Statistical X-ray tomography using empirical Besov priors

Simopekka Vänskä[1], **Matti Lassas**[2] **and Samuli Siltanen**[3]

[1]Rolf Nevanlinna Insitute
Department of Mathematics and Statistics
University of Helsinki
P.O.Box 68, 00014 UNIVERSITY OF HELSINKI, FINLAND
simopekka.vanska@helsinki.fi

[2]Department of Mathematics
Helsinki University of Technology
P.O.Box 1100, FI-02015 TKK, FINLAND
matti.lassas@tkk.fi

[3]Department of Mathematics
Tampere University of Technology
P.O.Box 553, 33101 TAMPERE, FINLAND
samuli.siltanen@tut.fi

ABSTRACT

Wavelet-based Besov space prior models for X-ray tomography are studied using the empirical Bayes approach. The hyperparameters for the prior models are estimated from statistical properties of the wavelet coefficients of measured X-ray projection images (which are related to the smoothness of the attenuation coefficient). Various statistical models for the wavelet coefficients are studied. Experiments using measured *in vitro* data suggest that the hyperparameters can be estimated with a simple method, leading to automated choice of prior parameters and improved tomographic reconstruction.

Keywords: tomography, wavelets, Besov space, prior.

2000 Mathematics Subject Classification: 65R32, 62P10.

1 Introduction

The aim of X-ray tomography is to reconstruct the inner structure of an object given a collection of projection images of the object. The measurement can be modeled with the equation

$$m = Af + \epsilon, \tag{1.1}$$

where f is a discrete representation of the X-ray attenuation coefficient, the matrix A comes from the pencil beam model, m is a vector containing all pixel values of the measured digital X-ray images, and ϵ is white noise with standard deviation $\sigma > 0$, (Natterer, 2001). For simplicity,

we study reconstructing two-dimensional objects from one-dimensional projections. However, the techniques we present apply for the three-dimensional case as well.

Depending on the number and geometry of the projection images, the inverse problem of reconstructing f from m is either mildly or severely ill-posed. Bayesian inversion is a flexible and powerful tool applicable to any collection of data. The solution of the inverse problem is the *posterior distribution* given by the Bayes formula:

$$\pi_{\text{post}}(f) \propto \pi(m \mid f)\pi_{\text{pr}}(f),$$

where the likelihood distribution $\pi(m \mid f)$ is derived from equation (1.1) using noise statistics:

$$\pi(m \mid f) = c \exp(-\frac{1}{2\sigma^2}\|Af - m\|_2^2).$$

All available *a priori* knowledge about f is modeled mathematically as the *prior distribution* $\pi_{\text{pr}}(f)$. This additional information complements the insufficient measurement data and enables stable solution of the ill-posed tomography problem by computing the *maximum a posteriori* estimate

$$f_{\text{MAP}} := \text{argmax } \pi_{\text{post}}(f). \tag{1.2}$$

See (Siltanen, Kolehmainen, Järvenpää, Kaipio, Koistinen, Lassas, Pirttilä and Somersalo, 2003) for a review and (Hanson and Wecksung, 1983; Sauer, James and Klifa, 1994; Frese, Bouman and Sauer, 2002; Kolehmainen, Siltanen, Järvenpää, Kaipio, Koistinen, Lassas, Pirttilä and Somersalo, 2003) for examples of the Bayesian methods in tomography.

We discuss empirical Bayes approaches for constructing priors for X-ray tomography: hyperparameters of the prior are estimated from data and then used in inversion. Prior distributions based on Besov norms are known to be effective for medical X-ray tomography (Rantala, Vänskä, Järvenpää, Kalke, Lassas, Moberg and Siltanen, 2006; Niinimäki, Siltanen and Kolehmainen, 2007; Lassas, Saksman and Siltanen, 2008), but the question of choosing the parameters $s \in \mathbb{R}$ and $1 \leq p, q \leq \infty$ of the space B^s_{pq} is largely open. Our new results can be used for choosing these parameters. See (Lee and Lucier, 2001) for a different wavelet vaguelette decomposition based approach which is applicable in the full angle dense data setting.

We make use of the fact that Besov space norms are conveniently expressed as weighted sums of wavelet coefficients. Hence, the finiteness of the Besov space norm can be considered by studying the statistics of the wavelet coefficients. Several distributional models for the wavelet coefficients have been proposed in the literature in the context of image and signal processing (Donoho and Johnstone, 1994; Donoho and Johnstone, 1995; Abramovich, Sapantinas and Silverman, 1998; Chang, Yu and Vetterli, 2000; Boubchir and Fadili, 2006), but it is unclear which distribution model fits to the X-ray images. In this paper, we consider three distribution models:

(a) **Threshold distribution model** (TDM) of Abramowitz, Sapatinas and Silverman (Abramovich et al., 1998). At each wavelet scale, a fixed percentage of wavelet coefficients vanish and the rest are normally distributed.

(b) **Stable distribution model** (SDM) in which the sparseness of the wavelet coefficients is captured with the long tailed α-stable distributions, (Achim, Bezerianos and Tsakalides, 2001; Boubchir and Fadili, 2006).

(c) **Decaying moments model** (DMM) in which the p-moments of random wavelet coefficients are assumed to decay at finer wavelet scales. No particular shape for the wavelet distributions is assumed.

The model parameters are assumed to decay at finer wavelet scales. We prove that for all these models the speed of the decay characterizes in which Besov space the corresponding function belongs to. (Our proof differs from the proof that is given for (a) in (Abramovich et al., 1998).) We estimate the decay parameters of the models from noisy projection data, which gives then the Besov smoothness for the projection images. The prior is chosen by using a relationship between the Besov smoothness of the object and its projections.

We demonstrate our empirical prior with both simulated and measured data. Our computations seem to give very good estimates for the Besov smoothness of the attenuation coefficient, solving the problem of choosing Besov priors and leading to high quality reconstructions.

2 Wavelet notation

Let ϕ, and ψ, be the scaling function, and the mother wavelet, respectively, that are $L^2(\mathbb{R})$ orthonormal and compactly supported in $\mathbb{R}$, (Daubechies, 1992; Meyer, 1992). On scale j, the corresponding wavelets are

$$\psi_j(x) = 2^{j/2}\psi(2^j x), \quad \phi_j(x) = 2^{j/2}\phi(2^j x),$$

and they are orthonormal as well.

In dimension n, there are 2^n different type of parent wavelets. Denote them by

$$\psi^\ell(x) \;=\; \prod_{i=1}^{n} \psi^{\ell_i}(x_i), \quad x \in \mathbb{R}^n, \tag{2.1}$$

where ℓ is a binary vector

$$\ell = (\ell_1, \ldots, \ell_n) \in \{0,1\}^n$$

and

$$\psi^0 = \phi, \qquad \psi^1 = \psi.$$

The parent of type $\ell = (0, \ldots, 0) = 0$ is called the scaling function (or the father wavelet), and the other $2^n - 1$ types are called mother wavelets (or the detail wavelets) of type ℓ, $\ell \neq 0$. On scale j, the wavelet of type ℓ is

$$\psi_j^\ell(x) \;=\; 2^{jn/2}\psi^\ell(2^j x). \tag{2.2}$$

Denote the transferred wavelets on scale j by

$$\psi_{j,k}^\ell(x) \;=\; 2^{jn/2}\psi^\ell(2^j x - k), \quad k \in \mathbb{Z}^n. \tag{2.3}$$

Let $D =]0,1[^n \subset \mathbb{R}^n$. Let $\mathcal{K}_{j,n}$ be the set of indices of those detail wavelets whose supports intersect with D on scale j. Now, the number of indices in $\mathcal{K}_{j,n}$,

$$K_{j,n} \quad = \quad \#\mathcal{K}_{j,n}, \tag{2.4}$$

is proportional to $(2^n - 1)2^{jn}$.

By the orthogonality of the wavelets, a function $f \in L^2(D)$ can be expressed with the wavelet expansion

$$f \quad = \quad \sum_{k \in \mathbb{Z}^n} w^0_{j_0,k} \psi^0_{j_0,k} + \sum_{j=j_0+1}^{\infty} \sum_{\ell \neq 0, k \in \mathbb{Z}^n} w^\ell_{j,k} \psi^\ell_{j,k}, \tag{2.5}$$

where $w^\ell_{j,k}$ are the wavelet coefficients,

$$w^\ell_{j,k} \quad = \quad \int_D f \, \psi^\ell_{j,k} \, dx. \tag{2.6}$$

In practice, the wavelet coefficients are computed with a discrete convolution with finite filters.

Assume that the wavelets are r-regular, which means that the wavelets have r vanishing moments and r continuous derivatives. If r is sufficiently large, then the Besov space norm in $B^s_{p,q}$ can be expressed with the wavelet coefficients by

$$\|f\|^q_{B^s_{p,q}} \quad = \quad \|(w^0_{j_0,k})\|^q_p + \sum_{j=j_0+1}^{\infty} 2^{jq(s+n(\frac{1}{2}-\frac{1}{p}))} \left(\sum_{\ell \neq 0, k} |w^\ell_{j,k}|^p \right)^{q/p}. \tag{2.7}$$

Since the finiteness of the Besov norm is determined from the tail $j \to \infty$, we drop the sum of the scaling coefficients away from the Besov norm expression from now on. Also, for simplicity, we consider the case $p = q$ and denote

$$B^s_p = B^s_{p,p},$$

see discussion in (Lee and Lucier, 2001).

3 Wavelet distribution models and Besov spaces

Assume that f is represented by its wavelet expansion (2.5). We assume that a wavelet coefficient $w^\ell_{j,k}$ is a realization of the random variable $W^\ell_{j,k}$ and that $W^\ell_{j,k}$ is distributed according to some probability density function $\pi^\ell_{j,k}$,

$$W^\ell_{j,k} \sim \pi^\ell_{j,k}.$$

In (Abramovich et al., 1998), the wavelet coefficients are distributed as

$$W^\ell_{j,k} \quad \sim \quad \tau_j \phi_{\sigma_j^2} + (1 - \tau_j)\delta_0 \tag{3.1}$$

where

$$\phi_{\sigma^2}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{t^2}{\sigma^2} \right)$$

is the normal distribution with variance σ^2. One may think this so that with probability $1-\tau_j$ the wavelets on scale j have been thresholded to zero, and the rest are normally distributed. This is the motivation for us to call this model as the *threshold distribution model*. The criteria for f in belonging to some Besov space is derived from the asymptotics of the distribution parameters σ_j and τ_j when the scale j grows. Our proof is in the n dimensional case and it is different from the proof of (Abramovich et al., 1998).

Theorem 3.1. *Assume that $W_{j,k}^{\ell}$ follow the threshold distribution model* (3.1)*, where the parameters satisfy*

$$\sigma_j = C_\sigma 2^{-\alpha j}, \qquad \tau_j = \min(1, C_\tau 2^{-\beta j}) \tag{3.2}$$

for $j \geq j_0$ with some constants C_σ, C_τ, α, and β, $\beta < n$. If f is given by the wavelet expansion (2.5)*, then $f \in B_p^s(D)$ almost surely if and only if*

$$s + \frac{n}{2} < \alpha + \frac{\beta}{p}, \tag{3.3}$$

Proof. Without loss of generality we may assume

$$s + \frac{n}{2} \quad \leq \quad \alpha + \frac{\beta}{p}, \tag{3.4}$$

because if $f \notin B_p^s(D)$ for $s = \alpha + \frac{\beta}{p} - \frac{n}{2}$, then $f \notin B_p^{s'}(D)$ for $s' > s$.

