projection interval of 8.5°. Whereas the MAP-estimate in the third column (projection interval 34°) provide useful information about the shape and size of the tooth, the fourth one (angular projection interval 51°) gives only a crude approximation for the size and shape of the target. Also, it can be seen that the MAP-estimates with the statistical model are less noisy than the reconstructions with the FBP-method. It is evident from Figure 13 that the statistical approach is more robust against large projection interval than the FBP method.

4.5. Local tomography from sparse projection data

As the last test case we consider a realistic example of extra-oral imaging using full-angle sparse projection data from a jaw bone phantom. This is an example of data type (B) with truncated projections. The cone beam measurement geometry for these experiments is illustrated in Figure 5. Using the experimental setup explained in section 2.4, we took 23 equally spaced projection images with total view angle of 187° (projection interval 8.5°) from the jaw bone phantom. Three of these projection images are shown in Figure 14. Referring forward to the results in Figure15, the projection image on left in Figure 14 was taken from the direction of positive x-axis (i.e, from right to left) with respect the reconstructed slice. The other two projections images in Figure 14 are from angles of 76.5° and 153° to counterclockwise direction with respect the positive x-axis in the reconstructed slice, respectively.

The left image on bottom row in Figure 14 shows one slice of this projection data in sinogram form. In the sinogram that is shown in the right image on the bottom row, only 12 projections with projection interval of 17° were used. As can be seen from Figure 14, this test case leads to local tomography problem with full-angle sparse projection data.

The results for this example are shown in Figures 15-16.

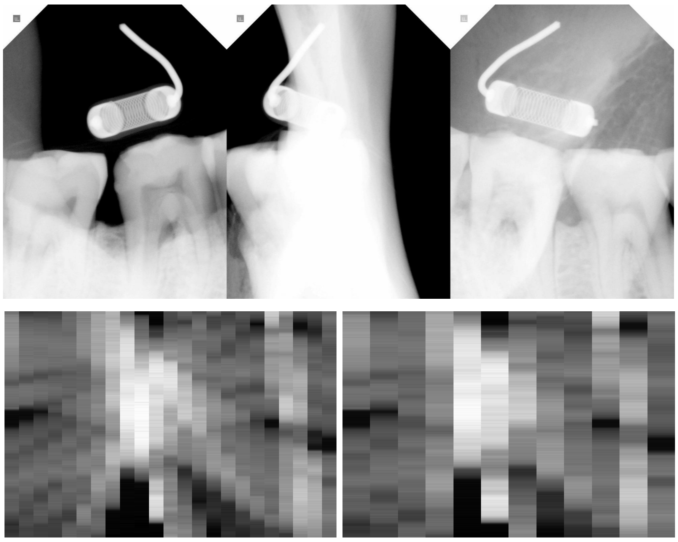

Figure 14. Top row: Three 872×664 projection images from the jaw phantom. The images are (approximately) from orthogonal directions. Bottom Row: Sinograms for one slice of the jaw phantom data with projection intervals of 8.5° (left) and 17° (right). The projection images and the sinograms reveal clearly the local tomography nature of the reconstruction problem.

In Figure 15 each image is a reconstruction of the same two-dimensional slice. The domain $\Omega \subset \mathbb{R}^2$ in the images is $78 \times 78\text{mm}^2$ square which was divided into a regular $498 \times 498 = 248004$ pixel lattice with pixel size $0.156 \times 0.156\text{mm}^2$. The data that was used for the reconstructions in the top row of Figure 15 was collected from a view-angle of 187° with regularly spaced projection intervals of 8.5° (total 23 projections, number of data $N = 15272$). In the bottom row, the total view-angle was the same but only 12 regularly spaced projections were used, leading to a projection interval of 17° ($N = 7968$).

The left column shows reconstruction with the Λ-tomography, middle column the tomosynthetic (backprojection) reconstructions and the right column MAP estimates with the statistical model that was explained in section 3 and used in earlier test cases. The parameters for the statistical method were the same that were used in Figures 7-8. The idea of Λ-tomography is discussed briefly in Appendix A.

