Nullspace regularization and MAP reconstruction in the ill-posed inverse imaging problem

Ivo W. Kwee*‡, Y. Tanikawa‡, S. Proskurin‡, Simon R. Arridge†, D.T. Delpy* and Y. Yamada‡
* Medical Physics & Bioengineering Department, University College London, UK.
† Department of Computer Science, University College London, UK.
‡ Biomechanics Division, Mechanical Engineering Laboratory, Tsukuba, JAPAN.

ABSTRACT

The ill-posed inverse problem is not an issue that is only restricted to optical tomography, but indeed a very common issue in image reconstruction problems in astronomy, geological surveying, and medical imaging in general. In this paper we investigate the consequences of ill-posed problems, and show that correct reconstruction is generally not possible using conventional linear inversion techniques because latter methods disregard contributions of the nullspace. We describe the rationale of a novel image reconstruction method that estimates the nullspace contibution using prior knowledge in a maximum-aposteriori-probability (MAP) framework. We illustrate our concept by an example of optical tomographic reconstruction from simulated and experimental data.

Keywords: Optical tomography, image reconstruction, maximum a posteriori probability, ill-posed problems, regularization, Bayesian probability, prior knowledge.

1. Introduction

The possibility of measuring the oxygenation state of biological tissues by near-infrared (NIR) light has recently received much attention. Equipment based on this effect has successfully been used to measure brain oxygenation in newborn infants.[1–3] It is expected that the use of NIR light can safely detect haematoma and breast tumors[4]; these lesions can be directly related to a locally increased oxygen level, and therefore a higher absorption of the probing NIR-light. *Optical Tomography* tries to map the optical properties of inner tissue by measuring the scattered light from the object.

The highly diffusive nature of the scattered photons in biological tissue complicates the image reconstruction. The measurements are highly dependent, and optical tomography belongs to the class of, so-called, *ill-posed* problems. The ill-posed problem is a common issue when we try to reconstruct images from limited number of data. This paper is concerned with the reconstruction of ill-posed problems, in particular how we can alleviate this problem with use of the *nullspace*.

The first few sections will briefly give an introduction to the general theory of linear inversion. Section 4. introduces the concept of nullspaces. Section 5. will deal with the issue of prior knowledge and Bayesian optimization. Section 6. and subsequent sections will describe our proposed algorithm using the previously mentioned concept of nullspaces and MAP reconstruction.

0-8194-2327-0/96/$6.00

2. Linearization of the Measurement Process

Light propagation in highly scattering media can be modelled in approximation by the diffusion theory as used in, e.g., linear transport theory.[5] It is frequently of interest to compute the change in observables caused by a small change of the parameters of the system. In this case we use the well-known *perturbation theory*[6] that relates changes between parameters and observables; linear perturbation expansions in the context of photon diffusion are described in literature..[7]

Using linear perturbation theory, the forward problem is approximated by a linear system, $\mathbf{A} \cdot \mathbf{x} = \mathbf{y}$, where $\mathbf{x}$ is the *image* vector containing the grey-values of pixelated image, $\mathbf{y}$ is the *data* vector containing the measurements, and $\mathbf{A}$ is the linearized measurement operator or *Jacobian*. If we perform M measurements and our digitized image has N pixels, then $\mathbf{A}$ is a $M \times N$ matrix, $\mathbf{x}$ a N-dimensional vector and $\mathbf{y}$ a M-dimensional vector.

The inverse imaging problem is to correctly resolve the unknown vector $\mathbf{x}$ when $\mathbf{y}$ and $\mathbf{A}$ are given. A solution can be found by minimizing a *data error* function, E, that measures the discrepancy of the guess from the measured data. Often the *chi-square* error norm is used for this, defined as

$$E = \chi^2 \equiv |\mathbf{W} \cdot (\mathbf{y} - \mathbf{A} \cdot \mathbf{x})|^2, \tag{1}$$

where diagonal matrix $\mathbf{W}$ contains the inverse of the standard deviations, σ, of the measurements. The "likelihood" of a solution is defined as "the probability of the data given the parameters", i.e.

$$P(\mathbf{y}|\mathbf{x}) \propto \exp(-\frac{1}{2}E). \tag{2}$$

Solutions that minimize the data error will maximize the likelihood, and vice versa; these solutions are also called "maximum likelihood solutions". For square and overdetermined systems ($M \geq N$) it is reasonable to believe that the "best" solution for $\mathbf{x}$ will minimize the data error E, but we will see that for underdetermined systems ($M < N$) this is not a sufficient condition.

