Publication I

P. Hiltunen, D. Calvetti, and E. Somersalo. An adaptive smoothness regularization algorithm for optical tomography. *Optics Express*, 16(24), 19957–19977, 2008.

© 2008 Optical Society of America. Reprinted with permission.

An adaptive smoothness regularization algorithm for optical tomography

P. Hiltunen¹, D. Calvetti² and E. Somersalo^{2*}

¹Helsinki University of Technology, Espoo, 02150, Finland ²Case Western Reserve University, Cleveland, OH 44106, USA erkki.somersalo@case.edu

Abstract: In diffuse optical tomography (DOT), the object with unknown optical properties is illuminated with near infrared light and the absorption and diffusion coefficient distributions of a body are estimated from the scattering and transmission data. The problem is notoriously ill-posed and complementary information concerning the optical properties needs to be used to counter-effect the ill-posedness. In this article, we propose an adaptive inhomogenous anisotropic smoothness regularization scheme that corresponds to the prior information that the unknown object has a blocky structure. The algorithm updates alternatingly the current estimate and the smoothness penalty functional, and it is demonstrated with simulated data that the algorithm is capable of locating well blocky inclusions. The dynamical range of the reconstruction is improved, compared to traditional smoothness regularization schemes, and the crosstalk between the diffusion and absorption images is clearly less. The algorithm is tested also with a three-dimensional phantom data.

© 2008 Optical Society of America

OCIS codes: (170.3010) Image reconstruction techniques; (170.6960) Tomography; (100.3010) Image reconstruction techniques; (100.3190) Inverse problems; (100.6890) Threedimensional image processing; (100.6950) Tomographic image processing

References and links

- 1. S. R. Arridge, "Optical Tomography in Medical Imaging," Inverse Probl. 15, R41-R93 (1999).
- J. Hebden, A. Gibson, R. Yusof, N. Everdell, E. Hillman, D. Delpy, S. Arridge, T. Austin, J. Meek, and J. Wyatt, "Three-dimensional optical tomography of the premature infant brain," Phys. Med. Biol. 47, 4155–4166 (2002).
- J. Hebden and T. Austin, "Optical tomography of the neonatal brain," European Radiology 17, 2926–2933 (2007).
 J. Culver, T. Durduran, D. Furya, C. Cheung, J. Greenberg, and A. Yodh, "Diffuse optical tomography of cerebral blood flow, oxygenation, and metabolism in rat during focal ischemia," J. Cerebral Blood Flow & Metabolism 23, 911–924 (2003).
- 5. J. Kaipio and E. Somersalo, Statistical and Computational Inverse Problems (Springer Verlag, 2004).
- J. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo, "Inverse problems with structural prior information," Inverse Probl. 15, 713–729 (1999).
- B. Pogue, T. McBride, J. Prewitt, U. sterberg, and K. Paulsen, "Spatially Variant Regularization Improves Diffuse Optical Tomography," Appl. Opt. 38, 2950–2961 (1999).
- A. Douiri, M. Schweiger, J. Riley, and S. Arridge, "Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information," Meas. Sci. Technol. 18, 87–95 (2007).
- B. Brooksby, S. Jiang, C. Kogel, M. Duyley, H. Dehgani, J. Weaver, S. Poplack, B. Pogue, and K. Paulsen, "Magnetic resonance guided near infrared tomography of the breast," Rev. Sci. Instrum. 75, 5262–5270 (2004).
- P. Yalavarthy, B. Pogue, H. Dehgani, C. Carpenter, S. Jiang, and K. Paulsen, "Structural information within regularization matrices improves near infrared diffuse optical tomography," Opt. Express 15, 8043–8058 (2007).
- D. Calvetti, F. Sgallari, and E. Somersalo, "Image inpainting with structural bootstrap priors," Image and Vision Comput. 24, 782–793 (2006).

(C) 2008 OSA

- D. Calvetti and E. Somersalo, "Microlocal sequential regularization in imaging," Inverse Problems and Imaging 1, 1–11 (2007).
- D. Calvetti and E. Somersalo, "Gaussian hypermodels and recovery of blocky objects," Inverse Probl. 23, 733– 754 (2007).
- D. Calvetti and E. Somersalo, "Hypermodels in the Bayesian imaging framework," Inverse Probl. 24, 034013 (20pp) (2008).
- D. Calvetti, J. P. Kaipio, and E. Somersalo, "Aristotelian prior boundary conditions," Int. J. Math. Comp. Sci. 1, 63–81 (2006).
- M. Schweiger, S. Arridge, M. Hiraoka, and D. Delpy, "The finite element method for the propagation of light in scattering media: Boundary and source conditions," Med. Phys. 22, 1779 – 1792 (1995).
- D. Calvetti and E. Somersalo, Introduction to Bayesian Scientific Computing Ten Lectures on Subjective Computing (Springer Verlag, 2007).
- L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D 60, 259–268 (1992).
- C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," IEEE Trans. Image Process. 7, 813–824 (1998).
- P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell. 12, 629–639 (1990).
- D. Calvetti and E. Somersalo, "Recovery of shapes: hypermodels and Bayesian learning," Proc. of the Applied Inverse Problems 2007: Theoretical and Computational Aspects. J. of Physics Conference Series (to appear).
- 22. "Automatically Tuned Linear Algebra Software (ATLAS)," http://math-atlas.sourceforge.net/ (19th June 2008).
- 23. Y. Saad, Iterative Methods for Sparse Linear Systems (Society for Industrial Mathematics, 2003).
- I. Nissilä, J. Hebden, D. Jennions, J. Heino, M. Schweiger, K. Kotilahti, T. Noponen, A. Gibson, S. Järvenpää, L. Lipiäinen, and T. Katila, "Comparison between a time-domain and a frequency-domain system for optical tomography," J. Biomed Opt. 11, 064015 (2006).
- T. Tarvainen, M. Vauhkonen, V. Kolehmainen, and J. P. Kaipio, "Finite element model for the coupled radiative transfer equation and diffusion approximation," Int. J. Numer. Meth. Eng. 63, 383–405 (2006).
- J. Heino, E. Somersalo, and J. Kaipio, "Statistical compensation of geometric mismodeling in optical tomography," Opt. Express 13, 296–308 (2005).
- S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen, "Approximation errors and model reduction with an application in optical diffusion tomography," Inverse Probl. 22, 175–195 (2006).
- J. P. Kaipio and E. Somersalo, "Statistical inverse problems: discretization, model reduction and inverse crimes," J. Comp. Appl. Math. 22, 493–504 (2006).
- C. Paige and M. Saunders, "LSQR: An Algorithm for Sparse Linear Equations And Sparse Least Squares," ACM Trans. Math. Software 8, 43–71 (1982).

1. Introduction

The diffuse optical tomography (DOT) is an emerging medical imaging modality in which the optical properties of the tissue are estimated from the measured scattering and transmission of near infrared light [1]. One of the potential uses of this modality is to estimate the local oxygen saturation level in the brain tissue of neonates [2, 3], conveying indirect information of, e.g., the cerebral metabolic rate of the oxygen [4].

