An interior-point method for large constrained discrete ill-posed problems

S. Morigi [a], L. Reichel [b,*,1], F. Sgallari [c,2]

[a] *Dipartimento di Matematica, Università degli Studi di Bologna, Piazza Porta S. Donato 5, 40127 Bologna, Italy.*

[b] *Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA.*

[c] *Dipartimento di Matematica - CIRAM, Università degli Studi di Bologna, Via Saragozza 8, 40123 Bologna, Italy.*

Dedicated to Bill Gragg on the occasion of his 70th birthday.

Abstract

Ill-posed problems are numerically underdetermined. It is therefore often beneficial to impose known properties of the desired solution, such as nonnegativity, during the solution process. This paper proposes the use of an interior-point method in conjunction with truncated iteration for the solution of large-scale linear discrete ill-posed problems with box constraints. An estimate of the error in the data is assumed to be available. Numerical examples demonstrate the competitiveness of this approach.

Key words: ill-posed problem, regularization, box constraint, truncated iteration, conjugate gradient method, interior point method.

1 Introduction

Bill Gragg has contributed significantly to the development of numerical analysis, both through his own work and through his encouragement of friends and

* Corresponding author.
Email addresses: `morigi@dm.unibo.it` (S. Morigi), `reichel@math.kent.edu` (L. Reichel), `sgallari@dm.unibo.it` (F. Sgallari).

[1] Supported by an OBR Research Challenge Grant.
[2] Supported by MIUR - PRIN 06 N.2006013187 - 004 and GNCS - INDAM grants.

colleagues. His work includes methods for the solution of ordinary differential equations, structured eigenvalue problems, and rational approximation; see, e.g., [1,5,13–16]. In addition, Gragg has discussed the solution of ill-posed and nonlinear problems; see [11,17]. This paper is concerned with the latter areas.

The discretization of Fredholm integral equations of the first kind with a smooth kernel, and in particular of deconvolution problems, gives rise to linear systems of equations

$$A\boldsymbol{x} = \boldsymbol{b}, \qquad A \in \mathbb{R}^{m\times n}, \qquad \boldsymbol{x} \in \mathbb{R}^{n}, \qquad \boldsymbol{b} \in \mathbb{R}^{m}, \tag{1}$$

with a matrix, whose singular values "cluster" at the origin. This makes the matrix severely ill-conditioned, possibly singular, and of ill-determined rank. Linear systems of equations with this kind of matrix are commonly referred to as linear discrete ill-posed problems. We allow $m \neq n$ in (1). If the linear system is inconsistent, then we consider (1) a least-squares problem.

Ill-posed problems often arise when one seeks to determine the cause of an observed effect. The latter is represented by the right-hand side $\boldsymbol{b}$, which is typically contaminated by a measurement error $\boldsymbol{e} \in \mathbb{R}^m$. We consider the situation when the norm of the error

$$\varepsilon = \|\boldsymbol{e}\|, \tag{2}$$

or an estimate thereof, are available. Here and throughout this paper $\|\cdot\|$ denotes the Euclidean vector norm.

Let $\hat{\boldsymbol{b}} \in \mathbb{R}^m$ denote the unavailable error-free vector associated with $\boldsymbol{b}$, i.e.,

$$\boldsymbol{b} = \hat{\boldsymbol{b}} + \boldsymbol{e}, \tag{3}$$

and consider the linear system of equations with the error-free right-hand

$$A\boldsymbol{x} = \hat{\boldsymbol{b}}. \tag{4}$$

This system is assumed to be consistent. We would like to compute its solution of minimal Euclidean norm, denoted by $\hat{\boldsymbol{x}}$. Since $\hat{\boldsymbol{b}}$ is not available, we seek to determine an approximation of $\hat{\boldsymbol{x}}$ by computing an approximate solution of the available linear system (1).

Let $A^\dagger$ denote the Moore-Penrose pseudo-inverse of A. Then $\hat{\boldsymbol{x}} = A^\dagger\hat{\boldsymbol{b}}$. However, due to the error $\boldsymbol{e}$ in $\boldsymbol{b}$ and the ill-conditioning of A, the vector $A^\dagger\boldsymbol{b}$ is generally severely contaminated by propagated error and is not a meaningful approximation of $\hat{\boldsymbol{x}}$.

The computation of a useful approximation of $\hat{\boldsymbol{x}}$ can be accomplished by replacing the linear system of equations (1) by a nearby system, whose solution is less sensitive to the error $\boldsymbol{e}$. This replacement is commonly referred to as regularization. The most popular regularization techniques are Tikhonov regularization and truncated iteration. The latter approach provides an implicit replacement of A by a matrix of low rank. For instance, consider the application of the minimal residual iterative method LSQR to the solution of (1) with initial iterate $\boldsymbol{x}_0 = \boldsymbol{0}$. Denote the computed iterates by $\boldsymbol{x}_j$, $j = 1, 2, 3, \ldots$, and let $\boldsymbol{x}_{j_\varepsilon}$ be the first iterate that satisfies the inequality

$$\|A\boldsymbol{x}_{j_\varepsilon} - \boldsymbol{b}\| \leq \eta\varepsilon, \tag{5}$$

where $\eta > 1$ is a user-chosen constant. Then $\boldsymbol{x}_{j_\varepsilon}$ is our computed approximate solution of (1) determined by truncated iteration. Truncated iteration can be thought of replacing the matrix A by a matrix of rank j_ε and determining the minimal-norm solution of the so-obtained linear system of equations. This replacement reduces the influence of the "tiny" singular values of A on the computed solution and allows the computation of a meaningful approximate solution of (1) in the presence of the error $\boldsymbol{e}$ in $\boldsymbol{b}$. Under suitable conditions $\boldsymbol{x}_{j_\varepsilon} \to \hat{\boldsymbol{x}}$ as $\varepsilon \searrow 0$; see, e.g., [10,18] for proofs and discussions in a Hilbert space setting.

