Published in final edited form as:

Comput Methods Programs Biomed. 2007 March ; 85(3): 247–256.

An Efficient Deconvolution Algorithm for Estimating Oxygen Consumption During Muscle Activities

Ranjan K. Dash^{1,2}, Erkki Somersalo⁴, Marco E. Cabrera^{1,2}, and Daniela Calvetti^{1,3}

¹Center for Modeling Integrated Metabolic Systems, Case Western Reserve University, Cleveland, OH–44106

²Department of Pediatrics, Case Western Reserve University, Cleveland, OH–44106 ³Department of Mathematics, Case Western Reserve University, Cleveland, OH–44106 ⁴Institute of Mathematics, PO Box 1100, FIN–02015 HUT, Helsinki University of Technology, Finland

Abstract

The reconstruction of an unknown input function from noisy measurements in a biological system is an ill-posed inverse problem. Any computational algorithm for its solution must use some kind of regularization technique to neutralize the disastrous effects of amplified noise components on the computed solution. In this paper, following a hierarchical Bayesian statistical inversion approach, we seek estimates for the input function and regularization parameter (hyperparameter) that maximize the posterior probability density function. We solve the maximization problem simultaneously for all unknowns, hyperparameter included, by a suitably chosen quasi-Newton method. The optimization approach is compared to the sampling-based Bayesian approach. We demonstrate the efficiency and robustness of the deconvolution algorithm by applying it to reconstructing the time courses of mitochondrial oxygen consumption during muscle state transitions (e.g., from resting state to contraction and recovery), from the simulated noisy output of oxygen concentration dynamics on the muscle surface. The model of oxygen transport and metabolism in skeletal muscle assumes an *in vitro* cylindrical structure of the muscle in which the oxygen from the surrounding oxygenated solution diffuses into the muscle and is then consumed by the muscle mitochondria. The algorithm can be applied to other deconvolution problems by suitably replacing the forward model of the system.

Keywords

Deconvolution; Bayesian inversion; Monte-Carlo simulation; Muscle oxygen uptake; Mitochondrial oxygen consumption; Oxygen transport and metabolism

1 Introduction

Consider a dynamical system in which an observable output is related to the not directly measurable input through a transfer function, which in the linear case is the system's response to a unit impulse function. The problem of reconstructing an input signal from the measured output in such a dynamical system is called deconvolution. In addition, the transfer function may depend on unknown parameters, and it is often of interest to estimate these model

Address for Correspondence: Daniela Calvetti, Ph.D., Department of Mathematics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH–44106, Email: dxc57@case.edu, Phone: (216) 368-2884, Fax: (216) 368-5163

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

parameters simultaneously with the linear input. When the transfer function is unknown or not fully determined, the problem is usually called blind deconvolution. Deconvolution problems, and blind deconvolution problems in particular, are typically ill-posed inverse problems, i.e., small errors in the data propagate, strongly amplified, to the estimated inputs unless some sort of regularization is used.

In the quantitative studies of complex physiological and pharmacokinetic systems, deconvolution allows reconstruction of important non-accessible input signals; see, e.g., [2,3,4,5,6,7,8,9,10,11,12]. The particular application that we consider in this paper is the estimation of the time course of mitochondrial oxygen consumption (nonmeasurable) in muscle tissue during muscle state transitions from rest to contraction and recovery, from the samples of its causally-related measurable effects, such as the dynamics of oxygen concentration in the surrounding medium [2,3]. Quantifying the time course of mitochondrial oxygen consumption during muscle activities is of great importance in the understanding of the dynamic regulation of oxidative phosphorylation and muscle energetics [3,4,5,13,14,15,16,17,18,19].

In a recent paper, Dash et al. [3] developed a computational model of oxygen transport and metabolism in a cylindrically-shaped muscle, immersed *in vitro* in an oxygenated chamber, and used a recently developed hierarchical Bayesian statistics-based parametric deconvolution algorithm [20] to obtain the estimate of the time course of mitochondrial oxygen consumption from the polarographic measurements of decayed oxygen concentration on the muscle surface (decayed oxygen partial pressure in the chamber) measured before, during and after an isometric twitch contraction of the muscle [19]. Although their analysis facilitated a better formal approach for estimating mitochondrial oxygen consumption in comparison to the previous analyses [5,18,19], their estimates were inaccurate and oscillatory at higher noise levels in the measured data. Furthermore, the algorithm was computationally expensive due to the empirical Bayes approach [6,9,10,20,21] of estimating the unknown input function and regularization parameter (hyperparameter).

In this paper, we propose a hybrid deconvolution algorithm for determining the mitochondrial oxygen consumption time course during muscle activities, which is found to be very efficient and robust. The algorithm is based on a Bayesian statistical framework and computes simultaneously the *maximum a posteriori* (MAP) estimate of the unknown input function parameters and the hyperparameter. The algorithm employs a sequential optimization (quasi-Newton) scheme that updates alternatively the model parameters and the hyperparameter to maximize the posterior probability density function. The algorithm requires solving a normal equation at each iteration of the optimization, which is computationally simpler than solving the ill-conditioned linear systems [22,23]. The performance of the algorithm is compared and validated with a Markov Chain Monte Carlo sampling-based analysis of the posterior probability density. Similar ideas can be found in the literature, see, e.g., [6,9,10,20,21]. We remark that the applicability of the proposed algorithm is not limited to the particular problem considered here. It can be applied to other deconvolution problems by suitably replacing the forward model of the system.

