Combining singular value decomposition and a non-negative constraint in a hybrid method for photothermal depth profiling

Wim Verkruysse
Beckman Laser Institute, University of California, Irvine, California 92612

Boris Majaron
Jožef Stefan Institute, Jamova 39, S1-1000, Ljubljana, Slovenia

Bernard Choi and J. Stuart Nelson
Beckman Laser Institute, University of California, Irvine, California 92612

(Received 19 July 2004; accepted 6 December 2004; published online 10 January 2005)

We present a method to solve the inverse problem in pulsed photothermal radiometry (PPTR) that exploits advantages of truncated singular value decomposition (T-SVD) while imposing a non-negativity constraint to the solution. The presented method is a hybrid in the sense that it expresses the solution vector as a linear superposition of right singular vectors, but with a non-negative constraint applied to it. The weights for the superposition are determined using an optimization algorithm. In one-dimensional PPTR simulation examples, the best reconstruction results are of comparable quality to those of the conjugate gradient method. Furthermore, the hybrid method exhibits a sharper knee in the L-curve and small susceptibility to over-iteration in presence of experimental noise, thus facilitating the regularization process. As a result, the reconstructed temperature profiles are more likely to be closer to the original initial profiles. [DOI: 10.1063/1.1851473]

I. INTRODUCTION

Pulsed photothermal radiometry (PPTR) is a noncontact technique that utilizes an infrared detector to measure the time dependent temperature increase in a test material.[1,2] Typically, the material is irradiated with pulsed laser light, which is selectively absorbed in subsurface chromophores. From the temporal behavior of the measured surface radiative emission, the laser-induced subsurface temperature profile can be reconstructed. PPTR has been used to determine the depth of chromophores in human skin such as epidermal melanin and hemoglobin in the vasculature of port wine stain (PWS) birthmarks.[3,4]

Determination of the spatial temperature distribution from the measured radiometric signals constitutes a severely ill-posed inverse problem which explains why a large number of inversion methods has been investigated. The earliest method to solve the PPTR inverse problem on human skin used a stepwise (layer by layer) least squares fitting approach.[3] Subsequently applied methods include least-squares truncated singular value decomposition (T-SVD),[4] Levenberg–Marquardt,[5,6] and conjugate gradient (CG) methods.[4] More recently, Xiao and Imhof[7] combined T-SVD with the maximum entropy method (MEM) and later just MEM,[8] to solve the inverse problem. The same group also demonstrated the application of a neural network as an alternative method.[9]

A thorough comparison of these methods for PPTR has not been done so it is difficult to rank their quality. However, several aspects that are important for an inversion method can be identified. First, the method must be able to implement a non-negativity constraint to the solution vector. This improves stability of the iteration and the ultimate solution is more accurate.[4] Second, due to severe ill-posedness of the PPTR inverse problem, regularization is required to avoid over-iteration (fitting of noise rather than real signal) which leads to unrealistic steep gradients in the predicted temperature profiles.

The CG method, currently used in our group, works directly with observable quantities (initial temperature rises at various depths), so a non-negative constraint is relatively straightforward to implement (NN-CG). Regularization can be done through "early termination" based on "L-curve" analysis.[10,11] However, in presence of experimental noise, the "knee" in the L-curve is often not very sharp, so the selected degree of regularization remains somewhat arbitrary. In contrast, T-SVD permits a more direct control over the degree of regularization (i.e., truncation parameter). In this approach, the solution only contains those spatial frequencies permitted by the appropriately truncated set of singular vectors. However, the solutions frequently contain unphysical negative temperatures since implementation of a non-negative constraint is inherently not possible. That latter was an important reason for Milner *et al.*[4] to prefer NN-CG over T-SVD for PPTR depth profiling of human skin.

In this paper, we demonstrate and test a novel method, which combines the advantages of T-SVD with a non-negative constraint to solve the PPTR inverse problem.