Linear discrete ill-posed problems (1) are numerically singular due to the clustering of the singular values of A at the origin. They therefore are numerically underdetermined. Because of this, numerical methods that impose known properties of the desired solution $\hat{\boldsymbol{x}}$, such as nonnegativity, on the computed approximate solution of (1), often are able to determine better approximations of $\hat{\boldsymbol{x}}$ than numerical methods that do not. This paper describes a new interior-point method for the solution of

$$\|A\boldsymbol{x} - \boldsymbol{b}\| \leq \eta\varepsilon, \qquad \boldsymbol{x} \geq \boldsymbol{0}, \tag{6}$$

where the inequality $\boldsymbol{x} \geq \boldsymbol{0}$ is understood component-wise. Thus, we are concerned with the computation of nonnegatively constrained approximate solutions of (1). Large linear discrete ill-posed problems with a nonnegative solution $\hat{\boldsymbol{x}}$ arise, for instance, in image restoration. In this application, the entries of $\hat{\boldsymbol{x}}$ correspond to pixel values, which are nonnegative. The numerical method of this paper easily can be modified to allow box constraints; however, for ease of discussion, we consider nonnegativity constraints only.

Our solution scheme is inspired by the interior-point method proposed by Rojas and Steihaug [28] for the solution of constrained minimization problems of the form

$$\min_{\substack{\|\boldsymbol{x}\|\leq\Delta\\ \boldsymbol{x}\geq\boldsymbol{0}}} \|A\boldsymbol{x}-\boldsymbol{b}\|, \tag{7}$$

where Δ is an available estimate of $\|\hat{\boldsymbol{x}}\|$. Rojas and Steihaug [28] apply Tikhonov regularization to determine an approximate solution of (7). An alternative implementation of this method, based on partial Lanczos bidiagonalization and the application of certain Gauss-type quadrature rules to compute bounds for certain pertinent functionals, is presented in [7].

Let for the moment $\Delta = \|\hat{\boldsymbol{x}}\|$. Then the solution of the minimization problem (7) typically is an accurate approximation of $\hat{\boldsymbol{x}}$ when ε, defined by (2), is small (compared to $\|\hat{\boldsymbol{b}}\|$). However, when ε is not small, say $\varepsilon = 0.1\|\hat{\boldsymbol{b}}\|$, the computed solution of (7) may be a poor approximation of $\hat{\boldsymbol{x}}$. This is illustrated by Example 5.2 of Section 5. The poor accuracy depends on the fact that the constraint Δ in (7) is chosen independently of the norm ε of the error $\boldsymbol{e}$. It is the purpose of the present paper to develop a variant of the method presented by Rojas and Steihaug [28] that delivers fairly accurate approximations of $\hat{\boldsymbol{x}}$ also when ε is not small.

Large-scale linear discrete ill-posed problems with constraints arise in many applications and several different approaches to their solution have been proposed; see, e.g., Bardsley [2], Bertero and Boccacci [3, Section 6.3], Calvetti et al. [6], Hanke et al. [19], Kim [22], Morigi et al. [23], Nagy and Strakoš [24], and references therein, in addition to the references already mentioned. There is presently not one best solution method for all constrained large-scale ill-posed problems. Interior-point methods have received considerable attention in the optimization literature. We therefore believe it to be of interest to develop numerical methods for constrained large-scale linear discrete ill-posed problems of the form (6) based on this approach.

This paper is organized as follows. Section 2 describes an adaption of the interior-point method discussed by Rojas and Steihaug [28] to the solution of (6). Section 3 discusses the approximate solution of the resulting linear systems of equations by partial Lanczos bidiagonalization, and Section 4 describes our solution method for (6). A few computed examples are presented in Section 5 and concluding remarks can be found in Section 6.

2 An interior-point method

This section describes a variant of the interior-point method proposed by Rojas and Steihaug [28] that is suited for the solution of (6). Define for $\boldsymbol{x} =$

$[\xi_1, \xi_2, \dots, \xi_n]^T > \mathbf{0}$ the function

$$f_\gamma(\boldsymbol{x}) = \frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \frac{\varphi(\gamma)}{2}\|\boldsymbol{x}\|^2 - \gamma \sum_{i=1}^{n} \log \xi_i, \tag{8}$$

where $\varphi(\gamma)$ is a positive increasing function of $\gamma \geq 0$ with $\varphi(0) = 0$. The last term in (8) is commonly referred to as a barrier function; we will refer to

$$\boldsymbol{x} \to \frac{\varphi(\gamma)}{2}\|\boldsymbol{x}\|^2 - \gamma \sum_{i=1}^{n} \log \xi_i$$

as a modified barrier function and to γ as a barrier parameter. The purpose of the term $\frac{\varphi(\gamma)}{2}\|\boldsymbol{x}\|^2$ is to secure that the function $f_\gamma(\boldsymbol{x})$ is strictly convex for any fixed $\gamma > 0$, even when A is not of full rank. In the computed examples, we let $\varphi(\gamma) = \gamma$; however, other choices of $\varphi(\gamma)$ are also possible.

We determine an approximate solution of the constrained minimization problem (6) by approximately solving a sequence of unconstrained minimization problems

$$\min_{\boldsymbol{x}} f_\gamma(\boldsymbol{x}) \tag{9}$$

for decreasing positive values of the barrier parameter $\gamma = \gamma_k$, $k = 0, 1, 2, \dots$, and an initial approximate solution in the first orthant. Since $f_\gamma(\boldsymbol{x})$ is strictly convex for any $\gamma > 0$, the minimization problem (9) has a unique solution.

We solve (9) for a given positive value of γ by a trust-region-like method. Introduce the quadratic model in $\boldsymbol{h}$ of f_γ in a neighborhood of $\boldsymbol{x} > \mathbf{0}$,

$$q_\gamma(\boldsymbol{x} + \boldsymbol{h}) = f_\gamma(\boldsymbol{x}) + \boldsymbol{h}^T \nabla f_\gamma(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{h}^T \nabla^2 f_\gamma(\boldsymbol{x}) \boldsymbol{h},$$

where

$$\begin{aligned}
\nabla f_\gamma(\boldsymbol{x}) &= (A^T A + \varphi(\gamma) I)\boldsymbol{x} - A^T \boldsymbol{b} - \gamma X^{-1} \boldsymbol{c}, \\
\nabla^2 f_\gamma(\boldsymbol{x}) &= A^T A + \varphi(\gamma) I + \gamma X^{-2}, \\
X &= \mathrm{diag}[\boldsymbol{x}] \in \mathbb{R}^{n \times n}, \\
\boldsymbol{c} &= [1, 1, \dots, 1]^T \in \mathbb{R}^n,
\end{aligned}$$

with I the identity matrix, and determine an approximate solution of the problem

$$\min_{\boldsymbol{h} \in \mathbb{R}^n} q_\gamma(\boldsymbol{x} + \boldsymbol{h}). \tag{10}$$

Trust-region methods for the solution of (9) minimize q_γ over a disc $\|\boldsymbol{h}\| \leq r$, where r is referred to as the trust-region radius, and then update q_γ. We will return to this aspect below.