2 Mathematical Model of Oxygen Transport and Metabolism

In this section, we briefly review the forward mathematical model simulating the oxygen uptake, transport and metabolism in an isolated skeletal muscle during muscle activities, i.e., during the resting, contraction and recovery periods. This model is presented in detail, including the experimental background, in Ref. [3].

The experimental protocol consists of an isolated skeletal muscle mounted in a glass chamber filled with a highly concentrated oxygenated solution. The muscle is electrically stimulated at a certain frequency for a time interval from the resting state to give a twitch contraction and

then allowed to return to the resting state. The decay of oxygen in the chamber is measured continuously through a polarographic electrode [3,19,5]. The physiological problem of estimating the time course of mitochondrial oxygen consumption and muscle oxygen uptake from the time course of oxygen decay in the chamber leads to a deconvolution problem considered in Ref. [3].

The mathematical model is based on the assumption of cylindrical geometry of the muscle, with circular radius R_{mu} much smaller than the length L_{mu} . The oxygen concentration $C_{\text{ch}} = C_{\text{ch}}(t)$ in the chamber decays due to the consumption by the muscle mitochondria in addition to the consumption by the polarographic electrode and leakage from the chamber, which is referred to as the apparatus baseline. The oxygen concentration $C_{\text{mu}} = C_{\text{mu}}(r, t)$ in the muscle varies due to the radial diffusion and mitochondrial oxygen consumption. The mitochondrial oxygen consumption is highly complex and depends, in general, on the concentrations of available oxygen, substrate dehydrogenase (NADH) and phosphates [14,15,16].

All the dependencies are lumped here into a simplified time-dependent flux expression. For numerical computations, we impose a positivity constraint on concentrations to avoid non-physiological solutions arising due to the simplified flux expression for mitochondrial oxygen consumption. The governing equations are given by

$$dC_{\text{ch}} = \left\{ -\frac{F_{\text{tot}}}{V_{\text{ch}}} - D \frac{2\pi R_{\text{mu}} L_{\text{mu}}}{V_{\text{ch}}} \left(\frac{\partial C_{\text{mu}}}{\partial r} \right) \right\}_{r=R_{\text{mu}}} dt, \quad C_{\text{ch}} \geq 0, \quad (1)$$

$$dC_{\text{mu}} = \left\{ D \left(\frac{\partial^2 C_{\text{mu}}}{\partial r^2} + \frac{1}{r} \frac{\partial C_{\text{mu}}}{\partial r} \right) - \frac{F_{\text{mr}} + F_{\text{ms}}(t)}{\pi R_{\text{mu}}^2 L_{\text{mu}}} \right\} dt, \quad C_{\text{mu}} \geq 0. \quad (2)$$

In Eq. (1), F_{tot} denotes the total rate (nmoles/sec) of oxygen decay in the chamber due to the apparatus baseline, V_{ch} denotes the chamber volume, and D is the diffusion coefficient, all of which are assumed to be known [3]. The second term in the right-hand side of Eq. (1) represents the muscle oxygen uptake through the muscle surface. In Eq. (2), F_{mr} and $F_{\text{ms}}(t)$ denote the fluxes (nmoles/sec) of mitochondrial oxygen consumption at resting state and during muscle stimulation above the resting state, respectively. The initial-boundary conditions for the system of diffusion-consumption equations (1) – (2) are given by

$$C_{\text{ch}}(t = 0) = C_{\text{ch},0}, \quad (3)$$

$$C_{\text{mu}}(r, t = 0) = C_{\text{mr}}(r), \quad (4)$$

$$C_{\text{mu}}(r = R_{\text{mu}}, t) = C_{\text{ch}}(t), \quad (5)$$

$$\frac{\partial C_{\text{mu}}}{\partial r}(r = 0, t) = 0. \quad (6)$$

The concentration $C_{\text{mr}}(r)$, which is obtained from the steady state version of the differential equation (2) and boundary conditions (5) – (6), has an explicit form:

$$C_{\text{mr}}(r) = \max \left\{ 0, C_{\text{ch},0} - \frac{F_{\text{mr}}}{4V_{\text{mu}}D} (R_{\text{mu}}^2 - r^2) \right\}, \quad (7)$$

Where $V_{\text{mu}} = \pi R_{\text{mu}}^2 L_{\text{mu}}$ is the volume of the cylindrical muscle. The numerical parameter values are as in Ref. [3].

The inverse problem that is addressed in this article can be stated as follows: *Based on the diffusion-consumption model (1) – (2) and the initial-boundary conditions (3) – (6), estimate $\{F_{mr}, F_{ms}(t) \mid 0 \leq t \leq T\}$ from the noisy observation of $\{C_{ch}(t) \mid 0 \leq t \leq T\}$, where T is duration of the experiment.*

The numerical solutions of the forward model is obtained using a stable explicit finite difference scheme (Euler-type) as presented in Ref. [3]. Briefly, we use a standard central difference in radial direction and forward difference in temporal direction to discretize the system. For the stability of the marching scheme, the spatial and temporal discretization steps Δr and Δt are chosen to satisfy the stability criterion $\Delta t \leq \Delta r^2 / 2D$. The positivity constraint for concentrations is implemented by a non-negativity projection at each time step. Observe that this projection renders is what makes this problem non-linear.

3 Estimation of the Input

In this section, we describe the deconvolution algorithm which is used in estimating the unknown parameter F_{mr} and the unknown input function $F_{ms}(t)$ in the diffusion-consumption equations (1) – (2) from the noisy output of decaying oxygen concentration $C_{ch}(t)$ in the chamber or on the muscle surface $C_{mu}(R_{mu}, t)$. This algorithm is based on a Bayesian statistical framework [24].

