

**Universidad Central de Venezuela  
Facultad de Ciencias  
Escuela de Computación**

***Lecturas en Ciencias de la Computación***  
*ISSN 1316-6239*

**Convex constrained optimization for  
large-scale generalized  
Sylvester equations**

A. Bouhamidi, K. Jbilou y M. Raydan

RT 2008-01

Centro de Cálculo Científico y Tecnológico de la UCV

CCCT-UCV

Caracas, Enero, 2008.

# Convex constrained optimization for large-scale generalized Sylvester equations

A. Bouhamidi,\* K. Jbilou \* and M. Raydan †

January 10, 2008

## Abstract

We propose and study the use of convex constrained optimization techniques for solving large-scale Generalized Sylvester Equations (GSE). For that, we adapt recently developed globalized variants of the projected gradient method to a convex constrained least-squares approach for solving GSE. We demonstrate the effectiveness of our approach on two different applications. First, we apply it to solve the GSE that appear after applying left and right preconditioning schemes to the linear problems associated with the discretization of classical PDE problems. Second, we apply the new approach, combined with a Tikhonov regularization term, to restore some blurred and highly noisy images.

**Key words:** Convex optimization; Spectral projected gradient method; Generalized Sylvester equation; Image restoration.

## 1 Introduction

Consider the generalized Sylvester matrix equation

$$\sum_{i=1}^q A_i X B_i = C, \quad (1)$$

where  $q$  is a positive integer,  $A_i \in \mathbb{R}^{n \times n}$ ;  $B_i \in \mathbb{R}^{p \times p}$ ; for  $i = 1, \dots, q$ ,  $C \in \mathbb{R}^{n \times p}$ , and  $X$  is the unknown matrix in  $\mathbb{R}^{n \times p}$ . If the matrices are large and ill-conditioned, then solving (1) requires in general the incorporation of a regularization strategy.

Equation (1) is of interest in many different applications, including eigendecomposition of matrix pencils [17], image restoration [8], and the numerical solution of implicit ordinary

---

\*L.M.P.A., Université du Littoral, 50 rue F. Buisson BP699, F-62228, Calais-Cedex, France (jbilou@lmpa.univ-littoral.fr; bouhamidi@lmpa.univ-littoral.fr)

†Departamento de Computación, Facultad de Ciencias, Universidad Central de Venezuela, Ap. 47002, Caracas 1041-A, Venezuela (mraydan@kuaimare.ciens.ucv.ve). Sponsored by the Center of Scientific Computing at UCV.

differential equations [20]. It is also of general interest since it includes as particular cases several classical and important linear problems in the space of matrices: commuting matrices, Block linear systems, standard Sylvester equation, and Lyapunov equation, among others. Another interesting feature related to equation (1) is that it allows to incorporate in a natural way left or right preconditioning strategies, for any of the previously mentioned linear matrix problems.

For solving special cases of the linear equation (1) several schemes have been proposed, including factorization techniques for small size problems (see e. g., [3, 24, 29]) and iterative schemes, for large-scale problems, based on projection methods that produce a low-dimensional linear equation that is then solved using direct methods (see e.g., [10, 19, 32, 33, 41]). For a novel iterative approach see [39], and for a complete review on iterative methods for large linear matrix equations, see [16]. More recently, global extensions of the well-known Krylov subspace methods (FOM and GMRES) have been proposed and analyzed to solve several large-scale linear matrix equations directly on the space of matrices (see e. g., [34, 35, 36]). However, all these previous ideas show difficulties when dealing with ill-conditioned large-scale problems.

In this work we propose to solve an equivalent constrained optimization problem instead. For that, we consider the operator  $\mathcal{A}$  defined as follows

$$\begin{aligned} \mathcal{A} : \mathbb{R}^{n \times p} &\longrightarrow \mathbb{R}^{n \times p} \\ X &\longrightarrow \sum_{i=1}^q A_i X B_i. \end{aligned}$$

Let  $\|Z\|_F = (\langle Z, Z \rangle_F)^{\frac{1}{2}}$  be the Frobenius matrix norm of a matrix  $Z$ ,  $\langle Z, Y \rangle_F = \text{tr}(Z^T Y)$  is the associated inner product of  $Z$  with a matrix  $Y$ , and  $\text{tr}(W)$  denotes the trace of a square matrix  $W$ . By using properties of the trace operator, we have that for any matrices  $W$ ,  $Y$ , and  $Z$ ,  $\langle W, YZ \rangle_F = \langle Y^T W, Z \rangle_F = \langle WZ^T, Y \rangle_F$ . Combining these properties we obtain that the transpose of the operator  $\mathcal{A}$  with respect to the inner product  $\langle \cdot, \cdot \rangle_F$  is defined from  $\mathbb{R}^{n \times p}$  onto  $\mathbb{R}^{n \times p}$  by

$$\mathcal{A}^T(X) = \sum_{i=1}^q A_i^T X B_i^T.$$

Instead of the generalized Sylvester equation (1), we consider the following constrained minimization problem

$$\text{Minimize } f(X) \quad \text{subject to } X \in \Omega, \quad (2)$$

where

$$f(X) = \|\mathcal{A}(X) - C\|_F^2. \quad (3)$$

The set  $\Omega$  could be a simple convex set (e. g., a sphere or a box) or the intersection of some simple convex sets. Specific cases that will be considered are

$$\Omega_1 = \{X \in \mathbb{R}^{n \times p} : L \leq X \leq U\} \quad (4)$$

and also

$$\Omega_2 = \{X \in \mathbb{R}^{n \times p} : \|X\|_F \leq \delta\}, \quad (5)$$

where  $L$  and  $U$  are given matrices and  $\delta > 0$  is a given scalar. Another option to be considered is  $\Omega = \Omega_1 \cap \Omega_2$ . In here,  $Y \leq Z$  means  $Y_{ij} \leq Z_{ij}$  for all possible entries  $ij$ . Choosing the matrices  $L$  and  $U$  and/or the positive parameter  $\delta$  in a suitable way, produces the desired regularization effect.

## 2 Projected gradient methods

In the ill-conditioned case, the linear equation (1) can be solved minimizing the function  $f$  given by (3) over a suitable convex set  $\Omega$ . Projected Gradient (PG) methods provide an interesting option for solving such large-scale ill-conditioned convex constrained problems. They are simple and easy to code, and avoid the need for matrix factorizations. Moreover, regularization can be imposed by choosing  $\Omega$  in a suitable way. Early references on PG methods can be traced back to Goldstein [23] and Levitin and Polyak [38], where constant step lengths are used. A modified and practical monotone backtracking line search is later introduced by Bertsekas [4] to the choice of step length. However, these early PG methods are frequently inefficient since their performance resembles the optimal gradient method (also known as the steepest descent method), which is usually very slow. Nevertheless, as we will discuss in this section, the effectiveness of PG methods can be greatly improved by incorporating recently developed choices of step length and nonmonotone globalization strategies.