The minimum $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{h}$ of (10) satisfies the linear system of equations

$$(A^T A + \varphi(\gamma) I + \gamma X^{-2})\boldsymbol{z} = A^T \boldsymbol{b} + 2\gamma X^{-1}\boldsymbol{c} \tag{11}$$

with a symmetric positive definite matrix. The solution of this system is discussed in Section 3. Here we just note that we solve (11) by a particular implementation of the conjugate gradient method and terminate the iterations, e.g., when an iterate that satisfies the inequality (5) has been determined. Thus, our solution scheme is a truncated iterative method. This solution method prevents the error $A^T\boldsymbol{e}$ in $A^T\boldsymbol{b}$ to be amplified and severely contaminate the computed approximate solution.

We remark that truncated iteration implicitly bounds the norm of the computed solution; see also the last paragraph of Section 3 for further comments. Our solution method for (11) therefore can be interpreted as a trust-region-like method with an implicitly defined trust-region radius. Truncated iteration has previously been applied in the context of trust-region computations for well-posed problems by Steihaug [29]. An extension of Steihaug's method, which exploits the connection between partial Lanczos tridiagonalization of a symmetric matrix (the Hessian) and the conjugate gradient method, is presented by Gould et al. [12]. The use of partial Lanczos tridiagonalization makes it possible to evaluate upper and lower bounds for certain relevant quantities inexpensively during the computations; see [8].

3 Solution of the quadratic model

We discuss the computation of an approximate solution of (11) by partial Lanczos bidiagonalization. Introduce for $\gamma > 0$ the positive definite diagonal matrix and vector

$$D_\gamma = (\varphi(\gamma) I + \gamma X^{-2})^{1/2}, \qquad \boldsymbol{g}_\gamma = 2\gamma(\varphi(\gamma) X^2 + \gamma I)^{-1/2}\boldsymbol{c},$$

and note that

$$\lim_{\gamma \searrow 0} D_\gamma = 0, \qquad \lim_{\gamma \searrow 0} \boldsymbol{g}_\gamma = \boldsymbol{0}. \tag{12}$$

Let

$$\breve{A} = \begin{bmatrix} A \\ D_\gamma \end{bmatrix} \in \mathbb{R}^{(m+n)\times n}, \qquad \breve{\boldsymbol{b}} = \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{g}_\gamma \end{bmatrix} \in \mathbb{R}^{(m+n)}. \tag{13}$$

Then (11) are the normal equations associated with the minimization problem

$$\min_{\boldsymbol{z}\in\mathbb{R}^n} \|\breve{A}\boldsymbol{z} - \breve{\boldsymbol{b}}\|. \tag{14}$$

We compute an approximate solution of (14) by partial Lanczos bidiagonalization of the matrix $\breve{A}$. The barrier parameter $\gamma > 0$ is kept fixed throughout this section. Its selection is discussed in Section 4.

Application of ℓ steps of Lanczos bidiagonalization to the matrix $\breve{A}$ with initial vector $\breve{\boldsymbol{b}}$ yields the decompositions

$$\breve{A}V_\ell = U_{\ell+1}\bar{C}_\ell, \qquad \breve{A}^T U_\ell = V_\ell C_\ell^T, \qquad U_\ell \boldsymbol{e}_1 = \breve{\boldsymbol{b}}/\|\breve{\boldsymbol{b}}\|, \tag{15}$$

where $V_\ell \in \mathbb{R}^{n\times\ell}$ and $U_{\ell+1} \in \mathbb{R}^{(m+n)\times(\ell+1)}$ satisfy

$$V_\ell^T V_\ell = I, \qquad U_{\ell+1}^T U_{\ell+1} = I,$$

and $U_\ell \in \mathbb{R}^{(m+n)\times\ell}$ consists of the first ℓ columns of $U_{\ell+1}$. Throughout this paper $\boldsymbol{e}_j$ denotes the jth axis vector. The matrix $\bar{C}_\ell \in \mathbb{R}^{(\ell+1)\times\ell}$ in (15) is lower bidiagonal with positive diagonal and subdiagonal entries and C_ℓ denotes the $\ell\times\ell$ leading submatrix of $\bar{C}_\ell$. We assume that ℓ is sufficiently small so that the decomposition (15) with the stated properties exist, see, e.g., Björck [4, Section 7.6] for more details on Lanczos bidiagonalization. Note that the range of V_ℓ is the Krylov subspace

$$\mathcal{K}_\ell(\breve{A}^T\breve{A}, \breve{A}^T\breve{\boldsymbol{b}}) = \mathrm{span}\{\breve{A}^T\breve{\boldsymbol{b}}, (\breve{A}^T\breve{A})\breve{A}^T\breve{\boldsymbol{b}}, \ldots, (\breve{A}^T\breve{A})^{\ell-1}\breve{A}^T\breve{\boldsymbol{b}}\}.$$

The LSQR method is an implementation of the conjugate gradient method applied to the normal equations (11) based on the decompositions (15); see, e.g., Björck [4, Section 7.6] or Paige and Saunders [26] for details on this method. Our solution method is a minor modification of LSQR; a few auxiliary quantities are computed to facilitate the evaluation of the stopping criteria (20) below.