II. THEORETICAL BACKGROUND

A. Constituting equations for PPTR profiling

In PPTR, the temperature field evolution $\Delta T(z,t)$ can be expressed in terms of the quasi-instantaneous laser-induced temperature profile $\Delta T(z,t=0)$, and the Green's function $G(z,z',t)$:

$$\Delta T(z,t) = \int_0^\infty G(z,z',t)\Delta T(z',0)dz'. \tag{1}$$

For the Green's function, we use the expression derived by Milner *et al.*[4] for a semi-infinite medium with a radiative/convective heat transfer at the air-tissue interface. The Green's function involves the thermal diffusivity ($D=1.1\times10^{-7}$ m^2 s^{-1}),[12] and the reduced heat transfer coefficient ($h=20$ m^{-1}).[13] Because the temperature increase due to the laser pulse is much smaller than the initial background temperature, the increase in radiometric signal $\Delta S(t)$ can be expressed as

$$\Delta S(t) = C\mu_{\mathrm{IR}}\int_{z=0}^\infty \Delta T(z,t)e^{-\mu_{\mathrm{IR}}z}dz. \tag{2}$$

Equation (2) reflects that the signals are collected from all depths z, attenuated according to Beer's law with the absorption coefficient μ_{IR} (2×10^4 m^{-1}, representative of tissue at wavelengths between 4–5 μm). The constant C accounts for optical properties of the sample surface and detection system specifics such as integration time and detectivity across the detector bandwidth.

Substitution of Eq. (1) into Eq. (2) results in a double integral which can be conveniently rewritten as

$$\Delta S(t) = \int_{z=0}^\infty \Delta T(z,0)K(z,t)dz \tag{3}$$

expressing $\Delta S(t)$ in terms of the initial temperature profile $\Delta T(z)$ and a kernel function $K(z,t)$. The explicit expression and guidelines to compute the function values $K(z,t)$ are provided in the Appendix. For a more detailed derivation of the above equations and underlying assumptions we refer to other papers.[4,13]

B. Singular value decomposition

The discrete form of Eq. (3) expresses the measured signal $\Delta S(t)$ and the initial temperature profile $\Delta T(z)$ as a known vector $\boldsymbol{S}$ of length n and unknown vector $\boldsymbol{T}$ of length m, respectively, while the kernel $\mathbf{K}$ is a n by m matrix with elements $K(z,t)\ \Delta z$:

$$\boldsymbol{S} = \mathbf{K}\cdot\boldsymbol{T}. \tag{4}$$

When m is equal to n, $\mathbf{K}$ is a square matrix and can be inverted. In theory, $\boldsymbol{T}$ may be calculated directly as $\boldsymbol{T}=\mathbf{K}^{-1}\boldsymbol{S}$. However, due to the nature of the kernel function, $\mathbf{K}$ is nearly singular (has a defect rank). Hence, $\mathbf{K}^{-1}$ is practically incalculable. Moreover, the result of the inversion is often useless because even the slightest noise on the measured signal $\boldsymbol{S}$ results in very large oscillations in the solution vector $\boldsymbol{T}$. When an inverse problem involves a near singular matrix $\mathbf{K}$, singular value decomposition (SVD) can be employed to rewrite $\mathbf{K}$ as

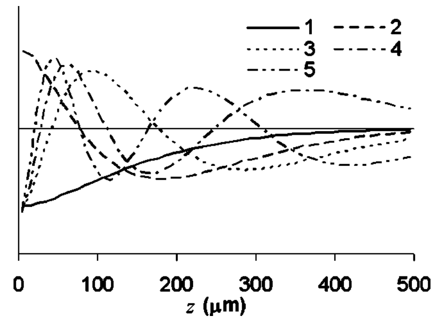

FIG. 1. The five lowest order right singular vectors calculated for the kernel matrix $\mathbf{K}$ used in this study.

$$\mathbf{K} = \sum_{i=1}^m \sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^T, \tag{5}$$

where we assume that the rank of $\mathbf{K}$ equals m, as in the case of PPTR for $m\leqslant n$. The vectors $\boldsymbol{u}_i$ are the left singular vectors which are of the same length as $\boldsymbol{S}$. The scalars σ_i are corresponding singular values. The corresponding right singular vectors $\boldsymbol{v}_i$ are orthonormal and of the same length as vector $\boldsymbol{T}$. The vectors with higher subscript i contain higher spatial frequencies. The five lowest order vectors $\boldsymbol{v}_i$ are shown in Fig. 1.

Truncated SVD of the kernel matrix $\mathbf{K}$ is a useful tool for inversion of PPTR signals. The solution estimate $\boldsymbol{T}^{(p)}$ can be written as a linear superposition of p right singular vectors:

$$\boldsymbol{T}^{(p)} = \sum_{i=1}^p c_i\boldsymbol{v}_i, \tag{6}$$

where c_1 to c_p are weights corresponding to the vectors $\boldsymbol{v}_1$ to $\boldsymbol{v}_p$.