There have been many different variations of the early PG methods. They all have the common property of maintaining feasibility of the iterates by frequently projecting trial steps on the feasible convex set. In particular, Birgin et al. [5, 6] combine the projected gradient method with recently developed ingredients in unconstrained optimization to propose an effective scheme that is known as the Spectral Projected Gradient (SPG) method. In our setting, the algorithm starts with  $X_0 \in \mathbb{R}^{n \times p}$ , and moves at every iteration  $k$  along the internal projected gradient direction  $D_k = P_\Omega(X_k - \alpha_k \nabla f(X_k)) - X_k$ , where  $\alpha_k$  is the spectral (also known as the Barzilai-Borwein [2]) choice of step length

$$\alpha_k = \frac{\langle S_{k-1}, S_{k-1} \rangle}{\langle S_{k-1}, Y_{k-1} \rangle},$$

$S_{k-1} = X_k - X_{k-1}$ ,  $Y_{k-1} = \nabla f(X_k) - \nabla f(X_{k-1})$ , and for  $Z \in \mathbb{R}^{n \times p}$ ,  $P_\Omega(Z)$  is the projection of  $Z$  onto  $\Omega$ . In the case of rejection of the first trial point,  $X_k + D_k$ , the next ones are computed along the same direction, i.e.,  $X_+ = X_k + \lambda D_k$ , using an extension of the Grippo, Lampariello and Lucidi (GLL) [28] nonmonotone line search to choose  $0 < \lambda \leq 1$  such that the following condition holds

$$f(X_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(X_{k-j}) + \gamma \lambda \langle D_k, \nabla f(x_k) \rangle,$$

where  $M \geq 1$  is a given integer and  $\gamma$  is a small positive number. As a consequence, the projection operation must be performed only once per iteration. More details can be found in [5] and [6].

As we mentioned before, the SPG method is related to the classical projected gradient method. However, some crucial differences make this method much more efficient than its gradient projection predecessors. The key issue is that the first trial step at each iteration is taken using the spectral step length introduced in [2] and later analyzed in [15], [21], [40] among others. The spectral step is a Rayleigh quotient related with an average Hessian matrix. For a review containing the more recent advances on this special choice of step length see [22].

The second important issue is the use of nonmonotone line search strategies to globalize the process. This feature seems to be essential to preserve the nice and nonmonotone behavior of the iterates produced by single spectral gradient steps. In this work, we enrich further the globalization technique by combining the GLL line search scheme with the recently proposed and analyzed globalization scheme of La Cruz *et al.* [37]. Roughly speaking our acceptance condition for the next iterate is

$$f(X_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(X_{k-j}) + \gamma \lambda \langle D_k, \nabla f(x_k) \rangle + \eta_k,$$

where  $\eta_k \geq 0$  is chosen such that

$$\sum_k \eta_k < \infty. \quad (6)$$

The terms  $\max_{0 \leq j \leq \min\{k, M-1\}} f(X_{k-j})$  and  $\eta_k$  are responsible for the sufficiently non-monotone behavior of  $f(X_k)$ .

For the sake of completeness, we now present the algorithm used in this work. It starts with  $X_0 \in \Omega$  and uses an integer  $M \geq 1$ ; a small parameter  $\alpha_{\min} > 0$ ; a large parameter  $\alpha_{\max} > \alpha_{\min}$ ; a sufficient decrease parameter  $\gamma \in (0, 1)$  and safeguarding parameters  $0 < \sigma_1 < \sigma_2 < 1$ . Initially,  $\alpha_0 \in [\alpha_{\min}, \alpha_{\max}]$  is arbitrary.

Given  $X_k \in \Omega$  and  $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ , our extended version of the SPG algorithm describes how to obtain  $X_{k+1}$  and  $\alpha_{k+1}$ , and when to terminate the process.

### Extended SPG

Step 1: Detect whether the current point is stationary. If

$$\|P_\Omega(X_k - \nabla f(X_k)) - X_k\|_F = 0,$$

stop, declaring that  $X_k$  is stationary.

Step 2: Backtracking

Step 2.1: Set  $D_k = P_\Omega(X_k - \alpha_k \nabla f(X_k)) - X_k$ , and  $\lambda \leftarrow 1$

Step 2.2: Set  $X_+ = X_k + \lambda D_k$

Step 2.3: If

$$f(X_+) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(X_{k-j}) + \gamma \lambda \langle D_k, \nabla f(x_k) \rangle + \eta_k \quad (7)$$

then  $\lambda_k = \lambda$ ,  $X_{k+1} = X_+$ ,  $S_k = X_{k+1} - X_k$ ,  $Y_k = \nabla f(X_{k+1}) - \nabla f(X_k)$ , go to Step 3.

If (7) does not hold, define  $\lambda_{new} \in [\sigma_1 \lambda, \sigma_2 \lambda]$ ,  $\lambda \leftarrow \lambda_{new}$  and go to Step 2.2.

Step 3: Compute  $b_k = \langle S_k, Y_k \rangle$ , if  $b_k \leq 0$ , set  $\alpha_{k+1} = \alpha_{max}$ , else, compute  $\alpha_k = \langle S_k, S_k \rangle$  and  $\alpha_{k+1} = \min(\alpha_{max}, \max(\alpha_{min}, \frac{a_k}{b_k}))$ .

For the calculation of  $\lambda_{new}$  in Step 2.3 we can use the one dimensional quadratic interpolation as described in Birgin et al [5]. If the set  $\Omega$  is a simple set (e. g., box or sphere), then projecting is an easy task as described in our next section. If  $\Omega$  is the intersection of several (finite) simple convex sets, then projecting requires the use of more advanced schemes. An interesting option is provided by Dykstra's alternating projection algorithm [9].

### 3 Ingredients for the optimization approach

The function  $f : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}$  defined in (3) is differentiable and its gradient can be obtained as follows. Consider the auxiliary function  $g : \mathbb{R} \rightarrow \mathbb{R}$ , given by

$$g(t) = f(X + tP),$$

for any arbitrary matrix  $P$ . From basic calculus we know that  $g'(0) = \langle \nabla f(X), P \rangle_F$ . After simple algebraic manipulations it follows that

$$g'(0) = 2\langle \mathcal{A}(X) - C, \mathcal{A}(P) \rangle_F = 2\langle \mathcal{A}^T(\mathcal{A}(X) - C), P \rangle_F, \quad (8)$$

and so the gradient of  $f$  is given by

$$\nabla f(X) = 2\mathcal{A}^T(\mathcal{A}(X) - C) = 2 \sum_{i=1}^q A_i^T(\mathcal{A}(X) - C)B_i^T.$$

Notice that computing the gradient of  $f$  at a given  $X$  requires  $O(p)$  matrix-matrix products.