The ℓth iterate determined by the LSQR method applied to (14) with initial approximate solution $\boldsymbol{z}_0 = \boldsymbol{0}$ is of the form

$$\boldsymbol{z}_\ell = V_\ell \boldsymbol{y}_\ell, \qquad \boldsymbol{y}_\ell \in \mathbb{R}^\ell, \tag{16}$$

and satisfies

$$\|\breve{A}\boldsymbol{z}_\ell - \breve{\boldsymbol{b}}\| = \min_{\boldsymbol{z}\in\mathcal{K}_\ell(\breve{A}^T\breve{A},\breve{A}^T\breve{\boldsymbol{b}})} \|\breve{A}\boldsymbol{z} - \breve{\boldsymbol{b}}\|, \qquad \boldsymbol{z}_\ell \in \mathcal{K}_\ell(\breve{A}^T\breve{A}, \breve{A}^T\breve{\boldsymbol{b}}). \tag{17}$$

Equivalently, $\boldsymbol{y}_\ell$ satisfies

$$\|\bar{C}_\ell \boldsymbol{y}_\ell - \boldsymbol{e}_1\|\breve{\boldsymbol{b}}\|\| = \min_{\boldsymbol{y}\in\mathbb{R}^\ell} \|\bar{C}_\ell \boldsymbol{y} - \boldsymbol{e}_1\|\breve{\boldsymbol{b}}\|\|. \tag{18}$$

Let $\boldsymbol{y}_{\ell-1} \in \mathbb{R}^{\ell-1}$ denote the solution of the minimization problem analogous to (18) with ℓ replaced by $\ell-1$. Because $\bar{C}_\ell$ is lower bidiagonal, the vector made up of the first $\ell-1$ entries of $\boldsymbol{y}_\ell$ is a multiple of $\boldsymbol{y}_{\ell-1}$; see, e.g., Björck [4, Section 7.6] or Paige and Saunders [26]. This implies that only a few of the most recently generated columns of the matrices $U_{\ell+1}$ and V_ℓ have to be stored simultaneously in order to be able to compute the approximate solution $\boldsymbol{z}_\ell$ of (17) from the already available approximate solution $\boldsymbol{z}_{\ell-1}$, and to be able to carry out another Lanczos bidiagonalization step. Hence, the storage requirement of the method is quite modest and bounded independently of the number of bidiagonalization steps ℓ. This makes it possible to solve large-scale problems also on fairly small computers. We note that our solution method is mathematically equivalent to applying the conjugate gradient method to (11).

We are interested in evaluating the norm of the residual error $A\boldsymbol{z}_\ell - \boldsymbol{b}$ for increasing values of ℓ in order to determine when to stop the iterations; cf. (5) and (6). Note that

$$\|\breve{A}\boldsymbol{z}_\ell - \breve{\boldsymbol{b}}\|^2 = \|A\boldsymbol{z}_\ell - \boldsymbol{b}\|^2 + \|D_\gamma \boldsymbol{z}_\ell - \boldsymbol{g}_\gamma\|^2. \tag{19}$$

Let $\bar{C}_\ell = Q_{\ell+1}\bar{R}_\ell$ be a QR-factorization, i.e., $Q_{\ell+1} \in \mathbb{R}^{(\ell+1)\times(\ell+1)}$ is orthogonal and $\bar{R}_\ell \in \mathbb{R}^{(\ell+1)\times\ell}$ has a leading $\ell\times\ell$ upper triangular submatrix and vanishing last row. We represent $Q_{\ell+1}$ as a product of ℓ Givens rotations. The left-hand side of (19) can be evaluated inexpensively by computing the right-hand side of

$$\|\breve{A}\boldsymbol{z}_\ell - \breve{\boldsymbol{b}}\|^2 = \|\bar{C}_\ell \boldsymbol{y}_\ell - \boldsymbol{e}_1\|\breve{\boldsymbol{b}}\|\|^2 = |\boldsymbol{e}_1^T Q_{\ell+1} \boldsymbol{e}_{\ell+1}|^2 \|\breve{\boldsymbol{b}}\|^2.$$

When $\boldsymbol{e}_1^T Q_\ell \boldsymbol{e}_\ell$ is available, the number of arithmetic floating point operations required to determine $\boldsymbol{e}_1^T Q_{\ell+1} \boldsymbol{e}_{\ell+1}$ is bounded independently of ℓ. This follows from the representation of Q_ℓ and $Q_{\ell+1}$ in terms of Givens rotations.

We turn to the evaluation of the rightmost term in (19). Assuming that $D_\gamma \boldsymbol{z}_{\ell-1}$ has been computed, we can evaluate $D_\gamma \boldsymbol{z}_\ell$ in only $\mathcal{O}(n)$ arithmetic floating point operations, with the number of operations bounded independently of ℓ, by using the representation (16) and the fact that the leading subvector of $\boldsymbol{y}_\ell$ of size $\ell - 1$ is a multiple of the vector $\boldsymbol{y}_{\ell-1}$. Thus, for each increase of ℓ by one, we can compute $\|A\boldsymbol{z}_\ell - \boldsymbol{b}\|$ in $\mathcal{O}(n)$ arithmetic floating point operations, with the number of operations bounded independently of ℓ.

The dominating computational work for large-scale problems for the method described is the evaluation of matrix-vector products; each Lanczos bidiagonalization step requires the evaluation of one matrix-vector product with A and one with A^T. Therefore, we would like to keep the number of Lanczos bidiagonalization steps ℓ small. We seek to achieve this by not increasing ℓ when this does not give a reduction in the norm of the residual error $\|A\boldsymbol{z}_\ell - \boldsymbol{b}\|$ or when a residual error of norm smaller than or equal to $\eta\varepsilon$ has been found. Thus, we let ℓ be the smallest integer, such that

$$\|A\boldsymbol{z}_\ell - \boldsymbol{b}\| \leq \|A\boldsymbol{z}_{\ell+1} - \boldsymbol{b}\| \qquad \text{or} \qquad \|A\boldsymbol{z}_\ell - \boldsymbol{b}\| \leq \eta\varepsilon. \tag{20}$$

The fact that the iterates $\boldsymbol{z}_\ell$ are computed by a minimal residual method (LSQR) applied to (14) and the properties (12) secure the existence of an iterate that satisfies the right-hand side inequality of (20) for all $\gamma > 0$ sufficiently small.

The norm of the iterates $\boldsymbol{z}_\ell$ determined by LSQR grows strictly with ℓ; see Steihaug [29] or Hestenes and Stiefel [21]. The termination criteria (20) are equivalent to requiring that $\|\boldsymbol{z}_\ell\| \leq r$ for a suitably chosen constant r. Our solution scheme for (14) and (11), therefore may be considered a trust-region-like method with an implicitly defined trust-region radius.