3.1 Posterior Probability Density

Consider a parameter-dependent mapping of a given input function, $f(t) \mapsto \Psi(\theta, f(t), t)$, $\theta = [\theta_1, \dots, \theta_K] \in \mathbb{R}^K$, modeling an idealized noiseless output of a dynamical system. In our application, $f(t) = F_{ms}(t)$, and $\theta = F_{mr} \in \mathbb{R}$, i.e., the model depends of only one parameter. The idealized output is oxygen concentration in the chamber, $\Psi(\theta, f(t), t) = C_{ch}(F_{mr}, F_{ms}(t), t)$ which is governed by the diffusion-consumption equations (1) – (2) and initial-boundary conditions (3) – (6). We denote by y_i the output data point at time t_i which is assumed to be corrupted by additive Gaussian noise e_i with mean zero and variance σ_i^2 . Then the noisy output y_i can be written as

$$y_i = \psi(\theta, f(t_i), t_i) + e_i \quad e_i \sim N(0, \sigma_i^2), \quad i = 1, \dots, M. \quad (8)$$

The unknown input function $f(t)$ is assumed to satisfy the conditions $f(0) = 0$ and $f(T) = 0$, since at the beginning and at the end the muscle is at rest. We approximate $f(t)$ by a truncated Fourier sine series over the interval $[0, T]$:

$$f(t) \approx f(\alpha, t) = \sum_{j=1}^N \alpha_j \phi_j \quad \phi_j = \sqrt{\frac{2}{T}} \sin(\beta_j t), \quad \beta_j = \frac{\pi j}{T}, \quad (9)$$

where $\alpha = [\alpha_1, \dots, \alpha_N] \in \mathbb{R}^N$ is the vector of amplitude coefficients and $\varphi(t) = [\varphi_1(t), \dots, \varphi_N(t)] \in \mathbb{R}^N$ is the vector of orthonormal basis functions. We denote the resulting approximation of the model by $\Psi(\theta, \alpha, t)$.

Let $y = [y_1, \dots, y_M]^T \in \mathbb{R}^M$ be the vector of noisy data, $e = [e_1, \dots, e_M]^T \in \mathbb{R}^M$ be the vector of Gaussian noise, $\Psi(\theta, \alpha) = [\Psi(\theta, \alpha, t_1), \dots, \Psi(\theta, \alpha, t_M)]^T \in \mathbb{R}^M$ be the vector of model predictions, and $S = \text{diag}[1/\sigma_1, \dots, 1/\sigma_M] \in \mathbb{R}^{M \times M}$ be a diagonal matrix. If we assume that the noise at time t_i is independent of the noise at time $t_j, j \neq i$, then the likelihood probability density of the data y conditioned on the parameter vectors θ and α is given by [24]

$$\pi(y \mid \theta, \alpha) \propto \exp \left(-\frac{1}{2} \| S[y - \Psi(\theta, \alpha)] \|^2 \right). \quad (10)$$

Here, “ $\propto$ ” means “equal up to a multiplicative constant”. Furthermore, if we assume a first order smoothness prior on the parameter vector α with variance $\lambda > 0$, then the probability density of α conditioned on λ is given by

$$\pi_{\text{pr}}(\alpha | \lambda) \propto \frac{1}{(2\pi\lambda)^{N/2}} \exp \left(-\frac{1}{2\lambda} \| f'(\alpha, \cdot) \|_L^2 \right) = \exp \left(-\frac{1}{2\lambda} \sum_{j=1}^N \alpha_j^2 \beta_j^2 - \frac{N}{2} \log(2\pi\lambda) \right), \quad (11)$$

where $f'(a, t)$ denotes the derivative of $f(a, t)$ with respect to t . Such prior indicates an *a priori* belief that the signal does not contain significant high frequency components that would result into large amplitude fast oscillations. From Eqs. (10) and (11), the joint probability density of y and α conditioned on θ and λ is then given by

$$\begin{aligned} \pi(y, \alpha | \theta, \lambda) &= \pi(y | \theta, \alpha) \pi_{\text{pr}}(\alpha | \lambda) \\ &\propto \exp \left(-\frac{1}{2} \| \mathcal{S}[y - \psi(\theta, \alpha)] \|^2 - \frac{1}{2\lambda} \sum_{j=1}^N \alpha_j^2 \beta_j^2 - \frac{N}{2} \log(2\pi\lambda) \right). \end{aligned} \quad (12)$$

In general, the values of θ and λ are not known *a priori*. However, since λ and θ must be a positive constant, we can use the *flat hyperprior*:

$$\pi_h(\theta, \lambda) \propto \pi_+(\theta, \lambda) = \begin{cases} 1 & \text{if } \lambda > 0 \text{ and } \theta > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

It follows from Bayes' formula [24] and Eqs. (12) and (13) that the joint probability density of parameters θ , α and λ conditioned on the data y is given by

$$\begin{aligned} \pi(\theta, \alpha, \lambda | y) &\propto \pi(y, \theta, \alpha, \lambda) = \pi(y, \alpha | \theta, \lambda) \pi_h(\theta, \lambda) \\ &\propto \exp \left(-\frac{1}{2} \| \mathcal{S}[y - \psi(\theta, \alpha)] \|^2 - \frac{1}{2\lambda} \sum_{j=1}^N \alpha_j^2 \beta_j^2 - \frac{N}{2} \log(2\pi\lambda) \right), \end{aligned} \quad (14)$$

subject to the constraints $\theta, \lambda > 0$. The conditional density in equation (14) is also known as the posterior probability density, and in statistical inverse problems can be considered as the solution of the inverse problem. In the following sections, we discuss various methods to explore the posterior density and how to calculate estimates from it.