We study the finiteness of the Besov norm (2.7) with Lemma A.1. Denote

$$X_i \quad = \quad |W_{j,k}^{\ell}|^p, \quad i = 1, 2, \ldots, \tag{3.5}$$

where the indeces are reordered as

$$j = j(i), \;\; k = k(i), \;\; \ell = \ell(i),$$

and satisfy

$$j(i_1) \leq j(i_2)$$

for $i_1 < i_2$. It holds

$$j(i) \to \infty \quad \text{as} \quad i \to \infty.$$

We check that the assumptions (A.1) and (A.2) are valid for

$$Y_i = 2^{jp(s+n(\frac{1}{2}-\frac{1}{p}))} X_i$$

with $q = 2$ and some r, $1 < r < 2$. Now,

$$\mathbb{E}(X_i) = \tau_j \sigma_j^p \nu_p, \quad \mathbb{E}(X_i^2) = \tau_j \sigma_j^{2p} \nu_{2p},$$

where

$$\nu_p = \int_{-\infty}^{\infty} |t|^p \, \phi_1(t) dt$$

is the p normal moment of the normal distribution. By substituting the threshold distribution model parameters, we get

$$\mathbb{E}(Y_i) \propto 2^{jp(s+\frac{n}{2}-\frac{n}{p}-\alpha-\frac{\beta}{p})}, \quad \mathbb{E}(Y_i^2) \propto 2^{j2p(s+\frac{n}{2}-\frac{n}{p}-\alpha)-\beta j}.$$

Now, the assumption (A.1) is valid by (3.4). To check (A.2), use (3.4) again to get

$$\frac{\mathbb{E}(Y_i^2)}{\mathbb{E}(Y_i)^r} \propto 2^{jp((2-r)(s+\frac{n}{2}-\frac{n}{p}-\alpha)-(1-r)\frac{\beta}{p})} \leq 2^{j((2-r)(\beta-n)-(1-r)\beta)} = 2^{j(\beta-2n+rn)},$$

which is bounded when

$$1 < r < 2 - \frac{\beta}{n}.$$

Hence, we can apply Lemma A.1 to get

$$\sum_i Y_i \quad < \quad \infty \tag{3.6}$$

almost surely if and only if

$$\sum_i \mathbb{E}(Y_i) < \infty.$$

Because at the wavelet scale j the number of terms in the sum is $K_{j,n} \propto 2^{jn}$, we get

$$\sum_i \mathbb{E}(Y_i) \propto \sum_j 2^{jn} 2^{jp(s+\frac{n}{2}-\frac{n}{p}-\alpha-\frac{\beta}{p})} = \sum_j 2^{jp(s+\frac{n}{2}-\alpha-\frac{\beta}{p})},$$

which is finite if and only if (3.3) holds. □

Remark. Because the convergence of (3.6) is a tail event, the Kolmogorov's zero-one law implies that either

$$P(f \in B_p^s(D)) = 1 \quad \text{or} \quad P(f \in B_p^s(D)) = 0$$

for the previous theorem type functions f. Hence, if (3.3) does not hold, then $f \notin B_p^s(D)$ almost surely.

In the stable distribution model, see (Boubchir and Fadili, 2006), the wavelet coefficients are independent random variables that are distributed as

$$W_{j,k}^{\ell} \quad \sim \quad f_{\alpha_j,c_j}, \tag{3.7}$$

where

$$f_{\alpha,c} \quad = \quad \mathcal{F}^{-1}\phi_{\alpha,c}, \tag{3.8}$$

$\mathcal{F}$ is the Fourier transform,

$$\mathcal{F}\phi(\xi) = \int_{\mathbb{R}} e^{-i\xi x}\phi(x)dx, \quad \mathcal{F}^{-1}\phi(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{i\xi x}\phi(\xi)d\xi,$$

and

$$\phi_{\alpha,c}(\xi) \quad = \quad e^{-|c\xi|^{\alpha}}. \tag{3.9}$$

Here, the parameter c describes the scale of the distribution and α tells about the decay of the tail. Remark that $f_{\alpha,c}$ is real because $\phi_{\alpha,c}$ is symmetric.

Note that in (A.4) there is no condition for the variances of stochastic terms in the series. This is important because the stable distributions (3.8) are heavy tailed and their variances are not finite for $\alpha \neq 2$ (Boubchir and Fadili, 2006).

Theorem 3.2. *Let $p < 2$. Assume that $W_{j,k}^{\ell}$ follow the stable distribution model* (3.7)*, where the parameters satisfy*

$$c_j = C_1 2^{-\gamma_1 j}, \qquad \alpha_j = 2 - C_2 2^{-\gamma_2 j} \tag{3.10}$$

for $j \geq j_0$ with some constants C_1, C_2, γ_1, and γ_2, $\gamma_2 < n$. If f is given by the wavelet expansion (2.5)*, then $f \in B_p^s(D)$ almost surely if and only if*

$$s + \frac{n}{2} \quad < \quad \gamma_1 + \frac{\gamma_2}{p}. \tag{3.11}$$

Proof. Let X_i and Y_i be as in the proof of Theorem 3.1. We may assume

$$s + \frac{n}{2} \quad \leq \quad \gamma_1 + \frac{\gamma_2}{p}. \tag{3.12}$$

Since

$$f_{\alpha,c}(t) \quad = \quad \frac{1}{c} f_{\alpha,1}(t/c), \tag{3.13}$$

we get

$$\mathbb{E}(X_i) = \int_{\mathbb{R}} |t|^p f_{\alpha_j,c_j}(t)dt = c_j^p \nu_{\alpha_j,1,p},$$

where

$$\nu_{\alpha_j,1,p} = \int_{\mathbb{R}} |t|^p f_{\alpha_j,1}(t)dt.$$

Asymptotically, as $|x| \to \infty$, (Samorodnitsky and Taqqu, 1994),

$$f_{\alpha_j,1}(x) \quad \asymp \quad \frac{\alpha_j(\alpha_j - 1)}{2\Gamma(2-\alpha_j)\cos(\frac{\pi}{2}(2-\alpha_j))|x|^{1+\alpha_j}} \quad \propto \quad \frac{2^{-\gamma_2 j}}{|x|^{1+\alpha_j}}. \tag{3.14}$$

So

$$\nu_{\alpha_j,1,p} \propto 2^{-\gamma_2 j} \int_1^{\infty} \frac{1}{|t|^{1+\alpha_j - p}} dt \quad \propto \quad 2^{-\gamma_2 j},$$

because

$$\epsilon < \alpha_j - p \leq 2 - p$$

for some $\epsilon > 0$ for large j. Here we need the condition for p. By substituting the decay model (3.10), we get

$$\mathbb{E}(Y_i) \propto 2^{jp(s + \frac{n}{2} - \frac{n}{p} - \gamma_1 - \frac{\gamma_2}{p})},$$

and so (A.1) is valid. Fix q so that $1 < q < 2/p$. Now,

$$\mathbb{E}(X_i^q) = c_j^{pq} \nu_{\alpha_j,1,pq}$$

and

$$\nu_{\alpha_j,1,pq} \propto 2^{-\gamma_2 j}$$

as above, because

$$pq < 2.$$

Hence,

$$\mathbb{E}(Y_i^q) \propto 2^{jpq(s+\frac{n}{2}-\frac{n}{p}-\gamma_1)-\gamma_2 j},$$

and

$$\frac{\mathbb{E}(Y_i^q)}{\mathbb{E}(Y_i)^r} \propto 2^{jp((q-r)(s+\frac{n}{2}-\frac{n}{p}-\gamma_1)-(1-r)\frac{\gamma_2}{p})} \leq 2^{j((q-r)(\gamma_2-n)-(1-r)\gamma_2)} = 2^{j(q\gamma_2-qn+rn-\gamma_2)},$$

which is bounded when

$$1 < r < \frac{\gamma_2}{n} + q - q\frac{\gamma_2}{n}.$$

Note that the upper bound satisfies

$$1 < \frac{\gamma_2}{n} + q - q\frac{\gamma_2}{n} < q.$$

By Lemma A.1,

$$\sum_i Y_i < \infty$$

if and only if

$$\sum_i \mathbb{E}(Y_i) \propto \sum_j 2^{jp(s+\frac{n}{2}-\gamma_1-\frac{\gamma_2}{p})} < \infty,$$

which is true if and only if (3.11) is satisfied. □

The inequalities (3.3) and (3.11) provide a tool to estimate the Besov space of a function. After computing the wavelet transform of the function, one can estimate α, β of (3.2), or γ_1, γ_2 of (3.10), and then the inequality gives the upper limit for s (for given p). The functional forms of the criteria inequalities are the same with the threshold distribution model parameters α, β and with the stable distribution model parameters γ_1, γ_2. Parameters α and γ_1 describe the decay of the scales of the density functions, and parameters β and γ_2 describe the change in the shape of the density functions as the wavelet scales get finer.

Always, there remains a question, if the true wavelet coefficients follow the chosen distribution model. However, it turns out that it is not necessary to know the explicit functional form of the wavelet distribution model. Only the decay of the p-moments in wavelet scales is needed.

We say that the wavelet coefficients follow the decaying moments model if

$$\sigma_j = C_1 2^{-\eta_1 j}, \qquad \nu_{p,j} = C_2 2^{-\eta_2 j}, \tag{3.15}$$

with some positive constants C_1, C_2, η_1, and η_2. Here

$$\sigma_j^2 = \mathbb{E}(|W_{j,k}^{\ell}|^2)$$

is the variance and

$$\nu_{p,j} = \mathbb{E}\left(\left|\frac{W_{j,k}^{\ell}}{\sigma_j}\right|^p\right)$$

the normalized p-moment. If the wavelet coefficients follow the decaying moments model, then the p-moments decay exponentially in scale,

$$m_{p,j} \quad = \quad \mathbb{E}(|W_{j,k}^{\ell}|^p) = C2^{-\gamma j} \tag{3.16}$$

with

$$\gamma = p\eta_1 + \eta_2,$$

because

$$m_{p,j} = \sigma_j^p \nu_{p,j}.$$

Note that always $\eta_2 \geq 0$ because

$$\nu_{p,j} \leq \mathbb{E}\left(\left|\frac{W_{j,k}^{\ell}}{\sigma_j}\right|^2\right)^{p/2} = 1$$

by Jensen's inequality.

Theorem 3.3. *Let $p < 2$. Assume that $W_{j,k}^{\ell}$ follow the decaying moments model* (3.15) *with*

$$\eta_2 \quad < \quad \frac{n(2-p)}{2}. \tag{3.17}$$

If f is given by the wavelet expansion (2.5)*, then $f \in B_p^s(D)$ almost surely if and only if*

$$s + \frac{n}{2} \quad < \quad \frac{\gamma}{p} \; = \; \eta_1 + \frac{\eta_2}{p}. \tag{3.18}$$

Proof. We can assume

$$s + \frac{n}{2} \leq \frac{\gamma}{p}.$$

Let X_i and Y_i be as in the proof of Theorem 3.1. Let $q = 2/p$. Now,

$$\mathbb{E}(X_i) = m_{p,j}, \quad \mathbb{E}(X_i^q) = \sigma_j^2.$$

The condition (A.1) is valid,

$$\mathbb{E}(Y_i) \propto 2^{jp(s+\frac{n}{2}-\frac{n}{p}-\frac{\gamma}{p})} \leq 2^{-j\frac{n}{p}}.$$

Given r, $1 < r < q$, we get

$$\frac{\mathbb{E}(Y_i^q)}{\mathbb{E}(Y_i)^r} \propto 2^{jp((q-r)(s+\frac{n}{2}-\frac{n}{p})-\frac{2\eta_1}{p}+r\frac{\gamma}{p})} \leq 2^{j((q-r)(\gamma-n)-2\eta_1+r\gamma)} = 2^{j(q\eta_2-qn+rn)},$$

which is bounded when

$$r < q(1 - \frac{\eta_2}{n}).$$

Note that for this bound it holds

$$1 < q(1 - \frac{\eta_2}{n}) < q$$

by the bound for η_2. Hence, the condition (A.2) is also valid. By Lemma A.1, $f \in B_p^s(D)$ almost surely if and only if

$$\sum_i \mathbb{E}(Y_i) \propto \sum_j 2^{jp(s+\frac{n}{2}-\frac{\gamma}{p})} < \infty,$$

which is valid if and only if the condition (3.18) holds. □

4 Besov prior for attenuation coefficients

In this section we study in which Besov space the attenuation coefficient belongs to when the space for the X-ray images is known. The idea is to look how the wavelet distribution parameters of the attenuation function and the X-ray images are connected. We study the connection with the stable distributions model case and with the decaying moments model case. The threshold distribution model is not invariant in summation so we do not consider it here. To be able to obtain the properties of the attenuation coefficient from a single X-ray image, an isotropy property is needed.