The weights can be determined either analytically through $c_i=\boldsymbol{u}_i^T\boldsymbol{S}/\sigma_i$, (e.g. Ref. 7) or by minimizing the Euclidian norm of the residual $f(c_1,c_2,c_3,\ldots,c_p)$ [Eq. (7)] using an optimization algorithm,[4]

$$f(c_1,c_2,c_3,\ldots,c_p) = \|\mathbf{K}\boldsymbol{T}^{(p)} - \boldsymbol{S}\|^2. \tag{7}$$

An advantage of T-SVD is that the truncation parameter p (the number of singular vectors used in the superposition) can be used as a robust regularization parameter. The "discrepancy principle"[14] relates the signal to noise ratio (SNR) of the PPTR signal $S(t)$ to the maximum value of p. Inclusion of higher-order vectors usually results in loss of convergence and temperature profiles with unphysical oscillating features. In practice, it may be difficult to determine the SNR and, consequently, the optimal value of p. A practical approach to determine p is by plotting an L-curve, the points of which represent the solution norm versus the residual norm obtained by gradually increasing the number of singular vectors. The "knee" of such a curve is typically sharp and indicates the optimal value for p (e.g., Refs. 4, 11, and 14).

III. METHOD

A. SVD based non-negative optimization

As mentioned before, the T-SVD method can predict negative values for $\boldsymbol{T}$, which is unrealistic for PPTR. Below, we present a method which also minimizes the residual, as in Eq. (7), but with a modified expression for $\boldsymbol{T}^{(p)}$.

We define $\boldsymbol{T}_{\mathrm{NN}}^{(p)}$ as a linear superposition of right singular vectors [as in Eq. (6)] and then apply the non-negativity constraint by setting the negative vector elements $\boldsymbol{T}_{\mathrm{NN}}^{(p)}(q)$ to 0:

$$\boldsymbol{T}_{\mathrm{NN}}^{(p)}(q) = \max\left\{ \sum_{i=1}^{p} c_i \boldsymbol{v}_i(q), 0 \right\}. \tag{8}$$

An optimization algorithm is used to find weights c_i that minimize the residual norm:

$$f_{\mathrm{NN}}(c_1, c_2, c_3, \ldots, c_p) = \|\mathbf{K}\boldsymbol{T}_{\mathrm{NN}}^{(p)} - \boldsymbol{S}\|^2. \tag{9}$$

Note that when one or more elements of $\boldsymbol{T}_{\mathrm{NN}}^{(p)}$ are modified (set to 0), $\boldsymbol{T}_{\mathrm{NN}}^{(p)}$ is no longer a linear superposition of right singular vectors. Therefore, the non-negativity constraint (8) constitutes a departure from SVD theory. Nevertheless, using the solution ansatz $\boldsymbol{T}_{\mathrm{NN}}^{(p)}$, (8) ensures that high order vector components, which cause instabilities in the solution, will not be present. We will refer to the presented method as SVD based non-negative optimization (SVD-NN).

As noted above, the SVD-NN method requires an algorithm to optimize the weights c_i. We apply a nonlinear optimization algorithm which is commercially implemented by Frontline Systems Inc. (Incline Village, NV) and available as add-in ("Solver") in Microsoft Excel®. This algorithm uses the generalized reduced gradient (GRG) method as implemented in an enhanced version of Lasdon and Waren's GRG2 code.[15–19]

Regularization of the solution with SVD-NN is similar as in T-SVD. The truncation parameter p, which limits the singular vector components and spatial frequencies allowed in the solution, is gradually increased until the iteration gets out of control.

B. Comparison of SVD-NN with NN-CG

We compare the SVD-NN approach to the more established CG algorithm, directly minimizing the residual norm $\|\mathbf{K}\boldsymbol{T} - \boldsymbol{S}\|$ by optimizing $\boldsymbol{T}$. We use a custom implementation (Fortran) of the CG developed earlier,[4,20] which includes the non-negative constraint to the solution.

To compare the methods, we used a simple initial temperature profile $\boldsymbol{T}_{orig}$ (see Fig. 2) and simulated a corresponding PPTR signal $\boldsymbol{S}$. The two layers in $\boldsymbol{T}_{orig}$, in which temperature is increased, represent a superficial absorbing layer (epidermis) and a deeper PWS blood vessel layer. For all calculations presented herein, the total time for $\boldsymbol{S}$ is 200 ms, discretized in 200 steps ($\Delta t = 1$ ms) and the total depth for $\boldsymbol{T}$ is 500 μm, discretized in 50 steps ($\Delta z = 10$ μm).