3.2 Exploring the Posterior Density

It is possible to calculate various estimates of the unknown based on the posterior density, or explore the density by Monte Carlo methods. We discuss here some of the estimation techniques which can be found in the literature and propose a numerically efficient optimization method for the parameter estimation.

3.2.1 Empirical Bayes Approach—A common practice when dealing with hierarchical Bayesian models is to marginalize the posterior density with respect to the parameters of primary interest, and then use the marginal density to estimate the hyperparameters [6,24,21, 9,8]. In the present setting, this amounts to calculating the marginal density

$$\pi(\lambda | y) = \int_{\mathbb{R}} N \int_{\mathbb{R}} \pi(\theta, \alpha, \lambda | y) d\theta d\alpha. \quad (15)$$

In general, there is no analytic formula for this integral and one has to resort to Monte Carlo integration techniques.

We write a partially linerized approximation α , i.e.,

$$\psi(\theta, \alpha) \approx \psi_0(\theta) + A_\theta \alpha, \quad (16)$$

where the mappings $\theta \mapsto \psi_0(\theta) \in \mathbb{R}^M$ and $\theta \mapsto A_\theta \in \mathbb{R}^{M \times N}$ are non-linear. Observe that the forward differential equation model without positivity constraints defines (1)–(6) a linear model, and the nonlinearities are due to the positivity projection. The above approximation is found reasonably accurate by numerical tests. Using this model, we may first integrate out analytically the variable α , using the Gaussian form of the posterior density. Notice that the dependency of A_θ of the parameter θ is due to the projection step in the forward solver. By writing

$$\|S[y - \psi(\theta, \alpha)]\|^2 + \frac{1}{\lambda} \sum_{j=1}^N \alpha_j^2 \beta_j^2 = \left\| \begin{bmatrix} SA_\theta \\ \lambda^{-1/2} I \end{bmatrix} \alpha - \begin{bmatrix} S(y - \psi_0(\theta)) \\ 0 \end{bmatrix} \right\|^2 = \|D\alpha - b\|^2, \quad (17)$$

where $L = \text{diag}(\beta) \in \mathbb{R}^{N \times N}$, we obtain, after some algebraic manipulations, the formula

$$\pi(\theta, \lambda | y) = \int_{\mathbb{R}^N} \pi(\theta, \alpha, \lambda | y) d\alpha = \left(\det(D^T D) \right)^{-1/2} \exp \left(-\frac{1}{2} b^T P b - \frac{N}{2} \log(2\pi\lambda) \right), \quad (18)$$

where $P \in \mathbb{R}^{(M+N) \times (M+N)}$ is the projection matrix

$$P = I - D(D^T D)^{-1} D^T. \quad (19)$$

Here the matrices D , P and the vector b depend on both parameters θ and λ . The integration with respect to θ has to be performed numerically, e.g., by MCMC techniques. The procedure of integrating out part of the variables analytically is sometimes referred to as Rao–Blackwellization (see, e.g., [25]), and it is known to produce estimates with smaller variance than a full MCMC integration.

In general, the MCMC runs that update the parameter θ are computationally intensive because for each new value of θ , the matrix A_θ and the vector b need to be recomputed. The computational cost decreases dramatically if we have a good estimate for the baseline parameter. Indeed, with θ fixed and A_θ precomputed, a sample can be generated by a block form of the Gibbs sampler via the following updating steps:

1. Given the current pair $(\lambda^\ell, \alpha^\ell)$, draw a new value $\alpha^{\ell+1}$ from the distribution

$$\pi(\alpha | \theta, \lambda^\ell, y) \propto \exp \left(-\frac{1}{2} \|D\alpha - b\|^2 \right), \quad D = D(\lambda^\ell), \quad (20)$$

by drawing $w_j \sim \mathcal{N}(0, 1)$, $1 \leq j \leq N + M$, and solving the system $D\alpha = b + w$ in the least squares sense;

2. Draw a new value $\lambda^{\ell+1}$ from the one-dimensional density

$$\pi(\lambda | \alpha^{\ell+1}, \theta, y) \propto \exp \left(-\frac{1}{2\lambda} \sum_{j=1}^N \beta_j^2 (\alpha_j^{\ell+1})^2 - \frac{N}{2} \log(2\pi\lambda) \right). \quad (21)$$

This updating is fast because it does not require the solution of the forward model.

3.2.2 Maximum A Posteriori Estimator—The Maximum A Posteriori (MAP) estimate maximizes the posterior probability density function $\pi(\theta, \alpha, \lambda | y)$, or, equivalently, minimizes the negative of its logarithm. Thus,

$$(\theta_{\text{MAP}}, \alpha_{\text{MAP}}, \lambda_{\text{MAP}}) = \arg\max \pi(\theta, \alpha, \lambda | y) = \arg\min \Phi(\theta, \alpha, \lambda), \quad \lambda > 0, \quad (22)$$

where

$$\Phi(\theta, \alpha, \lambda) = -\log[\pi(\theta, \alpha, \lambda | y)] = \frac{1}{2} \|S[y - \psi(\theta, \alpha)]\|^2 + \frac{1}{2\lambda} \sum_{j=1}^N \alpha_j^2 \beta_j^2 + \frac{N}{2} \log(2\pi\lambda). \quad (23)$$

The minimization of the objective function Φ is performed with a sequential algorithm, where one minimizes it with respect to the model parameters (θ, α) and with respect the prior parameter λ alternately. The algorithm can be described as follows.