Let $P_n f$ be the X-ray image of f in direction $e_n = (0, \ldots, 0, 1) \in S^n$,

$$P_n f(x') \quad = \quad \int_{\mathbb{R}} f(x', x_n) dx_n, \quad x = (x', x_n) \in \mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}. \tag{4.1}$$

Now, the X-ray image of the wavelet $\psi_{j,k}^{\ell}$ is

$$P_n \psi_{j,k}^{\ell}(x') = \psi_{j,k'}^{\ell'}(x') \int \psi_{j,k_n}^{\ell_n}(x_n) dx_n = \begin{cases} c 2^{-j/2} \psi_{j,k'}^{\ell'}(x'), & \ell_n = 0, \\ 0, & \ell_n = 1, \end{cases}$$

where

$$c = \int \phi(y) dy.$$

Hence, if $f \in L^2(D_n)$, $D_n =]0,1[^n$, is given by the wavelet expansion (2.5), then

$$P_n f(x') = \sum_j \sum_{k', \ell'} \left(c 2^{-j/2} \sum_{k_n} w_{j,k}^{(\ell', 0)} \right) \psi_{j,k'}^{\ell'}(x'),$$

and so the the wavelet coefficients of $P_n f$ are

$$w_{j,k'}^{\ell'} \quad = \quad c 2^{-j/2} \sum_{k_n} w_{j,k}^{(\ell', 0)}. \tag{4.2}$$

Note that the number of terms in the sum (4.2) is proportional to 2^j.

We say that f is *isotropic*, if the distribution of the wavelet coefficient $w_{j,k}^{\ell}$ is independent of ℓ. If f is isotropic, then its smoothness is determined from the distribution of $w_{j,k}^{(\ell', 0)}$ to which we have control through formula (4.2).

Theorem 4.1. *Let $p < 2$. Assume that the attenuation coefficient f is isotropic and $P_n f \in B_p^s(D_{n-1})$. If the wavelet coefficients $w_{j,k}^{\ell}$ and $w_{j,k'}^{\ell'}$* (4.2) *follow*

i) the decaying moments model (3.16) *and* (3.15)*,*

$$\sigma_j' = \sqrt{\mathrm{var}(w_{j,k'}^{\ell'})} = C 2^{-\eta_1' j}, \quad \nu_{p,j}' = \mathbb{E}\left(\left| \frac{w_{j,k'}^{\ell'}}{\sigma_j'} \right|^p \right) = C 2^{-\eta_2' j},$$

with $\eta_2' = 0$,

or,

ii) the stable distribution model (3.7)*,*

then $f \in B^{s-\frac{1}{2}}(D)$ almost surely.

Proof. i) The deviation of $w_{j,k'}^{\ell'}$ is

$$\sigma_j' = c2^{-j/2}\sqrt{\sum_{k_n} \mathsf{var}(w_{j,k}^{(\ell',0)})} = C\sigma_j = C2^{-\eta_1 j},$$

and so

$$\eta_1' = \eta_1.$$

Because $P_n f \in B_p^s(D_{n-1})$, Theorem 3.3 implies that

$$s + \frac{n-1}{2} < \frac{\gamma'}{p} = \eta_1' + \frac{\eta_2'}{p} \leq \eta_1 + \frac{\eta_2}{p} = \frac{\gamma}{p},$$

or

$$s - \frac{1}{2} + \frac{n}{2} < \frac{\gamma}{p}.$$

Hence, $f \in B_p^{s-\frac{1}{2}}(D_n)$ almost surely by Theorem 3.3.

ii) Now,

$$c2^{-j/2} w_{j,k}^{(\ell',0)} \sim f_{\alpha_j, c_j 2^{-j/2} c}.$$

Because the distribution density function of a sum is a convolution of the densities, and the Fourier transform transforms the convolution to products, we get

$$w_{j,k'}^{\ell'} \sim \mathcal{F}^{-1}\phi_{\alpha_j, c_j 2^{-j/2} c}^{2^j} = \mathcal{F}^{-1}\exp(-|2^{j/\alpha_j} c_j 2^{-j/2} c\xi|^{\alpha_j}) = f_{\alpha_j, d_j}, \quad d_j = c2^{-j/2 + j/\alpha_j} c_j.$$

Now,

$$d_j = c2^{-j/2+j/\alpha_j} c_j \asymp C2^{-j\gamma_1},$$

and so $w_{j,k'}^{\ell'}$ follows the stable distribution model with the same decay parameters γ_1 and γ_2 as $w_{j,k}^{\ell}$. By Theorem 3.2, and because $P_n f \in B_p^s(D_{n-1})$, we get

$$\gamma_1 + \frac{\gamma_2}{p} > s + \frac{n-1}{2} = s - \frac{1}{2} + \frac{n}{2}.$$

Hence, $f \in B_p^{s+\frac{1}{2}}(\mathbb{R}^{n-1})$ almost surely. □

Now, the Besov space smoothness s for an isotropic attenuation function can be chosen by determining the smoothness s' of an X-ray image with the wavelet distribution models, and by setting

$$s \;=\; s' - \frac{1}{2}. \tag{4.3}$$

The isotropicity condition is essential. Generally, one can not determine the smoothness of an attenuation function from a single X-ray image. Namely, by setting

$$f(x) = g(x_n), \quad x = (x', x_n),$$

the X-ray image in the x_n direction is a smooth function (constant) regardless how non-smooth the function g is. In practice, the attenuation coefficients may satisfy the isotropicity condition and the condition for η_2' only approximately.

5 Noise model

The Besov space results above are based on the asymptotics of the wavelet coefficient distributions when the wavelet scale $j \to \infty$. In practice, however, only a finite number of wavelet scales is available, and the wavelet coefficients are noisy. Hence, the asymptotics of the distributions are obtained rather from the decay of finite number of the scale distribution parameters than their asymptotics.

Consider a measurement situation. Today, X-ray imaging data is usually collected with digital sensors that count the number of photons that hit to each sensor pixel. The photon count data at detector point k is then converted to the pencil beam data m_k^ϵ (Siltanen et al., 2003). The density of the array in the detector sets a limit for the wavelet scales. The additive noise model for the measurement at detector point k is

$$m_k^\epsilon = m_k + \epsilon_k,$$

where $\epsilon_k \sim N(0, \sigma^2)$ are independent is a reasonable assumption (Bouman and Sauer, 1993; Siltanen et al., 2003).

Assume that we have a function f^ϵ whose approximation coefficients at the finest scale J are

$$w_{J,k}^{0,\epsilon} = w_{J,k}^0 + \epsilon_{J,k}^0,$$

where $\epsilon_{J,k}^0 \sim N(0, \sigma_\epsilon^2)$ are independent, as in the additive noise model. By linearity of the wavelet transform, the wavelet coefficients at every level have also additive noise,

$$w_{j,k}^{\ell,\epsilon} \quad = \quad w_{j,k}^\ell + \epsilon_{j,k}^\ell, \tag{5.1}$$

and $\epsilon_{j,k}^\ell \sim N(0, \sigma_\epsilon^2)$ are normally distributed with the *same* variance σ_ϵ^2 at every scale. Namely, every wavelet coefficient can be obtained from a finite number of filterings, and a filtering maintains the variance: If $\epsilon_k \sim N(0, \sigma_\epsilon^2)$ are independent, and $c = (c_k)$ is a filter corresponding to the wavelet transform, then the filtered

$$\sum_k c_k \epsilon_k \sim N(0, \sum_k c_k^2 \sigma_\epsilon^2) = N(0, \sigma_\epsilon^2),$$

because

$$\sum_k |c_k|^2 = 1$$

for orthonormal wavelets. Remark also that now

$$\text{var}(w_{j,k}^{\ell,\epsilon}) \quad = \quad \text{var}(w_{j,k}^\ell) + \sigma_\epsilon^2. \tag{5.2}$$

Assume that the wavelet decomposition for f^ϵ is

$$f^\epsilon = \sum_k w_{j_0,k}^{0,\epsilon} \psi_{j_0,k}^0 + \sum_{j=j_0+1}^{J} \sum_{\ell \neq 0,k} w_{j,k}^{\ell,\epsilon} \psi_{j,k}^\ell,$$

where the wavelet coefficients follow the additive noise model (5.1). Assume that the true coefficients $w_{j,k}^{\ell}$ follow the distribution model (3.1) or (3.7). In both wavelet distribution models, the parameters σ_j, c_j, tend to zero as scale j grows which means that noise becomes dominant at fine scales. We assume that for the finest scale $\sigma_J, c_J << \sigma_\epsilon$, and approximate σ_ϵ with

$$\sigma_\epsilon \quad = \quad \sqrt{\sum_{\ell \neq 0, k} |w_{J,k}^{\ell,\epsilon}|^2/(K_{J,n}-1)}. \tag{5.3}$$

The detail wavelets of the finest scale J are sacrificed to determine the noise level. The wavelet distribution models have to be estimated from the remaining scales $j_0, \ldots, J-1$. To simplify the notation, we assume from now on that wavelets are available up to scale J and the noise level σ_ϵ is known.

6 Reconstructing the attenuation coefficient

In this section, we recall how the reconstruction of the attenuation coefficient is done. For more details, we refer to, e.g., (Niinimäki et al., 2007).

Let $x = (x_i)$ be the discretized attenuation coefficient with wavelet coefficients

$$w = (w_{j_0,k}^0)(w_{j,k}^{\ell}),$$

where $w_{j_0,k}^0$ are the approximation coefficients and $w_{j,k}^{\ell}$ are the detail coefficients, $j \geq j_0, \ell \neq 0$. We write

$$w = Wx, \quad x = W^{-1}w,$$

for the discrete wavelet transform, and for the inverse of it, respectively. Assume that y is the data vector which is obtained from the X-ray images, and let A be the operator of X-ray imaging, ideally

$$y = Ax.$$

We use the Besov space prior, and also, we ask x to be a non-negative function. This non-negativity can be handled by setting

$$x = z \otimes z,$$

and by finding real values for $z = (z_i)$. Here,

$$\alpha \otimes \beta = ((\alpha \otimes \beta)_i), \quad (\alpha \otimes \beta)_i = \alpha_i \beta_i,$$

for $\alpha = (\alpha_i)$, $\beta = (\beta_i)$. Finding the MAP estimate for the posterior density $P(x|y)$ is now equivalent to minimizing the functional

$$F(z) \quad = \quad \frac{|Ax-y|^2}{2\sigma_\epsilon^2} + \frac{a}{\sigma_\epsilon^p}\|Wx\|_{B_p^s}^p, \quad x = z \otimes z. \tag{6.1}$$

Here, a is a weight for the smoothness prior. We use the conjugate gradient minimization method. The gradient of F is

$$\nabla F(z) = 2z \otimes \left[\frac{A^T(Ax-y)}{\sigma_\epsilon^2} + \frac{ap}{\sigma_\epsilon^p} W^{-1} |w_{j_0,k}^0|^{p-1} \mathrm{sign}(w_{j_0,k}^0) 2^{jp(s+n(1/2-1/p))} |w_{j,k}^{\ell}|^{p-1} \mathrm{sign}(w_{j,k}^{\ell}) \right],$$

where $x = z \otimes z$ and $w = Wx$. If one does a 3D reconstruction directly, then data consists of two dimensional X-ray images that are taken from different view angles, and then $n = 3$ in (6.1). When the 3D reconstruction is made by attaching two-dimensional slices, then data consists of one-dimensional slices of the X-ray images, and $n = 2$.

7 Estimating distribution parameters

In this section we describe how the distribution parameters are estimated from noisy wavelet coefficient data. We suppose that σ_ϵ is already determined. The scales are re-indexed so that $j_0 = 0$.

Decaying moments model

To determine the smoothness of the X-ray image we need parameter γ of Theorem 3.3. However, in Theorem 4.1 we assumed parameter η_2 of the X-ray image to be zero, see also (3.15). To test the accuracy of this assumption, we determine both γ and η_1 from which

$$\eta_2 = \gamma - p\eta_1$$

can be observed.

