

Selection of regularization parameter for optical topography

Teresa Correia

University College London
Department of Medical Physics and Bioengineering
Gower Street
London WC1E 6BT
United Kingdom
E-mail: tcorreia@medphys.ucl.ac.uk

Adam Gibson

University College London
Department of Medical Physics and Bioengineering
Gower Street
London WC1E 6BT
United Kingdom

Martin Schweiger

University College London
Department of Medical Physics and Bioengineering
Gower Street
London WC1E 6BT
United Kingdom
and
University College London
Department of Computer Science
Gower Street
London WC1E 6BT
United Kingdom

Jeremy Hebden

University College London
Department of Medical Physics and Bioengineering
Gower Street
London WC1E 6BT
United Kingdom

1 Introduction

1.1 Image Reconstruction and Regularization

Optical topography uses diffuse reflectance measurements on the surface of an object to derive its internal optical properties.^{1,2} It can be used to measure functional brain activity, because the optical properties depend on the concentration of oxyhemoglobin and deoxyhemoglobin. The first optical topography images, such as those obtained using the Hitachi ETG-100 system,³ were obtained by interpolating the measurements into the image space. However, it is becoming increasingly common for optical topography images to be reconstructed by solving the inverse problem. Boas et al.^{4,5} has shown that this provides improved spatial resolution and quantitative accuracy compared to interpolation. However, the inverse problem is ill-posed, which means that there is not a single, well-behaved solution. In order to obtain a meaningful

Abstract. The choice of the regularization parameter has a profound effect on the solution of ill-posed inverse problems such as optical topography. We review 11 different methods for selecting the Tikhonov regularization parameter that have been described previously in the literature. We test them on two trial problems, deblurring and optical topography, and conclude that the L-curve method is the method of choice, though in particularly ill-posed problems, generalized cross-validation may provide an alternative. © 2009 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.3156839]

Keywords: inverse problems; regularization; Tikhonov; optical topography; L-curve; generalized cross validation.

Paper 08368R received Oct. 14, 2008; revised manuscript received Apr. 15, 2009; accepted for publication May 5, 2009; published online Jun. 23, 2009.

solution, the problem must be regularized, for example, using Tikhonov regularization.

Tikhonov regularization requires a regularization parameter, here called λ . This is most often determined heuristically, by subjectively selecting a value of λ that appears by eye to give the "best" image. A number of more objective methods have been proposed for selecting λ in optical topography and other inverse problems. In this paper, we review 11 methods that have been proposed for selecting λ and apply them to the optical topography inverse problem. We concentrate on functional imaging of brain activity, where the measured changes are typically small and measurements are available before and after a small change in the optical properties. Under these conditions, the nonlinear image reconstruction problem can be linearized using the Rytov approximation. However, the problem remains highly ill-posed and underdetermined.

The above methods were initially applied to a deblurring problem, which is an ill-posed problem for which the solution

Address all correspondence to Teresa Correia, Department of Medical Physics and Bioengineering, University College London, Malet Place Engineering, Building-Gower Street-London, London WC1E 6BT United Kingdom; Tel: 004420767902641; Fax: 0044-20-7679-0255.

is known. Thereafter, the same methods were applied to experimental optical topography data.

1.2 Regularization of Inverse Problems

Deblurring and image reconstruction are both discrete ill-posed inverse problems of the form $\mathbf{Ax}=\mathbf{b}$, where \mathbf{b} is the data vector (length m), \mathbf{x} is the vector of unknown parameters (length n) and A is a matrix of size $m \times n$. For the deblurring problem, \mathbf{b} is the blurred image, \mathbf{x} is the original image, and A represents the blurring matrix. For the optical topography problem, A is the sensitivity matrix, which maps the changes in the measured data \mathbf{b} to the changes in the optical properties \mathbf{x} .

The least-squares solution $\hat{\mathbf{x}}$ is simply $\min_{\mathbf{x}}\|\mathbf{Ax}-\mathbf{b}\|_2$, but this is highly affected by noise and must therefore be regularized. A common method, and the one we choose to use here, is the zero-order Tikhonov regularization⁶⁻⁸

$$\begin{aligned} \mathbf{x}_\lambda &= \arg \min_{\mathbf{x}} \{ \|\mathbf{Ax}-\mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2 \}, \\ &= (\mathbf{A}^T \mathbf{A} + \lambda^2 \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}, \end{aligned} \quad (1)$$

where $\|\mathbf{Ax}-\mathbf{b}\|_2$ is a measure of the difference between the measured data \mathbf{b} and the data that would be obtained if the solution image was used to simulate data. It is sometimes called the least-squares error or the residual norm. We choose to call it the *data norm*. The norm $\|\mathbf{x}\|_2^2$ is a measure of the noise in the image and is sometimes called the regularized norm or solution norm. Here it is called the *image norm*.

If λ is increased, then the contribution of the image norm to the solution is increased and the solution becomes less sensitive to perturbations in the data. A smaller λ emphasises the contribution from the data norm, effectively assuming that the quality of the data is good; thus, the solution is allowed to conform more closely to the measured data. In the case of Tikhonov regularization, λ governs the level of smoothness enforced in the image.

A third norm, the *predictive norm*, is given by $\|\mathbf{Ax}_\lambda - \mathbf{Ax}_{\text{exact}}\|_2^2$, where $\mathbf{x}_{\text{exact}}$ is the exact solution. It requires knowledge of the exact solution, which in most real cases is unknown, or knowledge of the noise statistics.⁹

We briefly mention the singular value decomposition (SVD), a method for reducing a matrix into constituent parts, which also provides a way of analyzing the ill-posedness of a problem.^{8,10} The SVD of matrix $A=USV^T=\sum_{i=1}^n u_i \sigma_i v_i^T$, where U and V are square, orthonormal matrices and S is a $m \times n$ diagonal (though nonsquare) matrix. The components of S , σ_i , are known as the *singular values* and are arranged in order of decreasing magnitude. The problem can be regularized either by setting all $\sigma_i < \lambda = 0$, or by weighting them, for example, by a factor $f_i = \sigma_i^2 / (\sigma_i^2 + \lambda^2)$ which is equivalent to Tikhonov regularization.

The rate at which σ_i decreases is an indication of the ill-posedness of the problem. One measure of this is given by the discrete Picard condition (DPC).¹¹

DEFINITION 1.1. *The data vector \mathbf{b} satisfies the discrete Picard condition if the data space coefficients $|u_i^T \mathbf{b}|$ on average decay to zero faster than the singular values σ_i .*

If the DPC is violated for a given problem, then one should question the validity of the solution. In ill-posed problems, we find that the DPC holds initially and then fails at some point i_{DPC} , where the data become dominated by errors. If this is the case, and if the regularization parameter λ is accurately selected, then the regularized solution should provide a valid solution. Examining i_{DPC} provides a method of characterizing the ill-posedness of the problem.