1. Initialize $k = 0$, $(\theta, \alpha, \lambda) = (\theta_0, \alpha_0, \lambda_0)$.
2. Keeping $\lambda = \lambda_k$ fixed, update the model parameters θ and α as

$$(\theta_{k+1}, \alpha_{k+1}) = \arg \min (F_k(\theta, \alpha)), \quad F_k(\theta, \alpha) = \frac{1}{2} \|S[y - \psi(\theta, \alpha)]\|^2 + \frac{1}{2\lambda_k} \sum_{j=1}^N \alpha_j^2 \beta_j^2. \quad (24)$$

3. Keeping the model parameters $\theta = \theta_{k+1}$, $\alpha = \alpha_{k+1}$ fixed, update λ by

$$\lambda_{k+1} = \frac{1}{N} \sum_{j=1}^N \alpha_{k+1,j}^2 \beta_j^2. \quad (25)$$

4. Increment k and repeat from step 2 until convergence.

The updating formula for λ follows from the condition $\partial\Phi(\theta_{k+1}, \alpha_{k+1}, \lambda)/\partial\lambda = 0$. This updating is fast because it does not require the solution of the forward model. The second step, updating the model parameters, is calculated with a quasi-Newton algorithm. Writing $\xi_k = [\theta_k, \alpha_k]^T$, and $\delta\xi = [\delta\theta, \delta\alpha]^T$, we write the quadratic approximation

$$F_k(\xi_k + \delta\xi) \approx F_k(\xi_k) + G_k(\xi_k)\delta\xi + \frac{1}{2} \delta\xi^T H_k(\xi_k) \delta\xi, \quad (26)$$

where $G_k(\xi)$ and $H_k(\xi)$ are the gradient and Hessian of $F_k(\xi)$, respectively.

Approximating the gradient of F_k by the gradient of the quadratic approximation (26), and setting it equal to zero leads to the linear system

$$H_k(\xi_k)\delta\xi = -G_k(\xi_k), \quad (27)$$

whose solution we use to compute the update, $\xi_{k+1} = \xi_k + t\delta\xi$, where t is chosen as

$$t = \arg \min_{0 < s \leq 1} F_k(\xi_k + s\delta\xi). \quad (28)$$

For details on the selection of a suitable t , known in the optimization literature as backtracking algorithm, we refer to [22], Ch. 5. Details on the calculation of the gradient and the Hessian, and the solution of the associated linear system, can be found in the Appendix.

4 Computed Examples

In this section we test the algorithms on two different input functions. The first is a continuous function, starting at the resting value $F_{mr} = 0.005$ moles/min, increasing linearly to five-fold value, and, after staying there for a while, decreasing linearly back to the original resting value. We refer to this function as ramp input. The second input is similar, except that the peak value is reached discontinuously. This input is referred to as step input. Observe that the step input is in conflict with what we expect to see in light of the prior density, so the results may be of inferior quality than those obtained with the ramp input. Indeed, the Fourier coefficients of a

discontinuous input behave as $\alpha_j \sim 1/j$. We include this test just to demonstrate the robustness of the approach with respect to prior modeling.

In the first test, we generate the data using the forward model with using the ramp input and we add Gaussian noise with $\sigma_i = 0.15$. In particular, we want to compare the empirical Bayes reconstructions and the MAP estimate. The conditional mean for θ , calculated from the two-dimensional marginal distribution $\pi(\theta, \lambda | y)$, given by (18) and shown in Figure 1 (upper left), corresponds well to the true resting value $\theta = 0.005$ nmoles/min. After fixing the resting value θ , we perform a block Gibbs MCMC run to produce a sample $S = \{(\alpha^\ell, \lambda; \ell) \mid 1 \leq \ell \leq L\}$ of size $L = 2000$, that is distributed according to the conditional distribution $\pi(\alpha, \lambda | \theta, y)$. The chain is started at $\lambda = 1$, $\alpha_j = 0$, and a short burn-in sequence (of length 200) is removed from the beginning of the chain. It is our experience that the chain stabilizes very quickly, after a few iterations. Figure 1 (upper right) shows the sampling history of λ , indicating that the mixing appears to be efficient. The histogram and the conditional mean computed from this sample are shown in the same figure (lower left). We then calculated all the inputs corresponding to the sample S and form the envelopes containing 50% and 90% of the sample curves, respectively. The envelopes are also displayed in Figure 1 (lower right).

The corresponding results using the step input are plotted in Figure 2. The noise level was the same as for ramp input. Due to the fact that for discontinuous functions, the convergence of the Fourier expansion is slower, the modeling discrepancy between the data and the forward model used in the likelihood is larger than in the previous case. The effect is that the marginal variances are larger and the conditional mean deviates more from the true input.

We then applied our proposed algorithm to the same data, performing the optimization to find an approximation for the MAP estimate of the input as described in Section 3.2.2. The results are shown in Figure 3. As expected, the quality of reconstruction of the ramp input is better than for the step input. The optimal value of the prior parameter λ in this case is $\lambda_{\text{MAP}} = 9.7078 \times 10^{-8}$ for the ramp input and $\lambda_{\text{MAP}} = 1.0015 \times 10^{-7}$ for the step input. This should be compared to the conditional mean value $\lambda_{\text{CM}} = 3.0070 \times 10^{-7}$ for the ramp input and $\lambda_{\text{CM}} = 3.4526 \times 10^{-7}$ for the step input, found with the empirical Bayes approach. The parameters values found by the two different approaches are in the same range. Observe that larger λ corresponds to more oscillatory reconstructions. An effective estimation algorithm could be designed based on the maximization of the analytically computed marginal density $\pi(\theta, \lambda | y)$, but this relies on partial linearization of the model. Since the densities of λ are skewed to the right, the solution obtained with the parameter value found via maximization gives less oscillatory estimates. Estimating the conditional mean of all the parameters from an MCMC run with an exact model is significantly more expensive.