With general p there is no simple formula of type (5.2) connecting noisy and uncorrupted p-moments. We can get an approximation for $m_{p,j}$ numerically by *adding* noise to data as follows. To simplify the notation, consider a random variable Y. We want to get an estimate for

$$\mathbb{E}(|Y|^p)$$

from realizations of Y^ϵ,

$$Y^\epsilon = Y + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

Let

$$\begin{cases} Y^{i\epsilon} = Y^{(i-1)\epsilon} + \epsilon^i, & i = 1, 2, \ldots, \\ Y^{0\epsilon} = Y. \end{cases}$$

Here, $\epsilon^i \sim N(0, \sigma^2)$ are independent random noise, $i = 1, 2, \ldots$. Note that by adding $i - 1$ times noise $\epsilon^i \sim N(0, \sigma^2)$ to samples of Y^ϵ we get samples of $Y^{i\epsilon}$ for $i \geq 1$. Hence, we can approximate

$$\mathbb{E}(|Y^{i\epsilon}|^p) \approx \frac{1}{N} \sum_{j=1}^{N} |Y_j^{i\epsilon}|^p.$$

Now, we fit a polynomial $q = q(i)$ (e.g., of second degree) to points

$$i \mapsto \mathbb{E}(|Y^{i\epsilon}|^p), \quad i = 1, \ldots, M$$

and approximate

$$\mathbb{E}(|Y|^p) \quad \approx \quad q(0). \tag{7.1}$$

Remark that the variance could be approximated in this way, too, by putting $p = 2$.

Now, suppose that we have obtained $m_{p,j}$ approximately on each scale j. We get an estimate for γ by fitting a line in the least squares sense to points

$$(j, \log_2(m_{p,j})), \quad j = 0, \ldots, J.$$

The negative slope of the line is the approximation for γ.

An approximation for η_1 is obtained with the help of (5.2) from the sample variances

$$s_j^2 \;=\; \sum_{k,\ell \neq 0} |w_{j,k}^{\ell,\epsilon}|^2 / K_{j,n}. \tag{7.2}$$

We fit a line in the least squares sense to points

$$(j, \log_2(\sqrt{s_j^2 - \sigma_\epsilon^2})), \quad j = 0, \ldots, J.$$

The negative slope of the line is the approximation for η_1.

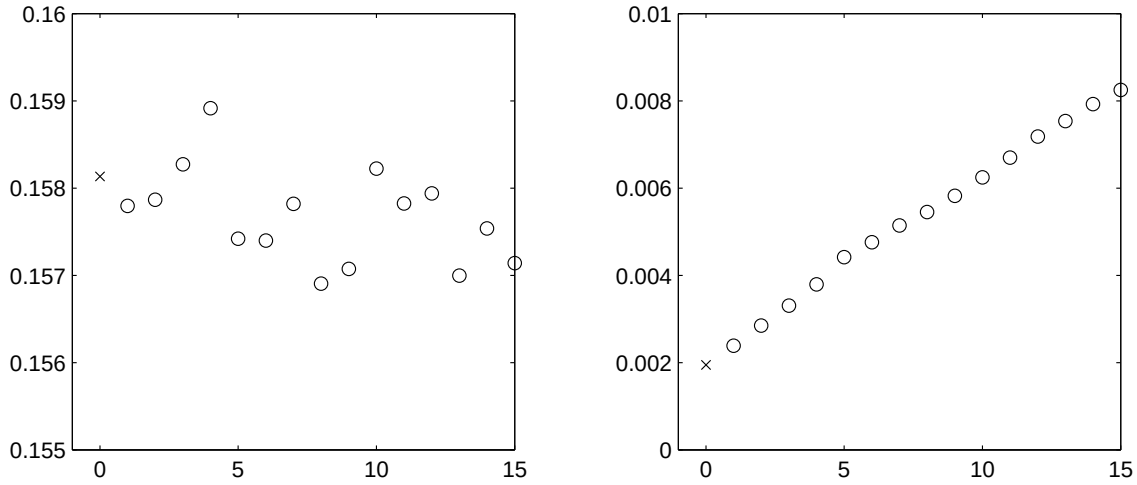

Figure 1: Estimating p-moments by adding noise. Here, x-axis is i and y-axis is $\mathbb{E}(|Y^{i\epsilon}|^p)$ ($\circ$ for $i \geq 1$, x for the estimate). Left image: Noise level is low. Right: Noise level is high.

Threshold distribution model

If $w_{j,k}^{\ell}$ follow the threshold distribution model (3.1), then the noisy threshold distribution model is

$$w_{j,k}^{\ell,\epsilon} \;\sim\; f_{\tau_j,\sigma_j}^{\epsilon} \;:=\; f_{\tau_j,\sigma_j} * \phi_{\sigma_\epsilon^2} \;=\; \tau_j \phi_{\sigma_j^2+\sigma_\epsilon^2} + (1-\tau_j)\phi_{\sigma_\epsilon^2}. \tag{7.3}$$

The goal is to get the posterior density for the decay parameters α, β of (3.2) as a marginal density

$$P_{post}(\alpha, \beta) = \int P_{post}(\alpha, \beta, C_\tau, C_\sigma) dC_\tau dC_\sigma.$$

The prior distribution for $(\alpha, \beta, C_\tau, C_\sigma)$ is assumed to be the uniform distribution in the computational domain, which is set with the help of the decay speed η_1 of σ_j. The computational

domain is

$$\begin{array}{rclcrcl} C_\tau &\in& [\,1/2\,,\,1\,], & & \beta &\in& [0, \beta_{max}], \\ C_\sigma &\in& \left[\,\sqrt{s_0^2-\sigma_\epsilon^2}\,,\,\sqrt{2(s_0^2-\sigma_\epsilon^2)}\,\right], & & \alpha &\in& \left[\,\frac{\eta_1}{q}-\frac{\beta_{max}}{2}\,,\,\eta_1 q\,\right]. \end{array} \tag{7.4}$$

Here, the limits for parameters C_σ and α are obtained from (5.2) by noting

$$C2^{-2\eta_1 j} \approx s_j^2 - \sigma_\epsilon^2 = \tau_j \sigma_j^2 \approx C_\tau C_\sigma^2 2^{-\beta j - 2\alpha j}.$$

With a relaxing parameter $q > 1$ we describe that the equations above hold only approximately.

By Bayes' formula,

$$P_{post}(\alpha, \beta, C_\sigma, C_\tau) = P(\alpha, \beta, C_\sigma, C_\tau \,|\, (w_{j,k}^{\ell,\epsilon})) \propto P((w_{j,k}^{\ell,\epsilon}) \,|\, \alpha, \beta, C_\sigma, C_\tau), \tag{7.5}$$

where the likelihood is

$$P((w_{j,k}^{\ell,\epsilon}) \,|\, \alpha, \beta, C_\sigma, C_\tau) = \prod_{j=j_0}^{J} P((w_{j,k}^{\ell,\epsilon})|\sigma_j, \tau_j)$$

Here σ_j, τ_j are given by (3.2). Note that the computational domain for $(\alpha, \beta, C_\tau, C_\sigma)$ determines the computational domain for (σ_j, τ_j) at each scale j. The likelihood at scale j is

$$P((w_{j,k}^{\ell,\epsilon}) \,|\, \sigma_j, \tau_j) \;\; = \;\; \prod_{k,\ell} f_{\sigma_j,\tau_j}^{\epsilon}(w_{j,k}^{\ell,\epsilon}). \tag{7.6}$$

In practice, the computation is done so that we compute first the likelihoods (7.6) on a two-dimensional (σ_j, τ_j) grid that is set on each scale j. Then we set a four dimensional grid onto the computational domain for $(\alpha, \beta, C_\tau, C_\sigma)$. Every grid point $(\alpha_i, \beta_i, C_{\tau,i}, C_{\sigma,i})$ determines the corresponding $(\sigma_{j,i}, \tau_{j,i})$ values, $j = 0, \ldots, J$, to which we interpolate the value for $P((w_{j,k}^{\ell,\epsilon})|\sigma_{j,i}, \tau_{j,i})$ from the previously computed (σ_j, τ_j) grid point values.

Stable distribution model

If $w_{j,k}^{\ell}$ follow the stable distribution model (3.7), then

$$w_{j,k}^{\ell,\epsilon} \;\; \sim \;\; f_{\alpha_j,c_j}^{\epsilon} \;:=\; f_{\alpha_j,c_j} * \phi_{\sigma_\epsilon^2} \;=\; \mathcal{F}^{-1}\left(\exp(-\frac{1}{2}|\sigma_\epsilon \xi|^2 - |c_j \xi|^{\alpha_j})\right). \tag{7.7}$$

In the stable distribution model, the parameters are estimated in the similar manner as in the threshold distribution model. In (7.6) the density function $f_{\tau_j,\sigma_j}^{\epsilon}$ of is replaced with $f_{\alpha_j,c_j}^{\epsilon}$ of (7.7). The values for $f_{\alpha_j,c_j}^{\epsilon}$ are computed by the inverse Fourier transform. The computational domain is set to be

$$\begin{array}{rclcrcl} C_1 &\in& [\,s_0/q\,,\,s_0 q\,], & & \gamma_1 &\in& \left[\,\frac{\eta_1}{q}\,,\,q\eta_1\,\right], \\ C_2 &\in& [C_{2,min}\,,\,C_{2,max}\,], & & \gamma_2 &\in& [\,0\,,\,\gamma_{2,max}\,]. \end{array} \tag{7.8}$$

8 Examples

The numerical examples have two purposes here. First, we illustrate the estimation of the smoothness parameters. The second purpose is to check how the smoothness parameter that is proposed by the method works when making the reconstructions. For this reason, we present some slice reconstructions of the attenuation coefficient with different smoothness parameters. We use dental X-ray data and the Shepp-Logan Phantom data. The Besov-space parameter p is 1.5.

We use Daubechies' 4-wavelets. When reasonable, we have computed the regression coefficient

$$R^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2},$$

where $\bar{Y} = \sum Y/N$ is the mean and $\hat{Y}$ is the prediction.

8.1 Estimating smoothness parameters

Shepp-Logan phantom case

We used the Shepp-Logan phantom in two ways. First, we estimate the smoothness of the 2D phantom itself. Secondly, we simulate one dimensional X-ray data of the phantom, and estimate the smoothness of the X-ray data. To increase the number of wavelet coefficients for better statistical properties, we have used seven view angles. In this way we can compare the smoothness of a known attenuation coefficient and its X-ray data. The results are collected in Table 1. We pick here some points:

- The phantom is 674×674 pixels. To the X-ray data we added 4% noise. See Figure 2.
- The p-moments follow the decay model with high reliability, see Figure 3.
- In Figures 4-5 we have illustrated the decay models by plotting the logarithms for the MAP estimates of the posterior densities $P(c_j, \alpha_j|(w_{j,k}^\ell))$ and $P(\sigma_j, \tau_j|(w_{j,k}^\ell)$ at different wavelet scales j. The regression coefficients for SDM and TDM are computed based on the least squares fit to these MAP estimates. For the stable distribution model, the decay tendency is not so evident, especially for the shape parameters α_j. The threshold distribution model seems to be valid better.
- The values for the decay parameters of SDM and TDM are the maximum of the posterior densities $P(\gamma_1, \gamma_2|(w_{j,k}^\ell))$ and $P(\alpha, \beta|(w_{j,k}^\ell))$, see Figure 6.

The threshold distribution model and the decaying moments model give approximately the same smoothness parameters for both the phantom and the X-ray slices. The difference in the smoothness between the X-ray slice and the phantom is 0.4–0.5 which is approximately what Theorem 4.1 is predicting. The stable distribution model seems to give too high smoothness in this case. An explanation for this is that the shape parameters α_j at available wavelet scales do not satisfy assumption $\alpha_j > p$, see the proof of Theorem 3.2.

	Model	DMM		SDM		TDM	
		γ	R^2	(γ_1, γ_2)	R^2	(α, β)	R^2
X-ray	parameters	1.66	.998	(1.92, .08)	(.854, .530)	(.72,.72)	(.944, .969)
	smoothness	0.6		1.47		0.7	
Phantom	parameters	1.74	.958	(2.56 .02)	(.982, .871)	(.71,.66)	(.953, .956)
	smoothness	0.2		1.5		0.2	

Table 1: Summary of the results in the Shepp-Logan phantom cases.