We now describe the projection onto the possible convex sets to be considered. One option is the *box*  $\Omega_1$  (bounds on the entries of  $X$ ) defined in (4). In that case, the matrices  $L$  and  $U$  are given by entry-wise lower and upper bounds. Some or all of the entries in the matrix  $L$  might have the value  $-\infty$ , and similarly some or all of the entries in the matrix  $U$  might have the value  $+\infty$ . Clearly, when  $L_{ij} = -\infty$  and  $U_{ij} = +\infty$  for all  $ij$ , then  $\Omega_1 = \mathbb{R}^{n \times p}$ , and we are dealing with the unconstrained minimization of  $f$ .

When the feasible region is  $\Omega_1$ , the projection (minimal distance) of a given matrix  $Z \in \mathbb{R}^{n \times p}$ ,  $P_{\Omega_1}(Z)$ , is obtained as the unique solution to the problem

$$\min_{X \in \Omega_1} \|X - Z\|_F,$$

whose  $ij$ -entry is given by

$$(P_{\Omega_1}(Z))_{ij} = \begin{cases} Z_{ij} & \text{if } L_{ij} \leq Z_{ij} \leq U_{ij} \\ U_{ij} & \text{if } Z_{ij} > U_{ij} \\ L_{ij} & \text{if } Z_{ij} < L_{ij}. \end{cases} \quad (9)$$

Another option to be considered is when the feasible region is the *sphere*  $\Omega_2$  defined in (5). In that case, the radius  $\delta > 0$  is given. As before,  $\delta$  might have the value  $+\infty$ , in which case  $\Omega_2 = \mathbb{R}^{n \times p}$ , and we are dealing with the unconstrained minimization of  $f$ .

When the feasible region is  $\Omega_2$ , the projection (minimal distance) of a given matrix  $Z \in \mathbb{R}^{n \times p}$ ,  $P_{\Omega_2}(Z)$ , is obtained as the unique solution to the problem

$$\min_{X \in \Omega_2} \|X - Z\|_F,$$

that is given by

$$P_{\Omega_2}(Z) = \begin{cases} Z & \text{if } \|Z\|_F \leq \delta \\ (\frac{\delta}{\|Z\|_F}) Z & \text{if } \|Z\|_F > \delta. \end{cases} \quad (10)$$

A third option is to consider  $\Omega = \Omega_1 \cap \Omega_2$  as the feasible convex region. In that case, we can use the inexact version of the SPG method, discussed in [7], that combines the SPG with Dykstra's alternating projection algorithm to find the projection onto the intersection inexactly (avoiding an excessive computational effort).

## 4 Standard Sylvester equations and preconditioning

We now consider the solution of standard Sylvester equations to illustrate the possibility of incorporating, in a natural way, left or right preconditioning strategies for solving several classical problems in the space of matrices, when dealing with the generalized matrix equation (1).

Consider the Sylvester equation

$$AX - XB = \widehat{C}, \quad (11)$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{p \times p}$ ,  $\widehat{C} \in \mathbb{R}^{n \times p}$ , and  $X$  is the unknown matrix in  $\mathbb{R}^{n \times p}$ . Equation (11) can be seen as a special case of (1) by setting  $q = 2$ ,  $A_1 = A$ ,  $B_2 = -B$ ,  $B_1 = I_p$ ,  $A_2 = I_n$ , and  $C = \widehat{C}$ .

Let us now assume that, as it happens frequently in applications, one of the dimensions ( $n$  or  $p$ ) is much larger than the other. For example, let us assume that  $n \gg p$ . Let us also assume that a suitable approximation, say  $W$ , to the inverse of  $A$  is available. In what follows the matrix  $W$  will play the role of a left (inverse) preconditioner for problem (11). By multiplying (11) from the left by  $W$  we clearly obtain a new special case of (1), given by

$$WAX - WXB = W\widehat{C}, \quad (12)$$

where  $q = 2$ ,  $A_1 = WA$ ,  $B_2 = -B$ ,  $B_1 = I_p$ ,  $A_2 = W$ , and  $C = W\widehat{C}$ . Hopefully (12) is easier to solve than (11). However, notice that (12) is no longer a standard Sylvester equation, and so classical schemes for solving (11) cannot be applied anymore. Nevertheless, in both cases we can apply the extended SPG algorithm described in Section 2, imposing regularity by choosing a suitable convex set.

Similarly, if  $p \gg n$  and a suitable approximation, say  $R$ , to the inverse of  $B$  is available, then  $R$  can be used as a right (inverse) preconditioner in (11) to obtain

$$AXR - XBR = \widehat{C}R, \quad (13)$$

where now  $q = 2$ ,  $A_1 = A$ ,  $B_2 = -BR$ ,  $B_1 = R$ ,  $A_2 = I_n$ , and  $C = \widehat{C}R$ . Once again, the extended SPG algorithm described in Section 2 can be applied to solve (13).

To illustrate the effectiveness of our approach we now present some numerical experiments. The Extended SPG method was implemented in Matlab 7.0 on an Intel Pentium workstation with about 16 significant decimal digits. The parameters in our extended SPG implementation are chosen in a standard way as follows:  $M = 10$ ,  $\eta_k = |f(X_0)|/k^{1.1}$ ,  $\alpha_{\min} = 1.D - 15$ ,  $\alpha_{\max} = 1.D + 15$ ,  $\gamma = 1.D - 4$ ,  $\sigma_1 = 0.1$ , and  $\sigma_2 = 0.9$ . The process was stopped when the norm of the residual satisfies

$$\|P_\Omega(X_k - \nabla f(X_k)) - X_k\|_F = tol,$$

for  $tol = 1.D - 5$ . We present the numerical results obtained for some experiments for which the matrices are obtained by discretizing classical PDE problems, introduced in [36], as follows. The matrix  $A$  in (11) is generated from the 5-point discretization of the operator

$$L_A(u) = \Delta u - f_1(x, y) \frac{\partial u}{\partial x} - f_2(x, y) \frac{\partial u}{\partial y} - f_3(x, y)u$$

on the unit square  $[0, 1] \times [0, 1]$  with homogeneous Dirichlet boundary conditions. We set  $f_1(x, y) = e^{x+y}$ ,  $f_2(x, y) = 100y$ , and  $f_3(x, y) = x$ . Similarly, the matrix  $B$  in (11) is generated from the 5-point discretization of the operator

$$L_B(u) = \Delta u - g_1(x, y) \frac{\partial u}{\partial x} - g_2(x, y) \frac{\partial u}{\partial y} - g_3(x, y)u$$

on the unit square  $[0, 1] \times [0, 1]$  with homogeneous Dirichlet boundary conditions. We set  $g_1(x, y) = \sin(x + 2y)$ ,  $g_2(x, y) = e^y$ , and  $g_3(x, y) = xy$ .