The optical tomography problem is very challenging from the point of view of instrumentation and data collection, as the transmission signal has a low signal to noise ratio and is corrupt by errors due to instabilities in the device. Moreover, the computational challenges are remarkable since the estimation problem is a non-linear severely ill-posed inverse problem.

The ill-posedness of the inverse problem can be compensated for by using regularization techniques, among which Tikhonov regularization [5] is the most widely popular. An alternative approach is to consider the inverse problem as one of statistical inference and to compensate the shortcomings of the data by complementary information, or prior information about the unknowns. Regularized solutions of inverse problems often have an interpretation in the Bayesian statistical framework. Conversely, while the statistical inverse problem is much richer than just one estimate, the regularization methods can be used as a guideline for designing reasonable prior densities. We make use of this interplay in this article.

A standard Tikhonov regularized solution based on a homogenous smoothness penalty pro-

duces typically a blurred diffuse image, thus failing to identify the boundaries of the internal structures of the target and possible inclusions. If prior information about the internal structures is available, e.g., from anatomical imaging modalities such as X-ray tomography or MRI, it is possible to construct structural penalties that allow abrupt changes across the structure boundaries [6, 7, 8, 9, 5, 10]. Often, however, such complementary information is not available, and a structure based on generic anatomical information alone may lead the algorithm astray. In the present article, we propose an iterative algorithm that updates dynamically the regularization functional based on the features in the current estimate. More precisely, starting with a homogenous smoothness penalty, the proposed algorithm uses the current estimate of an optical parameter as a *pilot image*, interpreted as a smooth approximation of the underlying structure, and iteratively improves the approximation, eventually finding a solution that is locally smooth but may have fast variations across the boundaries of the inclusions.

The idea of updating iteratively the smoothness penalty has previously been studied with structured grids and in the framework of linear inverse problems [11, 12, 13, 14]. In the present work, we extend these ideas to unstructured grids which makes them suitable for applications where the forward problem calls for finite element modelling. Moreover, the non-linearity of the forward map increases the complexity of the problem. Finally, the construction of the anisotropic smoothness penalty with undetermined boundary conditions is new, extending the idea of Aristotelian boundary conditions presented in [15]. The algorithm presented here is not limited to optical tomography but is also suitable for other linear or non-linear inverse problems where unstructured grid is used to represent distributed parameters in a discrete form.

2. Material and methods

2.1. Forward model and inverse problem

The incoherent light propagation in scattering medium can be described with the radiative transfer equation. Being an integro-differential equation, its numerical approximation leads easily to computational problems of prohibitive size. When the medium is strongly scattering, the radiative transfer equation is often replaced with the diffusion approximation.

Assuming that the light source is time harmonic with modulation frequency $\omega > 0$, the photon density $\phi = \phi(\mathbf{x})$ in the diffusion approximation satisfies the elliptic partial differential equation

$$-\nabla \cdot \left(\kappa(\mathbf{x}) \nabla \phi(\mathbf{x}) \right) + \left(\mu_a(\mathbf{x}) + \frac{\mathbf{i}\omega}{c} \right) \phi(\mathbf{x}) = 0, \quad \mathbf{x} \in \Omega,$$
(1)

where $\Omega \in \mathbb{R}^d$, d = 2,3, is a bounded domain representing the target, and the diffusion coefficient is defined in terms of the absorption coefficient $\mu_a(\mathbf{x})$ and reduced scattering coefficient $\mu'_s(\mathbf{x})$ as $\kappa(\mathbf{x}) = (d(\mu_a(\mathbf{x}) + \mu'_s(\mathbf{x})))^{-1}$. The appropriate boundary condition is a Robin condition,

$$\phi(\mathbf{x}) + 2\zeta \kappa(\mathbf{x}) \frac{\partial \phi}{\partial n}(\mathbf{x}) = q(\mathbf{x}), \quad \mathbf{x} \in \partial \Omega,$$
⁽²⁾

where q is the boundary source and the impedance coefficient ζ takes into account the refractive mismatch of the body and the outside air on the boundary. We use the semi-empirical model

$$\zeta = \frac{1+R}{1-R}, \quad R \approx -1.4399 n^{-2} + 0.7099 n^{-1} + 0.6681 + 0.0636n,$$

where *R* is the reflection coefficient of the boundary and *n* is the refraction index of the medium, and $n_{\text{air}} = 1.0$, see [16] for further details of the model. The modulation frequency of the source is $\omega = 2\pi f$, f = 100MHz and *c* is speed of light in the medium, assumed to be constant.

(C) 2008 OSA

The inverse problem of optical tomography is to estimate the optical properties of the body from the scattering and transmission of light measured on the surface of the body. In terms of the diffusion approximation model of light propagation, the problem is tantamount to estimating the absorption and the diffusion coefficient pair $\{(\mu_a(\mathbf{x}), \kappa(\mathbf{x})) \mid \mathbf{x} \in \Omega\}$ from the noisy observations of the outcoming photon densities corresponding to all applicable boundary sources. Given the boundary source *q*, the outcoming photon density on the boundary, or exitance, is defined as

$$\Gamma(\mathbf{x}) = \Gamma_q(\mathbf{x}) = -\kappa(\mathbf{x})\frac{\partial\phi}{\partial n}(\mathbf{x}) \quad \mathbf{x} \in \partial\Omega.$$
(3)

Hence, the inverse problem is to determine the pair (μ_a, κ) from all available pairs (q, Γ_q) , where the measured exitance Γ_q is corrupted by noise.

In the present formulation, the exitance is a complex quantity, which can be expressed in terms of the logarithm of the amplitude and the phase shift,

$$F(\mu_a, \kappa; q) = \begin{bmatrix} \operatorname{Re}\log\left(\Gamma_q(\mathbf{x})\right) \\ \operatorname{Im}\log\left(\Gamma_q(\mathbf{x})\right) \end{bmatrix} = \begin{bmatrix} \log|\Gamma_q(\mathbf{x})| \\ \arg\Gamma_q(\mathbf{x}) \end{bmatrix}.$$
(4)

In practice, we may apply a finite number of linearly independent boundary source patterns, q^{ℓ} , $1 \leq \ell \leq L$, and the data consists of the corresponding photon densities, denoted by Γ^{ℓ} , recorded only at few boundary points r_j , $1 \leq j \leq m$, where *m* is the number of the detectors at the boundary. Using the amplitude-phase representation and ignoring the measurement noise, the data set is thus

$$D = \left\{ \left(\log |\Gamma^{\ell}(r_j)|, \arg(\Gamma^{\ell}(r_j)) \right) \mid 1 \le \ell \le L, \ 1 \le j \le m \right\}.$$
(5)

2.1.1. Discretization

Assume that a triangular or tetrahedral finite element meshing of the domain $\Omega \in \mathbb{R}^d$, d = 2,3, is given, with vertices \mathbf{x}_i , $1 \le i \le N$, and assume that the unknown optical distributed parameters μ_a and κ are discretized and represented by the collocation values ($\mu_a(\mathbf{x}_i), \kappa(\mathbf{x}_i)$) at the nodal points. We denote by $p \in \mathbb{R}^{2N}$ the vector containing the *N* collocation values of both unknown parameters. Using finite elements, we may discretize the forward diffusion approximation model to obtain a discrete forward model. Let us denote by $y^{\ell} \in \mathbb{R}^{2m}$ the vector containing the boundary data (5) corresponding to the amplitude and phase of the exitance Γ^{ℓ} at the detector locations r_j , $1 \le j \le m$. By stacking all the observation vectors into a single vector *y*, the finite element model leads to a discrete forward model

$$y = f(p), \quad f: \mathbb{R}^{2N} \to \mathbb{R}^{2M}, M = Lm.$$
 (6)

2.1.2. Statistical inference

In the Bayesian statistical framework, all the variables are treated as random variables, the randomness being an expression of the lack of information about the values that they take on [5, 17]. The inverse problem in this framework is to infer on the probability density of the unknown vector of primary interest based on the realized values of the random variable representing the observable.