2 Methods for Selecting Regularization Parameter

2.1 Criteria

We are seeking a method for selecting the regularization parameter for optical topography. We begin by reviewing a number of methods that have been proposed in the literature, but reject some immediately because they are not suitable for optical topography. Our criteria are as follows:

1. The method should not require any subjective input from the user.
2. The method should only require knowledge that is available during clinical optical topography. For example, it should not require knowledge of the size of the feature being examined.
3. The method should not assume particular features in the image. For example, it should not assume there is a single, spatially isolated change.

2.2 Heuristic Method

The most straightforward—and most widely used—method for selecting λ is to examine solutions for a range of λ heuristically by eye and to select the one that results in the most acceptable reconstruction. This method is subjective and nonrepeatable.¹² A common variant is to take λ as being equal to the noise present in the data.¹³ We retain this method in our analysis as a measure against which to compare other, more objective methods.

2.3 Methods that Optimize Data and Image

2.3.1 L-curve

The L-curve is probably the most commonly employed objective method for finding the regularization parameter when solving ill-posed problems.^{8,14-16} It is a log-log plot, for different λ , of the image norm against the data norm. We take the value of λ , which corresponds to the point of maximum curvature on the graph, which is normally the point on the graph that is nearest to the origin and so mutually minimizes both the image norm and the data norm.

2.3.2 Fixed noise figure (NF)

The noise figure (NF) is the ratio of the signal-to-noise ratio in the measurements to the signal-to-noise ratio in the image.^{12,17,18} The regularization parameter is found by plotting NF as a function of λ , and we select the NF whose λ returns the most acceptable solution. This method replaces the selection of a regularization parameter as in the heuristic method by the selection of a fixed NF value. The ratio of the signal-to-noise ratios is more constant across different experimental setups than the heuristic method (i.e., when a fixed NF value is used the regularization parameter can be different for

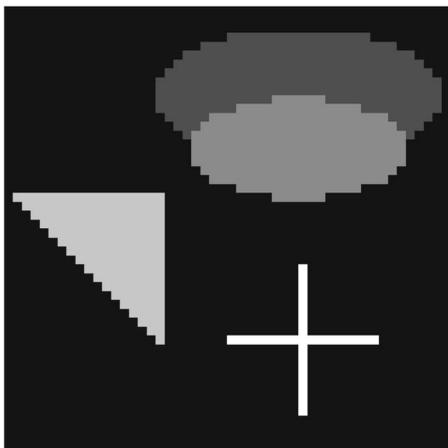


Fig. 1 Original test image.

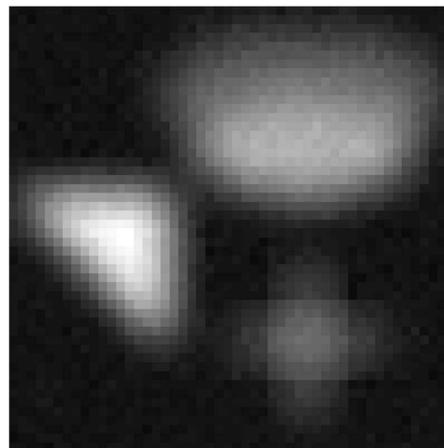


Fig. 2 Blurred image with 5% added Gaussian noise.

the different image reconstructions. However, we require a fully objective measure and therefore exclude this method from further analysis.

2.4 Methods that Optimize the Data

2.4.1 Generalized cross-validation (GCV)

GCV is based on the principle that, if a data point is omitted, then we should be able to estimate the missing data value from the regularized solution obtained from this reduced data set.^{19,20} We minimize $(GCV)(\lambda) = \frac{\|A\mathbf{x}_\lambda - \mathbf{b}\|_2^2}{[\text{trace}(I - AA_\lambda^\dagger)]^2}$, where A_λ^\dagger is the Tikhonov regularized pseudoinverse of A . The numerator is the data norm, and the denominator is inversely related to the number of singular values used in the regularized solution. Minimizing this then favors low data norms while penalizing solutions that require many singular values. GCV therefore finds the λ that provides a solution that can fit the data using the smallest possible number of parameters, thereby minimizing the contribution from small singular values.

2.4.2 Unbiased predictive risk estimator (UPRE)

The UPRE method seeks to minimize the predictive risk.⁹ The data noise is assumed to be random white noise of known variance. The accuracy of the method therefore depends on the accuracy of the estimate of noise. Furthermore, because the UPRE is an unbiased estimator, its expected value is the same as the expected value of the predictive risk, but it does not necessarily change with λ in the same way as the predictive risk. We cannot therefore guarantee that the solution error is small.

2.4.3 Discrepancy principle (DP)

This method selects the λ for which the data norm is equal to the data variance.⁹ Like the UPRE, it depends on our knowledge of the noise statistics: if the data variance is unknown and must be estimated, the method may not necessarily return the optimal regularization parameter. If we do have knowledge of the data variance, then we may see improved performance as this additional information is used.

2.4.4 Normalized cumulative periodogram (NCP)

This method favors the regularization parameter for which the residual vector resembles white noise.²¹ It is derived from the periodogram, which is the power spectrum of the residual and is obtained by taking the squares of the absolute values of the discrete Fourier transform for half the residual vector length. The NCP is the cumulative periodogram normalized by the sum of its elements. If the residuals are pure white noise, then the NCP is a straight line; hence, the selected regularization parameter is the one that minimizes the distance of the NCP to a straight line.²¹

2.5 Methods that Optimize the Image

2.5.1 F-slope

The f-slope is a plot of the image norm against $\ln(1/\lambda)$.²² We select the λ at the flattest part of the curve, which corresponds to the smallest difference between adjacent solution norms. This method only analyses the image norm and not the data norm.

2.5.2 Quasi-optimality criterion (QOC)

The regularization parameter is found by minimizing $Q_\lambda = \|\lambda^2 [d\mathbf{x}_\lambda / d(\lambda^2)]\|_2$.^{8,23} In an iterative method, this minimizes the difference between the current and previous solutions. In a noniterative approach, QOC minimizes the update to the initial guess.

2.5.3 Full width half maximum (FWHM)

The FWHM of the region of contrast is calculated for different regularization parameters.²⁴ This method is only applicable to images that contain a single isolated region, and thus, we reject it. Adler and Guardo¹⁷ proposed an alternative method of defining FWHM known as the blur radius as a method to select λ . This method has the same disadvantages as FWHM; thus, we reject it also.