The matrix  $\widehat{C}$  is chosen such that a preestablished  $X^*$  solves (11). We set  $X^*$  as a random matrix with entries in  $[0, 1]$ , and we start from  $X_0 = 0$  and  $\alpha_0 = 1/\|\nabla f(X)\|_2$ . The dimensions of  $A$  and  $B$  are  $n = n_0^2$  and  $p = p_0^2$  respectively, where  $n_0$  and  $p_0$  are the number of internal grid points in each direction. Our convex set is the box  $\Omega_1$  defined by  $L_{ij} = 0$  and  $U_{ij} = 1$  for all entries  $i, j$ .

We consider, without loss of generality, the case  $n \gg p$ , and use three possible options for preconditioning:  $W = I$  (no preconditioning),  $W = W_1 \equiv \widehat{L}^{-1}$ , and  $W = W_2 \equiv \widehat{U}^{-1}\widehat{L}^{-1}$ , where  $\widehat{L}$  and  $\widehat{U}$  come from an incomplete  $LU$  factorization of  $A$  given by the Matlab command  $[\widehat{L}, \widehat{U}] = luinc(A, 1e - 3)$ . Notice that  $\widehat{L}$  and  $\widehat{U}$  are sparse matrices. Notice also that, for approximating the inverse of  $A$ ,  $W_2$  is better than  $W_1$ . The results are reported in Table 1. We can observe the reduction in number of required iterations, for solving large-scale Sylvester equations, when the quality of the preconditioning strategy is improved.

Table 1: Performance of the extended SPG algorithm for solving (12) for different preconditioning strategies and different dimensions

Preconditioning strategy	$n = 100, p = 4$		$n = 900, p = 16$	
	Iter	Residual norm	Iter	Residual norm
$I$	1375	1.8D-7	> 5000	***
$W_1$	293	7.1D-8	2765	6.9D-7
$W_2$	33	8.6D-6	792	8.5D-7

## 5 Application to image restoration

Consider the linear discrete ill-posed problem

$$\min_{x \in \tilde{\Omega}} \|Hx - g\|_2, \quad (14)$$

where  $H \in \mathbb{R}^{M \times N}$ ,  $x \in \mathbb{R}^N$ ,  $g \in \mathbb{R}^M$  and  $M \geq N$ . The set  $\tilde{\Omega} \subset \mathbb{R}^N$  could be a simple convex set (e. g., a sphere or a box) or the intersection of some simple convex sets. The matrix  $H$  is of ill-determined rank, i.e.,  $H$  has many singular values of different orders of magnitude close to the origin. The matrix  $H$  is assumed to be very large so its factorization is undesirable.

The right-hand side vector  $g$  in (14) represents the available output and is assumed to be contaminated by an error (noise)  $\mathbf{n}$ , i.e.,  $g = \hat{g} + \mathbf{n}$ .

Such a system (14) arises in image restoration problems. The problem consists of the reconstruction of an original image that has been digitized and has been degraded by a blur and an additive noise. The matrix  $H$  represents the blurring matrix, the vector  $x$  to be approximated represents the original image, the vector  $\mathbf{n}$  represents the additive noise and the vector  $g$  represents the blurring and noisy (degraded) image. Additional details about image restoration can be found in the books by Chan and Shen [14], and Hansen, Nagy, and O’Leary [31].

The noise in the measurements, in combination with the ill conditioning of  $H$ , means that the exact solution of (14) has little relationship to the noise-free solution.

In the context of image restoration, when the point spread function (PSF) is separable, the blurring matrix  $H$  is given as a Kronecker product  $H = H_2 \otimes H_1$  of two blurring matrices where  $H_1$  and  $H_2$  are of size  $n \times n$  and  $p \times p$ , respectively, and  $\otimes$  denotes the Kronecker product (see, e.g., [31]). Using some properties of the Kronecker product, problem (14) is written as

$$\min_{X \in \tilde{\Omega}} \|\mathcal{A}(X) - G\|_F, \quad (15)$$

where  $\mathcal{A}(X) = H_1 X H_2^T$  with  $G$  and  $X$  such that  $\text{vec}(G) = g$ ,  $\text{vec}(X) = x$  where  $\text{vec}(X)$  is the vector obtained by stacking the columns of the matrix  $X$ . The set  $\tilde{\Omega}$  is such that

$$x = \text{vec}(x) \in \tilde{\Omega} \subset \mathbb{R}^N \Leftrightarrow X \in \Omega \subset \mathbb{R}^{n \times p},$$

and  $M = N = np$ .

## 5.1 Adding a Tikhonov regularization term

Consider, for a moment, the unconstrained linear discrete ill-posed problem

$$\min_x \|Hx - g\|_2. \quad (16)$$

One of the most popular regularization methods for solving (16) is due to Tikhonov [26]. In that case, problem (16) is replaced by the new one

$$\min_x (\|Hx - g\|_2^2 + \lambda^2 \|Lx\|_2^2), \quad (17)$$

where  $L$  is a regularization operator chosen to obtain a solution with desirable properties such as small norm or good smoothness. The minimizer of problem (17) is computed as the solution of the following linear system

$$H_\lambda x = H^T g, \text{ where } H_\lambda = (H^T H + \lambda^2 L^T L). \quad (18)$$

In some practical problems as in image processing, the solution of problem (18) must belong to some domain  $\tilde{\Omega}$ . Furthermore, problem (18) may be regarded as a minimization problem. Thus, instead of considering problem (18) we will consider hereafter the following minimization problem with constraints

$$\min_{x \in \tilde{\Omega}} \|H_\lambda x - H^T g\|_2. \quad (19)$$

We assume here that  $H = H_2 \otimes H_1$  and  $L = L_2 \otimes L_1$  where  $H_1, L_1$  are square matrices of dimension  $n \times n$  and  $H_2, L_2$  are square matrices of dimension  $p \times p$ . Then, problem (19) is written as

$$\min_{X \in \tilde{\Omega}} \|[ (H_2 \otimes H_1)^T (H_2 \otimes H_1) + \lambda^2 (L_2 \otimes L_1)^T ] \text{vec}(X) - (H_2 \otimes H_1)^T \text{vec}(G)\|_2, \quad (20)$$

where  $G$  and  $X$  are the matrices such that  $\text{vec}(G) = g$ ,  $\text{vec}(X) = x$ . Using some properties of the Kronecker product, problem (20) is then written as

$$\min_{X \in \tilde{\Omega}} \|\mathcal{A}_\lambda(X) - E\|_F, \quad (21)$$

where  $\mathcal{A}_\lambda(X) = AXD - \lambda^2 CXB$ , with  $A = H_1^T H_1$ ,  $B = L_2^T L_2$ ,  $C = -L_1^T L_1$ ,  $D = H_2^T H_2$  and  $E = H_1^T G H_2$ .