Figure 16 shows the central part (166×166 pixels) that includes the region of interest (ROI) for the respective reconstructions in Figure 15.

As can be seen from Figures 15-16, the MAP-estimates with the TV_β-prior are relatively good also in this difficult test problem of extraoral imaging. When using all the 23 projections, the structure of the teeth that are located in the region of interest was recovered with good accuracy. In the case of using only 12 projections with projection interval of 17°, the structure of the same teeth was recovered with almost as good accuracy. A notable and typical local tomography feature in the reconstructions are the "back projection artifact" type details outside the region of interest. These are evidently due to tissues that are visible possibly only in one or two projection images. Also, as can be seen from Figures 15-16 the images that are based on the statistical approach are better than the traditional backprojection

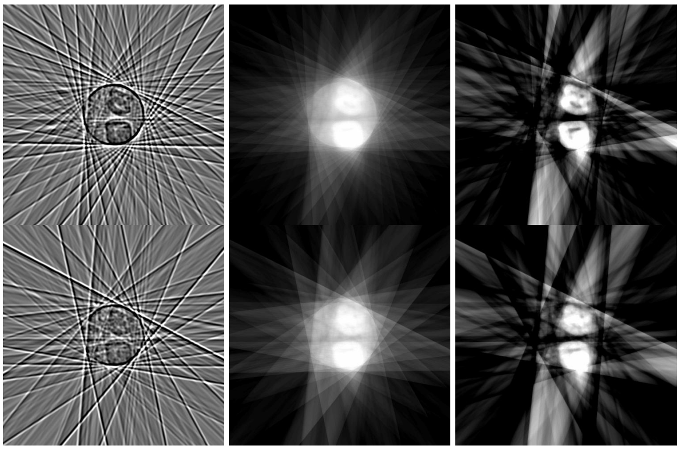

Figure 15. Reconstructions from local sparse projection data (total view-angle of 187°) from the jaw bone phantom. Top row: 23 projections with projection interval of 8.5°. Bottom row: 12 projections with projection interval of 17°. Columns from left to right: Λ-tomography, tomosynthesis (backprojection) and MAP-estimates with the TV_β-prior.

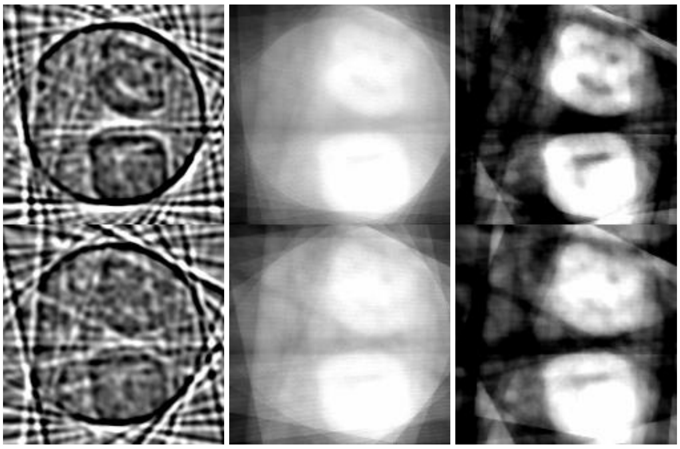

Figure 16. Regions of interest (ROI) from the local tomography reconstructions in Figure 15. Rows and columns are as in Figure 15.

reconstructions or the Λ-tomography reconstructions. Based on Figures 15-16, it seems that the statistical model would be useful in clinical studies of extraoral imaging.

5. Conclusions

Consider the following example of three-dimensional X-ray imaging. A dentist wants to know whether the roots of a certain tooth are close to the inferior dental canal. This is related to the risk of damaging nerves when removing the tooth. Often a single intraoral radiograph is not enough for answering this question due to overlapping of structures. So he takes, say, five digital intraoral projection radiographs using a X-ray source and a digital intraoral X-ray sensor choosing the directions of the images so that the images of the roots and the nerve canal are clearly separate in some of the images. The projection images together with knowledge of imaging geometry are given as input to a reconstruction algorithm. Resulting three-dimensional reconstruction is examined on computer screen and the diagnostic question answered.