2.5.4 Contrast-to-noise ratio (CNR)

The CNR is plotted as a function of the regularization parameter, where contrast is the ratio of the peak value of the image,

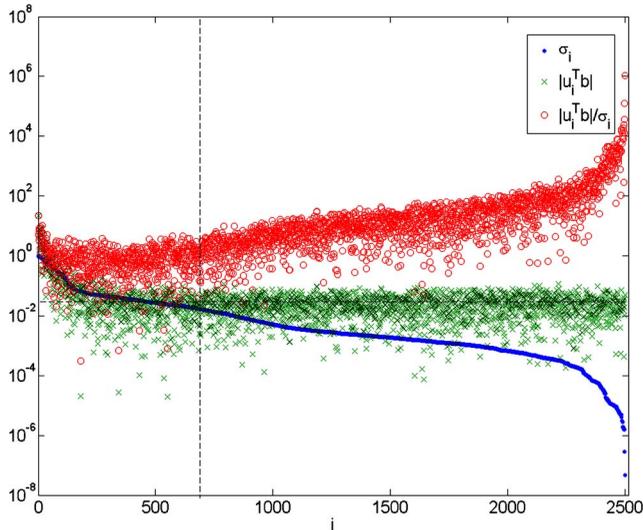


Fig. 3 Discrete Picard condition for the deblurring problem, where the vertical dashed line marks the beginning of $|U_i^T \mathbf{b}| < \sigma_i$ and the horizontal line represents the noise level. DPC is satisfied for $i_{\text{DPC}} \approx 695$.

after background subtraction, to the background value, and noise is the image norm.²⁴ We seek the regularization parameter that maximizes CNR.

Regińska²⁵ has shown that the minimum of $\Psi_\alpha(\lambda) = \|\mathbf{x}\|_2 \cdot \|\mathbf{A}\mathbf{x}_\lambda - \mathbf{b}\|_2^\alpha$, where $\alpha > 0$, is similar to the point that minimizes the L-curve. We propose a modification to the CNR method where, rather than maximizing CNR, we maximize CNR. Ψ_α^{-1} , which simultaneously optimizes both the image norm and the data norm.

3 Problem 1: Deblurring

3.1 Method

Deblurring is an example of an ill-posed inverse problem. The function *blur* from the Matlab package “Regularization Tools”²⁶ was used to generate the matrix A , the original image x (Fig. 1), with dimensions 50×50 pixels, and the corresponding blurred image b to which we added different levels of Gaussian noise, from 5 to 40% and for 500 noise realisations each (Fig. 2). The smoothing matrix A is chosen to have properties that make it computationally efficient to handle: it is the Kronecker product of a Toeplitz matrix T with itself: $A = T \otimes T$. The Toeplitz matrix contains in its diagonal elements of a Gaussian point-spread function with variance σ^2 , which models the blurring effect, and is a banded matrix, with band l that defines the number of diagonals, from the main diagonal, which are stored in the matrix T .^{9,27} For this test, we have set $\sigma=3$ and $\text{band}=5$.

One convenient property of matrix A is that its SVD depends only on the SVD of the initial Toeplitz matrices. Therefore, if the SVD of T is $T = U_b S_b V_b^T$, the regularized solution \hat{x} is given by

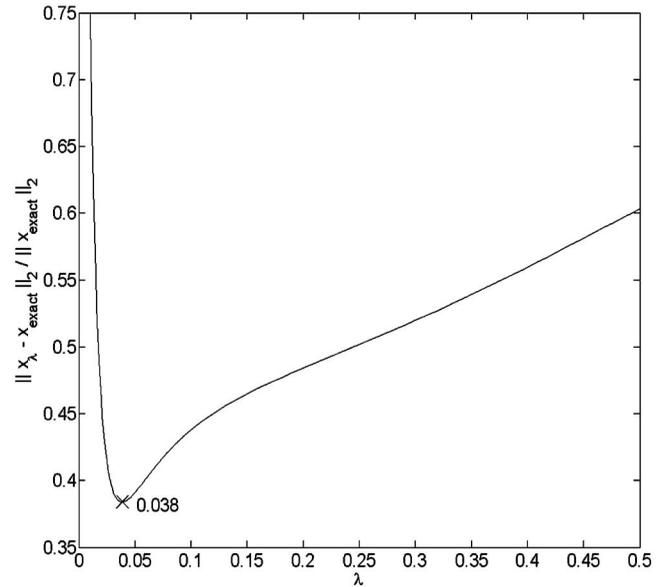


Fig. 4 Relative error for the regularised solution for a single realization of 5% noise. The λ_{opt} value is marked by the cross.

$$\hat{x} = V_b \frac{S_i(U_b^T b U_b)}{S_i^2 + \lambda^2} V_b^T, \quad (2)$$

where $S_i = \text{diag}(S_b) \text{diag}(S_b^T)$.

The discrete Picard condition is examined in Fig. 3. The value of i at which σ_i begins to decay more slowly than $|U_i^T \mathbf{b}|$ is shown by the vertical line. This point is emphasized by examining $|U_i^T \mathbf{b}|/\sigma_i$ whose gradient turns positive at the same point. This shows that the DPC is satisfied for $i_{\text{DPC}} \approx 695$. For higher values of i , the DPC is no longer satisfied and $|U_i^T \mathbf{b}|$ reaches the noise level. Because the DPC is at least partially satisfied, it means that we can expect to find a solution that approximately recovers the real solution.

The original image is known; thus, we can find the regularized solution \mathbf{x}_λ that is closest to the real solution $\mathbf{x}_{\text{exact}}$. This allows us to define the optimal regularization parameter λ_{opt} to be the one that minimizes the relative error

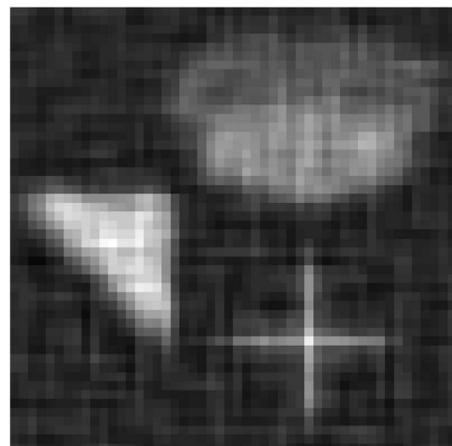


Fig. 5 Reconstructed image with $\lambda_{\text{opt}}=0.038$.