4 The solution method

We are in a position to discuss the choice of initial approximate solution, the updating of approximate solutions, the choice of barrier parameter values, and the stopping criteria. First we determine an approximate solution of (1) ignoring the nonnegativity constraint. Thus, we apply LSQR with initial iterate $\boldsymbol{x}_0 = \boldsymbol{0}$ to (1) and choose the first iterate, $\boldsymbol{x}_{j_\varepsilon}$, that yields a residual error of norm at most $\eta\varepsilon$ as our approximate solution; cf. (5). Let $\boldsymbol{x}^{(0)}$ denote the orthogonal projection of $\boldsymbol{x}_{j_\varepsilon}$ onto the first orthant, i.e., we obtain $\boldsymbol{x}^{(0)}$ from $\boldsymbol{x}_{j_\varepsilon}$ by setting all negative entries to zero. If $\boldsymbol{x}^{(0)}$ satisfies (6), then we are done; otherwise we proceed as follows.

Let δ be a fairly small positive constant in relation to the norm of the desired solution $\hat{\boldsymbol{x}}$. Determine the vector $\boldsymbol{x}^{(\delta)}$ from $\boldsymbol{x}^{(0)}$ by setting all entries smaller than δ in $\boldsymbol{x}^{(0)}$ to δ. Let

$$X = \mathrm{diag}[\boldsymbol{x}^{(\delta)}]. \tag{21}$$

We consider $\boldsymbol{x}^{(\delta)}$ the current approximate solution of (6), and determine a new approximate solution, $\boldsymbol{z}_\ell$, by application of LSQR to (14). The purpose of considering $\boldsymbol{x}^{(\delta)}$, rather than $\boldsymbol{x}^{(0)}$, the current approximate solution is to obtain an invertible matrix X and avoid computing with matrices X^{-1} of very large norm. Moreover, tiny entries in the available approximate solution can force the correction to be tiny, which may yield slow convergence of the interior-point method. This can be seen as follows. The computed approximate solution $\boldsymbol{z}_\ell$ of (14) satisfies (20) but may have negative components. Following Rojas and Steihaug [28], we define

$$\boldsymbol{h} = \boldsymbol{z}_\ell - \boldsymbol{x}^{(\delta)} \tag{22}$$

and compute the new candidate solution

$$\boldsymbol{x}^{(0)} = \boldsymbol{x}^{(\delta)} + \beta \boldsymbol{h} \tag{23}$$

of (6). The constant $\beta > 0$ is chosen so that $\boldsymbol{x}^{(0)} \geq \mathbf{0}$; we let

$$\beta = \min \left\{ 1, 0.9995 \min_{\substack{\boldsymbol{e}_j^T \boldsymbol{z}_\ell \leq 0 \\ 1 \leq j \leq n}} \frac{\boldsymbol{e}_j^T \boldsymbol{x}^{(\delta)}}{|\boldsymbol{e}_j^T \boldsymbol{h}|} \right\}. \tag{24}$$

It is clear that if $\boldsymbol{x}^{(\delta)}$ has a tiny component, then the correction $\beta \boldsymbol{h}$ might be tiny as well, and this may slow down convergence. We therefore should not choose $\delta > 0$ too small.

If the vector $\boldsymbol{x}^{(0)}$ given by (23) satisfies (6), then we are done; otherwise we repeat the computations described above, i.e., we set all components of $\boldsymbol{x}^{(0)}$ smaller than δ to δ in order to obtain the new vector $\boldsymbol{x}^{(\delta)}$, define the matrix X by (21), and compute a new candidate solution of (6) by solving (14) in the manner just described.

We determine the initial value of the barrier parameter from

$$\gamma = \sigma \frac{|(\boldsymbol{x}^{(0)})^T \boldsymbol{w}^{(0)}|}{n}, \tag{25}$$

where

$$\boldsymbol{w}^{(0)} = A^T \boldsymbol{b} - A^T A \boldsymbol{x}^{(0)}, \qquad \sigma = 0.01.$$

Rojas and Steihaug [28] present a heuristic justification for (25) in the context of their method. An analogous justification can be provided in the context of the minimization problem (6). For each subsequent solution of (14), we divide the available value of γ by 10. This has worked well for many examples.

Standard results for trust-region methods show convergence to the solution of (9) for fixed $\gamma > 0$ if a suitable sequence of trust-region radii is selected when solving (10), or equivalently, if appropriately many Lanczos bidiagonalization steps ℓ are used to solve each one of the sequence of linear systems of equations (11); see, e.g., [12,25,29]. Convergence results for interior-point methods and, in particular, barrier function methods, as $\gamma > 0$ decreases, show that these methods are able to determine a solution of (6). The existence of such a solution follows from equations (2)-(4), $\hat{\boldsymbol{x}} \geq 0$, and $\eta > 1$; see, e.g., [25] for results on interior-point and barrier function methods.

These results for trust-region and barrier function methods provide heuristic support for the scheme presented. However, a full theoretical justification remains to be developed. An analysis has to take into account that we do not want to compute an exact solution of (6). In fact, the residual error $\boldsymbol{b} - A\boldsymbol{z}_j$ for any approximate solution $\boldsymbol{z}_j$ generated during the computations must not be of norm significantly smaller than $\eta\varepsilon$, because this typically would cause severe propagation of the error $\boldsymbol{e}$ into the computed approximate solution. The contribution of the present paper is to present an algorithm that often is able to compute a meaningful solution of (6) with a fairly small number of matrix-vector product evaluations. We have achieved this, e.g., by using the stopping criteria (20) for the iterations, which have the effect that, generally, the computational effort decreases when ε increases. This is illustrated in Example 5.2 below.

Kim [22] recently presented a theoretical study of an iterative method for the solution of linear discrete ill-posed problems with a constraint. However, the computed examples reported in [22] show the method to require a large number of matrix-vector product evaluations. Therefore, the development of alternative approaches that require less computational effort is of interest.

The following algorithm summarizes the computations required by our method. The function $\max\{\boldsymbol{x}, \alpha\}$ for $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ yields an n-vector, whose jth entry is $\max\{x_j, \alpha\}$ for $1 \leq j \leq n$.

