Les Cahiers du GERAD ISSN: 0711–2440

The Projected Golub-Kahan Process for Constrained Linear Least-Squares Problems

D. Orban

G–2014–15

March 2014

The Projected Golub-Kahan Process for Constrained Linear Least-Squares Problems

Dominique Orban

GERAD & Department of Mathematics and Industrial Engineering
Polytechnique Montréal
Montréal (Québec) Canada, H3C 3A7

dominique.orban@gerad.ca

March 2014

Les Cahiers du GERAD
G–2014–15

Abstract: A preconditioned variant of the Golub and Kahan (1965) bidiagonalization process recently proposed by Arioli (2013) and Arioli and Orban (2013) allows us to establish that SYMMLQ and MINRES applied to least-squares problems in symmetric saddle-point form perform redundant work and are combinations of methods such as LSQR and LSMR. A well-chosen preconditioner allows us to formulate a projected variant of the Golub-Kahan process that forms the basis of specialized numerical methods for linear least-squares problems with linear equality constraints. As before, full-space methods such as SYMMLQ and MINRES applied to the symmetric saddle-point system defining the optimality conditions of such problems perform redundant work and are combinations of projected variants of methods such as LSQR and LSMR. We establish connections between numerical methods for least-squares problems, full-space methods and the projected and constraint-preconditioned Krylov methods of Gould, Orban, and Rees (2013).

Key Words: Linear least-squares, linear constraints, Golub-Kahan process, Lanczos process, Krylov methods, projected Krylov methods, nullspace methods, LSQR, LSMR, CRAIG, CRAIG-MR, CG, SYMMLQ, MINRES.

Résumé : Une variante préconditionnée du processus de bidiagonalisation de Golub and Kahan (1965) proposée récemment par Arioli (2013) et Arioli and Orban (2013) nous permet d'établir que SYMMLQ et MINRES appliquées à des problèmes aux moindres carrés sous forme de système augmenté effectuent des opérations redondantes et se décomposent suivant deux méthodes telles que LSQR et LSMR. Un préconditionneur bien choisi nous permet de formuler une variante projetée du processus de Golub-Kahan qui est à la base de méthodes numériques particularisées aux problèmes aux moindres carrés avec contraintes d'égalité linéaires. Comme auparavant, les méthodes travaillant dans l'espace entier telles que SYMMLQ et MINRES sont des combinaisons de variantes projetées de méthodes telles que LSQR et LSMR. Nous établissons des liens entre les méthodes numériques pour les problèmes aux moindres carrés, les méthodes travaillant dans l'espace entier, ainsi que les variantes projetées et avec préconditionneur par contraintes de Gould, Orban, and Rees (2013).

Acknowledgments: Research partially supported by an NSERC Discovery Grant. Paige (1974) and Saunders (1995) were constant sources of inspiration for this work. This paper also benefited from enlightening discussions with Mario Arioli and Michael Saunders.

1 Introduction

We consider the solution of constrained weighted linear least-squares problems of the form

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{M}^{-1}} \quad \text{subject to } \mathbf{E}\mathbf{x}=\mathbf{0}, \tag{1.1}$$

where $\mathbf{A}\in\mathbb{R}^{m\times n}$, $\mathbf{b}\in\mathbb{R}^m$, $\mathbf{M}\in\mathbb{R}^{m\times m}$ is symmetric and positive definite, and $\mathbf{E}\in\mathbb{R}^{p\times n}$. Typically, though not always, $m\geq n$. We assume that $p<n$ and that $\mathbf{E}$ has full row rank. Throughout the paper, we refer to (1.1) as a nullspace-constrained linear least-squares problem. Problems such as (1.1) arise in mixed regression-interpolation data fitting problems (Björck, 1996) and portfolio optimization (El Ghaoui, 2013). Such problems also arise frequently in the solution of smooth constrained optimization problems of the form