The parameter  $\lambda$  is a scalar to be determined. The L-curve criterion [11, 12, 18] and the Generalized cross-validation (GCV) method [25] are robust techniques for determining the optimal parameter  $\lambda$ .

When using the generalized cross-validation (GCV) method [25], the regularization parameter is chosen to minimize the GCV function

$$GCV(\lambda) = \frac{\|H\hat{x}_\lambda - \mathbf{g}\|_2^2}{[\text{tr}(I - HH_\lambda^{-1}H^T)]^2} = \frac{\|(I - HH_\lambda^{-1}H^T)\mathbf{g}\|_2^2}{[\text{tr}(I - HH_\lambda^{-1}H^T)]^2}$$

where  $H_\lambda = H^T H + \lambda^2 L^T L$ .

Let  $H = H_2 \otimes H_1$  and  $L = L_2 \otimes L_1$  where  $H_1, L_1$  and  $H_2, L_2$  are of size  $n \times n$  and  $p \times p$ , respectively and consider the Generalized Singular Value Decompositions (GSVD) [27] of the pairs  $(H_1, L_1)$  and  $(H_2, L_2)$ . Thus, there exist orthonormal matrices  $U_1, U_2, V_1, V_2$  and invertible matrices  $X_1, X_2$  such that

$$\begin{aligned} U_1^T H_1 X_1 &= C_1 = \text{diag}(c_{1,1}, \dots, c_{n,1}), \quad c_{i,1} \geq 0, \\ U_2^T H_2 X_2 &= C_2 = \text{diag}(c_{1,2}, \dots, c_{p,2}), \quad c_{i,2} \geq 0, \end{aligned}$$

and

$$\begin{aligned} V_1^T L_1 X_1 &= S_1 = \text{diag}(s_{1,1}, \dots, s_{n,1}), \quad s_{i,1} \geq 0, \\ V_2^T L_2 X_2 &= S_2 = \text{diag}(s_{1,2}, \dots, s_{p,2}), \quad s_{i,2} \geq 0. \end{aligned}$$

Then the GSVD of the pair  $(H, L)$  is given by

$$\begin{aligned} U^T H X &= C = \text{diag}(c_1, \dots, c_N), \quad c_i \geq 0, \\ V^T L X &= S = \text{diag}(s_1, \dots, s_N), \quad s_i \geq 0, \end{aligned}$$

where  $U = U_2 \otimes U_1, V = V_2 \otimes V_1, C = C_2 \otimes C_1, S = S_2 \otimes S_1$  and  $N = np$ . Therefore, one can show that the expression of the GCV function is given by

$$GCV(\lambda) = \frac{\sum_{i=1}^N \left( \frac{s_i^2 \tilde{g}_i}{c_i^2 + \lambda^2 s_i^2} \right)^2}{\left( \sum_{i=1}^N \frac{s_i^2}{c_i^2 + \lambda^2 s_i^2} \right)^2}, \quad (22)$$

where  $\tilde{g} = U^T g$ .

The second method is the L-curve criterion [30]. The method suggests to plot the L-curve  $(\|H_1 \hat{X}_\lambda H_2^T - G\|_F, \|L_1 \hat{X}_\lambda L_2^T\|_F)$ . Using the preceding GSVD decomposition of the pair  $(H, L)$ , it is not difficult to show the following relations

$$\begin{aligned} \|H_1 \hat{X}_\lambda H_2^T - G\|_F^2 &= \lambda^2 \sum_{i=1}^N \left( \frac{s_i^2 \tilde{g}_i}{c_i^2 + \lambda^2 s_i^2} \right)^2 \\ \|L_1 \hat{X}_\lambda L_2^T\|_F^2 &= \sum_{i=1}^N \left( \frac{s_i c_i \tilde{g}_i}{c_i^2 + \lambda^2 s_i^2} \right)^2. \end{aligned} \quad (23)$$

The best regularization parameter should lie on the corner of the L-curve. The L-curve method chooses the regularization parameter corresponding to the point on the curve with maximum curvature. For more details on the numerical comparison between different methods for the choice of the optimal regularization parameter see [26, 30]. In our numerical tests, we used a robust method called the method of Triangles [13] to determine the optimal regularization parameter.

## 5.2 Experimental results

Numerical tests in image restoration are presented to show the performance of our proposed method. Our codes were written with MATLAB 7.0 on an Intel Pentium workstation with about 16 significant decimal digits. We constructed our tests by taking a known image denoted by  $\widehat{X}$  that consists of  $256 \times 256$  grayscale pixel values in the range  $[0, 255]$ . We set  $n = p = 256$ . Let  $\widehat{x}$  denote the vector whose entries are the pixel values of the original image  $\widehat{X}$  and let  $H$  be the blurring matrix given as a Kronecker product  $H = H_2 \otimes H_1$ , where  $H_1$  and  $H_2$  are Toeplitz matrices of dimension  $256 \times 256$ . The vector  $\widehat{g} = H\widehat{x}$  represents the associated blurred and noise-free image. We generated a blurred and noisy image  $g = \widehat{g} + \mathbf{n}$ , where  $\mathbf{n}$  is a noise vector with normally distributed Gaussian random entries with zero mean and with variance chosen such that the Signal to Noise Ratio (*SNR*) has an appropriate (dB) value. We recall that the *SNR* is given by

$$SNR = 10 \log_{10} \left( \frac{\sigma_{\widehat{x}}^2}{\sigma_{\mathbf{n}}^2} \right),$$

where  $\sigma_{\mathbf{n}}^2$  and  $\sigma_{\widehat{x}}^2$  are the variance of the noise and the original image, respectively. The performances of the proposed algorithm are evaluated by computing the Improvement in Signal to Noise Ratio (*ISNR*) defined by

$$ISNR = 10 \log_{10} \left( \frac{\|\widehat{x} - g\|_2^2}{\|\widehat{x} - x\|_2^2} \right) = 10 \log_{10} \left( \frac{\|\widehat{X} - G\|_F^2}{\|\widehat{X} - X\|_F^2} \right),$$

where  $x = \text{vec}(X)$ ,  $g = \text{vec}(G)$  and  $X$  is the restored image. To evaluate the precision of the estimates, the following relative error is also computed

$$R_{er}(X) = \frac{\|\widehat{X} - X\|_F^2}{\|\widehat{X}\|_F^2}.$$

Now, we describe a way to choose the domains  $\Omega_1$  and  $\Omega_2$  given by (4) and (5). We have

$$\Omega_1 = \{Y \in \mathbb{R}^{n \times p} : L_b \leq Y \leq U_b\} \quad (24)$$

and also

$$\Omega_2 = \{Y \in \mathbb{R}^{n \times p} : \|Y\|_F \leq \delta\}. \quad (25)$$