3. Ill-Posedness, Pseudo-Inverses and the LSMN-solution

Diffusive imaging in optical tomography is severely restricted due to the *ill-posed* nature of the measurements. There are two main reasons to this: Firstly, the number of pixels that we want to reconstruct amounts typically to many more than the number of independent datapoints that is available. Secondly, due to the highly diffuse nature of the photons the measurements are highly dependent. This means that the system we want to solve for is effectively underdetermined, and in general there is no unique solution.

To force a solution we can utilize *pseudo-inverses*,[8] like the Moore-Penrose inverse that solves for the least-square minimum-norm (LSMN) solution. Mathematically, the LSMN-solution corresponds to the shortest vector that minimizes the least-squares error E. Arbitrarily large solutions are restricted by the Moore-Penrose pseudo-inverse because contributions of the, so-called, *nullspace* are implicitly set to zero.

Although the LSMN-solution seems to be a reasonable choice, it often does not correspond to the underlying physical image/object that we try to resolve. Moreover, because the nullspace vectors are disregarded we often obtain a "low-pass" or highly smoothed version of the original.

4. Orthogonal Spaces and Nullspace Vectors

The, so-called, *nullspace* consists of those vectors that are not measurable in the data; they are mapped to zero, $\mathbf{A} \cdot \mathbf{x} = 0$. It can be shown that a nullspace *always* exist for underdetermined systems. We can limit the nullspace by increasing the number of measurements as much as possible. However the physical shape of the object could restrict us to taking measurements only from a limited view; or in our case of optical tomography the measurements are highly dependent due to the diffusive nature of the photons. In those cases the number of *independent* measurements might not be increased proportionally by addition of measurements. Thus, in many practical situations we are dealing with incomplete or highly dependent data and therefore an inherently *underdetermined* or *ill-posed* system.

Image vectors that are accessible by a particular measurement setup are called *rowspace* vectors because they are row-entries of the matrix $\mathbf{A}$. The rowspace vectors only span a limited *subspace* in the case of an underdetermined system. An important point to notice is that if we have some measurements and use the pseudo-inverse operator to solve the system, then we can *only* obtain solutions that lie in the rowspace of $\mathbf{A}$: *only* the rowspace component of the real solution is recovered by the pseudo-inverse.

To obtain perfect reconstruction, we *must* extend the rowspace with vectors such that it spans the whole space. Fortunately, it happens that the nullspace contains just those vectors that the rowspace misses, and summed together they form a full set of orthogonal vectors. The nullspace is said to *complement* the rowspace, and vice versa. This means that in order to reconstruct the original object we *must* find a way to recover its nullspace component. The data error function as in Eq. 1 will not suffice because it is not affected by the nullspace. The next section will introduce so-called Bayesian "prior probability" functions as a way to incorporate the nullspace.

5. *A Priori* Information and Prior Probability Functions

The inability to measure the nullspace forms also the main difficulty in using nullspace vectors. If we can freely add any nullspace vector to a particular solution while still it does not affect the integrity with the measurements, how can we decide the right amount of nullspace vectors to correct the image? The answer is to use prior information of the (yet) unknown object. If we know that the object is mostly flat *a priori*, then we can add an amount of nullspace such that it conforms as much as possible to this prior information. If we know that the object is positive or has limited support *a priori*, then we can try to add just that amount of nullspace vectors to make the negative values positive, or to minimize the number of non-zero valued pixels. One approach is to introduce some kind of model-fit function that measures the closeness of our guess, $\mathbf{x}$, to a given prior model, $\mathbf{m}$, and maximize this function.

Instead of maximizing the likelihood probability function of Eq. 2 only, Bayesian optimization seeks the solution that maximizes the *a posteriori* probability defined as,

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})\, P(\mathbf{x}|\mathbf{m})}{P(\mathbf{y})} \tag{3}$$

where $P(a|b)$ means "the probability of an event a given b"; $P(\mathbf{y})$ is the prior probability of the data, and is mostly taken as a constant. Notice the similarity of the MAP formulation and the maximum likelihood (ML) estimation of Eq. 2. In fact ML can be regarded a special case of MAP where the prior probability $P(\mathbf{x}|\mathbf{m})$ is chosen to be constant or is—in many cases—carelessly neglected.