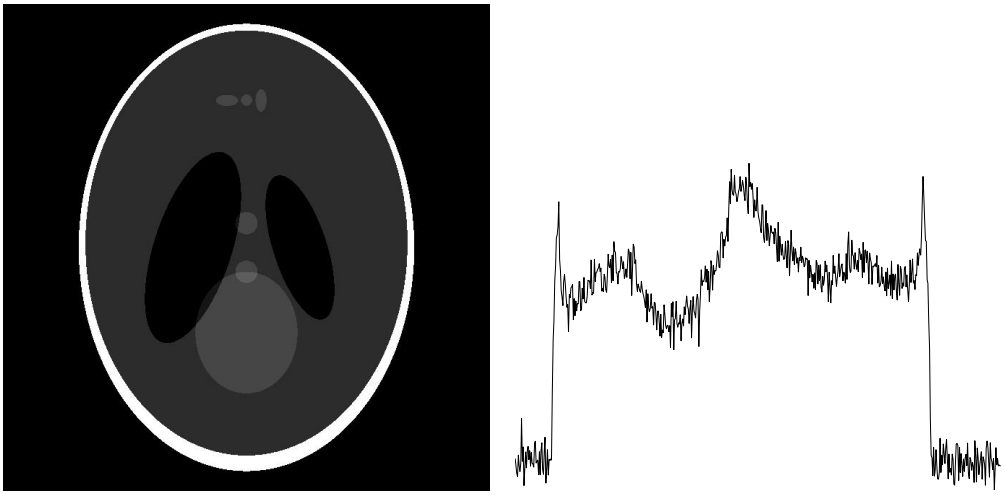

Figure 2: Shepp-Logan phantom (left) and one 1D X-ray slice of the phantom (right).

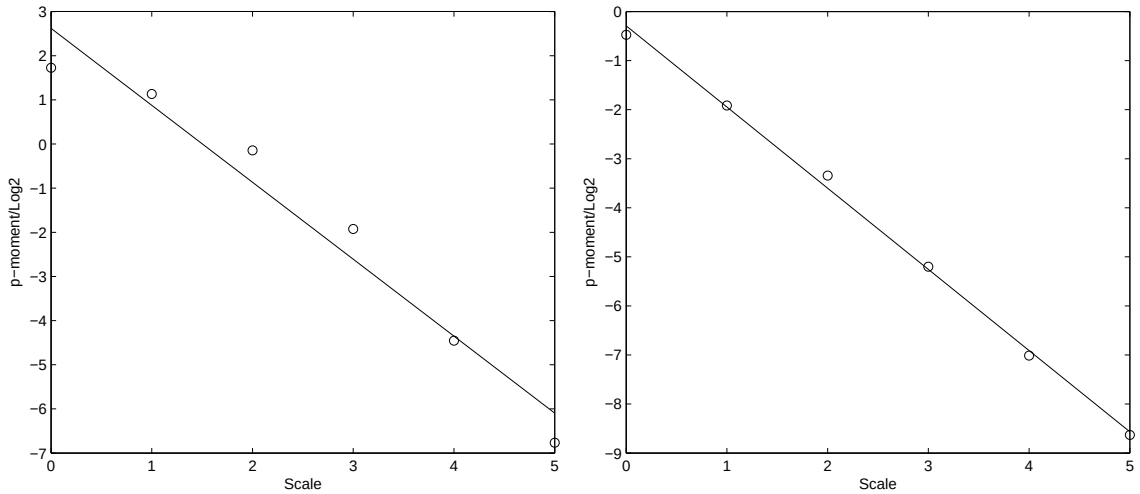

Figure 3: Least squares fit to points $(j, \log_2 m_{p,j})$. Left: Shepp-Logan phantom. Right: X-ray slices.

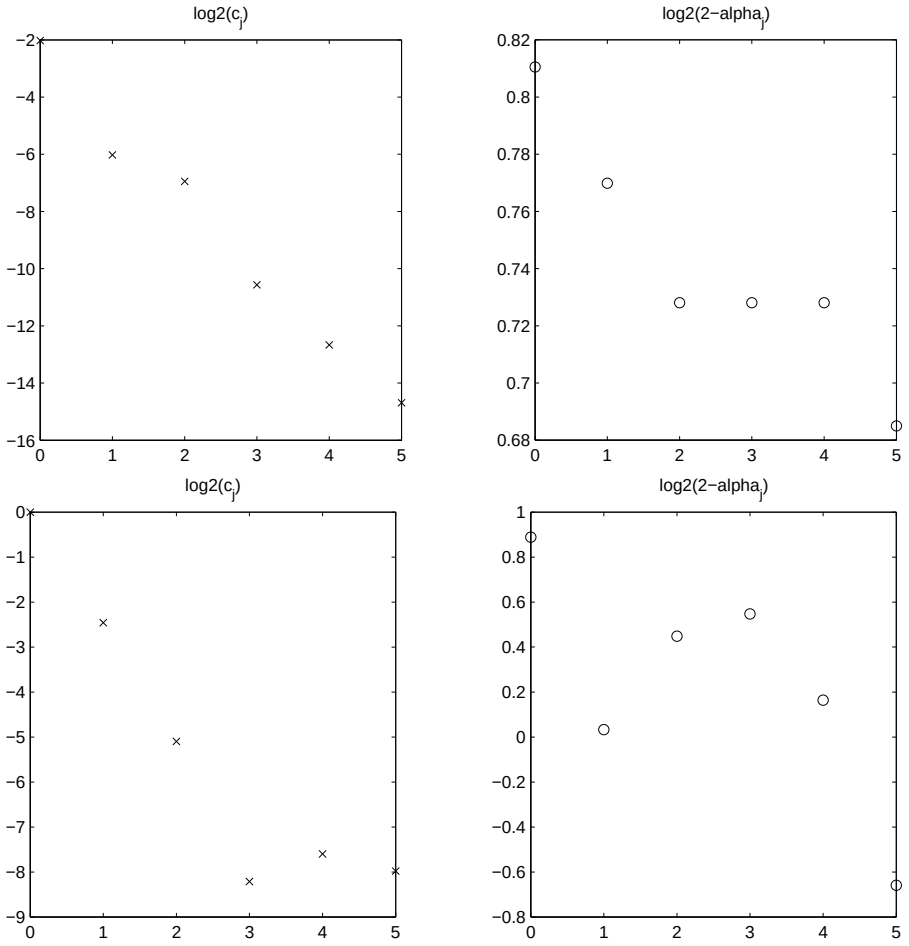

Figure 4: MAP estimates for the (c_j, α_j) of the stable distribution model in $\log_2$ scale. Top: Shepp-Logan phantom. Bottom: X-ray slice.

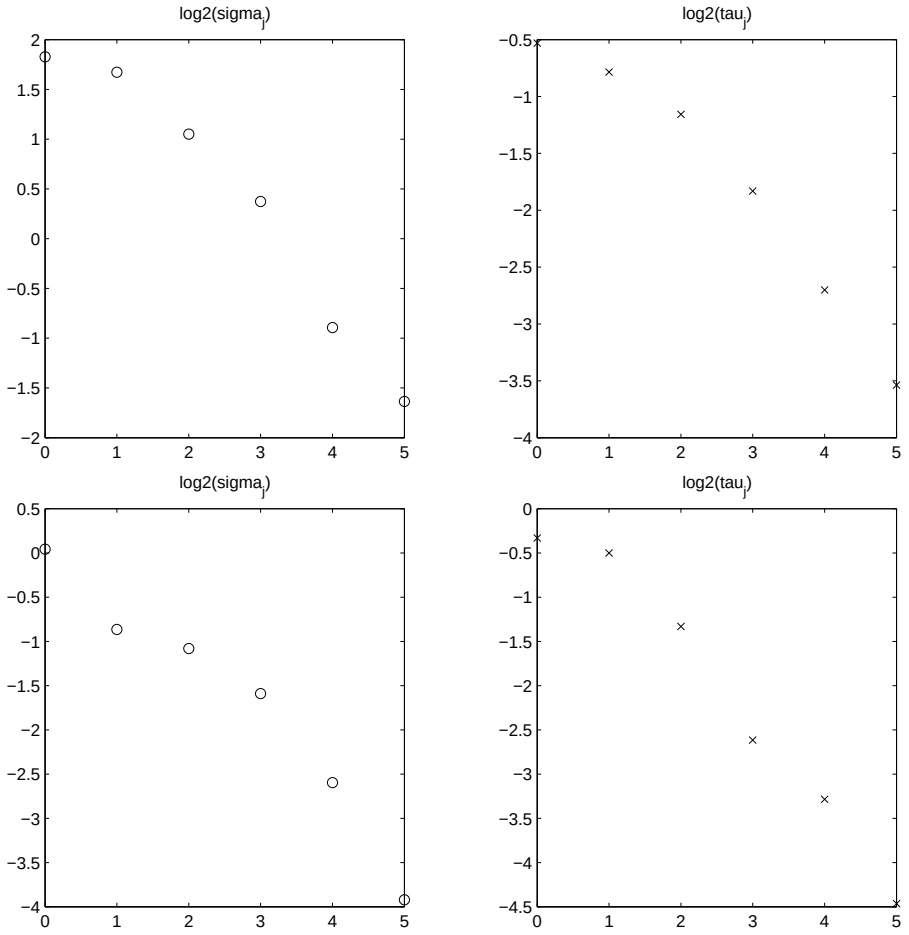

Figure 5: MAP estimates for the (σ_j, τ_j) of the threshold distribution model in $\log_2$ scale. Top: Shepp-Logan phantom. Bottom: X-ray slice.

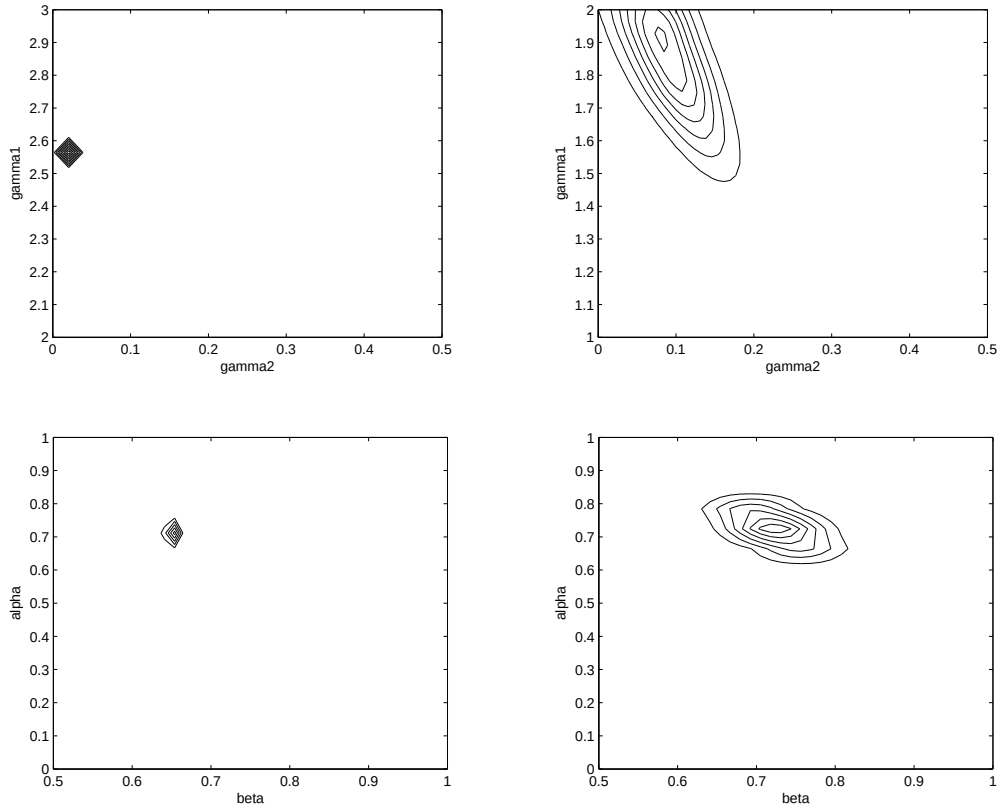

Figure 6: Posterior densities of the decay parameters. Left column: Shepp-Logan phantom. Right column: X-ray slice. Top: Stable distribution model (γ_1, γ_2). Bottom: Threshold distribution model (α, β).