First we investigate if SVD-NN reconstructed profiles are similar to those obtained with NN-CG. For this comparison we use the simulated signal $\boldsymbol{S}$ with no additional noise. Next, we add Gaussian noise to $\boldsymbol{S}$, equivalent to SNR=200, and investigate the robustness of SVD-NN in comparison with NN-CG. For each method we determine the reconstructed profile $\boldsymbol{T}$ that matches $\boldsymbol{T}_{orig}$ best. To quantify the match, we compute the Euclidian norm of the difference between the reconstructed and original temperature profiles: $\|\boldsymbol{T} - \boldsymbol{T}_{orig}\|$ and define $\boldsymbol{T}_{optimal}$ as the temperature profile with the smallest $\|\boldsymbol{T} - \boldsymbol{T}_{orig}\|$. In parallel, the L-curves will be analyzed to determine the optimal degree of regularization without knowledge of the actual profile. Finally, we consider the sensitivity of the solution with respect to over- and under-regularization.

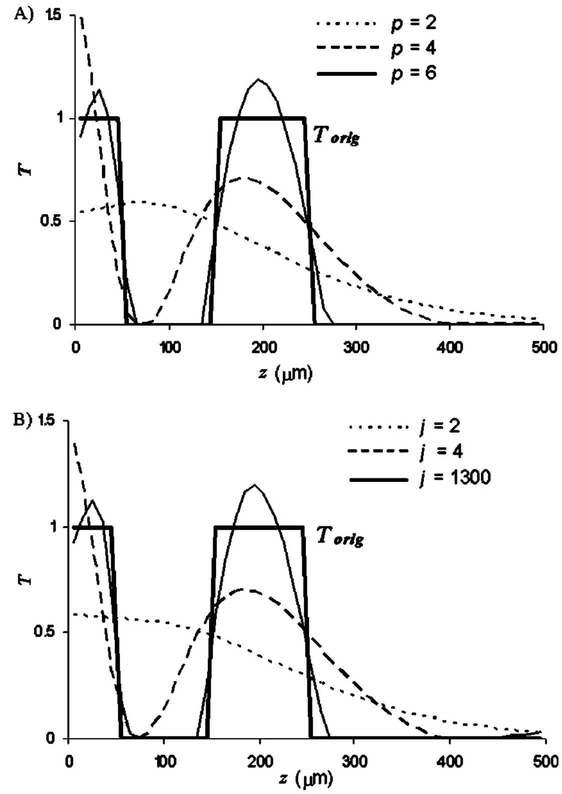

FIG. 2. (A) The original temperature profile $\boldsymbol{T}_{orig}$ and reconstructed profiles with SVD-NN method at $p=2$, 4, and 6 (indicated in chart) and (B) profiles reconstructed with NN-CG after iterations $j=2$, 4, and 1300.

IV. RESULTS

The original temperature profile $\boldsymbol{T}_{orig}$ and 3 reconstructed profiles obtained with SVD-NN (at $p=2$, 4, and 6) are shown in Fig. 2(A). These profiles were obtained starting the iteration at $p=1$ (optimizing the weight c_1). Subsequently, we added the second singular vector ($p=2$) and optimized c_1 and c_2 simultaneously, using the previously obtained value for c_1 (and $c_2=0$) as the initial values. We repeated this process until $p=6$, each time using the weight values obtained for the previous p as initial values. At each value of p, the iteration was automatically stopped when a pre-set convergence criterion (i.e., the relative improvement of the residual norm below 0.002) was achieved. The number of iterations required for convergence, and the corresponding residual norm values are listed in Table I. We also list the number of iterations required with NN-CG to reach the same residual norm values. Figure 2(B) shows the corresponding

TABLE I. The number of iterations (and total iterations) required to reach a preset convergence criterion with the SVD-NN method at each p-value, corresponding residual norm values and number of iterations required with the NN-CG method.

SVD-NN		NN-CG	
p	Iterations (total)	Iterations	$\|\mathbf{K}\boldsymbol{T}-\boldsymbol{S}\|$
1	1 (1)	1	0.366291
2	4 (5)	2	0.191721
3	7 (12)	3	0.068128
4	12 (24)	4	0.018759
5	62 (86)	300	0.000306
6	21 (107)	1300	7.6E−05

profiles for NN-CG. Comparing Figs. 2(A) and 2(B), it can be noted that the solutions obtained using both methods are very similar.

Figure 3 shows the residual norm $\|\mathbf{K}\boldsymbol{T}-\boldsymbol{S}\|$ plotted versus the total number of iterations. Initially, the SVD-NN method requires more iterations than NN-CG to reach the same residual norm values (>0.015). The opposite is true, however, when the optimization advances and solutions include higher spatial frequencies. With NN-CG, the decrease of the residual norm temporarily stalls at several stages in the iteration process, causing plateaus in the curve at, for example, residual norms of approximately 2×10^{-2} and 4×10^{-4}. Interestingly, these levels roughly coincide with those for $p=4$ and $p=5$ of the SVD-NN curve, indicating further similarity between the two methods.