We will choose for our prior probability function,

$$P(\mathbf{x}|\mathbf{m}) \propto \exp[-\frac{\lambda}{2\sigma^2}(\mathbf{m}-\mathbf{x})^2], \tag{4}$$

There is no real justification for the choice of this prior, only that it is congruent to the definition the χ^2 error.

We could proceed to solve for the MAP solution in the same manner as we did for solving for the ML estimator: take the logarithm and minimize its negative value. However, the quadratic function tends to let the optimization concentrate too much on larger values than values near our model. Instead of minimizing the negative logarithm of the prior probability we can optimize for the plain function $P(\mathbf{x}|\mathbf{m}) = \exp(-\frac{1}{2}\lambda S)$ itself; the exponential form suppresses the contribution of large values appropriately.

Let us recapitulate in words what the likelihood term and the Bayesian prior do in the MAP formulation: The former selects (one or more) solutions that are most *likely* because they "match" the measured data; the latter gives a measure of most *probable* solutions such that solutions that conform to *a priori* given information are more probable than ones that deviate from a prior model.

6. Formulation of a Nullspace-MAP Algorithm

Our proposed algorithm for solving ill-posed problems using nullspace consists of two main steps:

1. Firstly, we determined the linear solution of the underdetermined system while applying sufficient regularization to leave out the effects of noise.
2. Secondly, the solution is corrected using nullspace vectors by maximizing the prior probability function such that $\mathbf{x}$ becomes as close as possible to a prior model $\mathbf{m}$.

There are many available methods to solve for the initial linear solution. For underdetermined systems we can, for example, explicitly form the Moore-Penrose inverse and solve the normal equation; we can use iterative techniques, or we can use *singular-value-decomposition* (SVD). The optimal choice depends very much on the sort and scale of the problem; for smaller problems we prefer SVD not only for its elegance but because it explicitly yields the orthogonal nullspace which we needed for our algorithm.

Maximization of $P(\mathbf{x}|\mathbf{m})$ must be done in an iterative scheme because this function is inherently nonlinear. We suggest ascending iteratively along the upward gradient.

If the linear inversion step is part of a nonlinear optimization scheme, such as a Newton-Raphson solver, then one can iterate the two steps after recalculating the Jacobian matrix. The data-error function E is now defined in terms of *difference* vectors $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$, i.e.

$$E = |\mathbf{W} \cdot (\Delta\mathbf{y} - \mathbf{A} \cdot \Delta\mathbf{x})|^2 \quad . \tag{5}$$

The model prior of Eq. 4 is unchanged in the nonlinear case, and should still compare the *absolute* image vector $\mathbf{x}$ to the model $\mathbf{m}$. Each new guess is obtained by updating the current guess with the minimizing value of Eq. 5.

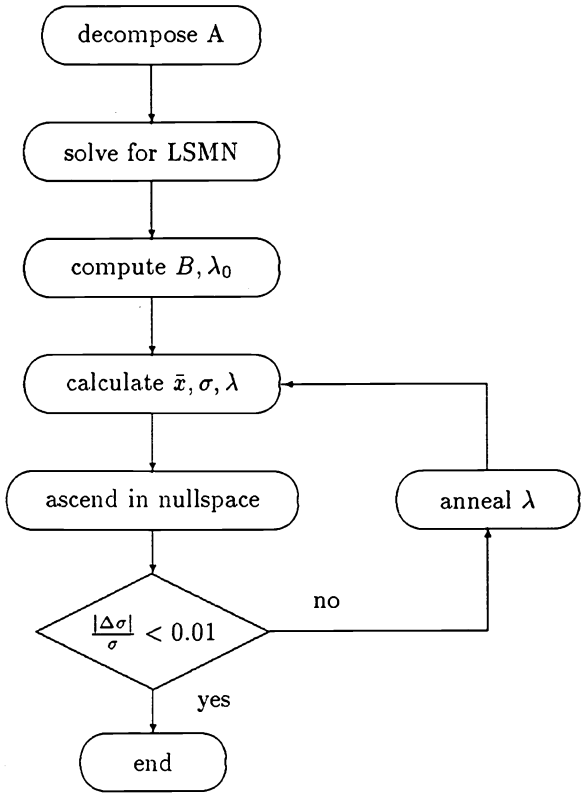

Figure 1: Flow diagram of the nullspace-MAP algorithm using SVD.