Table 1 Regularization parameters λ obtained using the selection methods for the deblurring problem with different noise levels.

Method	5%	10%	15%	20%
Optimal	0.039±2.6%	0.066±3.3%	0.090±4.0%	0.111±4.2%
Heuristic	0.050±60.0%	0.125±60.0%	0.130±53.8%	0.190±57.9%
L-curve	0.030±2.7%	0.065±1.7%	0.090±1.6%	0.110±1.5%
GCV	0.030±4.0%	0.047±5.1%	0.060±5.6%	0.071±5.8%
UPRE	0.030±3.3%	0.048±4.4%	0.062±4.4%	0.073±4.8%
DP	0.054±3.2%	0.087±3.2%	0.113±3.3%	0.133±3.1%
NCP	0.056±6.4%	0.088±7.6%	0.113±8.5%	0.134±8.7%
f-slope	0.070±1.7%	0.089±1.7%	0.104±1.8%	0.116±1.7%
QOC	0.121±1.5%	0.142±1.7%	0.160±1.8%	0.176±1.7%
CNR	0.014±35.0%	0.018±57.2%	0.025±60.0%	0.033±61.8%
CNR· Ψ^{-1}	0.022±3.6%	0.049±4.5%	0.077±13.2%	0.105±11.3%
Method	25%	30%	35%	40%
Optimal	0.131±4.5%	0.148±4.2%	0.164±4.3%	0.178±4.0%
Heuristic	0.190±57.9%	0.300±66.7%	0.300±66.7%	0.350±71.4%
L-curve	0.128±1.5%	0.145±1.5%	0.159±1.5%	0.172±1.3%
GCV	0.081±6.1%	0.090±5.7%	0.099±5.8%	0.108±5.0%
UPRE	0.083±5.1%	0.092±4.8%	0.101±4.9%	0.109±5.0%
DP	0.151±3.2%	0.167±3.1%	0.181±3.2%	0.195±3.1%
NCP	0.152±9.1%	0.168±9.6%	0.183±9.3%	0.195±9.7%
f-slope	0.127±1.8%	0.138±1.7%	0.147±1.8%	0.155±1.7%
QOC	0.189±1.7%	0.202±1.7%	0.213±1.6%	0.223±1.7%
CNR	0.039±55.4%	0.044±58.7%	0.049±53.4%	0.051±56.4%
CNR· Ψ^{-1}	0.129±9.7%	0.154±2.4%	0.183±2.8%	0.190±2.8%

$$\varepsilon = \frac{\|x_\lambda - x_{\text{exact}}\|_2}{\|x_{\text{exact}}\|_2}. \quad (3)$$

Figure 4 shows a plot of ε against λ , for one realization of 5% Gaussian noise. The minimum occurs for $\lambda_{\text{opt}}=0.038$. The corresponding deblurred and denoised image for λ_{opt} is shown in Fig. 5.

3.2 Results

The mean optimal regularization parameters, and corresponding standard deviations, for the ten selection methods tested are shown for different noise levels in Table 1. The corresponding errors calculated from Eq. (3) using the mean values in Table 1 are shown in Table 2 and in Fig. 6. The regulariza-

tion parameter was chosen from a set of 1000 logarithmically spaced points between 10^{-4} and $10^{-0.3}$.

3.3 Discussion

3.3.1 Heuristic method

It was not easy to identify a single λ with the heuristic method, and we instead selected a range of λ that provide acceptable results. In every case, the range included λ_{opt} . The corresponding ranges of errors are quoted in Table 2. Note that for better comparison with the other methods, we chose to display its central value and error, where the latter gives the range limits. The range of λ included λ_{opt} ; however, the regularization parameter errors are consistently higher than for

Table 2 Relative error for the deblurring problem with different noise levels.

Method	5%	10%	15%	20%
Optimal	0.387±1.0%	0.432±1.2%	0.454±1.1%	0.468±1.1%
Heuristic	0.393±0.8%	0.457±0.5%	0.463±0.7%	0.485±0.6%
L-curve	0.395±1.5%	0.432±1.2%	0.454±1.1%	0.468±1.1%
GCV	0.395±1.5%	0.433±1.2%	0.477±1.7%	0.499±1.8%
UPRE	0.395±1.5%	0.432±1.2%	0.473±1.7%	0.496±1.8%
DP	0.400±0.8%	0.438±0.9%	0.458±0.9%	0.471±0.9%
NCP	0.400±0.8%	0.439±0.7%	0.456±1.1%	0.471±0.9%
f-slope	0.414±0.5%	0.439±0.9%	0.456±0.9%	0.469±1.1%
QOC	0.451±0.2%	0.463±0.4%	0.473±0.5%	0.482±0.6%
CNR	0.578±2.8%	0.853±2.3%	0.878±2.4%	0.884±2.4%
CNR· Ψ^{-1}	0.403±1.5%	0.456±1.1%	0.457±0.9%	0.472±0.9%
Method	25%	30%	35%	40%
Optimal	0.479±1.0%	0.488±1.0%	0.495±1.2%	0.502±1.0%
Heuristic	0.489±0.6%	0.525±0.6%	0.525±0.6%	0.544±0.6%
L-curve	0.479±1.0%	0.488±1.0%	0.495±1.2%	0.502±1.2%
GCV	0.513±1.9%	0.526±1.6%	0.534±1.2%	0.541±1.9%
UPRE	0.510±2.0%	0.520±1.7%	0.528±1.9%	0.541±1.9%
DP	0.481±0.8%	0.489±0.8%	0.496±1.0%	0.503±1.0%
NCP	0.481±0.8%	0.490±0.8%	0.496±1.0%	0.503±1.0%
f-slope	0.480±1.3%	0.488±1.2%	0.497±1.2%	0.504±1.2%
QOC	0.489±0.6%	0.495±0.8%	0.501±0.8%	0.506±1.0%
CNR	0.895±2.6%	0.903±2.6%	0.935±2.7%	0.987±2.5%
CNR· Ψ^{-1}	0.481±0.8%	0.489±0.8%	0.498±1.0%	0.502±1.0%

most of the other methods, illustrating the irreproducibility of the heuristic method.

3.3.2 L-curve

The L-curve (plotted Fig. 7 for a 5% noise realization) did generally exhibit a single, easily identifiable point of maximum curvature. The predicted λ agree closely with λ_{opt} .

Hanke²⁸ showed that the L-curve method fails to find λ for very ill-posed solutions. When the singular values decay very rapidly to zero, λ_{opt} may occur before the data norm starts increasing, because a large number of singular values have to be included before the data norm increases significantly. Vogel²⁹ has shown another nonconvergence of λ , which occurs when the regularized solution fails to converge to the true solution as the dimensions of the problem tend toward infinity. We do not see evidence of either of these two concerns.