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ f(\mathbf{x}) \quad \text{subject to } c(\mathbf{x})=\mathbf{0},\ \mathbf{E}\mathbf{x}=\mathbf{d}, \tag{1.2}$$

involving a mixture of linear and nonlinear equality constraints, by way of a method related to the sequential quadratic programming paradigm, in which each step is computed as the sum of a normal and a tangential step (Nocedal and Wright, 2006). Numerical methods typically exploit linear constraints by generating iterates $\mathbf{x}_k$ satisfying $\mathbf{E}\mathbf{x}_k=\mathbf{d}$ at all times. A normal step then takes the form (1.1) where $\mathbf{A}$ is the Jacobian of c at $\mathbf{x}_k$ and $\mathbf{b}=-c(\mathbf{x}_k)$. In this optimization context, it is customary to add a trust-region constraint $\|\mathbf{x}\|_{\mathbf{N}}\leq\Delta$ where $\mathbf{N}=\mathbf{N}^{\mathsf{T}}$ is positive definite and $\Delta>0$ is a trust-region radius. As the next sections will show, we develop numerical methods for (1.1) without a trust-region constraint and are able to add this constraint afterwards at little extra cost.

The first and second-order optimality conditions of (1.1) may be stated as the symmetric saddle-point system

$$\begin{bmatrix}\mathbf{A}^{\mathsf{T}}\mathbf{M}^{-1}\mathbf{A} & \mathbf{E}^{\mathsf{T}}\\ \mathbf{E} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}=\begin{bmatrix}\mathbf{A}^{\mathsf{T}}\mathbf{M}^{-1}\mathbf{b}\\ \mathbf{0}\end{bmatrix} \tag{1.3}$$

or as the augmented symmetric saddle-point system

$$\begin{bmatrix}\mathbf{M} & \mathbf{A} & \mathbf{0}\\ \mathbf{A}^{\mathsf{T}} & \mathbf{0} & \mathbf{E}^{\mathsf{T}}\\ \mathbf{0} & \mathbf{E} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{r}\\ \mathbf{x}\\ \mathbf{y}\end{bmatrix}=\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \tag{1.4}$$

where $\mathbf{r}:=\mathbf{M}^{-1}(\mathbf{b}-\mathbf{A}\mathbf{x})$ is the residual. The system (1.3) is a standard symmetric saddle-point system with a positive semi-definite leading block and the projected conjugate gradient method of Gould, Hribar, and Nocedal (2001) is appropriate. The system (1.4) may be viewed as a standard symmetric saddle-point system of the form

$$\begin{bmatrix}\mathbf{Q} & \mathbf{J}^{\mathsf{T}}\\ \mathbf{J} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{q}\\ \mathbf{y}\end{bmatrix}=\begin{bmatrix}\mathbf{f}\\ \mathbf{0}\end{bmatrix} \quad \text{with} \quad \mathbf{Q}=\begin{bmatrix}\mathbf{M} & \mathbf{A}\\ \mathbf{A}^{\mathsf{T}} & \mathbf{0}\end{bmatrix}, \quad \mathbf{J}=\begin{bmatrix}\mathbf{0} & \mathbf{E}\end{bmatrix}, \quad \mathbf{f}=\begin{bmatrix}\mathbf{b}\\ \mathbf{0}\end{bmatrix}. \tag{1.5}$$

The projected Krylov methods of Gould, Orban, and Rees (2013), and in particular the projected MINRES algorithm, are then applicable. The latter requires a one-time factorization of a symmetric and indefinite matrix of the form

$$\begin{bmatrix}\tilde{\mathbf{Q}} & \mathbf{J}^{\mathsf{T}}\\ \mathbf{J} & \mathbf{0}\end{bmatrix} \quad \text{where} \quad \tilde{\mathbf{Q}}=\tilde{\mathbf{Q}}^{\mathsf{T}}\approx\mathbf{Q}$$

and matrix-vector products with $\mathbf{Q}$. However, we show in the next sections that it is possible to exploit the structure of (1.1) further and devise specialized iterative methods for linear least-squares problems. Our methods are especially applicable when it is possible to solve linear systems with coefficient matrices $\mathbf{M}$ and $\mathbf{N}$ efficiently, where $\mathbf{N}$ defines the appropriate norm for $\mathbf{x}$.