7. Implementation using SVD

Implementation of the nullspace-MAP algorithm is straightforward using SVD. The program takes data values $\mathbf{y}$ and the Jacobian $\mathbf{A}$ as input, and firstly computes the singular value decomposition of matrix $\mathbf{A}$ into $\mathbf{UDV}^T$. The rank is properly reduced and the regularized linear solution is computed. Most standard SVD routines—e.g. as given in[9]—give the nullspace readily as the last $N - M$ columns of $\mathbf{V}$; if not directly available, the nullspace can be computed explicitly by means of standard procedures such as Gramm-Schmidt orthogonalization. The nullspace is stored in a matrix $\mathbf{B}$, which rows contains the normalized complementary nullspace vectors.

In cases where an *a priori* model distribution is not available, we propose to use the *a posteriori* value of the median $\bar{x}$ and mean deviation σ of the intermediate solutions. Notice that this implicitly assumes that we expect the background to be reasonably flat around the median value.

Next, the program iteratively ascends towards the model prior. Each step is projected into nullspace with the projection operator $\mathbf{B}^T\mathbf{B}$; this ensures that the data error will not be tempered. The program terminates if a stopping condition—hereafter to be described—is met. A flow diagram of the nullspace-MAP algorithm is depicted in Fig. 1. Again, if the linear inversion step is a part of a nonlinear solver then we must iterate the algorithm. Some implementation details are given in the subsections below.

7.1. Determination of the effective rank

For the regularization using truncated-SVD we need to decide how many singular values we retain for forming the initial LSMN solution. If a (fast) exact forward model is available, iterative schemes, such as the Levenberg-Marquardt method, may be used for automatic parameter choice.

If a forward model is unavailable or unfavourable due to its computation requirements, we can refer to the "chi-square statistics for $N-M$ degrees of freedom". A rule of thumb is to proceed until the rest error $\chi^2 \approx N-M$ (see e.g., §15.1 in,[9] or [10]). The argument is rather *ad hoc* and we prefer to relate the required amount of regularization directly to the signal-to-noise ratio (SNR) of the measurements.

We can use Morozov's principle that relies on the observation that the estimated solution should not fit the data with an accuracy greater than the accuracy of the measurements.[11] Then the effective rank, R, is determined by solving,

$$\frac{\sum_{i=1}^{R} \mathbf{U}_i \cdot \mathbf{y}}{\sum_{i=R+1}^{M} \mathbf{U}_i \cdot \mathbf{y}} = \mathrm{SNR} \tag{6}$$

where the SNR is defined as the ratio of the average signal power and noise content in the measurements. $\mathbf{U}_i$ denotes the ith column of matrix $\mathbf{U}$ that contains the singular (dataspace) vectors. Singular values for $i > R$ are set to zero.

7.2. Statistical estimators $\bar{x}$ and σ

At each iteration we compute the median $\bar{x}$ and the *mean absolute deviation*, $\sigma = \frac{1}{N}\sum_{i=0}^{N} |x_i - \bar{x}|$, of the distribution. Both are needed to quantify the prior probability function in Eq. 4. We found that this choice was more robust than using the *average* and *variance* as statistical estimators.

7.3. Choice of the hyperparameter λ

Choice of the hyperparameter λ will determine the effective proportion of pixel values that are included in the nullspace correction process. A small value for λ will try to bring all pixel values towards the prior model $\mathbf{m}$,—here the median $\bar{x}$ of the image; a large value of λ will concentrate on values of $\mathbf{x}$ close to $\mathbf{m}$.

Tradionally, the hyperparameter λ is chosen rather *ad hoc* and the optimal value is often heuristically determined by visual comparison of several images.

However, we can obtain a good estimation for the maximum value of λ by considering a rationale similar to Morozov's principle: the estimated solution should not fit the model more than the number of nullspace vectors. Accordingly, we choose λ at each ascent step such that it affects only an average proportion of pixels equal to the proportion of the nullspace; i.e. using Eq. 4 we solve for λ_0 in,

$$\exp(-\frac{\lambda_0}{2}) = \frac{N-R}{N} \tag{7}$$

where R is the effective rank as determined from Eq. 6.

The estimation of λ_0 is rather an upperbound for the choice of λ and the optimal value depends much on the nature of the object. We found it helpful to "anneal" the hyperparameter gradually up to the value of λ_0, —in many cases the algorithm terminates before reaching this value.