Dental cases

We use dental X-ray data in two ways. Two-dimensional (2D) data is an X-ray image with 739×664 pixels that has been taken from a tooth phantom with an intra-oral sensor, see Figure 7. Each pixel value is the logarithm of the inverse of the photon count. One-dimensional (1D) data is a slice of the two-dimensional X-ray image. Often, the reconstructions are done a slice by slice and in that case data consists of one-dimensional functions. To increase the number of wavelet coefficients for better statistical properties, we have used seven one dimensional slices of the X-ray images that are taken from seven different view angles. We estimate the smoothness in both dimensions. The results are collected in Table 2. We pick here some points:

- In both dimensions, the noise parameter is estimated to be

$$\sigma_\epsilon \approx 0,01.$$

  Data is scaled so that the maximum value is one.

- Figure 8 illustrates estimating γ of the decaying moments model (3.16). A line is fitted in the least squares sense to the points $(j\,,\,\log_2(m_{p,j}))$.

- The validity of the assumption $\eta_2 = 0$ (or η_2') of Theorem 4.1 is studied by computing

$$\eta_2 = \gamma - \frac{\eta_1}{p}.$$

  In Figure 9 we have plotted points

$$\log_2 \nu_{p,j} = \log_2 \frac{m_{p,j}}{\sigma_j^p}.$$

  It seems that there is no decay tendency but the variation is merely from the computational approximation errors. This supports the assumption for η_2.

- In Figure 10 we have plotted the posterior densities $P(c_j, \alpha_j|(w_{j,k}^\ell))$ and $P(\sigma_j, \tau_j|(w_{j,k}^\ell))$ of the stable distribution and of the threshold distribution models on some different scales j. The decaying trends for the distribution parameters are evident. The MAP estimates of these distributions in each scale are presented in Figures 11-12. The regression coefficients for SDM and TDM are computed based on the least squares fit to these MAP estimates. Remark that the scale parameters σ_j and c_j seems to follow the decay models rather well (high R^2) but for the shape parameters τ_j and α_j the decay models are not so evident (lower R^2).

- The density functions for the wavelet coefficients corresponding to the MAP estimates (1D case) and the histograms of the absolute values of the wavelet coefficients are plotted in Figure 13. The bars of the histogram are getting wider for larger coefficients and they are scaled so that the overall area is one. The stable distribution model catches the tail of the wavelet distribution better whereas the exponential decay of the threshold distribution model seems to be too fast.

- The posterior densities $P(\gamma_1, \gamma_2|(w_{j,k}^\ell))$ and $P(\alpha, \beta|(w_{j,k}^\ell))$ of the decay parameters are presented in Figure 14.

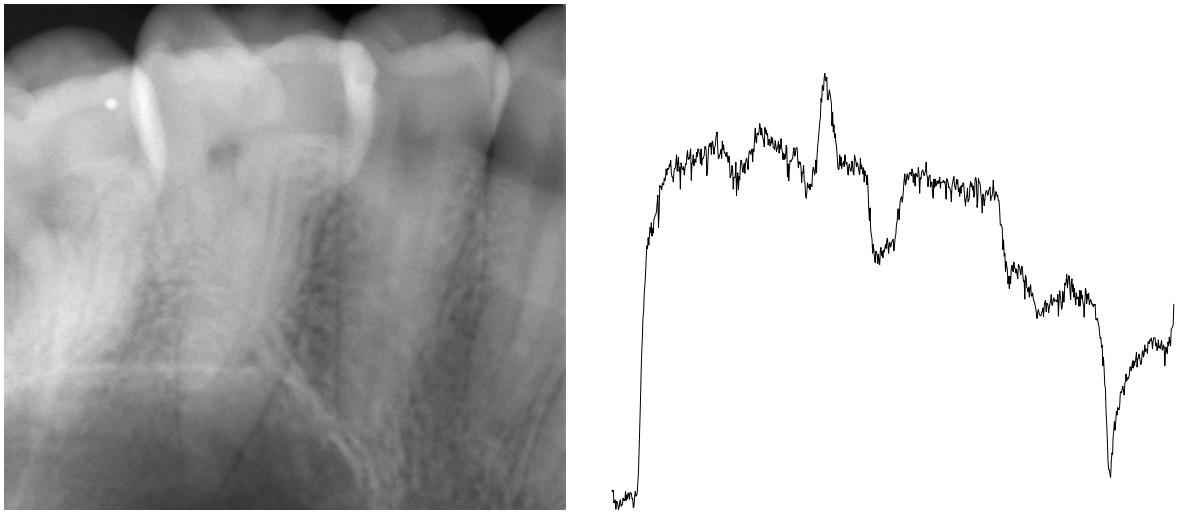

Figure 7: Dental X-ray data. Left: 2D data is a dental X-ray image. Right: One slice of 1D data.

	Model	DMM		SDM		TDM	
		γ	R^2	(γ_1, γ_2)	R^2	(α, β)	R^2
1D	parameters	1.92	.994	(1.3, .43)	(.998, .899)	(1.2, .24)	(.989, .947)
	smoothness	0.8		1.1		0.9	
2D	parameters	2.67	.997	(1.8, .67)	(.971, .583)	(1.7, .18)	(.967, .879)
	smoothness	0.8		1.2		0.8	

Table 2: Summary of the results in the dental cases.

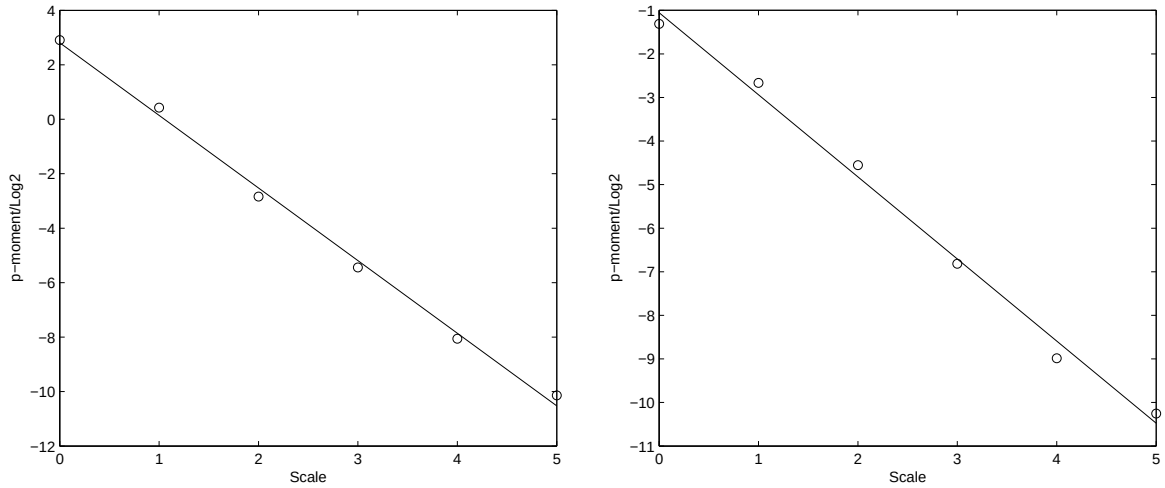

Figure 8: Least squares fit to points $(j, \log_2 m_{p,j})$. Left: 2D case. Right: 1D case.

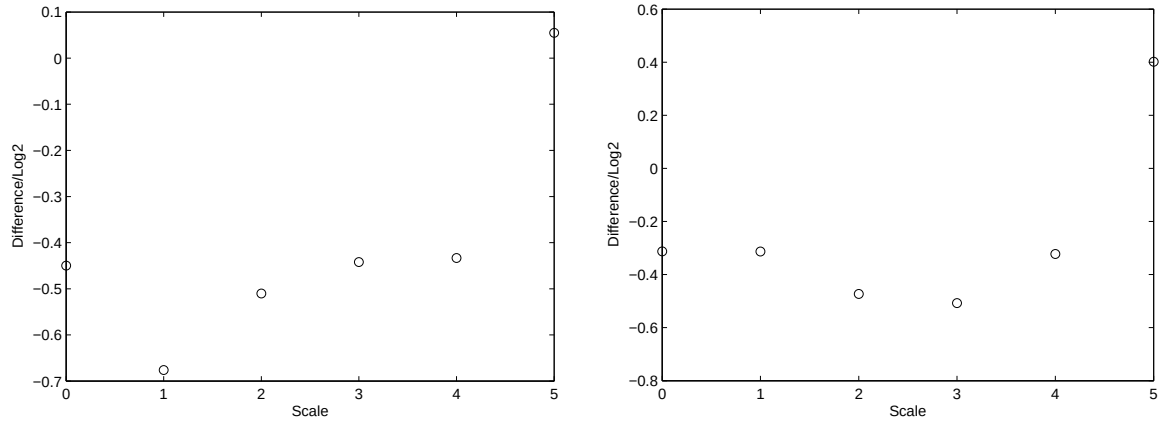

Figure 9: Points $(j, \log_2(m_{p,j}/\sigma_j^p))$. There is no descending tendency which supports the assumption $\eta_2 = 0$. Left: 2D case. Right: 1D case.

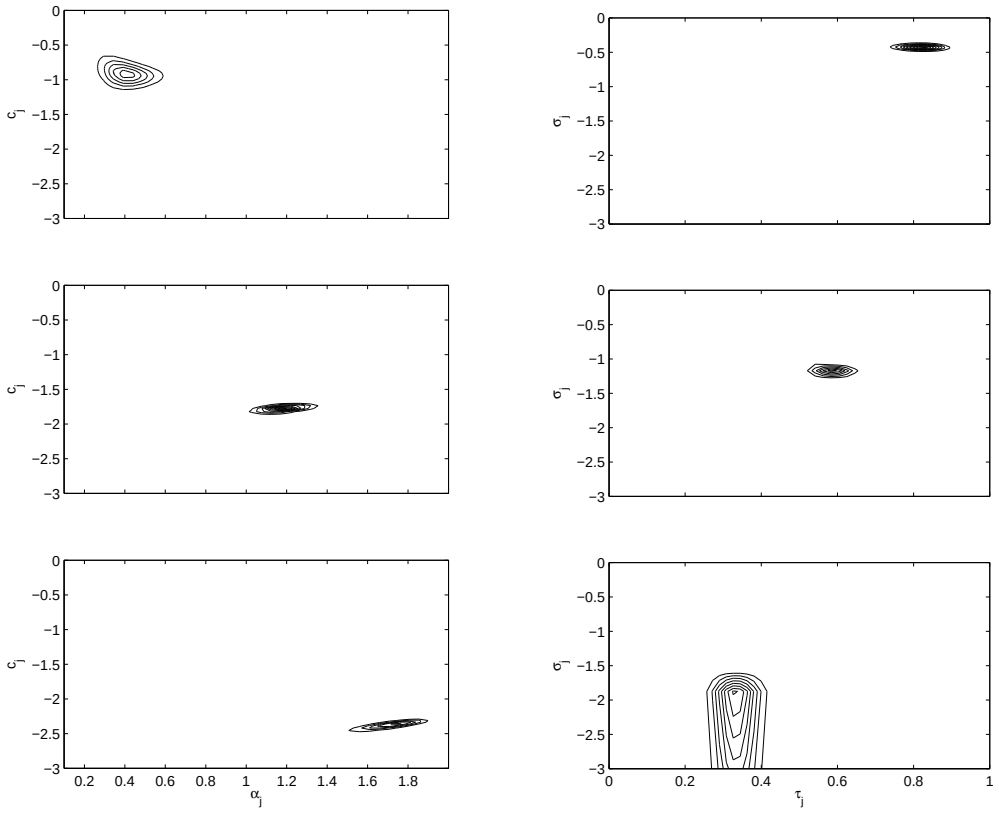

Figure 10: Likelihoods on three different scales ($j = 1, 3, 5$ from above) in the 1D case. The y-axis is in a logarithmic scale. Left: stable distribution model. Right: threshold distribution model.