The rest of the paper is organized as follows. The remainder of this section covers related research and sets the notation used throughout. Section 2 provides necessary background and basic results referred to in later sections. Section 3 lays the foundations by describing the preconditioned Golub-Kahan process, upon which preconditioned numerical methods for linear least squares are built. This process is used to construct a related Lanczos process in §4 and corresponding preconditioned full-space methods. A special case of the preconditioned Golub-Kahan process leads to methods for (1.1) in §5 and for (1.3) and (1.4) in §6. We present an application to optimization in §7 and conclude in §8.

Related Research

Björck (1996) mentions several approaches to solve (1.1). Among them, the method of direct elimination consists in computing the QR factorization of $\mathbf{E}$ to reduce the problem to an unconstrained linear least-squares problem. The nullspace method proceeds similarly using the LQ factorization of $\mathbf{E}$ to formulate an unconstrained linear least-squares problem defined in Null($\mathbf{E}$). The nullspace method is closely related to methods discussed in the present paper.

Van Loan (1985) proposes the method of weights, which is a penalty method consisting in solving the unconstrained regularized linear least-squares problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{M}^{-1}} + \tfrac{1}{2}\mu^2\|\mathbf{E}\mathbf{x}\|_2^2,$$

for well-chosen values of μ. The rationale is that as $\mu \to \infty$ the exact solution $\mathbf{x}(\mu)$ to this regularized problem converges to a solution of (1.1).

Benbow (1999) gives a formulation of LSQR for problems of the form (1.1) without constraints. His formulation is based on a preconditioned version of the Golub-Kahan process and serves as inspiration for some of the variants presented below.

Arioli and Orban (2013) give a full account of regularized problems, where the picture is more complete. As a result, we do not consider regularized least-squares problems here but we regularly refer to (Arioli and Orban, 2013) throughout the discussion to emphasize similarities and differences.

Gould, Orban, and Rees (2013) provide a framework in which projected Krylov methods may be derived from standard formulations of preconditioned Krylov methods. They establish an equivalence between the iterates generated by projected Krylov methods and the closely-related constaint-preconditioned Krylov methods.

Notation

We use a notation compatible with those of Arioli and Orban (2013) and Gould, Orban, and Rees (2013) and denote full-space quantities $\mathbf{G}$, $\mathbf{A}$, $\mathbf{x}$ and so forth, in upright boldface font. Quantities that are explicitly expressed in a basis of Null($\mathbf{E}$) appear in italicized Roman lightface font, e.g., r and x. Throughout, $\|\cdot\|_2$ denotes the Euclidian norm. If $\mathbf{Q}$ is any symmetric and positive-definite matrix, it determines the $\mathbf{Q}$-norm defined by $\|\mathbf{x}\|^2_{\mathbf{Q}} := \mathbf{x}^{\mathsf{T}}\mathbf{Q}\mathbf{x}$.

2 Background and Preliminary Results

For simplicity, we only consider nullspace constraints in (1.1) but we note that the more general equality-constrained problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{M}^{-1}} \quad \text{subject to } \mathbf{E}\mathbf{x}=\mathbf{c}$$

may be solved by first identifying $\mathbf{x}_0$ such that $\mathbf{E}\mathbf{x}_0 = \mathbf{c}$, solving the nullspace problem

$$\underset{\Delta\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{A}\Delta\mathbf{x}-(\mathbf{b}-\mathbf{A}\mathbf{x}_0)\|^2_{\mathbf{M}^{-1}} \quad \text{subject to } \mathbf{E}\Delta\mathbf{x}=\mathbf{0},$$

and finally setting $\mathbf{x} := \mathbf{x}_0 + \Delta\mathbf{x}$.

The problem (1.1) possesses a unique solution if and only if $\mathbf{E}$ has full row rank and Null($\mathbf{A}$)$\cap$Null($\mathbf{E}$) $= \{\mathbf{0}\}$ (Björck, 1996, Chapter 5).