The above type of three-dimensional imaging is not standard practice today. One reason for this is the lack of a flexible, fast, high-quality reconstruction algorithm for such imaging. It is evident from part I of this paper that such an algorithm should be able to use *a priori* information of the tissue to compensate for the incomplete information provided by few radiographs.

In this paper, we proposed a novel statistical model to three dimensional dental X-ray imaging with sparse projection data. In the model, the three dimensional reconstruction problem is approximated with a stack of two dimensional problems. Our model for *a priori* information includes total variation and positivity priors for each two dimensional slice, and the three dimensional nature of the problem is taken into account through a coupling L^1-prior between consecutive slices. A gradient-based optimization method was implemented for the computation of the MAP-estimates and a MCMC-algorithm for the computation of point estimates that necessitate integration. The performance of the model was evaluated based on *in vitro* projection data that was collected using a X-ray source and intraoral CCD-detector from a dentist's regular equipment. Reconstructions with traditional reconstruction methods were given as reference for the estimates with the statistical model. Four different test cases with sparse projection data were considered. It was seen that the statistical approach gave good results in all test cases. Furthermore, the statistical model gave improvement over traditional methods in all cases. Thus, the proposed statistical model seems promising for 3D imaging of dentomaxillofacial structures with sparse projection data.

In this study, the approximation of the three dimensional problem by a stack of two dimensional problems was made due to heavy computational demands of purely three dimensional case. The computation of the approximate 3D reconstructions that are shown in Figures 7-8 took approximately 6 hours using MatLab (version 5) on a PC computer with a 1GHz Pentium processor (number of unknowns for each 2D slice $M = 27556$). For the approximate 3D reconstruction in Figure 12 the respective time was approximately 8 hours (number of unknowns for each 2D slice $M = 64000$). However, it is our belief that these computational times can be reduced to class of a few minutes with an optimized implementation on a more basic level programming platform.

The computation of the purely three dimensional reconstructions (number of unknowns $M \gg 10^6$) with our current implementation and devices is not possible due to the excessive memory requirement of matrix A. However, the extension of the methods to purely 3D reconstruction is one of the main topics in the future work. This work is under way.

The development of more effective MCMC-codes is also a topic of future work. These are more likely to be realizable in cases in which the prior model and structure of the tissue can be well described in a lower dimensional parametric basis. The approximations of different tissues by low dimensional parametric models is one topic of future work.

We also wish to start *in vivo* tests with the proposed approach in near future. In addition to dental imaging, we believe that the proposed statistical model can also prove to be useful in other applications with "level set" type targets.

Acknowledgments

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

Appendix A. Λ-tomography

A traditional reconstruction method that has been developed for local tomography is the so-called Λ-tomography. As in the main body of this paper, let $s \in \Omega \subset \mathbb{R}^2$ denote the position vector and $x : \Omega \mapsto [0, \infty)$ denote the X-ray attenuation coefficient function. The idea of Λ-tomography is based on the result that it is possible to recover Λx, where Λ is a Calderón operator, instead of the attenuation coefficient x itself from continuous local full-angle data [29, 20, 9, 10].

The Calderón operator Λ is defined using the Fourier transform as

$$\widehat{\Lambda x}(\xi) = |\xi| \hat{x}(\xi),$$

where $\hat{x}(\xi) = \int_{\mathbb{R}^2} \exp(-is \cdot \xi) x(s) \mathrm{d}s$ for any $\xi \in \mathbb{R}^2$. This is satisfactory since Λ acts very much like a differential operator and enhances jumps (edges) of x. Further, it does not introduce sharp artifacts, only blurred ones. However, values of the attenuation coefficient x cannot be read from Λx, only the jumps are faithfully recovered.