Next, we compare both methods using the same simulated signal $\boldsymbol{S}$ with Gaussian noise added (SNR=200). As a result, once the residual norm reaches a value, corresponding to the noise level it is only marginally reduced by consecutive iterations (for NN-CG) or increase of p (for SVD-NN)—while the norm of the corresponding solutions $\|\boldsymbol{T}\|$ begins to increase. These effects combined form the knee in the corresponding L-curves. In Fig. 4(A), we show the L-curves for the NN-CG method and two different approaches for the SVD-NN optimization. The first approach (SVD-NN1) is the same as used for Figs. 2 and 3. In the second approach

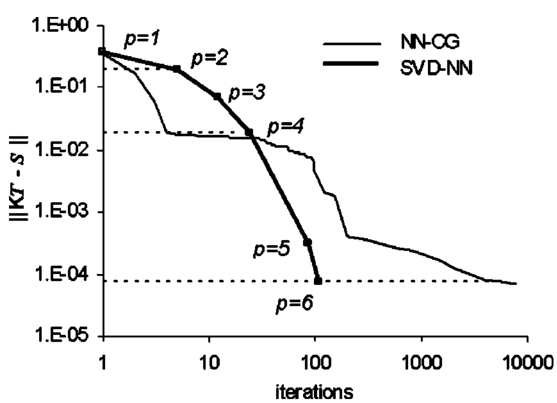

FIG. 3. The solution residual norm plotted versus the total number of iterations for the NN-CG and SVD-NN, respectively. The dashed lines are provided as a guide to help find the number of iterations required with the NN-CG to obtain the same solution residual norm level obtained for $p=2$, 4, and 6 with the SVD-NN method.

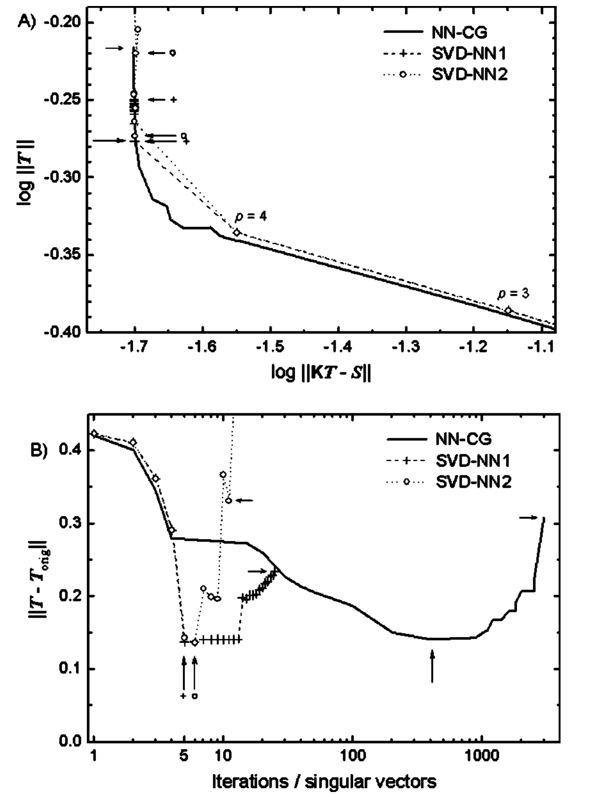

FIG. 4. The L-curves are plotted for NN-CG, and 2 different iteration strategies with the SVD-NN method. The arrows pointing towards the right indicate the iteration stages at which $\boldsymbol{T}_{optimal}$ is obtained. The three arrows pointing towards the left indicate iteration stages at which over-iteration resulted in some artifacts in $\boldsymbol{T}$. (B) shows $\|\boldsymbol{T}-\boldsymbol{T}_{\mathrm{orig}}\|$ versus the number of included singular vectors p for the two SVD-NN methods and the number of iterations for the NN-CG method. Vertical arrows indicate iteration stages at which $\boldsymbol{T}_{optimal}$ is achieved and horizontal arrows indicate overiterated stages.

(SVD-NN2), each time p is increased the initial values for all weights c_i to c_p are set to zero. In both approaches, we used p values of 1–25, providing 25 points for the L-curve.