The following result on the eigenvalues of an augmented matrix is a special case of (Saunders, 1995, Result 1). In particular, it implies that the spectrum of $\mathbf{K}$ is symmetric about $\frac{1}{2}$.

Proposition 2.1 (Saunders, 1995, Result 1) *Let $\bar{\mathbf{E}} \in \mathbb{R}^{p\times n}$ have row rank $p < n$ and let $\sigma_1, \ldots, \sigma_p$ be the nonzero singular values of $\bar{\mathbf{E}}$. The eigenvalues of*

$$\mathbf{K} := \begin{bmatrix} \mathbf{I} & \bar{\mathbf{E}}^{\mathsf{T}} \\ \bar{\mathbf{E}} & \mathbf{0} \end{bmatrix}$$

are

1. *$\lambda = 1$ with multiplicity $n - p$, and*
2. *$\lambda = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 + 4\sigma_k^2}$ for $k = 1, \ldots, p$.*

In addition, the eigenspace associated to $\lambda = 1$ is $\mathrm{Null}(\bar{\mathbf{E}}) \times \{\mathbf{0}\}$.

Several numerical methods for solving a generic linear least-squares problem

$$\underset{\bar{\mathbf{x}}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\bar{\mathbf{A}}\bar{\mathbf{x}} - \bar{\mathbf{b}}\|_2^2, \tag{2.1}$$

including those of Paige and Saunders (1982) and Fong and Saunders (2011), build upon the Golub and Kahan (1965) bidiagonalization process, stated as Algorithm 2.1.

Algorithm 2.1 Golub-Kahan Bidiagonalization Process for (2.1)

Require: $\bar{\mathbf{A}}$, $\bar{\mathbf{b}}$

$\beta_1\bar{\mathbf{u}}_1 = \bar{\mathbf{b}}$ with $\beta_1 > 0$ so that $\|\bar{\mathbf{u}}_1\|_2 = 1$
$\alpha_1\bar{\mathbf{v}}_1 = \bar{\mathbf{A}}^{\mathsf{T}}\bar{\mathbf{u}}_1$ with $\alpha_1 > 0$ so that $\|\bar{\mathbf{v}}_1\|_2 = 1$
for $k = 1, 2, \ldots$ **do**
  $\beta_{k+1}\bar{\mathbf{u}}_{k+1} = \bar{\mathbf{A}}\bar{\mathbf{v}}_k - \alpha_k\bar{\mathbf{u}}_k$ with $\beta_{k+1} > 0$ so that $\|\bar{\mathbf{u}}_{k+1}\|_2 = 1$
  $\alpha_{k+1}\bar{\mathbf{v}}_{k+1} = \bar{\mathbf{A}}^{\mathsf{T}}\bar{\mathbf{u}}_{k+1} - \beta_{k+1}\bar{\mathbf{v}}_k$ with $\alpha_{k+1} > 0$ so that $\|\bar{\mathbf{v}}_{k+1}\|_2 = 1$.

Upon defining $\bar{\mathbf{U}}_k := \begin{bmatrix}\bar{\mathbf{u}}_1 & \cdots & \bar{\mathbf{u}}_k\end{bmatrix}$, $\bar{\mathbf{V}}_k := \begin{bmatrix}\bar{\mathbf{v}}_1 & \cdots & \bar{\mathbf{v}}_k\end{bmatrix}$ as well as

$$\mathbf{L}_k := \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \end{bmatrix}, \quad \mathbf{B}_k := \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \\ & & & \beta_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_k \\ \beta_{k+1}\mathbf{e}_k^{\mathsf{T}} \end{bmatrix}, \tag{2.2}$$

the situation after k iterations of Algorithm 2.1 can be characterized by the identities

$$\bar{\mathbf{A}}\bar{\mathbf{V}}_k = \bar{\mathbf{U}}_k\mathbf{L}_k + \beta_{k+1}\bar{\mathbf{u}}_{k+1}\mathbf{e}_k^{\mathsf{T}} \quad = \bar{\mathbf{U}}_{k+1}\mathbf{B}_k, \tag{2.3a}$$

$$\bar{\mathbf{A}}^{\mathsf{T}}\bar{\mathbf{U}}_{k+1} = \bar{\mathbf{V}}_k\mathbf{B}_k^{\mathsf{T}} + \alpha_{k+1}\bar{\mathbf{v}}_{k+1}\mathbf{e}_{k+1}^{\mathsf{T}} = \bar{\mathbf{V}}_{k+1}\mathbf{L}_{k+1}^{\mathsf{T}}. \tag{2.3b}$$