7.4. Ascent in nullspace

The program ascends along the nullspace projected gradient of the Bayesian prior $P(\mathbf{x}|\mathbf{m})$, i.e

$$\begin{aligned} \Delta\mathbf{x} &= c_1\,\mathbf{B}^T\cdot\mathbf{B}\cdot\frac{\partial}{\partial\mathbf{x}}P(\mathbf{x}|\mathbf{m}) \\ &= -\ c_2\,\mathbf{B}^T\cdot\mathbf{B}\cdot\exp[-\frac{\lambda}{2\sigma^2}(\Delta\mathrm{m})^2]\,\Delta\mathbf{m} \end{aligned} \tag{8}$$

where the difference vector $\Delta\mathbf{m} = \mathbf{m} - \mathbf{x}$ represents the deviation of the current solution from the model. Notice the importance of the exponential weighting term that favours values close to the model and inhibits large values from changing; the weighting term would not have appeared if we would have used $\frac{\lambda}{2\sigma^2}(\mathbf{m}-\mathbf{x})^2$ rather that $\exp[-\frac{\lambda}{2\sigma^2}(\mathbf{m}-\mathbf{x})^2]$ as objective function. The *step length* is controlled by choosing $0 < c_2 \leq 1$; heuristically we found $c_2 \approx 0.6$ a good choice.

7.5. Program control and stopping condition

At each nullspace iteration, we recompute the median $\tilde{x}$ and average deviation σ. As previously mentioned, we apply an "annealing"-schedule that gradually increases λ towards λ_0, e.g., every 5%.

The algorithm is not complete unless we define a stopping condition. During proper execution σ will decrease monotonously. We terminate the program when the relative change of σ becomes less than a pre-defined value or reaches a maximum number of iterations. Heuristically we found that $|\Delta\sigma|/\sigma < 0.01$ is not difficult to reach within 20 iterations for most of our problems.

8. Reconstruction Examples

To illustrate our concept we will give two examples of optical tomographic reconstruction: one using simulated data, and one using experimental data.

For the reconstructions we have used the program `MMTOAST10` from the Medical Physics group at UCL, in which we have modified the central inversion step with our SVD-based nullspace-MAP algorithm. `MMTOAST10` is basically a Newton-Raphson (NR) nonlinear solver which recalculates the Jacobian by a finite-element-method (for details see[12,13]). Regularization of the linear SVD solution is automatically controlled using a Levenberg-Marquardt scheme.

8.1. Using simulated data

Our first example is that of "object1" as described in.[13] "Object1" is a pure-absorption case with two small objects (~3mm) of high contrast with absorption coefficient $\mu_a = 0.5\mathrm{mm}^{-1}$ in a homogeneous cylinder with

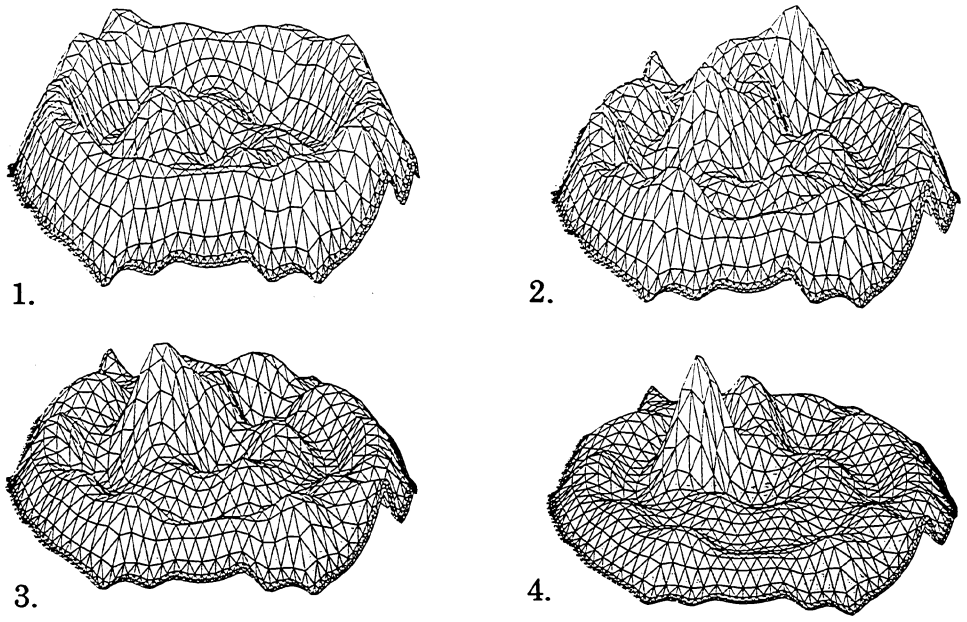

Figure 2: LSMN solution using SVD solver for the first 4 NR-iterations (simulated data). Starting error: 237.56; error residuals, respectively, 11.70, 0.23, 0.14 and 0.020. All normalized. Maximum value, respectively, 0.033, 0.050, 0.060 and 0.085 mm^{-1}