Based on the article [29], Kenrick Bingham [2, Formula (3.33)] wrote the reconstruction formula as

$$e * (\Lambda x) = C \int_0^{2\pi} \int_0^{2\pi} \Delta P_\theta e(E_\theta(s-a))(\mathcal{D}_a x(\theta) + \mathcal{D}_a x(-\theta))|a \cdot \vec{\theta}| \mathrm{d}\phi \mathrm{d}\theta, \qquad (A.1)$$

where $a = a(\phi) = R(\cos\phi, \sin\phi)$ is the location of the X-ray source, C is a constant, e is a smooth point spread function that approximates the Dirac delta function, Δ is the Laplace operator, P_θ is the parallel beam tomographic data (i.e., transformed projection radiograph), angle θ defines a unit vector $\vec{\theta} = (\cos(\theta), \sin(\theta))$ which in turn specifies the propagation direction of the X-rays (travelling from point a to direction $\hat{\theta}$), $\mathcal{D}_a$ is the divergent beam (or fan-beam) tomographic data, $\vec{\theta}$ is a unit vector parametrized by angle θ, E_θ is orthogonal projection onto $\vec{\theta}^\perp$ and $*$ denotes two-dimensional convolution.

The use of the point spread function e eliminates infinite values of Λf. Let $e : \mathbb{R}^2 \to \mathbb{R}$ be the radial function defined by

$$e(r) = \begin{cases} \frac{\pi}{5}(r+1)^4(r-1)^4 & \text{for } 0 \le r \le 1, \\ 0 & \text{for } 1 < r. \end{cases} \qquad (A.2)$$

Note that $\int e = 1$. The function $\Delta P_\theta e$ can be computed explicitly.

Appendix B. Reconstruction of singularities with the TV_β-prior

In this appendix we consider singularities appearing in the MAP-estimate with the TV_β-prior. The purpose is to show that the TV_β-prior does not destroy any of the singularities that are reconstructible based on the limited angle data alone.

We consider the continuous model, where $x = x(s)$ is the attenuation function defined in domain $\Omega \subset \mathbb{R}^3$ and the measured data corresponds to the line integrals

$$m(L) = \mathcal{A}x(L) + \epsilon(L), \quad \mathcal{A}x(L) = \int_L x(s)\mathrm{d}s.$$

Here the variable L is a line connecting a source to detector D, where D is assumed to be a subset of a plane. We denote by G the set of lines L along which we can do measurements and introduce coordinates on G by using the point $(s_1, s_2, s_3) \in D$ where L intersect D and two angles (α_1, α_2) related to the direction of the line L. Then $\mathcal{A}$ defines a continuous operator $\mathcal{A} : L^2(\Omega) \to L^2(G)$ where G has a measure $\mu = \mathrm{d}s_1\mathrm{d}s_2\mathrm{d}s_3\mathrm{d}\alpha_1\mathrm{d}\alpha_2$. This operator has adjoint $\mathcal{A}^* : L^2(G) \to L^2(\Omega)$. The composition of these operators defines the unfiltered backprojection operator $\mathcal{A}^*\mathcal{A}$: Indeed, using unfiltered backprojection algorithm with limited angle data $m = \mathcal{A}x$ corresponds to computation of $\mathcal{A}^*\mathcal{A}x$ [23].

It is well known that certain singularities of x can be seen in backprojection algorithm [26], see also the review in section 2, part I of this paper. For instance, assume that function x has a jump across surface S. If some line $L \in G$ is tangent to the surface S at point s, then the backprojection reconstruction $\mathcal{A}^*\mathcal{A}x$ is also singular at point s (in fact, the singularity is not so strong as the original singularity).

Mathematically speaking, we say that the pair of a point s and a direction ξ is in the wave-front set of function x if the function x is singular at point s in direction ξ. This is denoted by $(s, \xi) \in WF(x)$ (for presice definition, see [16]). For instance, if x is a piecewise smooth function that jumps across a surface S, then wave front set of x consist of pairs (s, ξ) where $s \in S$ and ξ is normal vector of S at s. The points s which do not have a neighborhood where x is infinitely differentiable are called singular points and their set is called singular support of x and denoted by $\mathrm{singsupp}(x)$.

Let now $H \subset \Omega$ be set of those points s for which there is $(s, \xi) \in WF(x)$ such that ξ is orthogonal to some line $L \in G$. In other words, $H \subset \mathrm{singsupp}(x)$ is set of those points where some measurement line is tangent to a "jump" of function x. We call H the set of the observable singularities.