To set up the statistical model, consider the forward model (6). Assuming that the observation is corrupted by additive noise that is independent of the optical properties of the body, we write

$$y = f(p) + e, \quad e \sim \pi_{\text{noise}}(e),$$

the notation above indicating that the probability density of the noise is π_{noise} . This model allows us to write the *likelihood density* of the measurement,

$$\pi(y \mid p) \propto \pi_{\text{noise}}(y - f(p)),$$

(C) 2008 OSA

where the symbol " \propto " stands for "equal up to a multiplicative constant". Hence, the data conditional on the true signal is distributed around the noiseless value f(p) as the noise is distributed around the origin.

Assume that complementary information concerning the unknown p is available, and this information is expressed in the form of a probability density $\pi_{prior}(p)$. This density, the *prior probability density*, expresses our belief about values that we expect, where $\pi_{prior}(p)$ is large, and on the other hand, it excludes values of p that we believe to be unlikely or impossible, where $\pi_{prior}(p)$ is small or vanishes. The *joint probability density* of the random variables p and y can be written as

$$\pi(p, y) = \pi(y \mid p)\pi_{\text{prior}}(p) \propto \pi_{\text{noise}}(y - f(p))\pi_{\text{prior}}(p)$$

Reversing the roles of the observable and unknown in the equation for the joint probability density, we may define the probability density of *p* conditioned on the observation. The result is Bayes' equation for the *posterior probability density*,

$$\pi_{\text{post}}(p) = \pi(p \mid y) = \frac{\pi(p, y)}{\pi(y)} \propto \pi_{\text{noise}}(y - f(p))\pi_{\text{prior}}(p), \quad y = y_{\text{measured}}.$$
 (7)

Above, $\pi(y)$ is the marginal density of *y*, obtained by integrating out the variable *p* from the joint probability density, and it is ignored here as it contributes only to normalizing of the posterior density. In the Bayesian framework, the posterior density is the solution of the inverse problem, since it is an expression of the information that is available about *p*, taking the prior belief, the noisy measurement and the model between the unknown and the data into account.

A particular instance of the posterior density is when the prior density and the likelihood are Gaussian. If

$$\begin{split} \pi_{\text{noise}}(e) &\propto & \exp\left(-\frac{1}{2}(e-e^*)^{\mathrm{T}}C_{\mathrm{n}}^{-1}(e-e^*)\right), \\ \pi_{\text{prior}}(p) &\propto & \exp\left(-\frac{1}{2}(p-p^*)^{\mathrm{T}}C_{\mathrm{p}}^{-1}(p-p^*)\right), \end{split}$$

where the noise and prior covariance matrices C_n and C_p are positive definite, then the posterior density assumes the form

$$\pi_{\text{post}}(p) \propto \exp\left(-\frac{1}{2}(y - e^* - f(p))^{\mathrm{T}}C_{\mathrm{n}}^{-1}(y - e^* - f(p)) - \frac{1}{2}(p - p^*)^{\mathrm{T}}C_{\mathrm{p}}^{-1}(p - p^*)\right).$$

By writing the symmetric (e.g., Cholesky) decompositions of the inverses of the covariance matrices,

$$C_{\rm n}^{-1} = S^{\rm T}S, \quad C_{\rm p}^{-1} = L^{\rm T}L,$$

the density can be expressed in the form

$$\pi_{\text{post}}(p) \propto \exp\left(-\frac{1}{2}\|S(y-e^*-f(p))\|^2 - \frac{1}{2}\|L(p-p^*)\|^2\right),$$

This equation is particularly useful for establishing the connection between regularization and statistics. We observe that the maximizer of the posterior density, known as the Maximum A Posteriori (MAP) estimate of p, denoted by p_{MAP} , satisfies

$$p_{\text{MAP}} = \operatorname{argmin}\left(\|S(y - e^* - f(p))\|^2 + \|L(p - p^*)\|^2\right),$$

24 November 2008 / Vol. 16, No. 24 / OPTICS EXPRESS 19961

(C) 2008 OSA

which is the Tikhonov regularized solution of the inverse problem with the penalty functional

$$J(p) = \|L(p - p^*)\|^2$$

and the regularization parameter equal to unity. Hence, an appropriate choice of a Gaussian prior density leads to the Tikhonov regularized solution as the MAP estimate, and conversely, a judiciously chosen penalty functional helps to design a Gaussian prior density. We use this well known interplay in both direction in the discussion to ensue.

2.1.3. Construction of the prior

In the Bayesian framework, the prior density comprises the prior information, or prior belief, concerning the unknown field, in the present problem the absorption and scattering coefficient. The challenge in constructing the prior density is how to translate qualitative information concerning the unknown into a quantitative form that is expressible as a probability distribution. The selection of a good prior model is by no means unique, and the quality can often be assessed only by its capacity of producing meaningful a posteriori estimates that respect the qualitative a priori constraints.

In the present context, we assume that the a priori information concerning the optical properties of the object is summarized in the following qualitative description: *the object consists of well defined simple subdomains with slowly varying optical properties, while across the subdomain boundaries, sharp variations may occur.* Such a description is often said to define a *blocky object*, and techniques for retrieving the parameter fields in such cases have been developed [18, 19, 20, 13, 12, 14]. In the present work, the goal is to extend the approach proposed in [12] to a strongly non-linear problem with a non-structured discretization. An extra computational challenge comes from the complexity of the forward model, requiring both effective numerical algorithms and use of memory.

The construction of the prior density consists of two steps. To allow the optical parameters to easily jump across the subdomain boundaries, we use inhomogenous direction sensitive smoothness priors with dynamic adaptation [11, 12]. To avoid boundary artifacts, we adjust the a priori variances in the smoothness prior model to be nearly constant throughout the body [15].

2.1.4. Structural interior prior

The first step towards constructing a Gaussian direction sensitive second order smoothness prior is the selection of a suitable penalty functional. Let $u : \Omega \to \mathbb{R}$ denote a twice differentiable function to be estimated from indirect observations. In the present application, u is either the diffusion coefficient or the absorption coefficient. Furthermore, let $\lambda : \Omega \to \mathbb{R}$ a given nonnegative weight function such that $\lambda \mid_{\partial\Omega} = 1$. We can then define an isotropic second order penalty functional with weight $\lambda \ge 0$ in the interior of Ω of the form

$$\mathscr{J}_{\rm int}(u;\lambda) = \int_{\Omega} (\nabla \cdot (\lambda \nabla u))^2 d\mathbf{x}.$$
(8)

The role of the penalty functional in the estimation problem is to favor solutions yielding small values. To understand better for which functions *u* the penalty takes on small values note that, when λ is fixed, the minimizer of $u \mapsto \mathscr{J}(u;\lambda)$ is a steady state diffusion field with the diffusion coefficient λ . In other words, the function λ controls which parts of the image *u* mix together via diffusion, thus conveying structural prior information to our model. Strategies for the choice of λ will be discussed later in this section.