The L-curve for the NN-CG shows a smoother knee, because the residual norm $\|\mathbf{K}\boldsymbol{T}-\boldsymbol{S}\|$ *decreases* gradually and the solution norm *increases* gradually as the iteration progresses through a large number of iteration steps. With both SVD-NN approaches, in contrast, the knees of the L-curves are sharper, indicating the optimal truncation parameter of $p=5$–6 [Fig. 4(A)].

In Fig. 4(B) we plot $\|\boldsymbol{T}-\boldsymbol{T}_{orig}\|$ versus the number of iterations for NN-CG and versus p for SVD-NN. We identify the number of iterations for NN-CG and the truncation parameter p for SVD-NN at which $\boldsymbol{T}_{optimal}$ is reached (longer, vertical arrows). These stages in the optimization procedure are also indicated in Fig. 4(A) (by the longer arrows). For SVD-NN, these points coincide with the sharp knee in the L-curve. This observation is important, because in practice, when the original object is not known, the curves as in Fig. 4(B) cannot be generated. Nevertheless, if the knee in the L-curve is distinct, the corresponding temperature profile $\boldsymbol{T}$ is likely to be close to $\boldsymbol{T}_{optimal}$.

The optimal reconstructed temperature profiles $\boldsymbol{T}_{optimal}$ for each method are shown in Fig. 5(A). To illustrate the effect of continued iteration we plot one arbitrarily chosen

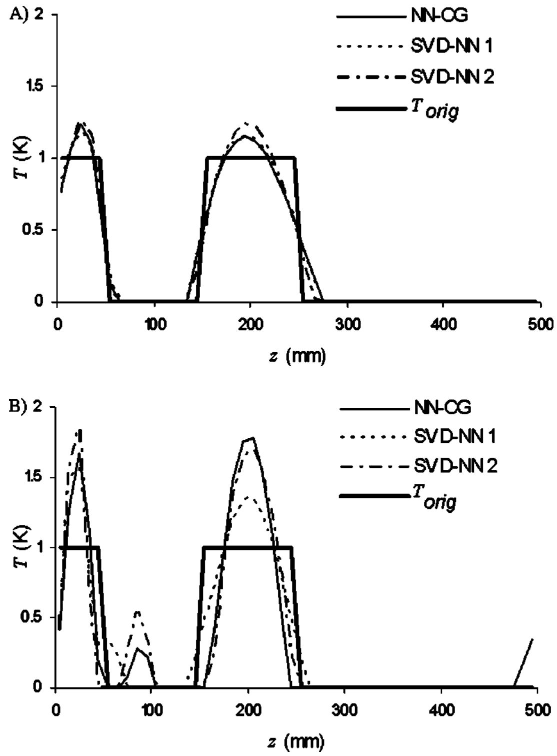

FIG. 5. (A) Temperature profiles $\boldsymbol{T}_{optimal}$, identified in Fig. 4(B), for each method and (B) examples of over-iterated profiles.

profile for each method in Fig. 5(B). The iteration stages for these three profiles are indicated in Figs. 4(A) and 4(B) by the shorter arrows.

V. DISCUSSION

To the best of our knowledge, the combination of a non-negative constraint with singular vectors with optimization of the SVD weights has not been reported before. We would like to emphasize that this hybrid approach is not a rigorous SVD method, but it exploits some important characteristics of T-SVD. The presented method takes advantage of the fact that singular vectors are ordered by relevance (i.e., the size of σ_i), which allows for a straightforward regularization through truncation, leading to effective rejection of noise in the inversion process.

In this study, we found that very similar reconstruction profiles can be achieved using different methods, requiring different numbers of iteration steps. For example, the profile in Fig. 2(A) (residual norm value of 7×10^{-7}) is achieved with 107 iterations at 6 optimization variables (p_1-p_6) with the SVD-NN method. Virtually the same profile is obtained with the NN-CG method involving 50 optimization variables (vector elements $\boldsymbol{T}_i$) after 1300 iteration steps [Fig. 2(B)]. Similarities or equivalence between optimization methods such as CG, T-SVD and regularization methods such as "Tikhonov"[21] have been noted before.[22] However, it should be noted that implementation of the NN constraint in a way violates the assumptions of the underlying theory not only for T-SVD but also for the CG algorithm (and calls for nontrivial augmentation of the latter). In that sense, it is reassuring that both constrained methods under test yield same, well behaved solutions. A rigorous investigation of these relations is beyond the scope of this paper.