It is known that the backprojection algorithm can reconstruct observable singularities,

$$H \subset \mathrm{singsupp}(\mathcal{A}^*\mathcal{A}x),$$

see e.g. [26]. Next we show that the same property is true for MAP-reconstruction with TV_β-prior. Assume that we have obtained (virtually errorless) measurements from an attenuation function $x_0 \in L^2(\Omega)$. This means that we are given $m = \mathcal{A}x_0$.

We recall that the MAP estimate is obtained from the minimization problem

$$\min_x F(x)$$

where x is compactly supported function in $\Omega \subset \mathbb{R}^3$ and

$$F(x) = \int_G (\mathcal{A}x(L) - m(L))^2 \, \mu(\mathrm{d}L) + \sum_{j=1}^{3} \int_\Omega h_\beta(\partial_j x(s)) \, \mathrm{d}s$$

where $\partial_j x = \frac{\partial x}{\partial s_j}$ are partial derivatives of $x(s)$ and h_β is defined by equation (10). Let x be the function which minimizes $F(x)$. Then the first variation of F, which we next compute, must vanish at x. Let v be a function which vanish in $\partial\Omega$. Then using integration by parts,

$$\begin{aligned}
&\lim_{t\to 0}\frac{F(x+tv)-F(x)}{t}\\
&=\int_G 2(\mathcal{A}x-m)(L)\cdot\mathcal{A}v(L)\,\mu(\mathrm{d}L)+\sum_{j=1}^{3}\int_\Omega h_\beta'(\partial_j x(s))\partial_j v(s)\,\mathrm{d}s\\
&=\int_\Omega\left(2\mathcal{A}^*(\mathcal{A}x-m)(s)-\sum_{j=1}^{3}\partial_j(h_\beta'(\partial_j x(s)))\right)v(s)\,\mathrm{d}s
\end{aligned}$$

where h_β' is the derivative of the function $h_\beta:\mathbb{R}\to\mathbb{R}$. Since x minimizes $F(x)$, the above integral has to vanish for any function v. Thus we see that minimizer x satisfies

$$2\mathcal{A}^*\mathcal{A}x(s)-\sum_{j=1}^{3}\partial_j\left(h_\beta'(\partial_j x(s))\right)=2\mathcal{A}^*m(s)=2\mathcal{A}^*\mathcal{A}x_0(s).$$

Assume that $s\in H$. Then the right hand side $\mathcal{A}^*\mathcal{A}x_0$ is not smooth. Thus, if the minimizer x would be smooth at s, we see that the left hand side should be smooth which would be a contradiction. This shows that $s\in H$ implies also $s\in\mathrm{singsupp}(x)$, that is,

$$H\subset\mathrm{singsupp}(x).$$

In other words, at least all the discontinuities that are seen with the standard backprojection method can be seen with approximative TV priors. Note that above computation does not say anything about possible artifact singularities that may appear.

References

[1] Barzilai J and Borwein J M 1988 Two point step size gradient method *IMA J. Numer. Anal.* **(8)** 141–148

[2] Bingham K 1998 Mathematics of local x-ray tomography *Master's thesis*, Helsinki University of Technology. (http://www.math.hut.fi/ kenny/opi/diplomityo/dtyo.html)

[3] Brocklebank L 1997 *Dental Radiology — Understanding the X-Ray Image* Oxford University Press, ISBN 0-19-262411-3

[4] Dobson D C and Santosa F 1994 An image enhancement technique for electrical impedance tomography. *Inv Probl*, 10:317–334.

[5] Dobson D C and Santosa F 1996 Recovery of blocky images from noisy and blurred data. *SIAM J Appl Math*, 56(4):1181–1198.

[6] Dobson D C and Vogel C R 1997 Convergence of an iterative method for total variation denoising. *SIAM J Numer Anal*, 34:1779–1791.

[7] Donoho D L, Johnstone I M, Hoch J C, and Stern A S 1992 Maximum entropy and the nearly black object. *J Roy Statist Ser B*, 54:41–81.