Assume that a finite element triangular or tetrahedral mesh over Ω with vertices \mathbf{x}_i is given. Let Ω_i denote a Voronoi cell containing an interior vertex \mathbf{x}_i , i.e.,

$$\Omega_i = \left\{ \mathbf{x} \in \Omega \mid |\mathbf{x} - \mathbf{x}_i| < |\mathbf{x} - \mathbf{x}_j|, \quad i \neq j \right\},\$$

(C) 2008 OSA



Fig. 1. Solid lines presents the triangular mesh. Dashed line is a voronoi cell Ω_i connected to node x_i .

see Fig. 1. To discretize the penalty functional, we write an approximation

$$\nabla \cdot (\lambda(\mathbf{x}_i) \nabla u(\mathbf{x}_i)) \approx \frac{1}{|\Omega_i|} \int_{\Omega_i} \nabla \cdot (\lambda \nabla u) d\mathbf{x} = \frac{1}{|\Omega_i|} \int_{\partial \Omega_i} \lambda \mathbf{n} \cdot \nabla u dS,$$

where the second identity follows from the divergence theorem, with *n* denoting the exterior normal vector of the boundary $\partial \Omega_i$. To approximate the integral in the rightmost term, we denote by \mathbf{x}_i an interior vertex and by $1 \le i \le N_{\text{int}}$, and $\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^p$ its neighboring nodes. Furthermore, we let $\mathbf{y}_i^j = (\mathbf{x}_i^j + \mathbf{x}_i)/2$ be the midpoint of the edge joining neighboring vertices and $\partial \Omega_i^j$ the facet of the polyhedron $\partial \Omega_i$ separating the vertices \mathbf{x}_i and $\mathbf{x}_i^j, 1 \le j \le p$. Then

$$\frac{1}{|\Omega_i|} \int_{\partial \Omega_i} \lambda \mathbf{n} \cdot \nabla u dS \approx \sum_{j=1}^p \frac{|\partial \Omega_i^j|}{|\Omega_i|} \lambda(y_i^j) (u(\mathbf{x}_i^j) - u(\mathbf{x}_i)).$$
(9)

Collecting in the vector *p* the values of the function *u* at the collocation points, $p_i = u(\mathbf{x}_i)$, the above approximation leads to a matrix representation of the second order differential in the interior points,

$$\nabla \cdot (\lambda(\mathbf{x}_i) \nabla u(\mathbf{x}_i)) \approx (L_{\text{int}} p)_i, \quad L_{\text{int}} \in \mathbb{R}^{N_{\text{int}} \times N}$$

where the matrix L_{int} , whose elements are computed using (9) depends on the weight function λ . The discrete counterpart of the interior penalty functional (8) is therefore

$$J_{\text{int}}(p;\lambda) = \|L_{\text{int}}p\|^2.$$

We partition the vector p into the vectors $p_{int} \in \mathbb{R}^{N_{int}}$ and $p_{bdry} \in \mathbb{R}^{N_{bdry}}$ containing the interior node and the boundary node values, respectively. Assuming, for the time being, that the boundary values are known, we may define the following probability density of the interior nodes

24 November 2008 / Vol. 16, No. 24 / OPTICS EXPRESS 19963

(C) 2008 OSA

conditional on the boundary nodes,

$$\pi(p_{\text{int}}|p_{\text{bdry}}) \propto \exp\left(-\frac{1}{2}\|L_{\text{int}}p\|^2\right).$$
(10)

To obtain a proper prior for the vector *p*, taking into account that the boundary values are also unknown, it suffices to construct a prior density for the boundary node values.

2.1.5. Variance adjustment at the boundary

Consider now the boundary collocation points $\mathbf{x}_i \in \partial \Omega$, $1 \le i \le N_{\text{bdry}}$. The smoothness penalty functional derived above involves the values of *u* at the boundary points, but in order to obtain a proper prior density, additional information concerning the regularity over the boundary is required. Our approach follows the reasoning in [15]. Here, for simplicity, we restrict the explanation to the case when the surface $\partial \Omega$ is smooth.

Assuming that along the boundary, the function *u* is second order smooth, we define a second order boundary penalty,

$$\mathcal{J}_{\rm bdry}(u) = \int_{\partial\Omega} (\Delta_{\tau} u)^2 dS, \tag{11}$$

where Δ_{τ} denotes the surface Laplacian. The mesh defined over Ω restricted to the boundary induces a meshing of the boundary. We discretize the surface Laplacian analogously with the interior Laplacian, that is,

$$\Delta_{\tau} u(\mathbf{x}_i) \approx \frac{1}{|\Omega_i \cap \partial \Omega|} \int_{\Omega_i \cap \partial \Omega} \Delta_{\tau} u dS = \frac{1}{|\Omega_i \cap \partial \Omega|} \int_{C_i} \mathbf{v} \cdot \nabla u d\ell, \tag{12}$$

where C_i is the boundary of $\Omega_i \cap \partial \Omega$ and ν is its normal vector tangent to $\partial \Omega$. If $\Omega \subset \mathbb{R}^2$, the calculation of the rightmost integral reduces to evaluations at two points. Approximating the normal derivatives with finite differences at neighboring collocation points, as in the case of interior points, we arrive at a discrete approximation of the boundary penalty functional,

$$J_{\rm bdry}(p_{\rm bdry}) = \|L_{\rm bdry}p_{\rm bdry}\|^2, \quad L_{\rm bdry} \in \mathbb{R}^{N_{\rm bdry} \times N_{\rm bdry}}$$

where L_{bdry} is the finite difference matrix which approximates the integral (12). We then define the Gaussian marginal smoothness density corresponding to this penalty function,

$$\pi(p_{\text{bdry}}) \propto \exp\left(-\frac{1}{2} \|\alpha L_{\text{bdry}}p\|^2\right),$$
(13)

where $\alpha > 0$ is a scaling constant, whose role will be discussed below.