Algorithm 1 (Interior point method)

Input: A, b, ε, δ, η;
Output: approximate solution $\tilde{\boldsymbol{x}}$ *of (1);*
Determine the unconstrained solution $\boldsymbol{x}_{j_\varepsilon}$ *of (5) by* j_ε *steps of LSQR;*
$\boldsymbol{x}^{(0)} := \max\{\boldsymbol{x}_{j_\varepsilon}, 0\}$;
Define γ *according to (25);*

while $\boldsymbol{x}^{(0)}$ *does not satisfy (6) do*

  $\boldsymbol{x}^{(\delta)} := \max\{\boldsymbol{x}^{(0)}, \delta\}$;

  Compute the solution $\boldsymbol{z}_\ell$ *of (17) by application of* ℓ *steps of LSQR;*

  $\boldsymbol{h} := \boldsymbol{z}_\ell - \boldsymbol{x}^{(\delta)}$;

  $\boldsymbol{x}^{(0)} := \boldsymbol{x}^{(\delta)} + \beta\boldsymbol{h}$, *where* β *is computed by (24);*

  $\gamma := \gamma/10$;

end while

$\tilde{\boldsymbol{x}} := \boldsymbol{x}^{(0)}$;

5 Computed examples

This section illustrates the performance of our interior-point method. The matrices in all examples are of ill-determined rank. All computations are carried out in Matlab with about 16 significant decimal digits.

Example 5.1. Consider the Fredholm integral equation of the first kind

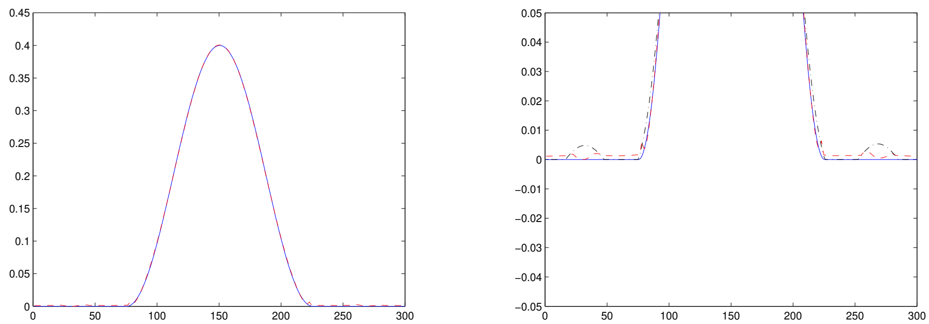
$$\int_{-6}^{6} \kappa(\tau,\sigma)x(\sigma)d\sigma = b(\tau), \qquad -6 \leq \tau \leq 6, \tag{26}$$

discussed by Phillips [27]. Its solution, kernel, and right-hand side are given by

$$\begin{aligned} x(\sigma) &= \begin{cases} 1+\cos(\frac{\pi}{3}\sigma), & \text{if } |\sigma| < 3, \\ 0, & \text{otherwise,} \end{cases} \\ \kappa(\tau,\sigma) &= x(\tau-\sigma), \\ b(\tau) &= (6-|\tau|)(1+\frac{1}{2}\cos(\frac{\pi}{3}\tau)) + \frac{9}{2\pi}\sin(\frac{\pi}{3}|\tau|). \end{aligned} \tag{27}$$

We discretize the integral equation (26) with the Matlab code phillips from the program package Regularization Tools by Hansen [20]. Discretization by a Galerkin method with 300 orthonormal box functions as test and trial functions yields a symmetric indefinite matrix $A \in \mathbb{R}^{300\times 300}$. The code also provides a discretization of the solution (27). We consider this discretization the exact solution $\hat{\boldsymbol{x}} \in \mathbb{R}^{300}$ and compute the right-hand side vector $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ of (4). An error vector $\boldsymbol{e}$ with normally distributed zero-mean random entries is added to $\hat{\boldsymbol{b}}$, cf. (3), to give the right-hand side $\boldsymbol{b}$ of (1). The relative error

$$\rho = \frac{\|\boldsymbol{e}\|}{\|\hat{\boldsymbol{b}}\|} \tag{28}$$



(a) (b)

Fig. 1. Example 5.1: (a) Computed solution $\tilde{\boldsymbol{x}}$ (dashed red graph) and solution $\hat{\boldsymbol{x}}$ of the error-free system (4) (continuous blue graph), (b) blow-up of $\tilde{\boldsymbol{x}}$ (dashed red graph), $\hat{\boldsymbol{x}}$ (continuous blue graph), and of the approximate solution $\boldsymbol{x}^{(0)}$ obtained by projecting the computed unconstrained solution of (5) onto the first orthant (dash-dotted black graph).

is referred to as the noise level. In the present example, the error vector $\boldsymbol{e}$ is scaled to yield the noise level $\rho = 5 \cdot 10^{-3}$. We assume that the noise level is known accurately and therefore choose the parameter η in (6) close to unity; specifically, we let $\eta = 1.02$.

We first solve the unconstrained problem and terminate the iterations as soon as the inequality (5) is satisfied. This yields the vector $\boldsymbol{x}_{j_\varepsilon}$, which has negative components. Setting these to zero gives $\boldsymbol{x}^{(0)}$.

Let $\delta = 1 \cdot 10^{-3}$ and determine the vector $\boldsymbol{x}^{(\delta)}$ from $\boldsymbol{x}^{(0)}$. A correction of $\boldsymbol{x}^{(\delta)}$ is obtained by solving (14). This gives a new vector, which we correct in the manner described until a solution of (6) has been found. We denote this solution by $\tilde{\boldsymbol{x}}$. It has relative error $\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\| = 7.67 \cdot 10^{-3}$. Figure 1 displays $\tilde{\boldsymbol{x}}$, $\boldsymbol{x}^{(0)}$, and $\hat{\boldsymbol{x}}$. The figure shows $\tilde{\boldsymbol{x}}$ to be a better approximation of $\hat{\boldsymbol{x}}$ than $\boldsymbol{x}^{(0)}$. The computation of $\tilde{\boldsymbol{x}}$ requires the evaluation of 52 matrix-vector products with the matrices A or A^T. $\Box$

ρ	mat.-vec. prod.	$\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$
$1 \cdot 10^{-1}$	22	$7.76 \cdot 10^{-2}$
$1 \cdot 10^{-2}$	34	$1.43 \cdot 10^{-2}$

Table 1
Example 5.2: Relative errors and number of matrix-vector product evaluations required by the method of the present paper for two noise levels.