For the deblurring problem, an image reconstructed with 5% noise was slightly undersmoothed, with a visible noise component. However, all the results were acceptable, as confirmed by the small relative errors (Table 2) and by the small regularization parameter error (Table 1). The latter reflects the small sensitivity of this method to perturbations of similar nature in the data.

3.3.3 GCV

The minimum value for this function was easily identifiable, although it was located on a relatively flat region of the curve. The selected λ was similar to λ_{opt} at low noise levels but was rather low at higher noise levels.

The GCV method has been shown to perform well in different situations. However, it can sometimes be difficult to locate its minimum as the function may be flat near λ_{opt} or

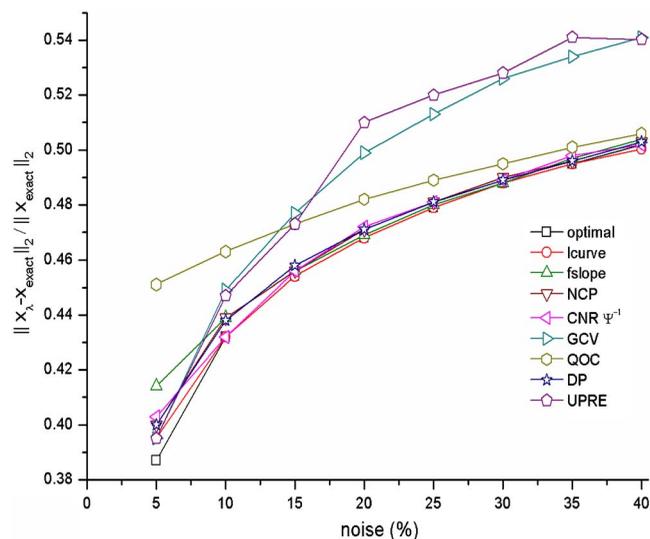


Fig. 6 Mean relative error of λ_{opt} (\square), L-curve (\circ), f-slope (\triangle), NCP (∇), $C \cdot \Psi^{-1}$ (\triangleleft), GCV (\triangleright), QOC (\diamond), DP (\star), and UPRE (\diamond) methods for different noise levels.

display multiple minima.³⁰ Hansen and O’Leary¹⁵ presented an example where GCV failed to find λ due to the flatness of the curve. However, we were able to locate a single minimum for all our test cases from 5 to 40% noise.

3.3.4 UPRE

The UPRE function did exhibit an easily identified minimum. However, to calculate UPRE it is necessary to find the SVD of the blur matrix, which, although it was a reasonable calculation in this case due to the carefully chosen blurring matrix, in general, could be prohibitively computationally intensive. The performance of UPRE appears to be similar to that of GCV, with good estimates of λ at low noise levels but an underestimated λ at higher noise levels. Vogel⁹ used a two-dimensional deblurring problem with added white noise to compare these two methods, and found that the UPRE and GCV methods were similar.

3.3.5 DP

This method appeared to overestimate values of λ , confirming the results of Thompson et al.³¹ and Hansen.⁸

3.3.6 NCP

This method and the DP have similar performances, as expected, because they both select the regularization parameter such that the data norm is equal to the noise variance. However, this method seems to be more sensitive to the different noise realizations, in particular for noise levels above 25%. Hansen et al.²¹ found that the NCP gives better results than the GCV method, which is confirmed here at higher noise levels.

3.3.7 f-Slope

The f-slope curve clearly shows a flat part, which is less sensitive to perturbations. This method appears to overestimate λ at low noise levels but selects λ close to λ_{opt} at higher noise levels.

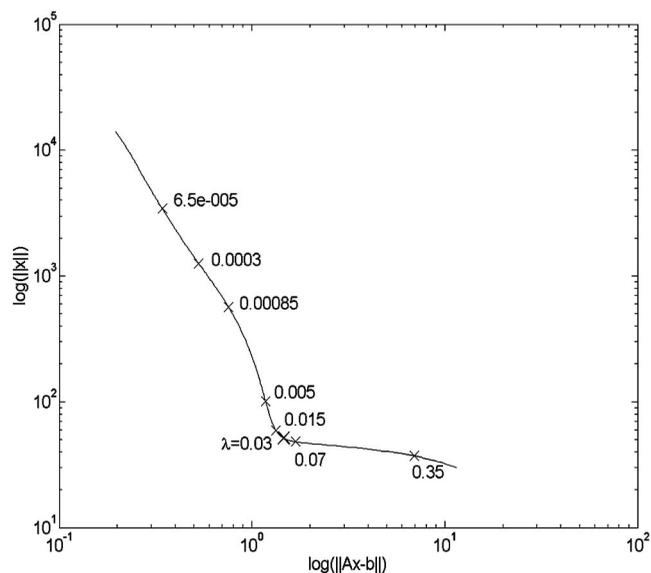


Fig. 7 L-curve for deblurring with 5% noise.

It has been claimed²² that the f-slope method can perform better than the L-curve method, particularly when little regularization is needed. However, this work mainly looked at much smaller λ than in our deblurring problem (as low as 10^{-15}) and it is unclear how the results translate to our more ill-posed problem. Using f-slope, we were able to reconstruct an image that is close to the true image when the noise in the data is equal to or higher than 15% but for smaller noise levels, it tended to overregularize the solution.

3.3.8 QOC

The regularization parameter λ selected by this method was invariably higher than λ_{opt} . However, the performance of QOC appeared to improve at higher noise levels.

Hansen⁸ compared the L-curve, the GCV, and the QOC methods by applying them to six test problems with different data perturbations but with the same noise level. He found that the QOC method generally oversmooths and that there may be a problem with local minima. We saw the same over-smoothing effect but did not observe local minima.

3.3.9 CNR

There was a single clear global maximum. However, for all noise levels, this method found $\lambda < \lambda_{\text{opt}}$, and generated much greater relative errors than for any of the other methods, and the errors associated with λ are as large as those corresponding to the heuristic method.

3.3.10 $CNR \cdot \Psi^{-1}$

The addition of the Ψ parameter led to λ which were closer to λ_{opt} than for CNR alone. The method still tended to underestimate λ for low noise levels, but for higher noise levels, the selection of λ was good. Note that the regularization parameter uncertainty is particularly large for noise in the data between 15 and 25%.

Table 3 Regularization parameters for experimental data using different selection methods.