Note that the total reconstruction time is also dependent on the computation time required per iteration step. The FORTRAN implementation of CG algorithm running on a PC platform, is much faster (<2 s for 1000 iterations) than the SVD-NN method with Solver as an optimizer (~15 s for 100 iterations). This may be an unfair comparison because both the optimization algorithms and implementations are different. In preliminary experiments, we compared our SVD-NN method with straightforward optimization of 50 elements of the vector $\boldsymbol{T}$, both using the same optimizing algorithm (Solver, using CG algorithm). In this case, SVD-NN was about 3 times faster per iteration step and more than 100 times faster to obtain solutions corresponding to $p=5$ or higher (see Fig. 3). However, in a Matlab® implementation using the nonlinear optimizing function "lsqcurvefit" (choosing the subspace trust region method based on interior-reflective Newton method[23,24] over the Levenberg–Marquardt[6,25,26] and Gauss–Newton[27] line search methods), we found that both approaches took about 1 s to obtain solutions corresponding to $p=5$ or higher. It should also be noted that the SVD-NN method involves the numerical decomposition of the kernel matrix $\mathbf{K}$. For $n=200$ and $m=50$, this time is negligible (<20 ms) in comparison to the iteration time. Moreover, for a given n, m, Δt, and Δz the decomposition must be done only once, after which the right singular vectors can be stored and reused for subsequent PPTR profile reconstructions. A more systematic investigation would be necessary to conclude on possible advantages of SVD-NN over straightforward optimization with regard to computational speed.

The L-curve characteristics of SVD-NN [Fig. 4(A)] are very similar to those shown by Milner *et al.*[4] for unconstrained T-SVD. The sharp knee, a direct result of using the number of singular vectors allowed in the solution (p) as the regularization parameter, permits a more objective and conclusive regularization than the smoother knee obtained with NN-CG regularized by early termination. In the latter approach, namely, higher spatial frequencies are introduced into the solution more gradually, over a larger number of iteration steps. SVD-NN also allows a more direct control over the regularization degree: The number of singular vectors allowed in the solution can be linked to the SNR via the discrepancy principle.

We considered two different iteration strategies for SVD-NN and found a significant difference between the two in terms of susceptibility to over-iteration [Fig. 4(B)]. The robustness of the SVD-NN1 approach arises from the fact that the most significant weights c_i are always biased by the values determined in the preceding step, which is by definition less ill-posed.

In summary, the hybrid SVD-NN method combines reconstruction stability and quality of the non-negativity constraint with more direct control over the regularization degree (from T-SVD). The observed quality of reconstructions and robustness with respect to over-iteration make it promising for PPTR photo-thermal depth profiling.

ACKNOWLEDGMENTS

This work was supported by the following grants at the National Institute of Health: AR 47551 and AR48458.

APPENDIX

The kernel $K(z,t)$ of Eq. (3) is expressed as

$$K(z,t) = (C\mu_{\mathrm{IR}}\Delta z/2)e^{-z^2/4Dt}\Bigg\{\mathrm{erfcx}(u_-) + \mathrm{erfcx}(u_+) - \frac{2h}{h-\mu_{\mathrm{IR}}}[\mathrm{erfcx}(u_+) - \mathrm{erfcx}(u_1)]\Bigg\}, \tag{A1}$$

where u_+, u_-, and u_1 are defined as

$$u_\pm = \mu_{\mathrm{IR}}\sqrt{Dt} \pm z/\sqrt{4Dt}, \quad u_1 = h\sqrt{Dt} + z/\sqrt{4Dt}. \tag{A2}$$

For a correct numerical implementation of the Kernel matrix **K** in Matlab® and to guarantee smooth corresponding singular vectors, care must be taken.

First, we have found it essential to use a numerical approximation[28] for the error function $\mathrm{erf}(x)$ instead of the Matlab® implementation for this function or the related function $\mathrm{erfcx}(x)$,

$$\mathrm{erf}(x) \cong 1 - (a_1t + a_2t^2 + a_3t^3 + a_4t^4 + a_5t^5)e^{-x^2}, \tag{A3}$$

where $t=1/(1+0.3275911x)$, $a_1=0.254829592$, $a2=-0.284496736$, $a_3=1.421413741$, $a_4=-1.453152027$, $a_5=1.061405429$. Since this approximation starts to oscillate for values larger than approximately 5, the related $\mathrm{erfcx}(x)$ also becomes oscillatory. For large values of x, this latter function is correctly approximated by

$$\mathrm{erfcx}(x) \cong 1/(x\sqrt{\pi}) \quad \text{for } x \gg 1. \tag{A4}$$

At some value x_T, the transition between the approximations (A3) and (A4) must be made which introduces some discontinuity. We have found for our application that the transition is smoothest when $x_T=5.6$ and the resulting singular vectors in the matrix K are smooth.