Example 5.2. We modify the above example by increasing the norm of the error $\boldsymbol{e}$ in the right-hand side $\boldsymbol{b}$. Thus, let A and $\hat{\boldsymbol{b}}$ be the same as in Example 5.1, and let the error vector $\boldsymbol{e}$ have normally distributed zero-mean random entries, scaled to give noise levels 0.1 or 0.01; cf. (28). The right-hand side $\boldsymbol{b}$

ρ	mat.-vec. prod.	$\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$
$1 \cdot 10^{-1}$	267	$1.07 \cdot 10^{-1}$
$1 \cdot 10^{-2}$	118	$2.29 \cdot 10^{-2}$

Table 2
Example 5.2: Relative errors and number of matrix-vector product evaluations required by the method described in [7] for two noise levels.

of (1) is defined by (3).

Table 1 reports the performance of the method of the present paper. The error in the computed solution decreases with the noise level ρ. The table also displays the number of matrix-vector product evaluations with A or A^T required to determine the computed solution. This number is seen to increase when ρ decreases. This is typical, because the smaller the value of ρ, the more the residual error has to be reduced during the solution of (14); compare also to results for the noise level $\rho = 5 \cdot 10^{-3}$ reported in Example 5.1.

Table 2 shows the performance of the method described in [7]. This method does not assume knowledge of an estimate of the norm of the error in the right-hand side of (1). Instead, one assumes that an estimate of the norm of the desired solution is available, i.e., we supply $\Delta = \|\hat{\boldsymbol{x}}\|$ and solve the constrained minimization problem (7). The table shows the relative error in the computed approximate solutions and the number of matrix-vector product evaluations required for two noise levels. The method in [7] is seen to determine worse approximations of $\hat{\boldsymbol{x}}$ and require the evaluation of more matrix-vector products than the method of the present paper.

This example is typical for large noise levels. A related example, with similar performance, when the nonnegativity constraint is ignored in (7) is presented in [9, Example 4.1]. We conclude that the methods described in [7,28] should not be applied when the error $\boldsymbol{e}$ in the right-hand side $\boldsymbol{b}$ is not small. However, we note that the computed examples reported in [7] show the method discussed there to perform well when the error $\boldsymbol{e}$ is small, and the method is found to require fewer matrix-vector product evaluations than the method in [28]. □

Example 5.3. We are concerned with the restoration of images that have been contaminated by noise and blur. We would like to determine approximations of the noise- and blur-free image shown in Figure 2. This image is assumed not to be available. It is represented by 128×128 pixel values, which when ordered lexicographically define the vector $\hat{\boldsymbol{x}} \in \mathbb{R}^{128^2}$.

The blur in the contaminated images is generated by the blurring operator $A \in \mathbb{R}^{128^2 \times 128^2}$ defined by the Matlab function blur in [20]. This function determines a banded block Toeplitz matrix with banded Toeplitz blocks, which models blurring by a Gaussian point spread function. The variance of the point

Fig. 2. Example 5.3: noise- and blur-free image.

(a)

(b)

Fig. 3. Example 5.3: (a) contaminated image with noise level $1 \cdot 10^{-2}$ and parameter values band = 3 and sigma = 1 for the function blur, (b) restored image.

spread function is sigma^2, where sigma is a user-specified parameter for the function blur. The parameter band specifies the half-bandwidth for the Toeplitz blocks. The larger the parameters band and sigma, the more the original image is smeared out by the blurring operator. The vector $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ represents the (unavailable) blurred noise-free image. The right-hand side of (1), defined by (3), represents the available image contaminated by blur and noise. The noise $\boldsymbol{e}$ is normally distributed with zero mean and scaled to correspond to the different noise levels specified in Table 3.

We let $\eta = 1.001$ and $\delta = 1 \cdot 10^{-3}$. Table 3 displays results achieved for different blurring operators and noise levels. The corresponding contaminated and restored images are shown in Figures 3-6. The restored images are represented by the computed solutions $\tilde{\boldsymbol{x}}$ of (6). In all examples, the norm of the error in the computed solutions $\tilde{\boldsymbol{x}}$ is smaller than the norm of the error in

(a) (b)

Fig. 4. Example 5.3: (a) contaminated image with noise level $1 \cdot 10^{-2}$ and parameter values band = 3 and sigma = 3 for the function blur, (b) restored image.

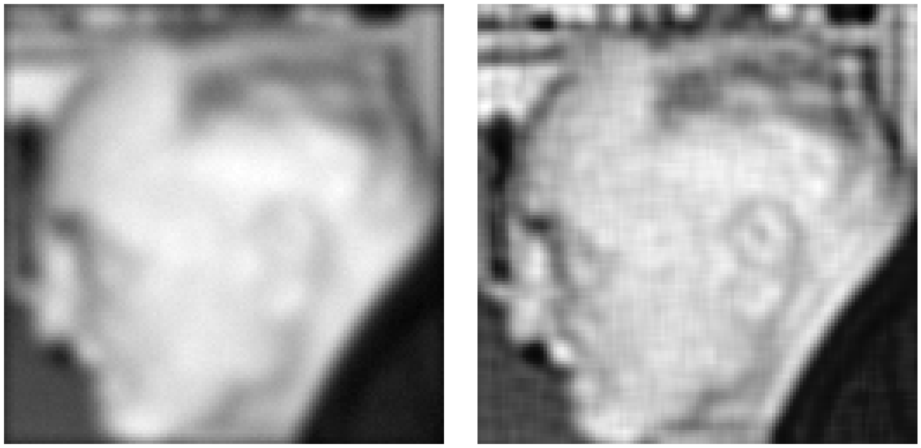
(a) (b)

Fig. 5. Example 5.3: (a) contaminated image with noise level $1 \cdot 10^{-1}$ and parameter values band = 3 and sigma = 3 for the function blur, (b) restored image.

the approximate solutions $\boldsymbol{x}^{(0)}$ determined by projecting the computed unconstrained solution $\boldsymbol{x}_{j_\varepsilon}$ of (5) onto the first orthant. The table also shows the Peak Signal-to-Noise Ratio (PSNR) for the computed solutions $\tilde{\boldsymbol{x}}$ defined by

$$\mathrm{PSNR} = 20 \log_{10} \frac{255}{\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|} \,\mathrm{dB}. \tag{29}$$

This quantity often is used to measure image quality. Large PSNR-values indicate better restoration. □



(a) (b)

Fig. 6. Example 5.3: (a) contaminated image with noise level $1 \cdot 10^{-2}$ and parameter values band = 5 and sigma = 3 for the function blur, (b) restored image.