The main purpose of this comparison is to verify that the MAP solution computed by our algorithm falls well within the support of the probability density. In fact, since the MAP solution only returns one estimate it not convey any information about stability. This is a well know shortcoming of the MAP approach. On the other hand, the estimation of the model parameters and the hyperparameter using the proposed sequential MAP estimate algorithm is very efficient and robust. The iterations converge to the optimal parameter values extremely fast irrespective of the choice of the number of basis function, the tolerance levels for the linear system solver and quasi-Newton iterations, and the initial guess for the parameter values. The computation time using MATLAB is less than 5 minutes for a typical data set with standard setting of computational parameter values. The passage from the linear system (27) to the normal equation (see Appendix) means that we use for its solution standard iterative linear system solver, like the Conjugate Gradient Least Squares (CGLS) or MINimum RESidual (MINRES) methods, [22,23]. The use of an iterative solver in our case is not dictated by the dimensionality of the problem, but rather by the need of enforcing the nonnegativity of some of the parameters;

see [1] for details. The overall scheme is quite fast since the updating of hyperparameter does not require the solution of the forward model. While it is also possible to estimating all the parameters at once, without resorting to an alternating procedure, the convergence rate of the resulting quasi-Newton method in that case appears to be quite sensitive to organization of the computation, thus making the algorithm more difficult to implement. Therefore, the proposed sequential optimization algorithm has the computational advantage of being both quite fast and rather straightforward to implement over other optimization algorithms proposed in the literature for the computation of the MAP estimate [2,3,6,9,8,10].

A comment on the selection of the parameter N in the basis function representation of the signal is now in order. Clearly, if N is too small, the basis functions may not be able to adequately represent the input signals, and one might believe that choosing N large could make the problem unstable, in view of its inherent ill-conditioning. While a systematic statistical study of an optimal choice of N is beyond the scope of this article, to test the robustness of the proposed MAP estimator algorithm with respect to N , we solved the problem repeatedly for various values of N . The results, shown in Figure 3, indicate that the solution remains quite stable as N increases, thus suggesting that the truncation index should be always chosen large enough to well represent the input signal. On the other hand, since choosing N too large would only add to the computational load without improving the results, the selection of the truncation index should take into consideration both the computational complexity and representation power of the basis functions.

Finally, we applied the algorithm to the analysis of experimental data published in [3]. The estimated mitochondrial oxygen consumption during muscle state transitions and the corresponding output of the forward model versus the data are shown in Figure 4. The estimates are in good agreement with the expected behavior of the true mitochondrial oxygen consumption. The actual resting oxygen consumption was estimated at 0.005 nmoles/min from the data, and increased to a five-fold value during the peak muscle stimulation. For details of the significance of this result, we refer the interested readers to [3]. We remark that since in the real data, the initial oxygen concentration in the chamber as well as the actual noise level in the data were not precisely known, the estimate of the actual initial oxygen concentration in the chamber was the average of the readings in the first 15 seconds of the experiment. The value of the noise level in the data was also based on the standard deviation the readings in the first 15 seconds.

5 Discussion and Conclusion

The article proposes an optimization algorithm for solving deconvolution problems with the impulse response depending on unknown parameters. The approach is therefore applicable to blind deconvolution problems as well. The optimization step estimates also the hyperparameters of the prior density. A comparison of the computed result with the solution obtained by using sampling-based approaches which are more time-consuming are also discussed. Although the number of unknowns to be estimated is not very high, MCMC based algorithms become easily prohibitively time-consuming since each evaluation of the forward map is tantamount to solving a partial differential equation describing the oxygen diffusion. In our comparison we use a block form of the Gibbs sampler, and to keep the computation time for the simulation reasonable, we used a partial linearization of the problems which could be used also as an effective proposal move for a Metropolis-Hastings sampling algorithm. Since the focus in this work is in the optimization approach, this approach is not discussed further here. We remark that since the proposed optimization algorithm does not require the derivatives in analytic form, it does not require any such approximation. The optimization algorithm is applied to estimating the mitochondrial oxygen consumption in a numerically simulated *in vitro* setup.

Our algorithm is able to produce less oscillatory solutions than approaches described earlier in the literature and at considerably smaller computational cost, see, e.g., [3] and references therein.

The present algorithm parametrizes the solution via a truncated Fourier series. However, the Fourier representation is not necessary for the algorithm and other basis functions can be chosen instead. The algorithm can be applied to other deconvolution problems simply by replacing the forward model of the system and providing the derivatives of the model with respect to the parameters of interest needed to calculate the Hessian, as explained in the Appendix. Extensions of this work should include statistical modeling of the truncation error in the construction of the likelihood, which would improve the performance of the approach and diminish the dependency of the model on the truncation parameter N , e.g., see [24,26].

Acknowledgements

This research was supported by the grant GM-66309 from the National Institute for General Medical Science (NIGMS) of the National Institute of Health (NIH). The work of Erkki Somersalo was supported by the Academy of Finland, project 204753. We are thankful to Prof. Matrin J. Kushmerick for permitting us to use his experimental data on muscle oxygenation analyzed in Figure 4. We are grateful to the reviewers for their useful comments and suggestions.