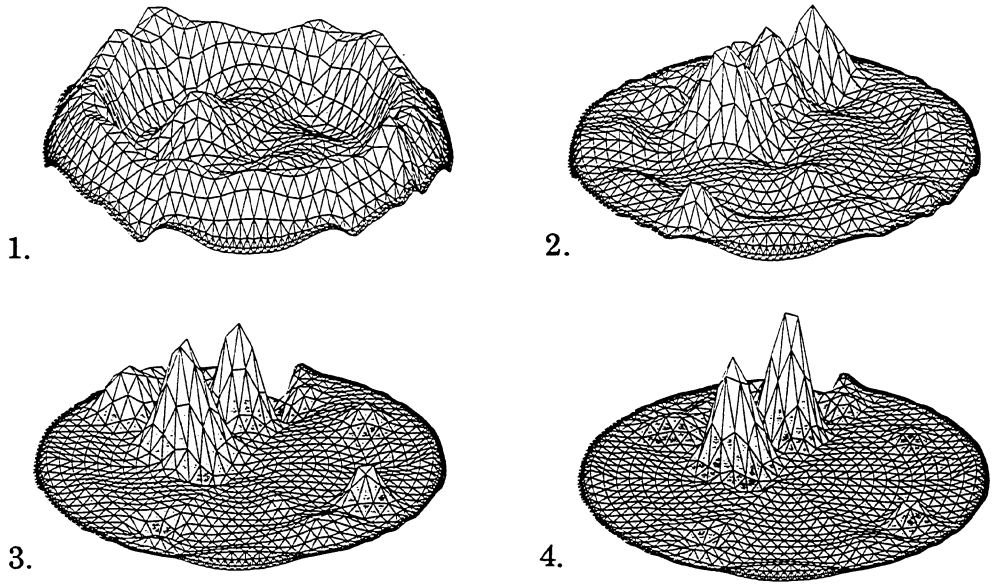

Figure 3: MAP solution using nullspace-MAP for the first 4 NR-iterations (simulated data). Starting error: 237.56; error residuals, respectively, 10.02, 0.78, 0.40 and 0.21. All normalized. Maximum value, respectively, 0.034, 0.047, 0.094 and 0.105 mm^{-1}

diameter 50mm. Homogeneous background values are: $\mu_a = 0.025\text{mm}^{-1}$ and scattering coefficient $\mu_s' = 2.0\text{mm}^{-1}$. In contrast to the original example we have diminished the number of measurements to 64 (8 sources, 8 detectors) and increased the grid resolution to 24×24 with 484 pixels contributing to the Jacobian; the problem is now strongly underdetermined. Simulated data are created by forward calculation of the photon densities; we have only used *mean time-of-flight* information for the reconstruction. No noise was added. The reconstruction was started from homogeneous values $\mu_a = 0.010\text{mm}^{-1}$ and $\mu_s' = 2.0\text{mm}^{-1}$.

Figure 2 shows the "raw" LSMN-solution obtained by `MMTOAST10`'s standard SVD iterative solver, and Fig. 3 the result obtained by using our algorithm. Notice how our nullspace-MAP algorithm flattens the background values correctly in favour of the absolute values of the peaks.

8.2. Using experimental data

The experimental data is obtained using a 68mm diameter highly scattering solid epoxy-resin cylinder with a 20mm diameter removable absorbing rod inside. The phantom is irradiated with NIR-light at 4 successive source positions and the scattering intensity is measured around the circumference at 16 fiber positions; i.e. we have a total of $4 \times 16 = 64$ measurements. The reconstruction is performed on a circularly cropped 24×24 image grid, which totaled a number of 484 pixels. Note that this problem is again highly underdetermined.

Each measurement is an intensity difference of the phantom measured with the absorbing rod and that with a neutral rod. In this example the optical properties of the background are presumed to be known: reduced scattering coefficient $\mu_s' = 0.52\text{mm}^{-1}$ and absorption coefficient $\mu_a = 0.0042\text{mm}^{-1}$. The absorption coefficient of the absorbing rod was measured to be around 0.017mm^{-1}.