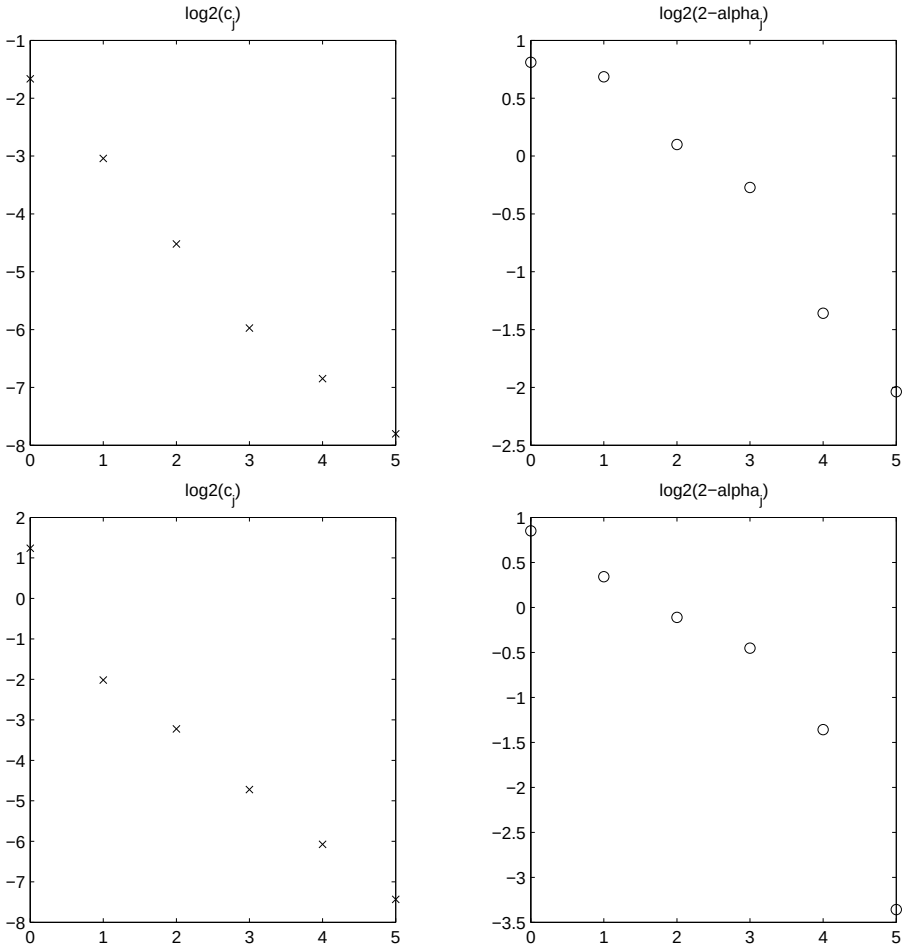

Figure 11: MAP estimates for the (c_j, α_j) of the stable distribution model in $\log_2$ scale. Top: 1D case. Bottom: 2D case.

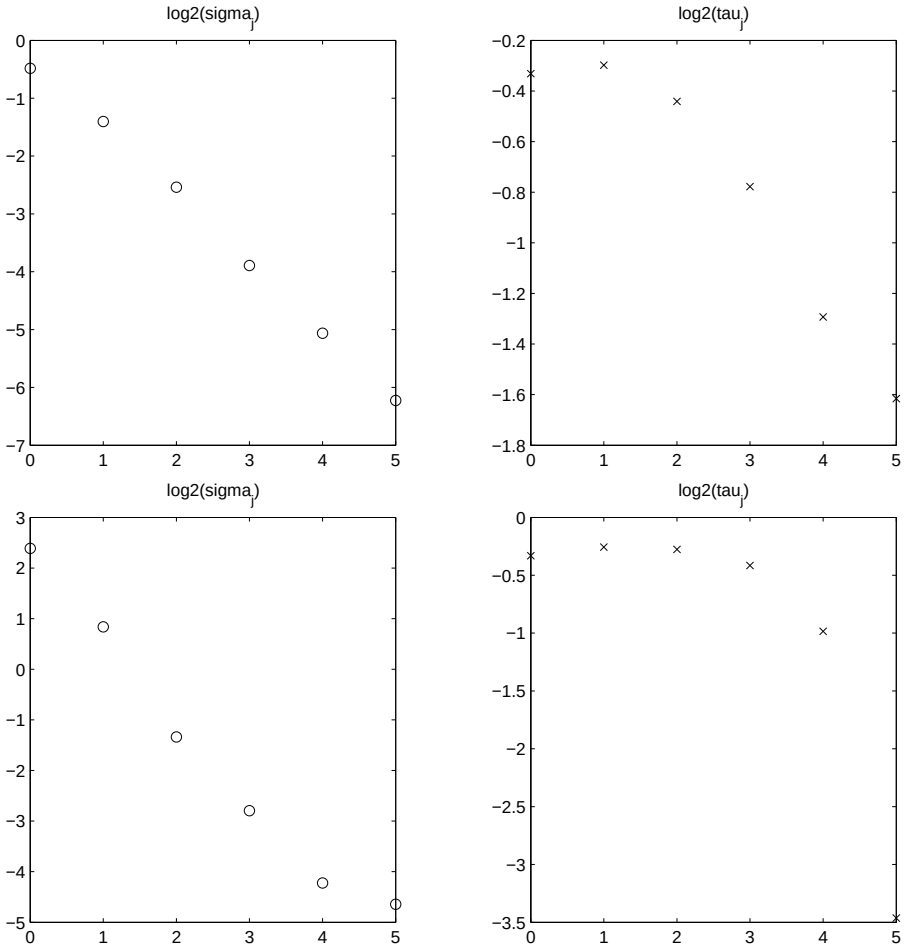

Figure 12: MAP estimates for the (σ_j, τ_j) of the threshold distribution model in $\log_2$ scale. Top: 1D case. Bottom: 2D case.

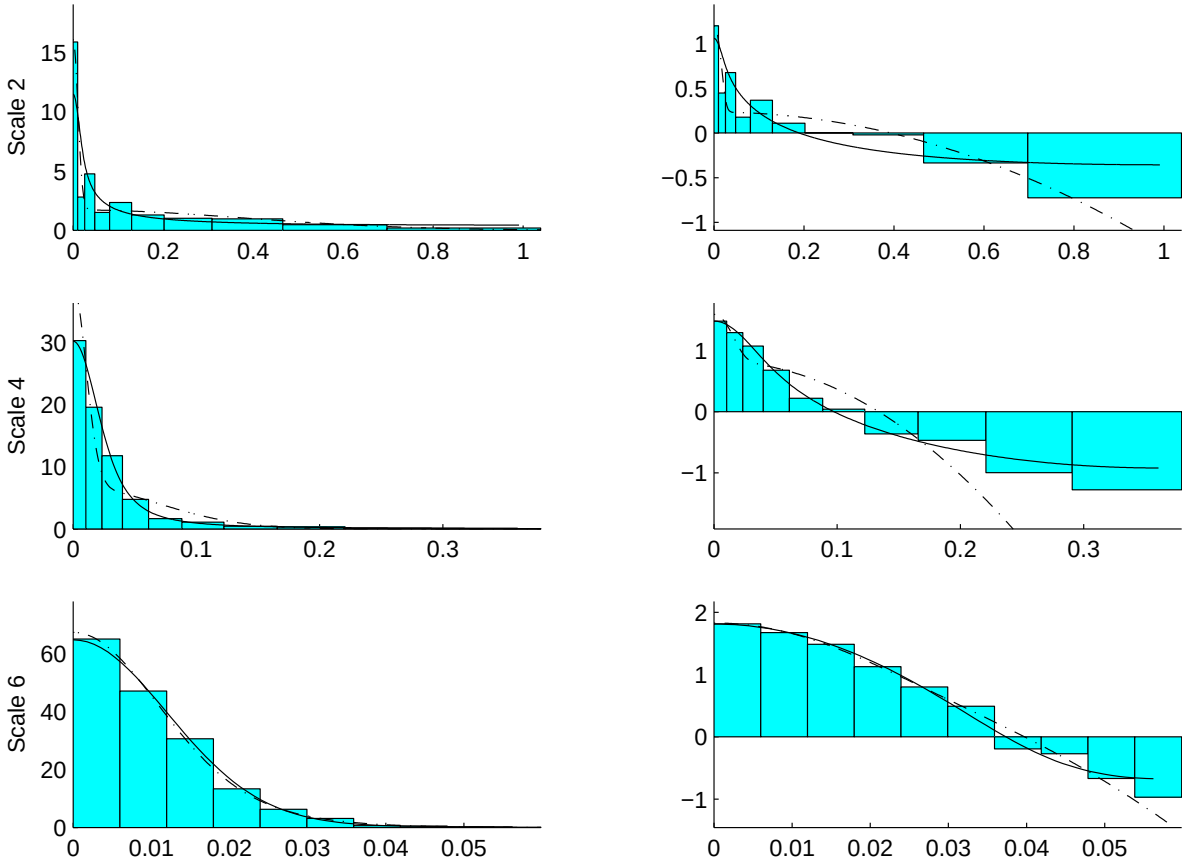

Figure 13: Left column: Histogram of the absolute values of the wavelet coefficients in 1D case and estimated model density functions on three different wavelet scales (rows). Solid line is the stable distribution model, dashed line is the stable distribution model. Right column is the logarithm of the left column.

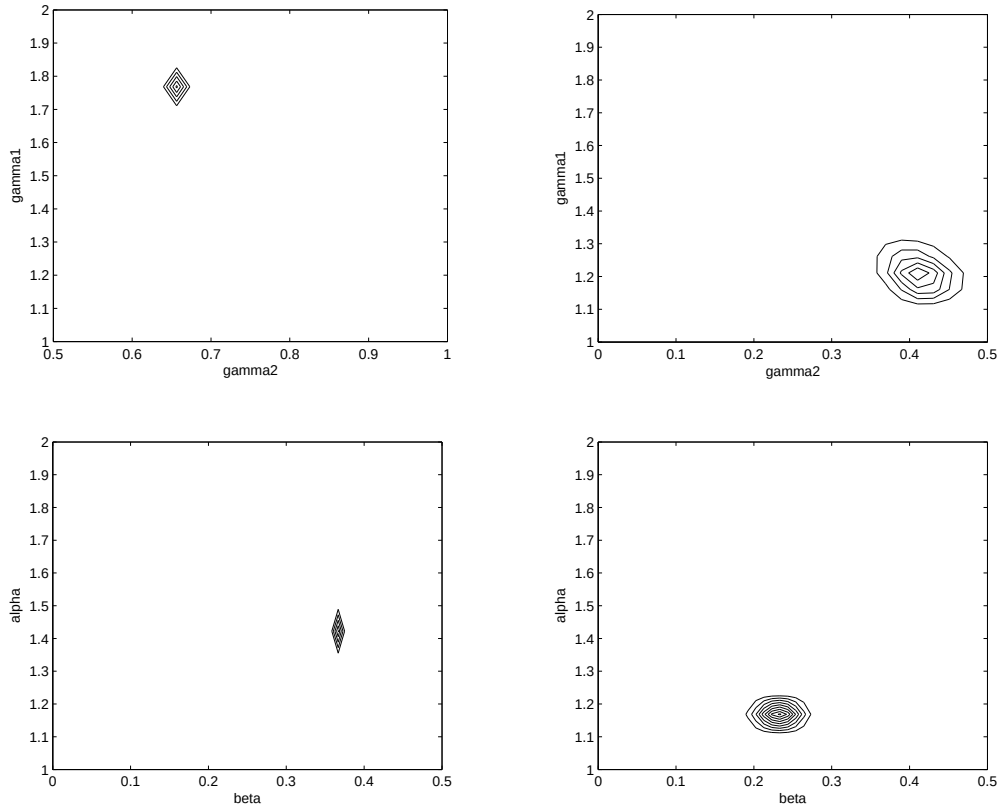

Figure 14: Posterior densities of the decay parameters. Left column: two dimensional case. Right column: one-dimensional case. Top: Stable distribution model (γ_1, γ_2). Bottom: Threshold distribution model (α, β).

8.2 Effect of the choice of the smoothness parameter

In Figures 16-15, we have demonstrated how the choice of the smoothness parameter effects to the local tomography reconstructions.