Finally, the product of the conditional density $\pi(p_{int}|p_{bdry})$ and the marginal density $\pi(p_{bdry})$ yields the prior density

$$\pi_{\text{prior}}(p) = \pi(p_{\text{int}}|p_{\text{bdry}})\pi(p_{\text{bdry}}).$$

If we partition the matrix L_{int} according to the partitioning of p into interior and boundary values, $L_{\text{int}} = [L'_{\text{int}} L''_{\text{int}}], L'_{\text{int}} \in \mathbb{R}^{N_{\text{int}} \times N_{\text{bdry}}}, \text{we can express the prior density in the form}$

$$\begin{aligned} \pi_{\text{prior}}(p) &\propto & \exp\left(-\frac{1}{2} \left\| \begin{bmatrix} L_{\text{int}}' & L_{\text{int}}'' \end{bmatrix} \begin{bmatrix} P_{\text{int}} \\ P_{\text{bdry}} \end{bmatrix} \right\|^2 - \frac{1}{2} \|\alpha L_{\text{bdry}}p\|^2 \right) \\ &= & \exp\left(-\frac{1}{2} \|Lp\|^2\right), \end{aligned}$$

(C) 2008 OSA

where L is the matrix

$$L = \begin{bmatrix} L'_{\text{int}} & L''_{\text{int}} \\ 0 & \alpha L_{\text{bdry}} \end{bmatrix}.$$
 (14)

To choose the value of the parameter α , we require that when $\lambda = 1$, the variance of components p_i in the interior and at the boundary are of the same order of magnitude, recalling that the variance of a component p_i can be calculated by the equation

$$\operatorname{var}(p_i) = (L^{-1}L^{-T})_{ii} = ||L^{-T}e_i||^2,$$

where e_i is the *i*th unit coordinate vector. The details can be found in Appendix A.

2.1.6. Structural prior

In this section we discuss the choice of the function λ in the definition of the smoothness penalty functional and an adaptive updating scheme for it. From Eq. (9) it follows that $\lambda(\mathbf{y}_i^j)$ can be viewed as a coupling constant between the neighboring collocation values $u(\mathbf{x}_i)$ and $u(\mathbf{x}_i^j)$. The coupling should be small if we *believe* that the function value is likely to change significantly between these two collocation points, while it should be large if the collocation values are believed to be close to each other.

If we have available a smooth approximation \hat{u} of the unknown function *u* to be estimated, which we will refer to as *pilot function*, we can define λ to be

$$\lambda(\mathbf{y}_i^j) = \frac{1}{1 + |\tau \mathbf{D}_{\mathbf{y}_i^j} \widehat{u}(\mathbf{y}_i^j)|^k},\tag{15}$$

where $\tau > 0, k \in \mathbb{N}$ are shape parameters of the function λ , and $D_{\mathbf{v}_i^j}$ denotes the directional derivative in the direction \mathbf{v}_i^j ,

$$\mathbf{v}_i^j = \frac{1}{h_i^j} (\mathbf{x}_i^j - \mathbf{x}_i), \quad h_i^j = |\mathbf{x}_i^j - \mathbf{x}_i|.$$

This selection of λ makes the smoothness prior direction sensitive. The rational for this choice is that when the pilot function has a strong gradient in the given direction \mathbf{v}_i^j , the penalty for the unknown function *u* for changing rapidly while passing form \mathbf{x}_i to \mathbf{x}_i^j should be small, to allow a rapid variation. The matrix obtained by modifying *L* using the pilot image \hat{u} , denoted by $L_{\hat{u}}$ can be formed knowing only the values of the pilot image at the interpolation points y_i^j .

2.2. Adaptation

We now focus our attention to the construction of the structural priors for the optical tomography problem that we are considering. Assume that $\hat{\kappa}$ and $\hat{\mu}_a$ are the pilot functions of the diffusion and absorption coefficients, respectively. Since these coefficients are estimated simultaneously from the data, we define the matrix

$$W = W_{\hat{\kappa},\hat{\mu}_{a}} = \begin{bmatrix} \gamma_{1}L_{\hat{\kappa}} & \\ & \gamma_{2}L_{\hat{\mu}_{a}} \end{bmatrix} \in \mathbb{R}^{2N \times 2N},$$
(16)

where $\gamma_1, \gamma_2 > 0$ are scaling constants. Similarly as in [11, 12], we propose an adaptive algorithm for updating iteratively the penalty matrix (16) based on the current information about the optical parameters.

1. Initialize $\hat{\kappa} = 1$, $\hat{\mu}_a = 1$, set j = 1 and prescribe a maximum number of iterations n_{iter} .

(C) 2008 OSA

- 2. Using the current pilot images $\hat{\kappa}$ and $\hat{\mu}_a$, form the matrix *W* according to (16).
- 3. Calculate $p = (\mu_a(\mathbf{x}_i), \kappa(\mathbf{x}_i)) \in \mathbb{R}^{2N}$ which satisfies

$$p = \underset{p \in \mathbb{R}^{2N}}{\arg\min} \left\{ \|S(y - e^* - f(p))\|^2 + \delta \|W(p - p^*)\|^2 \right\}, \quad \delta > 0.$$
(17)

4. If $j = n_{\text{iter}}$, stop, otherwise

(a) Update the pilot images $\hat{\kappa}$ and $\hat{\mu}_a$ using the current estimate of *p*, normalized by the maximum values,

$$\widehat{\kappa} = rac{1}{\max \kappa(\mathbf{x}_i)}\kappa, \quad \widehat{\mu}_{\mathrm{a}} = rac{1}{\max \mu_{\mathrm{a}}(\mathbf{x}_i)}\mu_{\mathrm{a}}.$$

Compute the values at the interpolation points y_i^j , according to the equations

$$\widehat{\kappa}(\mathbf{y}_i^j) = \frac{1}{2}(\widehat{\kappa}(\mathbf{x}_i) + \widehat{\kappa}(\mathbf{x}_i^j)), \quad \widehat{\mu}_{a}(\mathbf{y}_i^j) = \frac{1}{2}(\widehat{\mu}_{a}(\mathbf{x}_i) + \widehat{\mu}_{a}(\mathbf{x}_i^j)),$$

(b) Increase *j* by one and continue from 2.

The selection of the value of Tikhonov regularization parameter $\delta > 0$ in (17) will be discussed in the next section where we address various issues concerning the practical implementation of the algorithm and the details of the numerical experiments. The solution of the minimization subproblem (17) is computed with a damped Gauss-Newton algorithm.

We point out that the updating the penalty functional is tantamount to updating the prior density based on the data, which may look dubious in the classical Bayesian paradigm. Although not rigorously analyzed in the present form, the updating of the prior may be justified by in the context of hierarchical models: see [13, 14, 21].

3. Results

In this sections we illustrate the performance of the proposed algorithm, when applied to both synthetic and real phantom data. The algorithm has been implemented in C, so as to take advantage of the shared memory parallelism and to take advantage of automatically tuned BLAS libraries [22] for linear algebraic operations.

3.1. Simulated data

We start by testing our algorithm with simulated data in two dimensions. The target is a circle of diameter 7 cm, with three circular inclusions, each of diameter 1 cm. Two of the inclusions have a scattering coefficient different from the constant background and two of them have absorbtion coefficients different from the constant background, therefore one of the inclusions has both scattering and absorption coefficients different from the background.

Two different simulations are carried out. In the first one, the optical parameters of the inclusions are constants, hence we refer to it as the "blocky inclusion model". In the second one, which we will refer to as the "smooth inclusion model", the optical parameter functions within the inclusions are Gaussian humps. The optical properties of the background and the inclusions are listed in the Table 1, and the targets are plotted in Fig. 2. When the smooth inclusion model is used, the tabulated values refer to the peak values of the Gaussian humps. The refraction index in the body for this simulation is n = 1.4 assumed known, and 17 optical fibers are uniformly spaced along the circumference of the target. The data are simulated by sequentially letting one of the fibers act as a source and the remaining ones as detectors, leading to

Table 1. Optical parameters of the simulated object. The numbering of the inclusion is in agreement with Fig. 2.