Second, for small times t and relatively large depths z, a problem may arise when the term $\mathrm{erfcx}(u_-)$ is calculated. When u_- is smaller than -30 (e.g., for $t<0.5$ ms, $z>450$ μm), the terms $\exp(u_-^2)$ in $\mathrm{erfcx}(u_-)$ become extremely large ($>10^{300}$). Straightforward evaluation of $\mathrm{erfcx}(u_-)$ leads to an error in most computations which can be completely avoided by rewriting Eq. (A1) as

$$K(z,t) = e^{-z^2/4Dt}\Bigg\{\mathrm{erfcx}(u_+) - \frac{2h}{h-\mu_{\mathrm{IR}}}[\mathrm{erfcx}(u_+) - \mathrm{erfcx}(u_1)]\Bigg\} + e^{\mu_{\mathrm{IR}}(\mu_{\mathrm{IR}}Dt-z)}\,\mathrm{erfc}(u_-). \tag{A5}$$

Note that a similar problem does not occur for evaluation of $\mathrm{erfcx}(u_+)$) with large positive values for u_+ because of the application of Eq. (A4).

[1] A. C. Tam and B. Sullivan, Appl. Phys. Lett. **43**, 333 (1983).
[2] R. E. Imhof, D. J. S. Birch, F. R. Thornley, J. R. Gilchrist, and T. A. Strivens, J. Phys. E **17**, 521 (1984).
[3] S. L. Jacques, J. S. Nelson, W. H. Wright, and T. E. Milner, Appl. Opt. **32**, 2439 (1993).
[4] T. E. Milner, D. M. Goodman, B. S. Tanenbaum, and J. S. Nelson, J. Opt. Soc. Am. A **12**, 1479 (1995).
[5] U. S. Sathyam and S. A. Prahl, J. Biomed. Opt. **2**, 251 (1997).
[6] J. J. Moré, *The Levenberg–Marquardt Algorithm: Implementation and Theory*, in *Numerical Analysis (Proceedings of the 7th Biennial Conference, University of Dundee, Dundee, 1977)*, edited by G. A. Watson, Lecture Notes in Mathematics 630 (Springer, Berlin, 1978).
[7] P. Xiao and R. E. Imhof, Proc. SPIE **3601**, 340 (1999).
[8] P. Xiao, S. F. Gull, and R. E. Imhof, Anal. Sci. **17**, S394 (2001).
[9] Y. X. Cui, P. Xiao, R. E. Imhof, and C. Glorieux, Rev. Sci. Instrum. **74**, 368 (2003).
[10] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problem* (Prentice–Hall, Englewood Cliffs, 1974).
[11] P. C. Hansen, SIAM Rev. **34**, 561 (1992).
[12] F. Duck, *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, London, 1990).
[13] B. Majaron, W. Verkruysse, B. S. Tanenbaum, T. E. Milner, and J. S. Nelson, Phys. Med. Biol. **47**, 1929 (2002).
[14] C. W. Groetsch, *The Theory of Tichonov Regularization for Fredholm Equations of the First Kind* (Pitman, Marshfield, 1984).
[15] D. Fylstra, L. S. Lasdon, J. Watson, and A. D. Waren, Interfaces **28**, 29 (1998).
[16] L. S. Lasdon, R. L. Fox, and M. W. Ratner, Rev. Fr. Inform. Rech. Oper. **8**, 73 (1974).
[17] L. S. Lasdon and A. D. Waren, Comput. Chem. Eng. **7**, 595 (1983).
[18] L. S. Lasdon, A. D. Waren, A. Jain, and M. Ratner, ACM Trans. Math. Softw. **4**, 34 (1978).
[19] M. Ratner, L. S. Lasdon, and A. Jain, J. Optim. Theory Appl. **26**, 253 (1978).
[20] M. R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand. **49**, 409 (1952).
[21] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Wiley, New York, 1977).
[22] P. C. Hansen, Inverse Probl. **8**, 849 (1992).
[23] T. F. Coleman and Y. Li, SIAM J. Control Optim. **6**, 418 (1996).
[24] T. F. Coleman and Y. Li, Math. Program. **67**, 189 (1994).
[25] K. Levenberg, Q. Appl. Math. **2**, 164 (1944).
[26] D. Marquardt, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **11**, 431 (1963).
[27] J. E. J. Dennis, "Nonlinear Least-Squares," in *State of the Art in Numerical Analysis*, edited by D. Jacobs (Academic, New York, 1977).
[28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965).