In addition, the identities $\bar{\mathbf{U}}_k^{\mathsf{T}}\bar{\mathbf{U}}_k = \mathbf{I}_k$ and $\bar{\mathbf{V}}_k^{\mathsf{T}}\bar{\mathbf{V}}_k = \mathbf{I}_k$ are satisfied in exact arithmetic. Algorithm 2.1 is convenient because it allows to approximate a solution of (2.1) by a sequence of solutions of linear least-squares problems whose operator is bidiagonal.

It is well known that the Golub-Kahan process is intimately connected to the Lanczos (1952) process. Consider a generic symmetric linear system $\mathbf{Hw} = \mathbf{d}$. The Lanczos process applied to $\mathbf{H}$ and $\mathbf{d}$ generates vectors $\mathbf{w}_k$ according to the three-term recurrence

$$\omega_1\mathbf{w}_1 = \mathbf{d}, \quad \omega_{k+1}\mathbf{w}_{k+1} = \mathbf{H}\mathbf{w}_k - \chi_k\mathbf{w}_k - \omega_k\mathbf{w}_{k-1}, \quad \text{where} \quad \chi_k := \mathbf{w}_k^{\mathsf{T}}\mathbf{H}\mathbf{w}_k. \tag{2.4}$$

Each $\omega_k > 0$ is chosen so that $\|\mathbf{w}_k\|_2 = 1$. In exact arithmetic, the vectors $\mathbf{w}_k$ are orthonormal. Numerous Krylov methods build upon the Lanczos process, e.g., the conjugate gradient method of Hestenes and Stiefel (1952), SYMMLQ and the minimum residual method of Paige and Saunders (1975), and others—see (Saad, 2003) for more details. Suppose now that the vectors $\mathbf{w}_k$ should be measured in $\mathbf{Y}$ norm, where $\mathbf{Y}$ is symmetric

and positive definite. If we write the Lanczos process (2.4) for the symmetrically-preconditioned system $\bar{\mathbf{H}}\bar{\mathbf{w}} = \bar{\mathbf{d}}$, where $\bar{\mathbf{H}} := \mathbf{Y}^{-\frac{1}{2}}\mathbf{H}\mathbf{Y}^{-\frac{1}{2}}$ and $\bar{\mathbf{d}} := \mathbf{Y}^{-\frac{1}{2}}\mathbf{d}$, and apply the change of variable $\bar{\mathbf{w}} := \mathbf{Y}^{\frac{1}{2}}\mathbf{w}$, we obtain the *preconditioned* Lanczos process

$$\omega_1 \mathbf{Y}\mathbf{w}_1 = \mathbf{d}, \quad \omega_{k+1}\mathbf{Y}\mathbf{w}_{k+1} = \mathbf{H}\mathbf{w}_k - \chi_k \mathbf{Y}\mathbf{w}_k - \omega_k \mathbf{Y}\mathbf{w}_{k-1}; \quad \chi_k := \mathbf{w}_k^T \mathbf{H}\mathbf{w}_k, \tag{2.5}$$

and where this time the scalars $\omega_k > 0$ are chosen so that $\|\mathbf{w}_k\|_{\mathbf{Y}} = 1$. Note that the computation of each $\mathbf{w}_k$ requires the solution of a linear system with coefficient $\mathbf{Y}$. After k iterations, the situation may be summarized as

$$\mathbf{H}\mathbf{W}_k = \mathbf{Y}\mathbf{W}_k\mathbf{\Omega}_k + \omega_{k+1}\mathbf{Y}\mathbf{w}_{k+1}\mathbf{e}_k^T, \tag{2.6}$$

where

$$\mathbf{W}_k := \begin{bmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_k \end{bmatrix}, \qquad \text{and} \qquad \mathbf{\Omega}_k := \begin{bmatrix} \chi_1 & \omega_2 & & \\ \omega_2 & \chi_2 & \ddots & \\ & \ddots & \ddots & \omega_k \\ & & \omega_k & \chi_k \end{bmatrix}.$$