In our tests, the domain  $\Omega_2$  was chosen such that  $\delta = \|G\|_F$ , where  $G$  is the degraded image. Since an image consists of grayscale pixel values in the range  $[0, 255]$ , a first option to choose the domain  $\Omega_1$  is to define the lower bound-matrix  $L_b$  to be the zero matrix and the upper bound-matrix as  $U_b = 255 \times \mathbf{1}$ , where  $\mathbf{1}$  is the matrix whose entries are all equal to 1. In order to define local smoothing constraints, another option to choose the domain  $\Omega_1$  is to determine the bound matrices  $L_b$  and  $U_b$  from the parameters that describe the local properties of an image. In our tests, we used the local variance for local activity and local maximum intensity value, see [1]. For the degraded image  $G$ , the local mean matrix  $\overline{G}$  and the local variance  $\sigma_G^2$  are measured over a  $3 \times 3$  window given by

$$\overline{G}(i, j) = \frac{1}{9} \sum_{l=i-3}^{i+3} \sum_{k=j-3}^{j+3} G(l, k), \quad \text{and} \quad \sigma_G^2(i, j) = \frac{1}{9} \sum_{l=i-3}^{i+3} \sum_{k=j-3}^{j+3} [G(l, k) - \overline{G}(l, k)]^2.$$

The maximum local variance over the entire image  $G$ , denoted by  $\sigma_{max}^2$  is given by

$$\sigma_{max}^2 = \max_{1 \leq i, j \leq 256} \sigma_G^2(i, j).$$

Let  $\beta > 0$  be a positive constant, the matrices  $L_b$  and  $U_b$  defining the domain  $\Omega_1$  are given by

$$L_b(i, j) = \max(\overline{G}(i, j) - \beta \frac{\sigma_G^2(i, j)}{\sigma_{max}^2}, 0), \quad \text{and} \quad U_b(i, j) = \overline{G}(i, j) + \beta \frac{\sigma_G^2(i, j)}{\sigma_{max}^2}. \quad (26)$$

The constant  $\beta$  controls the tightness of the bounds. In the following examples, the domain  $\Omega_1$  with the lower bound-matrix  $L_b$  and upper bound-matrix  $U_b$ , specified by (26), is denoted by  $\Omega_{1, \beta}$  and the domain  $\Omega_1$  with the lower bound-matrix  $L_b = 0$  and the upper bound-matrix is  $U_b = 255 \times \mathbf{1}$  still denoted by  $\Omega_1$ .



Figure 1: Original images: "Lena" (left) and "Cameraman" (right).

**Example 1:** In the first example, the original image is the "lena" image shown on the left side of Figure 1. The blurring matrix  $H$  is given by  $H = H_2 \otimes H_1 \in \mathbb{R}^{256^2 \times 256^2}$ , where  $H_1 = I_{256}$  is the identity matrix and  $H_2 = [h_{ij}]$  is the Toeplitz matrix of dimension  $256 \times 256$  given by

$$h_{ij} = \begin{cases} \frac{1}{2r-1}, & |i - j| \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

The blurring matrix  $H$  models a uniform blur. In our example we set  $r = 3$ . A white random Gaussian noise of a specific variance is added to produce a blurred and noisy image  $G$ , with  $SNR = 5dB$ , and is shown on the left side of Figure 3. The restoration of the image from the degraded image is obtained in this example by solving the minimization problem (21) using the *SPG* algorithm. The regularization matrix  $L = L_2 \otimes L_1 \in \mathbb{R}^{256^2 \times 256^2}$  is

chosen such that  $L_1 = I_{256}$  and  $L_2$  is the tridiagonal matrix of size  $256 \times 256$  given by

$$L_2 = \begin{pmatrix} 2 & 1 & & & \\ 1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & 1 \\ & & & 1 & 2 \end{pmatrix}.$$

The projected domain in the minimization problem (21) was  $\Omega = \Omega_{1,\beta}$  with  $\beta = 0.001$ . In this example, we used the L-curve criterion to compute the optimal value of the parameter  $\lambda$ . The corner of the L-curve is localized by using the Triangle method [13]. Figure 2 shows the behavior of the L-curve and the localization of its corner. The value of the optimal value of  $\lambda$  was  $\lambda_{opt} = 0.04$ . The obtained approximation  $X$  represents the

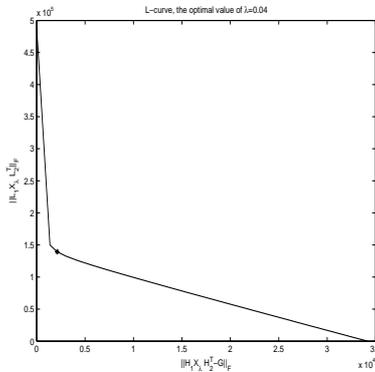


Figure 2: The L-curve with the optimal value located at the \* point.

restored image and is represented on the right side of Figure 3. The relative error was  $R_{er}(X) \simeq 1.1217 \times 10^{-1}$  and the  $ISNR \simeq 2.9646$ . We tested the proposed algorithm for various values of the band  $r$  of the matrix  $H_2$ , for various values of the  $SNR$  and for various values of the parameter  $\beta$  by choosing successively the projected domain  $\Omega$  in the minimization problem (21) equal to  $\Omega_1$ ,  $\Omega_2$  and  $\Omega_1 \cap \Omega_2$  where  $L_b$ ,  $U_b$  and  $\delta$  are as specified, previously. We also tested the problem of the image restoration by solving the minimization problem (15) and by using the Extended  $SPG$  algorithm with different domains presented above. Some results are reported in Table 5.2. According to our numerical tests, there is a slight advantage when adding the regularization term (21) and we noticed that the quality of the restoration is highly dependent on the parameter  $\beta$ .

**Example 2:** In the second example, the original image is the "cameraman" image from Matlab and is shown on the right side of Figure 1. The blurring matrix  $H$  is given by  $H = H_2 \otimes H_1 \in \mathbb{R}^{256^2 \times 256^2}$ , where  $H_1 = I_{256}$  and  $H_2 = [h_{ij}]$  with  $[h_{ij}]$  is the Toeplitz

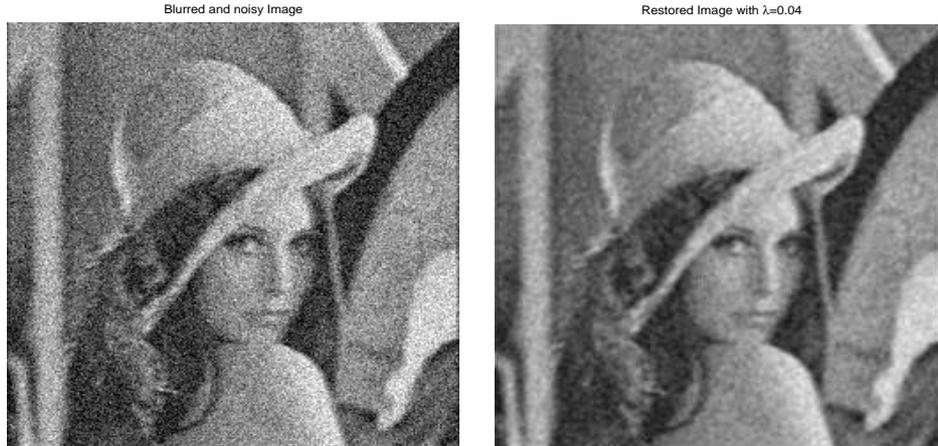


Figure 3: Degraded image (left) and restored image (right).