Fig. 4 shows the final image using our nullspace-MAP after 5 NR-iterations. For comparison we also show the expected ideal image of the object, and the LSMN solution using `MMTOAST10`'s standard ART-like iterative solver (also 5 NR-iterations). The total time for both reconstructions was around 20 minutes on a Sun Ultra 1 workstation.

For this case of a relatively large inhomogeneity, the standard solver seems to perform quite well. Using the value of λ_0 as in 7, the nullspace-MAP solution underestimates the size of the inhomogeneity and consequently overestimates the absorption values. Fig 5 shows how the final nullspace-MAP solution alters if we manually adjust the hyperparameter and take a smaller value.

9. Discussion and Conclusions

We have shown that use of nullspace vectors is not only helpful but a prerequisite for correct inversion of ill-posed problems. In fact we would like to argue that *any regularization* should be restricted in to nullspace. The use of the data-error function, which measures the discrepancy between the measurements and the simulated response of our guess, is correct in well-posed problems but insufficient for underdetermined systems. Bayesian theory gives a framework to introduce a secondary objective function that measures the closeness/misfit of our guess from a prior model. An advantage of explicit separation of the solution space in measurement– and nullspace is that we can minimize for the data-error and model-error (or prior probabiltiy) functions independently.

In case a prior model distribution is not available, we have shown that we can use the *median* value as the Bayesian prior; it implicitly assumes a normal distribution around a flat constant value. For images which show objects that have strong limited support this should perform well, but for extended and more complex objects a "local" median value could perform better as model. It seems that our nullspace-MAP algorithm inherently underestimates larger objects because it tries to "adjust" pixel values as much as possible to the prior model

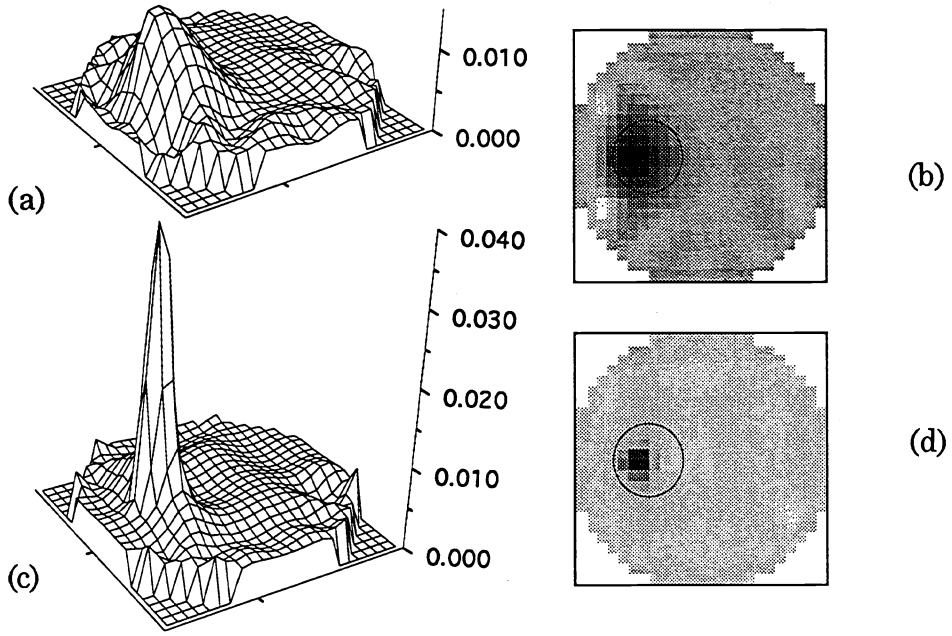

Figure 4: Experimental images. (a) and (b): LSMN solution using MMTOAST10's ART-like solver after 5 NR iterations; (c) and (d): MAP solution using the nullspace-MAP algorithm. The diameter of the original object is superposed for comparison.