	$\mu_{a} [{ m mm}^{-1}]$	κ [mm]
Background	0.01	0.33
Inclusion 1	0.02	0.16
Inclusion 2	0.01	0.16
Inclusion 3	0.02	0.33

 $16 \times 17 = 272$ measured amplitudes and phases. Due to the wide dynamical range, amplitudes are routinely expressed in the logarithmic scale, corresponding to the model of Section 2.1.

To avoid the "inverse crime" of using the same computational forward and inverse problem, the simulated data are produced with a FEM model using 3790 triangular elements with quadratic basis functions, while the model used in the solution of the inverse problem uses 3566 triangular elements with quadratic basis functions.

The optical parameters are represented in a third triangular mesh of 2190 elements and are interpolated to the FEM meshes by piecewise linear interpolants. In our simulation we add zero mean Gaussian noise with standard deviation $\sigma_{\log|\Gamma|} = 10^{-3} \max |\log|\Gamma|| = 0.0228$ for the the component added to the logarithm of the amplitude, and of $\sigma_{\varphi} = 10^{-3} \max |\varphi| = 0.0016$ for the component added to the phase shift. Therefore the Cholesky factor of inverse of the noise covariance matrix is the diagonal matrix

$$S = \begin{bmatrix} \sigma_{\log|\Gamma|}^{-1} I & 0\\ 0 & \sigma_{\varphi}^{-1} I \end{bmatrix}.$$
 (18)

In the construction of the penalty functional the shape parameters in (15) to be k = 2 and $\tau = 50$. The scaling coefficients γ_1 and γ_2 appearing in the definition (16) of W are set in the first iteration round, when the pilot images are flat and the penalty is a homogenous second order smoothness penalty, and never changed afterwards. The criterion used for the selection of the scaling is that we no pixel values of the simulated inclusions should not deviate from the background by more than two standard deviations. This adjustment can be done once, together with the boundary variance adjustment. In a realistic situation, the standard deviation needs to be estimated based on the prior belief of the dynamical range about the coefficients.

The update of the value of Tikhonov regularization parameter δ should be be adjusted at each iteration sweep, e.g., using the Morozov discrepancy principle, but this strategy would increase the computational burden significantly. In our computed examples, set the value of *delta* to 10^{-3} once and for all. The maximum number of iteration steps was set to $n_{\text{iter}} = 20$, and up to 15 Gauss-Newton iteration steps were allowed for the solution of the minimization problem. Within each Gauss-Newton step, the new search direction is computed by solving the corresponding linear system with the GMRES algorithm [23].

The reconstructions the blocky inclusion model are displayed in Fig. 3. The top left panel a) corresponds to the absorption coefficient μ_a and the top right one b) to the diffusion coefficient κ . The image in the left of each panel displays the reconstructions after the first outer iteration before the smoothness penalty adaptation, while the image on the right shows the final iteration ($n_{\text{iter}} = 20$). The corresponding profiles of the reconstructions along the circular curve indicated in Fig. 2, traversed in counterclockwise direction, are shown underneath.

The advantage of using the penalty adaptation in this example is quite dramatic: after the adaptation, the contrast of the inclusions is sharp and the dynamical range improved. Interest-



Fig. 2. Images of the absorption coefficient μ_a (top) and diffusion coefficient κ (bottom) of the two dimensional simulated targets. The left hand side images correspond to the blocky inclusion model and the right hand side images to the smooth inclusion model. The parameter values are as in Table 1.

ingly, the adaptation is also capable of removing almost completely the crosstalk between the absorption and diffusion coefficient images.

Figure 4 is the counterpart of Fig. 3 for the smooth inclusion model. Although the prior belief that the target is blocky is in conflict with the model, the reconstruction is reasonable and in line with what one would expect. Notice that also in this case, the adaptation is able to reduce the crosstalk very effectively.

(C) 2008 OSA



Fig. 3. Reconstructions of optical parameters μ_a (a) and κ (b) from simulated data from the blocky inclusions model. The reconstructions on the left of each panel are those obtained after the first Gauss-Newton iteration, thus corresponding to a homogenous smoothness penalty without adaptation, while on the right are those obtained after 20 iteration sweeps. Panels (c) and (d) show the respective profiles along the circular contour indicated in Fig. 2, traversed counterclockwise. The solid line is the true target, the dotted line the solution after the first iteration and the dashed line the final reconstruction.

(C) 2008 OSA



Fig. 4. Reconstructions of optical parameters μ_a (a) and κ (b) from simulated data from the smooth inclusions model. The reconstructions on the left of each panel are those obtained after the first Gauss-Newton iteration, thus corresponding to a homogenous smoothness penalty without adaptation, while on the right are those obtained after 20 iteration sweeps. Panels (c) and (d) show the respective profiles along the circular contour indicated in Fig. 2, traversed counterclockwise. The solid line is the true target, the dotted line the solution after the first iteration and the dashed line the final reconstruction.

(C) 2008 OSA

3.2. Reconstructions from phantom data

In this section we test the efficiency of the adaptive regularization approach with measured phantom data. The details about the instrumentation utilized for the data collection can be found in [24].

The two data sets available used in the computed experiments, two homogenous phantoms, with the same background optical properties, one without and one with two inclusions, are encoded in the vectors y_h and y_{nh} , respectively. To suppress reconstruction artifacts due to the diffusion approximation model, we consider the difference data translated around the computed response of the homogenous background. In other words, we define

$$y \equiv y_{\rm nh} - y_{\rm h} + f(p_{\rm h}),\tag{19}$$

where $f(p_h)$ is the 3D FEM solution computed on a grid of 21207 tetrahedral elements with quadratic basis functions, assuming that the homogeneous optical parameters are known. The absorption and scattering coefficient of the inclusions in the non-homogeneous phantom are approximately twice the values for the background. We estimated absorption coefficient at $\mu_{a,bg} \approx 0.0097 \text{ mm}^{-1}$ for the scattering coefficient $\mu'_{s,bg} \approx 1.04 \text{ mm}^{-1}$ and at n = 1.56 for the refraction index. A schematics of the non homogeneous phantom is shown in Fig. 5.

The data was collected by positioning 16 source and 16 detector fibers around the cylindrical surface in two rings (see Fig. 5). Since the diffusion approximation is known to be a poor model for light propagation near a singular source, measurements from source–detector pairs less than 30 mm apart are omitted. Thus the actual data used in this example consist of 192 amplitude and phase measurements.

A preliminary study of the distribution of the data suggests that the noise in the logamplitude/phase representation can be well modelled with a Gaussian distribution, although the variance of the data depends on the mean. Fig. 6 shows the histograms of the log-amplitude and phase of the homogeneous phantom data from the first source–detector pair and normal probability densities with the means and variances equal to the empirical means and variances of the sample. Empirical analysis of the data reveals that the standard deviation of the noise at detectors 30 mm or more away from the source increases with the amplitude of the data, and can be fairly well estimated using a linear regression model. This justifies an additive noise model of the form

$$y = f(p) + e$$
, $e \sim \mathcal{N}(0, (\operatorname{diag}(y))^2)$.