Problem (21) with regularization				Problem (15) without regularization			
Domain $\Omega$	$\beta$	$ISNR$	$R_{er}(X)$	Domain $\Omega$	$\beta$	$ISNR$	$R_{er}(X)$
$\Omega_{1,\beta}$	0.1	2.6308	0.1210	$\Omega_{1,\beta}$	0.1	2.7977	0.1515
$\Omega_{1,\beta}$	2.5	0.2651	0.2094	$\Omega_{1,\beta}$	2.5	0.7961	0.2406
$\Omega_{1,\beta}$	10	0.4665	0.1993	$\Omega_{1,\beta}$	10	1.2526	0.2167
$\Omega_1$	---	0.5161	0.1969	$\Omega_1$	---	1.3026	0.2137
$\Omega_2$	---	1.4030	0.1608	$\Omega_2$	---	1.6851	0.1958
$\Omega_{1,\beta} \cap \Omega_2$	0.01	2.7215	0.1188	$\Omega_{1,\beta} \cap \Omega_2$	0.01	2.4000	0.1661
$\Omega_{1,\beta} \cap \Omega_2$	10	1.3440	0.1627	$\Omega_{1,\beta} \cap \Omega_2$	10	1.7560	0.1928

Table 2:  $SNR = 5dB$ .

matrix of dimension  $256 \times 256$  given by

$$h_{ij} = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(i-j)^2}{2\sigma^2}\right), & |i-j| \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

The blurring matrix  $H$  models a blur arising in connection with the degradation of digital images by atmospheric turbulence blur. In our example we set  $\sigma = 3$  and  $r = 2$ . As in Example 1, a white random Gaussian noise of a specific variance is added to produce a blurred and noisy image  $G$  with  $SNR = 5dB$ . The blurred and noisy image is shown on the left side of Figure 5. The regularization matrix  $L$  is similar to the one given in Example 1. The restoration of the image from the degraded one is obtained by solving the minimization problem (21) using the Extended *SPG* algorithm and the domain  $\Omega = \Omega_{1,\beta}$ , with  $\beta = 0.1$ . In this example, we used the GCV method to find an estimation of the optimal value of the parameter  $\lambda$ . The curve of the GCV is plotted in Figure 4 and the optimal value of the parameter  $\lambda$  is  $\lambda_{opt} = 0.253$ . The restored image shown on the right

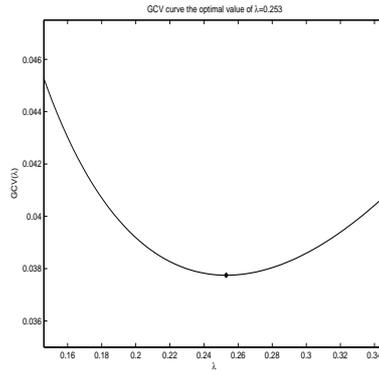


Figure 4: The GCV curve with the optimal value located at the \* point.

side of Figure 4 which was obtained with  $ISNR = 2.8605$ . The relative error in this case is  $R_{er} \simeq 0.1430$ .

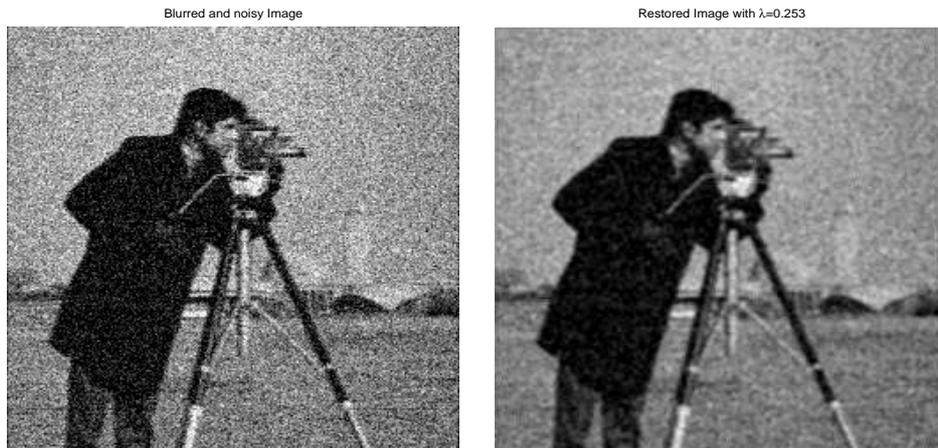


Figure 5: Degraded image (left) and restored image (right).

## 6 Concluding remarks

We have proposed a global convex constrained optimization technique for solving large-scale ill-conditioned Generalized Sylvester Equations (GSE). Our approach takes advantages of the fast behavior of the SPG method and the robustness imposed by a combined nonmonotone line search strategy.

We illustrate the advantages of our approach by solving some GSE that appear when incorporating, in a natural way, left or right preconditioning strategies for solving several

classical PDE problems.

We also show the effectiveness of the new approach by restoring noisy and blurred images. For this important application, we report experimental results with and without using a Tikhonov regularization term. These results indicate that the Tikhonov regularization term produces only a slight advantage when combined with the new technique, proving that indeed our approach is suitable for solving ill-conditioned problems, as the ones related to the presence of highly noised images. Concerning this application, our approach shows some limitations when restoring highly blurred images. In the future, we plan to study how to enrich the optimization strategy to deal effectively with highly blurred images like the ones that appear, for example, in out-of-focus deblurring problems.