References

1. Calvetti D, Landi G, Reichel L, Sgallari F. Nonnegativity and Iterative methods for ill-posed problems. *Inverse Problems* 2004;20:1747–1758.
2. Calvetti D, Dash RK, Somersalo E, Cabrera ME. Local Regularization Method Applied to Estimating Oxygen Consumption During Muscle Activities. *Inverse Problems* 2006;22:229–244.
3. Dash RK, Bell BM, Kushmerick MJ, Vicini P. Estimating in vitro mitochondrial oxygen consumption during muscle contraction and recovery: a novel approach that accounts for diffusion. *Ann Biomed Eng* 2005;33:343–355. [PubMed: 15868725]
4. van Beek JHGM, Westerhof N. Response time of cardiac mitochondrial oxygen consumption to heart rate steps. *Am J Physiol Heart Circ Physiol* 1991;260:613–625.
5. Mast F, Elzinga G. Time course of aerobic recovery after contraction of rabbit papillary muscle. *Am J Physiol Heart Circ Physiol* 1997;253:325–332.
6. De Nicolao G, Sparacino G, Cobelli C. Nonparametric input estimation in physiological systems: problems, methods and case studies. *Automatica* 1997;33:851–870.
7. De Nicolao G, Liberati D, Sartorio A. Deconvolution of infrequently sampled data for the estimation of growth hormone secretion. *IEEE Trans Biomed Eng* 1995;42:851–870.
8. Pilonetto G, Bell BM. Deconvolution of non-stationary physical signals: A smooth variance model for insulin secretion rate. *Inverse Problems* 2004;20:367–383.
9. Pilonetto G, Sparacino G, Cobelli C. Handling non-negativity in deconvolution of Physiological Signals: A nonlinear stochastic approach. *Ann Biomed Eng* 2002;30:1077–1087. [PubMed: 12449768]
10. Sparacino G, Cobelli C. A stochastic deconvolution method to reconstruct insulin secretion rate after a glucose stimulus. *IEEE Trans Biomed Eng* 1996;43:512–529. [PubMed: 8849464]
11. Levitt DG. The use of physiologically based pharmacokinetic model to evaluate deconvolution measurements of systemic absorption. *BMC Clin Pharmacol* 2003;3:1–29.
12. Verotta D. Concepts, properties and applications of linear systems to describe distribution, identify inputs, and control endogenous substances and drugs in biological systems. *Crit Rev Biomed Eng* 1996;24:73–139. [PubMed: 9108983]
13. Lai N, Dash RK, Nasca MM, Saidel GM, Cabrera ME. Relating pulmonary oxygen uptake to muscle oxygen consumption at exercise onset: in vivo and in silico studies. *Eur J Appl Physiol* 2006;97:380–394. [PubMed: 16636861]
14. Beard DA. A biophysical model of the mitochondrial respiratory system and oxidative phosphorylation. *PLoS Comput Biol* 2005;1(4e36):252–264.
15. Gnaiger E. Bioenergetics at low oxygen: dependence of respiration and phosphorylation on oxygen and adenosine diphosphate supply. *Respir Physiol* 2001;128:277–297. [PubMed: 11718759]

16. Korzeniewski B. Regulation of ATP supply in mammalian skeletal muscle during resting state → intensive work transition. *Biophys Chem* 2000;83:19–34. [PubMed: 10631477]
17. van Beek JHGM, Tian X, Zuurbier CJ, de Groot B, van Echteld CJ, Eijgelshoven MH, Hak JB. The dynamic regulation of myocardial oxidative phosphorylation: analysis of the response time of oxygen consumption. *Mol Cell Biochem* 1998;184:321–344. [PubMed: 9746328]
18. Elzinga G, Langewouters GJ, Westerhof N, Wiechmann AHCA. Oxygen uptake of frog skeletal muscle fibers following titanic contraction at 18° C. *J Physiol London* 1984;346:365–377. [PubMed: 6607988]
19. Kushmerick MJ, Paul RJ. Aerobic recovery metabolism following a single isometric tetanus in frog sartorius muscle at 0° C. *J Physiol London* 1976;254:693–709. [PubMed: 1082933]
20. Bell BM, Pilonetto G. Estimating parameters and stochastic functions of one variable using nonlinear measurement models. *Inverse Problem* 2004;20:627–646.
21. Magni P, Bellazzi R, De Nicolao G. Bayesian function learning using MCMC methods. *IEEE Pattern Anal Machine Intel* 1998;20:1319–1331.
22. Dennis, JE., Jr; Schnabel, RB. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. SIAM; Philadelphia: 1996.
23. Saad, Y. *Iterative Methods for Sparse Linear Systems*. 2. SIAM; Philadelphia: 2003.
24. Kaipio, JP.; Somersalo, E. *Statistical and Computational Inverse Problems*. Applied Mathematics Series 160; Springer-Verlag: 2004.
25. Liu, JS. *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag; 2003.
26. Kaipio JP, Somersalo E. Statistical inverse problems: discretization, modelling error and inverse crimes. *J Comp Appl Math*. (in press)

Appendix: Gradient and Hessian

This appendix gives the necessary details for calculating the gradient and Hessian approximation that are used in the quasi-Newton algorithm for finding the MAP estimate. For the sake of completeness, we assume that the parameter vector θ may have dimension $K \geq 1$. To simplify the notation, we suppress the subindex k from F_k and its derivatives.