Dental data is a slice data of 23 X-ray images taken in total 187^o view angle. Only the region of interest (ROI) is drawn. Data is the same as in (Niinimäki et al., 2007) in which one can find more detailed descriptions. However, compared to (Niinimäki et al., 2007), we have used the full wavelet tree without sparsening outside of ROI. The dimension of the reconstruction is $n = 2$. The Besov space parameters are $p = 1.5$ and we have varied the smoothness parameter from $s = 0$ up to $s = 1.2$. The weight for the prior in (6.1) is $a = 0.05$. Noise has been estimated to be $\sigma_\epsilon = 0.01$ when the data vector has been normalized to be with maximum one. The resolution of the reconstruction images is 128×128. By the previous computations, the estimate for the smoothness of the 1D dental slices of the X-ray images are from $s \approx 0.8$ to $s \approx 1.1$, depending on the wavelet distribution model. This suggest that the smoothness parameter should be about $s \approx 0.4$ in the 2D reconstruction slice.

The second example is done with simulated data from the Shepp-Logan phantom. The X-ray projections were computed with the same view angles as the dental data. We added 4%

random noise to the measurement ($\sigma_\epsilon = 0.04$) and also 0.36^o random noise to the view angles (geometry parameters). Other parameters were as in the dental case.

Reconstructions with varying smoothness prior parameter s are drawn in Figures 16-15. It is obvious that the for small smoothness parameters s the over fitting to noisy data gets more role, and that for larger smoothness the reconstruction gets blurred. It is remarkable that the parameter which is suggested by the method gives also visually the best reconstruction.

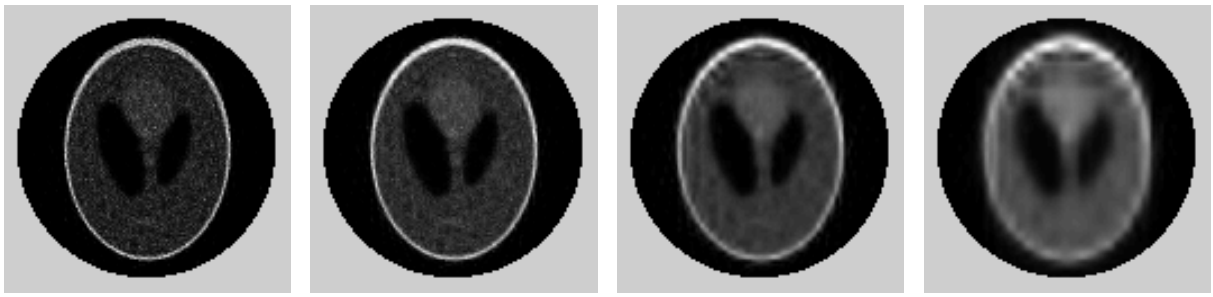

Figure 15: Reconstructions from simulated Shepp-Logan phantom data. From left: $s = 0$, $s = 0.4$, $s = 0.8$, $s = 1.2$.

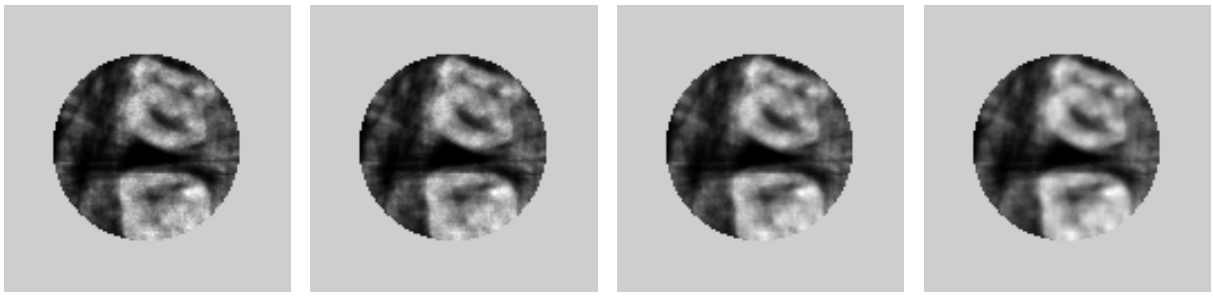

Figure 16: Reconstructions from in vitro dental data. From left: $s = 0$, $s = 0.4$, $s = 0.8$, $s = 1.2$.

9 Conclusions

The quality of X-ray tomography reconstructions depends strongly on the choice of the smoothness parameter s in a Besov prior. One way to choose the feasible smoothness parameter is to compute the reconstruction with several smoothness levels and then choose the reconstruction that is visually the best. In this article, we have studied how the suitable smoothness level could be chosen automatically when p is given.

The smoothness for the attenuation coefficient is determined by estimating the smoothness of the X-ray image. The analysis is based on the wavelet coefficients' distributions that are related to the Besov space smoothness. To determine the smoothness of a function, its wavelet coefficients are assumed to follow some distribution at each wavelet scale. The decay speed of the distributional parameters as the scale gets finer then determines in which Besov space the corresponding function belongs to.

We studied two distributional models for the wavelet coefficients, the threshold distribution model and the stable distribution model. For the threshold distribution model, we generalized to the n dimensional case the result of (Abramovich et al., 1998) connecting the Besov space and the decay parameters, Theorem 3.1. For the stable distribution model, we proved a similar result, Theorem 3.2. In both of these models, there is a distributional scale parameter and a shape parameter that are assumed to have exponential decay in the wavelet scales. We estimated the decay of these parameters from noisy wavelet coefficients. At fixed wavelet scales, the stable distribution model seems to catch the heavy tail of the wavelet distributions better. In both models, the posterior distributions for the decay parameters are rather concentrated. This can be interpreted so that the parameters are well determined when the model is fixed. However, there remains some uncertainty for the models since there is quite large variation in the decay of the shape parameters, especially with the stable distribution model. The decay of the scale parameters are following the models well.

The proofs for Theorems 3.1-3.2 show that the decay of the p-moments is essential in determining the Besov space. In fact, there is no need to know or model the shape of the wavelet coefficient distribution, see Theorem 3.3 (decaying moments model). At each wavelet scale, we approximated the p-moment from noisy data with a method involving deliberately added noise. The p-moments at each wavelet scale seem to follow the decay model very well.

Remarkably, all three models are in give reasonably close estimates for the smoothness parameter s in the case of measured data. (Somewhat surprisingly, the estimates deviate more for the simulated data.) As a conclusion we suggest that the Besov space smoothness of the X-ray images (or slices of them) should be approximated with the decaying moments method which is simple and well supported by data. The smoothness parameter value that is obtained in this way seems to be the one that gives also visually the best reconstructions, see Figure 16.

Acknowledgments

The authors would like to thank doctors Petri Koistinen and Mikko Sillanpää for useful discussions. Dental X-ray data is in courtesy of PaloDEx Group. This work was supported by the Finnish Funding Agency for Technology and Innovation (TEKES) under Contract 206/03, PaloDEx Group, and the Finnish Centre of Excellence in Inverse Problems Research.

A On convergence of positive random series

Lemma A.1. *Let $Y_j \geq 0$, $j = 1, 2, \ldots$, be independent random variables. Assume that there are $q > 1$ and $C_1, C_2 > 0$ so that*

$$\mathbb{E}Y_j \quad \leq \quad C_1 \tag{A.1}$$

and

$$\mathbb{E}(Y_j^q) \quad \leq \quad C_2(\mathbb{E}Y_j)^r \tag{A.2}$$

with some $r \geq 1$. Then

$$\sum_{j=1}^{\infty} Y_j \quad < \quad \infty \tag{A.3}$$

almost surely if and only if

$$\sum_{j=1}^{\infty} \mathbb{E}(Y_j) \quad < \quad \infty. \tag{A.4}$$

Proof. Note first that by (A.1)

$$\mathbb{E}(Y_j^q) \leq C_2 C_1^{r-1} \mathbb{E} Y_j,$$

so we can assume $r = 1$.

We use (Kallenberg, 1997, Proposition 3.14) which states that (A.3) is equivalent to

$$\sum_j \mathbb{E}\left(\min(Y_j, M)\right) \quad < \quad \infty \tag{A.5}$$

for any $0 < M < \infty$. For any M we have

$$Y_j \geq \min(Y_j, M) \geq Y_j - M^{1-q} Y_j^q,$$

or by (A.2),

$$\mathbb{E}(Y_j) \geq \mathbb{E}(\min(Y_j, M)) \geq \mathbb{E}(Y_j) - M^{1-q} C_3 \mathbb{E}(Y_j).$$

By letting

$$M = (2C_3)^{q-1},$$

we obtain

$$\mathbb{E}(Y_j) \geq \mathbb{E}(\min(Y_j, M)) \geq \frac{1}{2}\mathbb{E}(Y_j).$$

Hence, (A.4) is equivalent to (A.5). □

Remark. One does not need condition (A.2) for the "if" direction. Namely, if (A.4) is satisfied, then

$$\mathbb{E}\left(\sum_{j=1}^{\infty} Y_j\right) < \infty$$

by (Rudin, 1987, Theorem 1.27), and hence, (A.3) holds. The condition is needed for the opposite direction. The ratio

$$\frac{\mathbb{E}(Y^q)}{\mathbb{E}(Y)^r}$$

describes about the weight of the tail of the distribution. Hence, condition (A.2) can be interpreted to be a uniform bound for the tail of the distributions Y_j.

References

Abramovich, F., Sapantinas, T. and Silverman, B. 1998. Wavelet thresholding via a Bayesian approach, *J. Roy. Statist. Soc. B* **60**: 725–749.

Achim, A., Bezerianos, A. and Tsakalides, P. 2001. Novel Bayesian multi-scale method for speckle removal in medical ultrasound images, *IEEE Trans. Med. Imag.* **20**: 772–783.

Boubchir, L. and Fadili, J. 2006. A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate α-stable prior, *Pattern Recognition Letters* **27**: 1370–1382.

Bouman, C. and Sauer, K. 1993. A generalized Gaussian image model for edge-preserving MAP estimation, *IEEE Trans. Image Process.* **2**(3): 296–310.

Chang, S., Yu, B. and Vetterli, M. 2000. Adaptive wavelet thresholding for image denoising and compression, *IEEE Trans. Image Processing* **9**(9): 1532–1546.

Daubechies, I. 1992. *Ten Lectures on Wavelets*, SIAM.

Donoho, D. and Johnstone, I. 1994. Ideal spatial adaptation via wavelet shrinkage, *Biometrika* **81**: 425–455.

Donoho, D. and Johnstone, I. 1995. Adapting to unknown smoothness via wavelet shrinkage, *Journal of the American Statistical Assoc.* **90**: 1200–1224.

Frese, T., Bouman, C. and Sauer, K. 2002. Adaptive wavelet graph model for Bayesian tomographic reconstruction, *IEEE Trans. Image Process.* **7**: 756–770.

Hanson, K. M. and Wecksung, G. W. 1983. Bayesian approach to limited-angle reconstruction in computed tomography, *J. Opt. Soc. Am.* **73**: 1501–1509.

Kallenberg, O. 1997. *Foundations of Modern Probability*, Springler-Verlag.

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

Lassas, M., Saksman, E. and Siltanen, S. 2008. Discretization invariant Bayesian inversion and Besov space priors, *Submitted, preprint* .

Lee, N. and Lucier, B. 2001. Wavelet methods for inverting the Radon transform with noisy data, *IEEE Trans. Image Processing* **10**(1): 79–94.

Meyer, Y. 1992. *Wavelets and Operators*, Cambridge University Press.

Natterer, F. 2001. *The mathematics of computerized tomography*, SIAM.

Niinimäki, K., Siltanen, S. and Kolehmainen, V. 2007. Bayesian multiresolution method for local tomography in dental x-ray imaging, *Physics in Medicine and Biology* **52**: 6663–6678.

Rantala, M., Vänskä, S., Järvenpää, S., Kalke, M., Lassas, M., Moberg, J. and Siltanen, S. 2006. Wavelet-based reconstruction for limited angle X-ray tomography, *IEEE Trans. Medical Imaging* **25**(2): 210–217.

Rudin, W. 1987. *Real and complex analysis*, McGraw-Hill Book Company.

Samorodnitsky, G. and Taqqu, M. 1994. *Stable Non-Gaussian Random Processes*, Chapman and Hall, London.

Sauer, K., James, S. J. and Klifa, K. 1994. Bayesian estimation of 3-D objects from few radiographs, *IEEE Trans. Nucl. Sci.* **41**: 1780–1790.

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