## References

- [1] G. L. ANDERSON AND A. N. NETRAVALI, Image restoration based on a subjective criterion. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(1976), pp.845–856.
- [2] J. BARZILAI AND J.M. BORWEIN, *Two point step size gradient methods*, *IMA Journal of Numerical Analysis*, 8(1988), pp. 141–148.
- [3] R. H. BARTELS AND G. W. STEWART, *Solution of the matrix equation  $AX + XB = C$* , *Comm. ACM*, 15(1972), pp. 820–826.
- [4] D. P. BERTSEKAS, *On The Goldstein-Levitin-Polyak gradient projection method*, *IEEE Transactions on Automatic Control*, 21(1976), pp. 174–184.
- [5] E. G. BIRGIN, J. M. MARTÍNEZ AND M. RAYDAN, *Nonmonotone spectral projected gradient methods on convex sets*, *SIAM J. Opt.*, 10(200), pp. 1196–1211.
- [6] E. G. BIRGIN, J. M. MARTÍNEZ AND M. RAYDAN, *Algorithm 813: SPG - software for convex-constrained optimization*, *ACM Transactions on Mathematical Software*, 27(2001), pp. 340–349.
- [7] E. G. BIRGIN, J. M. MARTÍNEZ AND M. RAYDAN, *Inexact spectral gradient method for convex-constrained optimization*, *IMA Journal on Numerical Analysis*, 23(2003), pp. 539–559.
- [8] A. BOUHAMIDI AND K. JBILOU, *Sylvester Tikhonov-regularization methods in image restoration*, submitted (2006).
- [9] J. P. BOYLE AND R. L. DYKSTRA, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, *Lecture Notes in Statistics*, 37(1986), pp. 28–47.
- [10] D. CALVETTI, N. LEVENBERG AND L. REICHEL, *Iterative methods for  $X - AXB = C$* , *J. Comput. Appl. Math.*, 86(1997), pp. 73–101.

- [11] D. CALVETTI, G.H. GOLUB AND L. REICHEL, *Estimation of the L-curve via Lanczos bidiagonalization*, BIT, 39(1999), pp. 603–619.
- [12] D. CALVETTI, B. LEWIS AND L. REICHEL, *GMRES, L-curves and discrete ill-posed problems*, BIT, 42(2002), pp. 44–65.
- [13] J.L. CASTELLANOS, S. GÓMEZ AND V. GUERRA *The triangle method for finding the corner of the L-curve*, Appl. Num. Math., 43(2002), pp. 359–373.
- [14] T. F. CHAN AND J. SHEN, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, 2005.
- [15] Y. H. DAI AND L. Z. LIAO, *R-linear convergence of the Barzilai and Borwein gradient method*, IMA Journal on Numerical Analysis, 22(2002), pp. 1–10.
- [16] B. N. DATTA, *Numerical Methods for Linear Control Systems Design and Analysis*, Elsevier Press, 2003.
- [17] J. W. DEMMEL AND B. KÅGSTRÖM, *Accurate solutions of ill-posed problems in control theory*, SIAM J. Matrix Anal. Appl., 9(1988), pp. 126–145.
- [18] M. HANKE AND P.C. HANSEN, *Regularization methods for large-scale problems*, Surveys Math. Indust., 3(1993), pp. 253–315.
- [19] A. EL GUENNOUNI, K. JBILOU AND A. J. RIQUET, *Block Krylov subspace methods for solving large Sylvester equations*, Numer. Alg., 29(2002), pp. 75–96.
- [20] M. A. EPTON, *Methods for the solution of  $AXD - BXC = E$  and its application in the numerical solution of implicit ordinary differential equations*, BIT, 20(1980), pp. 341–345.
- [21] R. FLETCHER, *Low storage methods for unconstrained optimization*, Lectures in Applied Mathematics (AMS), 26(1990), pp. 165–179.
- [22] R. FLETCHER, *On the Barzilai-Borwein method*, In: Optimization and Control with Applications, (L.Qi, K. L. Teo, X. Q. Yang, eds.) Springer, 2005, pp. 235–256.
- [23] A. A. GOLDSTEIN, *Convex Programming in Hilbert Space*, Bull. Amer. Math. Soci., 70(1964), pp. 709–710.
- [24] G. H. GOLUB, S. NASH AND C. VAN LOAN, *A Hessenberg-Schur method the problem  $AX + XB = C$* , IEEE Trans. Autom. Control, AC 24(1979), pp. 909–913.
- [25] G.H. GOLUB, M. HEATH, G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, , Technometrics 21(1979), pp. 215–223.
- [26] G.H. GOLUB, U. VON MATT, *Tikhonov regularization for large scale problems*, in: G.H. Golub, S.H. Lui, F. Luk, R. Plemmons (Eds.), Workshop on Scientific Computing, Springer, New York, 1997, pp. 3–26.

- [27] G.H. GOLUB AND C. VAN LOAN, *Matrix Computations*, The Johns Hopkins University Press, 1996, Third edition.
- [28] L. GRIPPO, F. LAMPARIELLO AND S. LUCIDI, *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis, 23(1986), pp. 707–716.
- [29] S. J. HAMMARLING, *Numerical solution of the stable, nonnegative definite Lyapunov equations*, IMA Journal of Numerical Analysis, 2(1982), pp. 303–323.
- [30] P. C. HANSEN *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Rev., 34(1992), pp. 561–580.
- [31] P. C. HANSEN, J. G. NAGY, AND D. P. O'LEARY, *Deblurring Images: Matrices, Spectra, and Filtering*, SIAM, Philadelphia, 2006.
- [32] D. Y. HU AND L. REICHEL, *Krylov subspace methods for the Sylvester equation*, Linear Algebra Appl., 174(1992), 283–314.
- [33] I. M. JAIMOUKHA AND E. M. KASENALLY, *Krylov subspace methods for solving large Lyapunov equations*, SIAM J. Matrix Anal. Appl., 31(1994), pp. 227–251.
- [34] K. JBILOU, A. MESSAOUDI AND H. SADOK, *Global FOM and GMRES algorithms for matrix equations*, Appl. Num. Math., 31(1999), pp. 49–63.
- [35] K. JBILOU AND A. RIQUET, *Projection methods for large Lyapunov matrix equations*, Linear Algebra Appl., 415(2006), pp. 344–358.
- [36] K. JBILOU, *Low rank approximate solutions to large Sylvester matrix equations*, Applied Mathematics and Computation, 177(2006), pp. 365–376.
- [37] W. LA CRUZ, J. M. MARTÍNEZ AND M. RAYDAN, *Spectral residual method without gradient information for solving large-scale nonlinear systems of equations*, Mathematics of Computation, 75(2006), pp. 1449–1466.
- [38] E. S. LEVITIN AND B. T. POLYAK, *Constrained Minimization Problems*, USSR Computl. Math. Mathl. Physics, 6(1966), pp. 1–50.
- [39] M. MONSALVE, *Block linear method for large scale Sylvester equations*, Computational and Applied Mathematics, in press (2008).
- [40] M. RAYDAN, *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA J. Numer. Anal., 13(1993), pp. 321–326.
- [41] V. SIMONCINI, *On the numerical solution of  $AX - XB = C$* , BIT, 36(1996), pp. 814–830.