[8] Ekestubbe A, Gröndahl K and Gröndahl H-G 1997 The use of tomography for dental implant planning. *Dentomaxillofacial Radiology* 26,206–213

[9] Faridani A, Ritman E L, and Smith K T 1992 Local tomography. *SIAM J. Appl. Math.* **52**(2) 459–84; Examples of local tomography **52**(4) 1193–1198 (A reorganization of the examples which became disorganized while the article was in press.)

[10] Faridani A, Finch D V, Ritman E L, and Smith K T 1997 Local tomography II *SIAM J. Appl. Math.*, **57**(4) 1095–1127

[11] Fiacco A V and McCormick G P 1990 Nonlinear programming: sequential unconstrained minimization techniques. SIAM.

[12] Gamerman D 1997 *Markov chain Monte Carlo - Stochastic simulation for Bayesian inference.* Chapmann & Hall.

[13] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell*, 6:721–741.

[14] Gilks W R, Richardson S, and Spiegelhalter D J 1996 *Markov Chain Monte Carlo in Practice.* Chapmann & Hall.

[15] Grant D G 1972 Tomosynthesis: A three-dimensional radiographic imaging technique. *IEEE Transactions on Biomedical Imaging*, 19(1):20–28

[16] Hörmander L *The analysis of linear partial differential operators. I.* Springer Study Edition. Springer-Verlag, Berlin, 1990. xii+440 pp.

[17] Kaipio J P, Kolehmainen V, Somersalo E, and Vauhkonen M 2000 Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems*, 16:1487–1522.

[18] Kak A C and Slaney M 1988 *Principles of Computerized Tomographic Imaging.* IEEE Press.

[19] Kolehmainen V 2001 *Novel Approaches to Image Reconstruction in Diffusion Tomography.* PhD thesis, University of Kuopio, Kuopio, Finland.

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

[21] Mozzo P, Procacci C, Tacconi A, Tinazzi Martini P and Bergamo Andreis I A 1998 A new volumetric CT machine for dental imaging based on the cone-beam technique: preliminary results *European Radiology*, **8**, 1558–1564

[22] Nair M K 1998 Tuned Aperture Computed Tomography and Detection of Recurrent Caries. *Caries Research*, **32**(1), 23–30

[23] Natterer F 1986 *The Mathematics of Computerised Tomography.* Wiley, New York.

[24] Persson M, Bone D and Elmqvist H 2001 Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine and Biology*, **46** 853–866

[25] 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 Maxillofacial Radiology* **93**, 341–349.

[26] 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

[27] Ruttimann U E, Groenhuis R A J, and Webber R L 1984 Restoration of digital multiplane tomosynthesis by a constrained iteration method. *IEEE transactions on medical imaging*, 3(3):141–148.

[28] Schafer R W, Merserau R M, and Richards M A 1981 Constrained iterative restoration algorithms. *Proc. IEEE*, 69:432–450.

[29] Smith K T and Keinert F 1985 Mathematical foundations of computed tomography. *Applied Optics* **24**(23) 3950–7

[30] Sukovic P, Brooks S, Perez L and Clinthorne N 2001 DentoCAT - A Novel Design of a Cone Beam CT Scanner for Dentomaxillofacial Imaging: Introduction and Preliminary Results In *CARS 2001, editors H U Lemke, M W Vannier, K Inamura, A G Farman and K Doi*, 659–664.

[31] Vassilevski P S and Wade J G 1997 A comparison of multilevel methods for total variation regularization. *El. Trans Num Meth*, 6:255–270.

[32] 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.

[33] 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 *Dentomaxillofacial Radiology* **26**, 53–62

[34] Webber R L, Horton R A, Underhill T E, Ludlow J B, and Tyndall D A: Comparison of film, direct digital, and tuned aperture computed tomography images to identify the location of crestal defects around endosseous titanium implants, *Oral Surgery Oral Medicine Oral Pathology* **81**(4),480–490, 1996

[35] 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 Surgery Oral Medicine Oral Pathology* 88,239–47.

[36] Ziedses des Plantes 1932: Eine neue Method zur Differenzierung der Röntgenographie, *Acta Radiol.* **(13)** 182–192