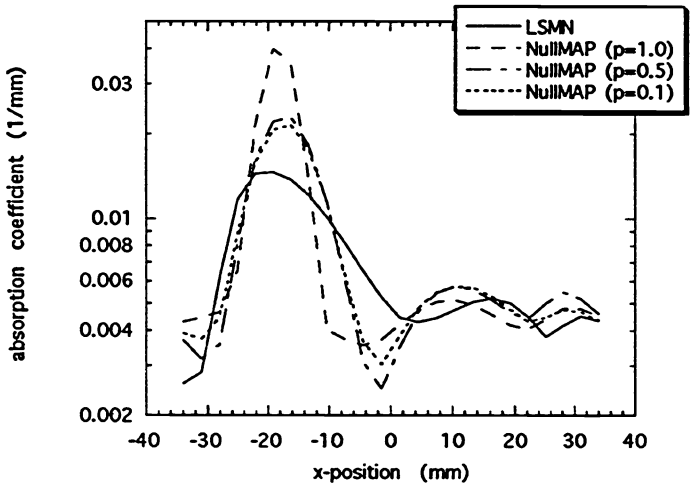

Figure 5: Plot of the cross-section through the inhomogeneity versus hyperparameter using values λ_0, $0.5\lambda_0$ (p=0.5), $0.1\lambda_0$ (p=0.1). The LSMN solution is also given for comparison.

value. However, if an upperbound of the solution might be available (in our case this might correspond to 100% oxygen saturation of the blood) then we can try to "bring down" excessive values,—but again *only* in nullspace.

Further investigation is needed for determining the optimal choice of the hyperparameter, λ, as this significantly affects the effectiveness of the algorithm. We have suggested that "annealing" can help the algorithm to find the optimal amount of nullspace correction. Improvement might be gained by an optimal annealing schedule, possibly including some statistical features —such as in *simulated annealing*—that might help the algorithm to escape from local minima.

A drawback of the current SVD-based algorithm is that the nullspace has to be computed explicitly and that decomposition of larger matrices requires a significant amount of computation time. We are currently considering a iterative methods that find the MAP solution without explicit calculation of the nullspace.

Acknowledgements

We thank Martin Schweiger from UCL for his excellent programming of the FEM library. This research has been performed while I.K. was a research fellow under a grant from the Science & Technology Agency in Japan.

10. REFERENCES

[1] J. E. Brazy, D. V. Lewis, Mitnick M. H., F. F. Jobsis, and van der Vliet, "Noninvasive monitoring of cerebral oxygenation in preterm infants: Preliminary observations", *Paediatrics*, , no. 75, pp. 217–225, 1985.

[2] J. S. Wyatt, M. Cope, D. T. Delpy, S. Wray, and E. O. R. Reynolds, "Quantitation of cerebral oxygenation and haemaodynamics in sick newborn infants by near infrared spectroscopy", *Lancet*, vol. 2, pp. 1063–1066, 1986.

[3] A. D. Edwards, J. S. Wyatt, C. E. Richardson, D. T. Delpy, M. Cope, and E. O. R. Reynolds, "Cotside measurement of cerebral blood flow in ill newborn infants by near infrared spectroscopy", *Lancet*, vol. 2, pp. 770–771, 1988.

[4] Jeremy C. Hebden, David J. Hall, Michael Firbank, and David T. Delpy, "Time-resolved optical imaging of a solid tissue-equivalent phantom", *Appl. Opt.*, vol. 34, no. 34, pp. 8038–8047, December 1995.

[5] James J. Duderstadt and Louis J. Hamilton, *Nuclear Reactor Analysis*, John Wiley & Sons, New York, 1976, 103–232.

[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York, 1953.

[7] S. R. Arridge, P. van der Zee, M. Cope, and D. T. Delpy, "Reconstruction methods for infra-red absorption imaging", *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1431, pp. 204–215, 1991.

[8] Adi Ben-Israel and Thomas N.E. Greville, *Generalized Inverses: Theory and Applications*, "Pure and Applied Mathematics". John Wiley & Sons, New York, 1973.

[9] W. H. Press and et al., *Numerical Recipes in C: The art of scientific computing*, Cambridge University Press, New York, 2nd edition, 1992.

[10] S. F. Gull, "Developments in maximum entropy data analysis", in *"Maximum Entropy and Bayesian Methods"*, J. Skilling, Ed., pp. 53–71. Kluwer Academic, 1989.

[11] M. Defrise, "Possible criteria for choosing the number of iterations in some iterative reconstruction methods", in *"Mathematics and Computer Science in Medical Imaging"*, M. A. Viergever and A. Todd-Pokropek, Eds., pp. 293–303. Springer-Verlag, Berlin-Heidelberg, 1988.

[12] Simon R. Arridge and M. Schweiger, "Sensitivity to prior knowledge in optical tomographic reconstruction", *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 2389, pp. 378–388, 1995.

[13] M. Schweiger, S. R. Arridge, and D. T. Delpy, "Application of the finite-element-method for the forward and inverse models in optical tomography", *Journ. Math. Imag. Vision*, vol. 3, pp. 263–283, 1993.