Method	λ
Heuristic	[0.004, 0.015]
L-curve	0.0066
GCV	0.0085
UPRE	0.0071
DP	0.0061
NCP	0.0032
f-slope	—
QOC	—
CNR	0.0051
$\text{CNR} \cdot \Psi^{-1}$	0.0068
$\lambda_{\epsilon\mu_a}$	0.0071

4 Problem 2: Optical Topography

4.1 Method

Data were obtained with the University College of London (UCL) optical topography system with light sources at 770 nm³² using a test object consisting of absorbing targets within a tank filled with a solution with tissue-equivalent optical properties (the absorption coefficient μ_a was 0.01 mm⁻¹ and the reduced scattering coefficient μ'_s was 1 mm⁻¹). A cylindrical absorbing target (radius 5 mm and height 10 mm) was made with the same μ'_s as the solution but with twice the background absorption ($\mu_a=0.02$ mm⁻¹).

The optical topography array consisted of eight sources and eight detectors from which 64 measurements were made.³³ The array was placed on one of the walls of the tank, which is made of epoxy resin with the same optical properties as the solution, and is 2 mm thick. The target was positioned in the center of the array, which should provide higher sensitivity and resolution, and at a depth of 10 mm. Data were collected for 20 s and averaged to reduce the noise, and a baseline measurement was acquired with no target present.

The software package TOAST (temporal optical absorption and scattering tomography), which has been developed by Prof. S. R. Arridge and Dr. M. Schweiger at the UCL, models the propagations of light in highly scattering media and was used to generate the sensitivity matrix A . Finally, images were reconstructed using a 3D linear model.¹³

4.2 Results

The results for all the methods are summarized in Table 3. The λ values used are the same as before.

4.2.1 DPC

Figure 8 shows that the DPC was partially satisfied, and thus, regularization should be able to give a stable solution. The

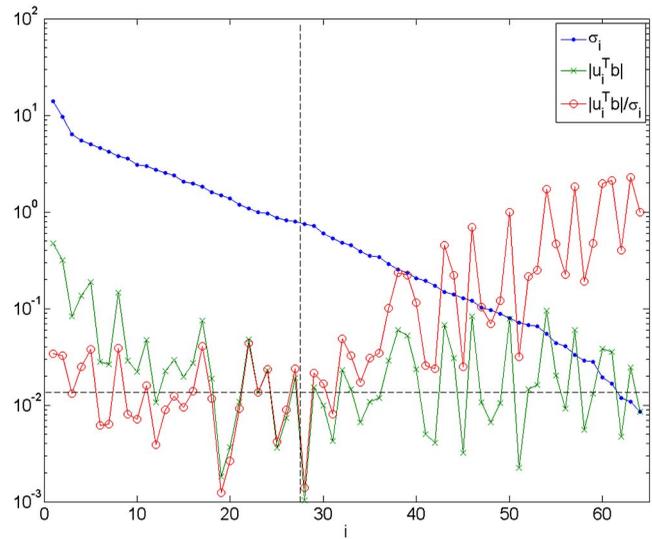


Fig. 8 Discrete Picard condition. There are 64 singular values σ_i . The horizontal dashed line represents the noise level, and the vertical line $i_{\text{DPC}} \approx 27$ for which $i > i_{\text{DPC}}$ no longer satisfies the DPC.

vertical line represents i_{DPC} , where for $i < i_{\text{DPC}}$ the data space coefficients $|u_i^T \mathbf{b}|$ decay faster than the singular values σ_i , and for $i > i_{\text{DPC}}$ the data space coefficients reach a level determined by the perturbations in the data; thus the DPC is no longer satisfied.

4.3 Discussion

4.3.1 Heuristic method

The range of λ that generated acceptable images, which were relatively noise-free, with adequate spatial localization, and which placed the target in the middle of the image was $\lambda \in [0.004, 0.015]$.

For this phantom study, we know the optical properties and target position with an accuracy of 5–10%. Figure 9 shows the absorption coefficient error $\epsilon\mu_a$, which is the difference between the maximum μ_a , calculated from the reconstructed images for regularization parameters found with the heuristic method in steps of 0.001, and the target μ_a . We refer to the point where the error is zero, $\lambda_{\epsilon\mu_a}=0.0071$, as the optimal regularization parameter. An image reconstructed using $\lambda = \lambda_{\epsilon\mu_a}=0.0071$, at 10 mm depth, is shown in Fig. 10. The target size in the image, measured by the FWHM, is $10 \pm 3.5 \times 14 \pm 3.5$ nm, which is not too different from the real target dimensions. The target size is approximately constant for the selected range of regularization parameters.

The values of λ found by the other selection methods are compared to $\lambda_{\epsilon\mu_a}$ below. Table 3 shows all λ values.

4.3.2 L-curve

The L-curve did not exhibit a pronounced corner, but it was still possible to calculate the point of maximum curvature. The failure to find a sharp corner was due to the high ill-posedness of this problem (as shown by the exponential decay of singular values in Figure 8). However, λ at the point of maximum curvature of the L-curve was closer than any of the other methods to $\lambda_{\epsilon\mu_a}$. The L-curve method has been used

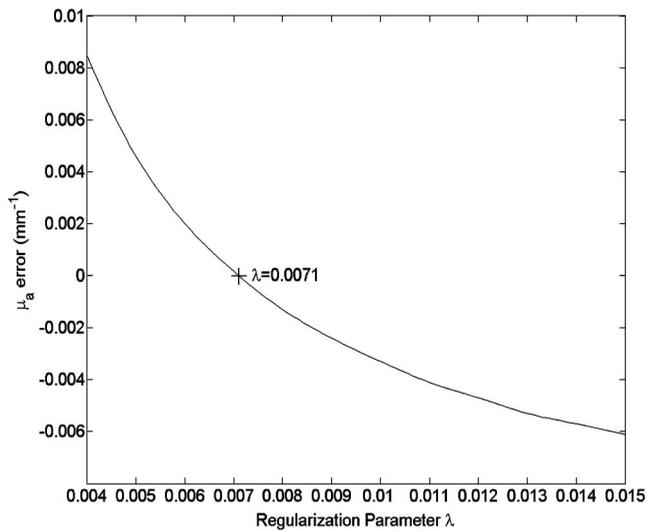


Fig. 9 Absorption coefficient error for the heuristic method regularization parameters. The cross indicates the point where the error is zero.

previously in optical imaging, for simulated and real data, and its results are considered to generate acceptable images.^{34–36}

4.3.3 GCV

The value for λ found using GCV was the highest of all the methods. Hansen¹⁶ found that the GCV method fails to compute a useful solution when errors are highly correlated. In the presence of uncorrelated errors, this method gives a slightly overregularized λ . Our result could be explained by a low presence of correlated errors.