While at first this empirical noise model might be surprising, it becomes easier to understand after recalling that both the absolute value of the logarithm of the amplitude and the phase shift increase when the receiver is moved away from the source. Observe that we the constant of proportionality has been omitted from the noise level, with the understanding that it can be accounted for in the Tikhonov regularization parameter.

The Cholesky factor of the inverse of the noise covariance C_n^{-1} for this model is

$$S = \text{diag}(1/|y_1|, 1/|y_2|, \dots, 1/|y_{2M}|).$$

The construction of the prior proceeds as in the synthetic data case, with the optical parameters discretized in a mesh of 25924 tetrahedral elements with linear basis functions. The Cholesky factor *L* of the inverted prior covariance matrix C_p^{-1} and its parameters are as in Section 3.1, except that now $\tau_i = 150$. The Tikhonov regularization parameter is held at $\delta = 10^{-6}$.

The adaptive algorithm is stopped after 8 outer iterations, with a maximum of 15 inner Gauss-Newton iteration steps allowed.

Figure 7 displays the projection of the reconstructed optical parameter distributions through the plane z = 0, see Fig. 5. The scattering coefficient image is calculated from the reconstructed diffusion and absorption coefficient images.

(C) 2008 OSA



Fig. 5. The geometry of the phantom with the diameter of 69.25 mm and height of 110 mm. The 16 sources (marked with \times) and 16 detectors (marked with \circ) are located in two rows on the cylindrical surface. A two dimensional slice of the reconstructions are shown along the plane of intersection at z = 0 (red). The locations of the inclusions are projected on the bottom.

We define the *contrast* of the inclusion C_{μ_a} to be the ratio

$$C_{\mu_{\rm a}} = \frac{\overline{\mu_{\rm a}}_{\rm inside \ inclusion}}{\overline{\mu_{\rm a}}_{\rm outside \ inclusion}},\tag{20}$$

of the means $\overline{\mu_a}$ over nodal values of μ_a inside and outside the inclusion. The contrast $C_{\mu_s'}$ is defined in a similar manner. Here, the boundary of the inclusions is assumed to be known. The contrasts in the reconstruction with the smoothness prior are $C_{\mu_a} = 1.12$ and $C_{\mu'_s} = 1.11$, respectively, and after eight adaptive iterations, they attain the values $C_{\mu_a} = 1.57$ and $C_{\mu'_s} = 1.21$. The algorithm does not remove completely the crosstalk in the absorption image: in fact, in the location of the inclusion with the same absorption coefficient as the background, the mean value of μ_a is 0.0102 mm^{-1} before adaptation and $\mu_a = 0.0106 \text{ mm}^{-1}$ after the adaptation. However, since the adaptation improves considerably the dynamical range of the reconstruction, the relative crosstalk is diminished, and it is visually less prominent.

4. Discussion

The article describes a new adaptive regularization algorithm for diffuse optical tomography. In contrast to the traditional smoothness penalties that keeps the cost of variations in the reconstructed optical parameters fixed, the proposed algorithm updates iteratively the smoothness

(C) 2008 OSA



Fig. 6. Histograms of the logarithm of the amplitude (left) the phase (right) from a repeated measurement with one source/detector pair. The red line shows a Gaussian with the same mean and variance as the sample. The values of the abscissae are arbitrary, as the data are not calibrated.

penalty, using the current estimate as a structural pilot image to decrease the penalty where the possibility of strong contrasts is suggested. The iterations can thus be thought of as a learning process, where the current iterate provides prior information for the new update. A similar approach has been previously applied to linear inverse problems with structural meshes, while in the present application, the mesh is unstructured and the forward map is strongly non-linear. The adaptive algorithm is expected to work best with data that correspond to a target with a blocky structure, the optical parameters changing abruptly across the inclusion boundaries. The process of relaxing locally the smoothness penalties is likely to diminish or remove typical overshooting phenomena creating disturbing artifacts in the reconstruction images.

The algorithm has been tested with both synthetic data and with phantom data. The synthetic data corresponds to a two-dimensional target, and both blocky and smoothly varying optical parameter models are considered. The forward model in the synthetic data simulation is the diffusion approximation. In this case the algorithm was able to get rid almost completely of the crosstalk between the absorption and diffusion coefficients, whereas traditional smoothness penalized algorithms are corrupted by artifacts. The reconstructed inclusions exhibit sharp boundaries, even if the data was generated with a smoothly varying model. This result is not surprising as it is in line with the prior belief but it should be taken into account when the user is in doubt about the blocky structure of the target. The absence of boundary artifacts in the reconstruction is the result of implementing the smoothness penalty using a statistical analysis of the prior image variances.

A second set of tests was performed using phantom data corresponds to a cylindrical target with blocky inclusions, and a three-dimensional diffusion approximation model. To suppresses the effects of the modelling errors, difference data was used. Although the optical fibre contacts may change between the two targets and therefore not cancel in the difference data, the computed reconstructions show little boundary effects. With real data crosstalk is not completely removed, but since the adaptive iterative algorithm considerably improves the contrast between the inclusions and the background without increasing the crosstalk artifacts, the latter becomes



Fig. 7. Cross sections of the reconstructed absorption (top) and scattering (bottom) coefficient through the plane z=0 (see Fig. 5). The reconstructions on the left are before adaptation, corresponding to a homogenous smoothness penalty, while those on the right are after eight adaptation sweeps.

visually less prominent. The remaining crosstalk artifacts indicate that the coupling between the scattering and absorption coefficients is in the higher order terms of the harmonic expansions of the radiative transfer equation and therefore not captured by the lowest order, or diffusion, approximation. Hybrid methods that use the radiative transfer equations in the regions where the diffusion approximation is not expected to be valid have been suggested [25]. Altenatively, the modelling error due to the low order approximation can be taken into account in the Bayesian statistical framework, see [5, 26, 27, 28].

The choice of a rather coarse grid for the discretization of the optical parameters, which has

(C) 2008 OSA

a regularization effect on the solution, cannot resolve well sharp edges accurately, thus causing some artifacts near the boundary of the inclusions.

The modelling of the noise in the optical tomography measurements is not a trivial issue, in particular when the data is in the amplitude-phase format. Here a Gaussian approximation with a semi-empirical variance model was used, and the scaling of the noise covariance was absorbed by the regularization parameter when phantom data was used. Proper modelling of the noise is particularly important when absolute rather than difference data need to be used. A detailed analysis will be the topic of future work.

A. Variance adjustment

In this appendix, we discuss the details of the construction of the smoothness penalty and the variance adjustment at the boundary.