The first K components of $G(\xi)$ are the partial derivatives of $F(\xi)$ with respect to $[\xi_1, \dots, \xi_K] = [\theta_1, \dots, \theta_K]$. Since only the first term in the right hand side of Eq. (23) is a function of $\theta_1, \dots, \theta_K$, we have

$$G_i(\xi) = \frac{\partial}{\partial \xi_i} F(\xi) = - \left[\frac{\partial}{\partial \theta_i} \psi(\theta, \alpha) \right]^T S^2[y - \psi(\theta, \alpha)], \quad i = 1, \dots, K. \quad (29)$$

The partial derivatives of $F(\xi)$ with respect to $[\xi_{K+1}, \dots, \xi_{K+N}] = [\alpha_1, \dots, \alpha_N]$ are

$$G_{K+i}(\xi) = \frac{\partial}{\partial \xi_{K+i}} F(\xi) = - \left[\frac{\partial}{\partial \alpha_i} \psi(\theta, \alpha) \right]^T S^2[y - \psi(\theta, \alpha)] + \frac{1}{\lambda_k} \alpha_i \beta_i^2, \quad i = 1, \dots, N. \quad (30)$$

Similarly, the Hessian $H \in \mathbb{R}^{(K+N) \times (K+N)}$ is

$$\begin{aligned} H_{i,j}(\xi) &= \frac{\partial^2 F}{\partial \xi_i \partial \xi_j}(\xi) = \left[\frac{\partial \psi}{\partial \xi_i}(\xi) \right]^T S^2 \left[\frac{\partial \psi}{\partial \xi_j}(\xi) \right] - \frac{1}{2} \left[\frac{\partial^2 \psi}{\partial \xi_i \partial \xi_j}(\xi) \right]^T S^2[y - \psi(\xi)] \\ &+ \frac{\partial^2}{\partial \xi_i \partial \xi_j} \left(\frac{1}{2\lambda_k} \sum_{i=1}^N \alpha_i^2 \beta_i^2 \right). \end{aligned} \quad (31)$$

The computation of the second order partial derivatives of the model $\Psi(\xi)$ with respect to the parameters ξ can be expensive. In view of the fact that, when a good fit of the model to the data can be achieved, the residual $y - \Psi(\xi)$ becomes small, instead of the full Hessian, we use an approximation obtained by ignoring the second term in Eq. (31). Hence, we write

$$H(\xi) \approx \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} + (1/\lambda_K)B^2 \end{bmatrix}, \quad (32)$$

where the blocks R_{ij} are given as

$$\begin{aligned} R_{11,ij} &= \left[\frac{\partial \psi}{\partial \theta_i} \right]^T S^2 \left[\frac{\partial \psi}{\partial \theta_j} \right], \quad 1 \leq i, j \leq K, \\ R_{12,ij} = R_{21,ji} &= \left[\frac{\partial \psi}{\partial \theta_i} \right]^T S^2 \left[\frac{\partial \psi}{\partial \alpha_j} \right], \quad 1 \leq i \leq K, \quad 1 \leq j \leq N, \\ R_{22,ij} &= \left[\frac{\partial \psi}{\partial \alpha_i} \right]^T S^2 \left[\frac{\partial \psi}{\partial \alpha_j} \right], \quad 1 \leq i, j \leq N, \end{aligned}$$

and $B = \text{diag}(\beta)$. We remark that since the matrices R_{11} and R_{22} are symmetric and $R_{21} = R_{12}^T$ the matrix R is symmetric.

It is easy to show that the gradient vector G and Hessian matrix H can be factorized into

$$G = J^T V \text{ and } H = J^T J, \quad (33)$$

where

$$J = - \begin{bmatrix} S \left[\frac{\partial \psi}{\partial \theta} \right] & S \left[\frac{\partial \psi}{\partial \alpha} \right] \\ 0 & \frac{B}{\sqrt{\lambda}} \end{bmatrix} \text{ and } V = \begin{bmatrix} S[y - \psi] \\ \frac{B\alpha}{\sqrt{\lambda}} \end{bmatrix}. \quad (34)$$

Thus the Hessian matrix H is symmetric and positive definite. Also the associated linear system (27) for the quasi-Newton step $\delta \xi_c$ forms a normal equation which can be solved very fast and efficiently using the Conjugate Gradient Least Square (CGLS) or MINimum RESidual (MINRES) algorithm [22,23].

Figure 1.

Results of the empirical Bayes approach with the ramp input. Top left: marginal distribution $\pi(\theta, \lambda | y)$. The CM values are marked by the hair-cross. Top right: MCMC scatter plot of the hyperparameter drawn from the distribution $\pi(\alpha, \lambda | \theta, y)$. Bottom left: histogram of the hyperparameter based on the same sample. Bottom right: 50% and 90% confidence envelopes based on the MCMC run. The yellow curve represents the estimated conditional mean. In all figures, the simulated data was corrupted with additive Gaussian white noise with standard deviation $\sigma = 0.15$.

Figure 2.
The corresponding results as in the previous figure with a step input. The noise level is the same as before.

True and Estimated Fluxes of Mitochondrial Oxygen Consumption

Decay of Oxygen in the Chamber with Muscle Stimulated

Figure 3.

Estimated flux of mitochondrial oxygen consumption and corresponded model fit for ramp and step input for different number of basis functions and noise level $\varepsilon = 0.15$.

Figure 4. Estimated flux of mitochondrial oxygen consumption and corresponded model fit to real data in Ref. [3].