ρ	sigma	band	$\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|/\|\hat{\boldsymbol{x}}\|$	mat.-vec. prod.	PSNR	figure
$1 \cdot 10^{-2}$	1	3	$3.48 \cdot 10^{-2}$	35	24.71	3
$1 \cdot 10^{-2}$	3	3	$5.61 \cdot 10^{-2}$	9	24.65	4
$1 \cdot 10^{-1}$	3	3	$9.92 \cdot 10^{-2}$	15	23.59	5
$1 \cdot 10^{-2}$	3	5	$8.16 \cdot 10^{-2}$	15	25.18	6

Table 3
Example 5.3: Noise level in the available contaminated images represented by $\boldsymbol{b}$ and shown in Figures 3(a)-6(a), parameters sigma and band for the function blur, relative error in the restored images represented by $\tilde{\boldsymbol{x}}$ and shown in Figures 3(b)-6(b), the number of matrix-vector product evaluations required for the restoration, and PSNR-values for the restored images.

6 Conclusion

This paper presents an interior-point method designed for problems for which an estimate for the error in the right-hand side $\boldsymbol{b}$ is available. The computed examples show the method to require a fairly small number of matrix-vector product evaluations. The storage requirement is modest, just a few vectors in addition to the storage needed to represent A and A^T. The computational effort decreases with increasing noise level.

Acknowledgments

We would like to thank the referees for comments. Work by LR was carried out during visits to the University of Bologna and the Technical University of Berlin. LR would like to thank Fiorella Sgallari and Volker Mehrmann for making these visits possible and enjoyable.

References

[1] G. S. Ammar, W. B. Gragg, and L. Reichel, On the eigenproblem for orthogonal matrices, in: Proceedings of the 25th IEEE Conference on Decision and Control, IEEE, Piscataway, 1986, pp. 1963–1966.

[2] J. M. Bardsley, A nonnegatively constrained trust region algorithm for the restoration of images with an unknown blur, Electron. Trans. Numer. Anal., 20 (2005) 139–153.

[3] M. Bertero and P. Boccacci, Introduction to Inverse Problems in Imaging, Institute of Physics Publishing, Bristol, 1998.

[4] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.

[5] C. F. Borges and W. B. Gragg, A parallel divide and conquer algorithm for the generalized real symmetric definite tridiagonal eigenproblem, in: L. Reichel, A. Ruttan, and R. S. Varga (Eds.), Numerical Linear Algebra and Scientific Computing, de Gruyter, Berlin, 1993, pp. 10–28.

[6] D. Calvetti, G. Landi, L. Reichel, and F. Sgallari, Nonnegativity and iterative methods for ill-posed problems, Inverse Problems 20 (2004) 1747–1758.

[7] D. Calvetti, B. Lewis, L. Reichel, and F. Sgallari, Tikhonov regularization with nonnegativity constraint, Electron. Trans. Numer. Anal. 18 (2004) 153–173.

[8] D. Calvetti and L. Reichel, Gauss quadrature applied to trust region computations, Numer. Algorithms 34 (2003) 85–102.

[9] D. Calvetti and L. Reichel, Tikhonov regularization with a solution constraint, SIAM J. Sci. Comput. 26 (2004) 224–239.

[10] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.

[11] J. W. Evans, W. B. Gragg, and R. J. LeVeque, On the least squares exponential sum approximation, Math. Comp. 34 (1980) 203–211.

[12] N. I. M. Gould, S. Lucidi, M. Roma, and Ph. L. Toint, Solving the trust-region subproblem using the Lanczos method, SIAM J. Optim. 9 (1999) 504–525.

[13] W. B. Gragg, On extrapolation algorithms for ordinary initial value problems, SIAM J. Numer. Anal. 2 (1965) 384–403.

[14] W. B. Gragg, The Padé table and its relation to certain algorithms of numerical analysis, SIAM Review 14 (1972) 1–62.

[15] W. B. Gragg, The QR algorithm for unitary Hessenberg matrices, J. Comput. Appl. Math., 16 (1986), pp. 1–8.

[16] W. B. Gragg and L. Reichel, A divide and conquer method for unitary and orthogonal eigenproblems, Numer. Math. 57 (1990) 695–718.

[17] W. B. Gragg and R. A. Tapia, Optimal error bounds for the Newton-Kantorovich theorem, SIAM J. Numer. Anal. 11 (1974) 10–13.

[18] M. Hanke, Conjugate Gradient Type Methods for Ill-Posed Problems, Longman, Essex, 1995.

[19] M. Hanke, J. Nagy, and C. Vogel, Quasi-Newton approach to nonnegative image restorations, Linear Algebra Appl. 316 (2000) 223–236.

[20] P. C. Hansen, Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems, Numer. Algorithms 6 (1994) 1–35. Software is available in Netlib at `http://www.netlib.org`.

[21] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards B49 (1952) 409–436.

[22] B. Kim, Numerical Optimization Methods for Image Restoration, Ph.D. thesis, Department of Management Science and Engineering, Stanford University, 2002.

[23] S. Morigi, L. Reichel, F. Sgallari, and F. Zama, An iterative method for linear discrete ill-posed problems with box constraints, J. Comput. Appl. Math. 198 (2007) 505–520.

[24] J. Nagy and Z. Strakoš, Enforcing nonnegativity in image reconstruction algorithms, in D. C. Wilson et al. (Eds.) Mathematical Modeling, Estimation and Imaging, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE), vol. 4121, The International Society for Optical Engineering, Bellingham, WA, 2000, pp. 182–190.

[25] J. Nocedal and S. J. Wright, Numerical Optimization, Springer, New York, 1999.

[26] C. C. Paige and M. A. Saunders, LSQR: An algorithm for sparse linear equations with sparse least squares, ACM Trans. Math. Software 8 (1982) 43–71.

[27] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, J. ACM 9 (1962) 84–97.

[28] M. Rojas and T. Steihaug, An interior-point trust-region-based method for large-scale non-negative regularization, Inverse Problems 18 (2002) 1291–1307.

[29] T. Steihaug, The conjugate gradient method and trust regions in large scale optimization, SIAM J. Numer. Anal. 20 (1983) 626–637.