4.3.4 UPRE

UPRE found a very shallow minimum at $\lambda = \lambda_{\epsilon\mu_a}$. The noise variance was assumed to be the point at which $i = i_{DPC}$ in Fig. 8. At this point, $u_i^T b$ reaches the noise floor and no longer decreases.

4.3.5 DP

The L-curve, GCV, DP, and UPRE methods have previously been compared for simulated, phantom, and clinical data in electrical impedance tomography.³⁷ All the methods were successful for simulated and clinical data, whereas for data acquired on a test phantom, the DP and UPRE methods failed to converge. Overall, the preferred method was the GCV. For optical topography data, both DP and UPRE methods converged and the predicted λ agreed with the values found heuristically.

4.3.6 NCP

It is necessary to preprocess data so that all the variables have zero mean and unit variance. In optical topography, we use the difference imaging approach, where the baseline is subtracted from the data; hence, all the measurements should have zero mean, but different variance. Whitening could be accomplished by dividing each measurement by the respective standard deviation. This method returns a very low λ , which is not included in the values found heuristically. If we set the Kolmogorov–Smirnov limits³⁸ to a significance level

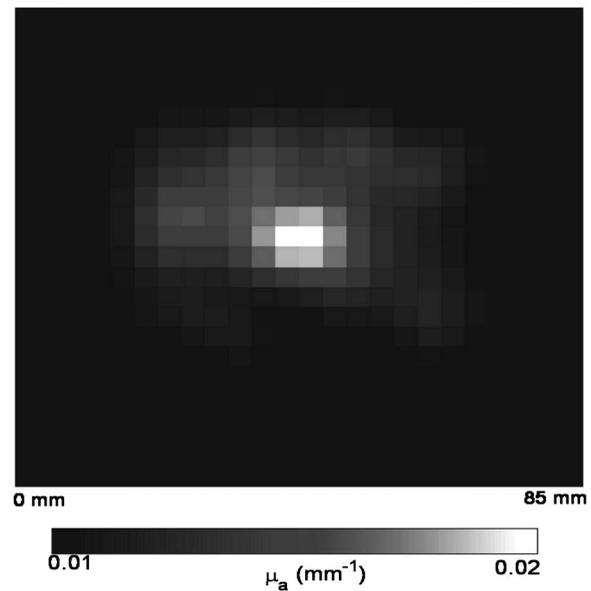


Fig. 10 Reconstructed optical topography image, using $\lambda_{\epsilon\mu_a} = 0.0071$.

of 5%, which is equivalent to a 95% confidence level, and choose the largest regularization parameter, then we obtain $\lambda = 0.0056$, which is a more reasonable result.

4.3.7 f-slope

The solution norm monotonically increased with $\ln(1/\lambda)$; thus, it was not possible to identify the minimum slope. To our knowledge the f-slope method has not been tested previously on real data, only for testing models where it has a good performance.²² This method failed when applied to our optical topography problem, probably due to its ill-posedness.

4.3.8 QOC

The QOC monotonically decreased with $\log(\lambda)$ and failed to show a reliable minimum, and consequently, this method failed.

4.3.9 CNR

To calculate the contrast, it is necessary to know $\mu_{a_{\max}}$ and $\mu_{a_{\text{bkg}}}$. The value of $\mu_{a_{\max}}$ was taken to be the mean of the pixel of maximum intensity in the image and its eight nearest neighbors in the xy plane. The value of $\mu_{a_{\text{bkg}}}$ was the average of 24 pixels furthest from the peak in the image. This method gave the lowest λ of all the methods, and artifacts were present in reconstructed images.

4.3.10 CNR Ψ^{-1}

As before, including Ψ gave a much more realistic estimate of λ than CNR alone, which was very close to $\lambda_{\epsilon\mu_a}$.

The regularization parameter selection methods were tested on two further data sets, obtained using the same liquid phantom, but targets with different absorption coefficient and the results were consistent with those shown here.

5 Conclusion

Optimizing λ is critical when reconstructing diffuse optical images because it controls the smoothness of the regularized solution and balances the influence of the noise present in the image against its accuracy. We have examined a number of methods for selecting λ . Diffuse optical imaging is a challenging ill-posed and underdetermined problem, and the correct target image is unknown, making it difficult to validate the image quality. We therefore initially tested the methods by applying them to a simpler ill-posed problem, the deblurring problem, where we can control the amount of blur and noise applied to the data and where the exact solution is known (and so λ_{opt} can be found). We believe that if a method fails to produce a good regularized solution for the deblurring problem, we cannot rely on it for optical topography. However, if a good solution is found for the deblurring problem, then that method will not necessarily be reliable for the more demanding optical problem.

In Sec. 2.1, we list three criteria that a selection method must meet. On the basis of these criteria, we reject the heuristic method, the fixed noise figure method, and methods related to optimizing the FWHM of the image. All the remaining methods performed acceptably well for the deblurring problem. However, the f-slope and QOC failed to converge for the more demanding optical topography problem.

Of the remaining methods, the L-curve consistently demonstrated the lowest error on the deblurring problem (see Tables 1 and 2). It is easy to implement and simultaneously minimizes both the data norm and the image norm. On the other hand, the DP and UPRE methods require knowledge of the noise variance, which may not always be available. However, the DPC seems to provide a good estimate of the noise level in optical topography. The NCP method has the advantage of selecting λ automatically; hence, it does not directly require an estimate of the noise variance. Nevertheless, this method is sensitive to error fluctuations and only gives reasonable results for the optical topography problem under certain assumptions. DP, UPRE, NCP, and GCV only consider the data norm. The CNR method gave poor results but was much more successful when Ψ was minimized simultaneously. The use of Ψ in this way, as proposed by Regińska²⁵ and developed further here, could be applied to other methods and may be worth further examination. However, here we conclude that L-curve is the optimal selection method for optical topography.

Thus far, we have studied ideal or almost ideal cases, whereas *in vivo* studies suffer further sources of error. These include motion artifacts, changes in the contact between the optodes and the skin, which can result in intensity fluctuations in the collected data, and detection of light that has not passed through the investigated medium. All these effects produce correlated errors, which are not necessarily apparent in the raw data. The L-curve has been shown to perform well in the presence of these type of errors.^{15,16} However, as mentioned previously, Hanke²⁸ showed that the L-curve method may fail to find a good regularization parameter when the solutions are very smooth. Under these circumstances, GCV should be investigated as an alternative to L-curve.