Consider the block partitioning (14) of the matrix *L*. Since the block $L'_{int} \in \mathbb{R}^{N_{int} \times N_{int}}$ corresponds to a discrete Laplacian with homogenous Dirichlet data, it is invertible, while the matrix $L_{bdry} \in \mathbb{R}^{N_{bdry} \times N_{bdry}}$ has a one-dimensional null space spanned by the unit vector $1 = [1, 1, ..., 1]^T \in \mathbb{R}^{N_{brdy}}$. Therefore it follows that the null space of *L* is also one-dimensional, spanned by the unit vector $1 \in \mathbb{R}^N$. Consequently, given Lp, it is possible to recover the vector *p* only up to an undetermined additive constant which can be specified by, e.g., specifying the mean value μ over the boundary values p_{bdry} . We therefore define the folding matrix

$$C = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & -1 \end{bmatrix} \in \mathbb{R}^{N_{\text{bdry}} \times (N_{\text{bdry}} - 1)},$$

whose range is an $(N_{bdry} - 1)$ -dimensional subspace in $\mathbb{R}^{N_{bdry}}$ perpendicular to the null space of L_{bdry} , and we reparameterize p_{bdry} as

$$p_{\text{bdry}} = Cz + \mu 1, \quad z \in \mathbb{R}^{N_{\text{bdry}} - 1}$$

With this representation,

$$Lp = \begin{bmatrix} L'_{\text{int}} & L''_{\text{int}} \\ \alpha L_{\text{bdry}} \end{bmatrix} \begin{bmatrix} p_{\text{int}} \\ Cz + \mu 1 \end{bmatrix} = \begin{bmatrix} L'_{\text{int}} & L''_{\text{int}}C \\ \alpha L_{\text{bdry}}C \end{bmatrix} \begin{bmatrix} p_{\text{int}} \\ z \end{bmatrix} + \begin{bmatrix} \mu L''_{\text{int}}1 \\ 0 \end{bmatrix}$$
$$= K_{\alpha}p' + r,$$

where

$$K_{\alpha} = \begin{bmatrix} L'_{\text{int}} & L''_{\text{int}}C \\ \alpha L_{\text{bdry}}C \end{bmatrix} = \begin{bmatrix} I \\ \alpha I \end{bmatrix} \begin{bmatrix} L'_{\text{int}} & L''_{\text{int}}C \\ L_{\text{bdry}}C \end{bmatrix} = \begin{bmatrix} I \\ \alpha I \end{bmatrix} K_{1}.$$
(21)

Without loss of generality, we may assume here that $\mu = 0$, i.e, r = 0.

Consider the Gaussian density

$$\pi(p \mid \mu = 0) \propto \exp\left(-\frac{1}{2}\|Lp\|^{2}\right) = \exp\left(-\frac{1}{2}\|K_{\alpha}p'\|^{2}\right),$$
(22)

where the matrix $K_{\alpha} \in \mathbb{R}^{N \times (N-1)}$ has full column rank. We are interested in the covariance of the components of the vector p with respect to this density. Consider an interior node value p_j , expressed in terms of p' as

$$p_j = e_j^{\mathrm{T}} p_{\mathrm{int}} = \begin{bmatrix} e_j \\ 0 \end{bmatrix}^{\mathrm{T}} p',$$

(C) 2008 OSA

where $e_j \in \mathbb{R}^{N_{int}}$ is *j*th unit coordinate vector. Denoting by E the expectation with respect to the probability density (22), the variance of the component is

$$\operatorname{var}(p_j) = \operatorname{E}\left\{p_j^2\right\} = \begin{bmatrix} e_j \\ 0 \end{bmatrix}^{\mathrm{T}} \operatorname{E}\left\{p'(p')^{\mathrm{T}}\right\} \begin{bmatrix} e_j \\ 0 \end{bmatrix}$$
$$= \begin{bmatrix} e_j \\ 0 \end{bmatrix}^{\mathrm{T}} \left(K_{\alpha}^{\mathrm{T}}K_{\alpha}\right)^{-1} \begin{bmatrix} e_j \\ 0 \end{bmatrix} = \|v\|^2,$$

where v is the least squares solution of the linear system

$$K_{\alpha}^{\mathrm{T}}v = \begin{bmatrix} e_j \\ 0 \end{bmatrix} \quad \text{or} \quad v = \left(K_{\alpha}^{\mathrm{T}}\right)^+ \begin{bmatrix} e_j \\ 0 \end{bmatrix}, \tag{23}$$

where $(K_{\alpha}^{\mathrm{T}})^+$ is the pseudo-inverse of K_{α}^{T} . Here we used the identity $(K_{\alpha}^{\mathrm{T}}K_{\alpha})^{-1} = K_{\alpha}^+(K_{\alpha}^+)^{\mathrm{T}}$ and $(K_{\alpha}^+)^{\mathrm{T}} = (K_{\alpha}^{\mathrm{T}})^+$. The solution is found by solving the normal equations,

$$K_{\alpha}K_{\alpha}^{\mathrm{T}}v=K_{\alpha}\begin{bmatrix}e_{j}\\0\end{bmatrix},$$

or, in view of the equation (21),

$$K_1 K_1^{\mathrm{T}} \begin{bmatrix} 1 & \\ & \alpha 1 \end{bmatrix} v = K_1 \begin{bmatrix} e_j \\ 0 \end{bmatrix},$$

implying that the solution is

$$\mathbf{v} = \begin{bmatrix} 1 & \\ & \alpha^{-1}1 \end{bmatrix} (K_1^{\mathrm{T}})^+ \begin{bmatrix} e_j \\ 0 \end{bmatrix} = \begin{bmatrix} v_1 \\ \alpha^{-1}v_2 \end{bmatrix},$$

and therefore,

$$\operatorname{var}(p_j) = \|v_1\|^2 + \frac{1}{\alpha^2} \|v_2\|^2.$$

Similarly, we compute the variance of a boundary node value. Denoting by $\mathbf{n}_k \in \mathbb{R}^{N_{\text{bdry}}}$ a unit vector with a single non-zero element, the corresponding boundary node value can be expressed as

$$p_k = n_k^{\mathrm{T}}(Cz) = \begin{bmatrix} 0\\ C^{\mathrm{T}}n_k \end{bmatrix}^{\mathrm{T}} p',$$

and the variance of the value is found to be

$$\operatorname{var}(p_k) = \|u\|^2,$$

where \mathbf{u} is the least squares solution of the system

$$K_{\alpha}^{\mathrm{T}} u = \begin{bmatrix} 0 \\ C^{\mathrm{T}} n_k \end{bmatrix}$$

or, equivalently,

$$u = \begin{bmatrix} 1 & \\ & \alpha^{-1}1 \end{bmatrix} (K_1^{\mathrm{T}})^+ \begin{bmatrix} 0 \\ C^{\mathrm{T}}n_k \end{bmatrix} = \begin{bmatrix} u_1 \\ \alpha^{-1}u_2 \end{bmatrix},$$

leading to the equation

$$\operatorname{var}(p_k) = \|u_1\|^2 + \frac{1}{\alpha^2} \|u_2\|^2.$$

(C) 2008 OSA

The condition $var(p_j) = var(p_k)$ therefore leads to the condition

$$\alpha^{2} = \frac{\|u_{2}\|^{2} - \|v_{2}\|^{2}}{\|u_{1}\|^{2} - \|v_{1}\|^{2}}.$$

In practice, the linear systems are solved in the least squares sense using the LSQR algorithm [23, 29].

Acknowledgments

The authors would like to acknowledge D.Sc. Ilkka Nissilä for providing data, and CSC – Scientific Computing Ltd. for computer resources. The research of Petri Hiltunen was supported by CSC – Scientific Computing Ltd and Jenny and Antti Wihuri Foundation.