Acknowledgments

The work has been supported by a scholarship awarded to T.C. by Fundação para a Ciência e a Tecnologia, Portugal. Thanks to Dr. Simon Arridge, from the Department of Computer Science, for useful discussions.

References

1. J. C. Hebden, "Advances in optical imaging of the newborn infant brain," *Psychophysiology* **40**(4), 501–510 (2003).
2. M. Wolf, M. Ferrari, and V. Quaresima, "Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications," *J. Biomed. Opt.* **12**(6), 062104 (2007).
3. Y. Yamashita, A. Maki, and H. Koizumi, "Measurement system for noninvasive dynamic optical topography," *J. Biomed. Opt.* **4**(4), 414–417 (1999).
4. D. A. Boas, T. Gaudette, G. Strangman, X. Cheng, J. J. Marota, and J. B. Mandeville, "The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics," *Neuroimage* **13**(1), 76–90 (2001).
5. D. A. Boas, K. Chen, D. Grebert, and M. A. Franceschini, "Improving the diffuse optical imaging spatial resolution of the cerebral hemodynamic response to brain activation in humans," *Opt. Lett.* **29**(13), 1506–1508 (2004).
6. A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, Winston, New York (1977).
7. G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.* **21**(1), 185–194 (1999).
8. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia (1998).
9. C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia (2002).
10. G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore (1996).
11. P. C. Hansen, "The discrete Picard condition of discrete ill-posed problems," *BIT Numer. Math.* **30**(4), 658–672 (1990).
12. B. M. Graham and A. Adler, "Objective selection of hyperparameter for EIT," *Physiol. Meas* **27**, S65–S79 (2006).
13. A. P. Gibson, T. Austin, N. L. Everdell, M. Schweiger, S. R. Arridge, J. H. Meek, J. S. Wyatt, D. T. Delpy, and J. C. Hebden, "Three-dimensional whole-head optical tomography of passive motor evoked responses in the neonate," *Neuroimage* **30**, 521–528 (2006).
14. P. C. Hansen, "The L-curve and its use in the numerical treatment of inverse problems," in *Computational Inverse Problems in Electrocardiology*, P. Johnston, Ed., Advances in computational bioengineering, pp. 119–142 (2001).
15. P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput. (USA)* **14**(6), 1487–1503 (1993).
16. P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.* **34**(4), 561–580 (1992).
17. A. Adler and R. Guardo, "Electrical impedance tomography: regularized imaging and contrast detection," *IEEE Trans. Med. Imaging* **15**(2), 170–179 (1996).
18. A. Adler, "Accounting for erroneous electrode data in electrical impedance tomography," *Physiol. Meas* **25**, 227–238 (2004).
19. G. Wahba, *Spline Models for Observational Data*, SIAM, Philadelphia (1990).
20. G. Golub, M. Heath, and G. Wahba, "Generalized-cross validation as a method for choosing a good ridge parameter," *Technometrics* **21**(2), 215–223 (1979).
21. P. C. Hansen, M. E. Kilmer, and R. H. Kjeldsen, "Exploiting residual information in the parameter choice for discrete ill-posed problems," *BIT Numer. Math.* **46**, 41–59 (2006).
22. L. Wu, "A parameter choice method for Tikhonov regularization," *Electron. Trans. Numer. Anal.* **16**, 107–128 (2003).
23. T. Kitagawa, S. Nakata, and Y. Hosoda, "Regularization using QR factorization and the estimation of the optimal parameter," *BIT Numer. Math.* **41**(5), 1049–1058 (2001).
24. J. P. Culver, R. Choe, M. J. Holboke, L. Zubkov, T. Durduran, A. Slemple, V. Ntziachristos, B. Chance, and A. G. Yodanis, "Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: evaluation of a hybrid frequency domain/

- continuous wave clinical system for breast imaging," *Med. Phys.* **30**(2), 235–247 (2003).
25. T. Regińska, "A regularization parameter in discrete ill-posed problems," *SIAM J. Sci. Comput. (USA)* **17**(3), 740–749 (1996).
 26. P. C. Hansen, "Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms* **6**(1–2), 1–35 (1994).
 27. P. C. Hansen, "Deconvolution and regularization with Toeplitz matrices," *Numer. Algorithms* **29**(56), 323–378 (2002).
 28. M. Hanke, "Limitations of the L-curve method in ill-posed problems," *BIT Numer. Math.* **36**(2), 287–301 (1996).
 29. C. Vogel, "Nonconvergence of the L-curve regularization parameter selection method," *Inverse Probl.* **12**(4), 535–547 (1996).
 30. A. M. Thomson, J. W. Kay, and D. M. Titterton, "A cautionary note about crossvalidatory choice," *J. Stat. Comput. Simul.* **33**(4), 199–216 (1989).
 31. A. M. Thomson, J. C. Brown, J. W. Kay, and D. M. Titterton, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(4), 326–339 (1991).
 32. N. Everdell, A. Gibson, I. Tullis, T. Vaithianathan, J. Hebden, and D. Delpy, "A frequency multiplexed near infrared topography system for imaging functional activation in the brain," *Rev. Sci. Instrum.* **76**(9), 093705.1-5 (2005).
 33. A. Blasi, S. Fox, N. C. Everdell, A. Volein, L. Tucker, G. Csibra, A. P. Gibson, J. C. Hebden, M. Johnson, and C. Elwell, "Depth dependent changes in cerebral haemodynamics during face perception in infants," *Phys. Med. Biol.* **52**, 6849–6864 (2007).
 34. R. J. Gaudette, D. A. Boas, D. H. Brooks, C. A. DiMarzio, M. E. Kilmer, and E. L. Miller, "Comparison of linear reconstruction techniques for 3D DPDW imaging of absorption coefficient," *Proc. SPIE, The International Society for Optical Engineering*, pp. 55–66 (2000).
 35. C. Zhou, G. Yu, F. Daisuke, J. H. Greenberg, A. G. Yodh, and T. Durduran, "Diffuse optical correlation tomography of cerebral blood flow during cortical spreading depression in rat brain," *Opt. Express* **14**, 1125–1144 (2006).
 36. Q. Zhao, L. Ji, and T. Jiang, "Improving performance of reflectance diffuse optical imaging using a multicentered mode," *J. Biomed. Opt.* **11**(6), 064019 (2006).
 37. J. P. Abascal, S. Arridge, R. Bayford, and D. Holder, "Comparison of methods for optimal choice of the regularization parameter for linear electrical impedance tomography of brain function," *Physiol. Meas.* **29**, 1319–1334 (2009).
 38. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed., Springer, New York (1991).