In exact arithmetic, the vectors $\mathbf{w}_k$ are such that $\mathbf{W}_k^T\mathbf{Y}\mathbf{W}_k = \mathbf{I}_k$, i.e, $\mathbf{Y}^{\frac{1}{2}}\mathbf{W}_k$ is orthogonal. At iteration k, a numerical method based on (2.5) seeks an approximate solution of the form $\mathbf{W}_k w_k$ for a vector of coefficients $w_k \in \mathbb{R}^k$ chosen so that the approximate solution satisfies a certain optimality condition.

In the next section, we perform a similar conversion of Algorithm 2.1 to a different metric.

We close this section with a representation of the various operators involved in (1.1) as operators between appropriately defined Hilbert spaces. Arioli and Orban (2013) introduce a formal Hilbert space setting but we keep details to a minimum here for clarity. Let $\mathbb{M} = (\mathbb{R}^m, \mathbf{M})$ be the Hilbert space $\mathbb{R}^m$ endowed with the norm defined by $\mathbf{M}$. Similarly, let $\mathbb{N} = (\mathbb{R}^n, \mathbf{N})$ and their duals $\mathbb{M}^\star = (\mathbb{R}^m, \mathbf{M}^{-1})$ and $\mathbb{N} = (\mathbb{R}^n, \mathbf{N}^{-1})$. Operators $\mathcal{M}$ and $\mathcal{M}^{-1}$ are represented by $\mathbf{M}$ and $\mathbf{M}^{-1}$ in the canonical bases of $\mathbb{M}$ and $\mathbb{M}^\star$. The operators $\mathcal{N}$ and its inverse are defined similarly. The matrix $\mathbf{A}$ represents an operator $\mathcal{A}$ from $\mathbb{N}$ into $\mathbb{M}^\star$ and we denote its adjoint $\mathcal{A}^\star$. The latter operator is represented by $\mathbf{A}^T$. Finally, $\mathbf{E}$ represents an operator $\mathcal{E}$ between $\mathbb{N}$ and the dual of another Hilbert space, say $\mathbb{P}$. Finally, let $\mathbb{Z}$ denote the Hilbert subspace of $\mathbb{N}$ defined as the nullspace of $\mathcal{E}$ and equipped with the induced norm. In §5, we introduce a matrix $\mathbf{Z}$ whose columns form a basis for Null($\mathbf{E}$). This $\mathbf{Z}$ and its transpose represent corresponding operators $\mathcal{Z}$ and $\mathcal{Z}^\star$ between $\mathbb{Z}$, $\mathbb{N}$ and their duals. Figure 2.1 represents the situation in the form of a commutative diagram. Throughout the rest of this paper, we refer to this diagram to understand what norm should be used to measure a certain residual or error.

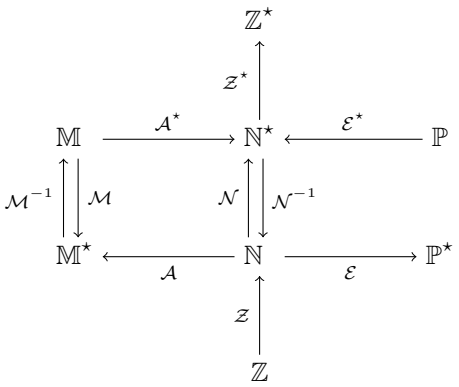

Figure 2.1: Commutative diagram between the relevant Hilbert spaces.

3 The Preconditioned Golub-Kahan Bidiagonalization Process

Consider now the preconditioned linear-least-squares problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{M}^{-1}}, \tag{3.1}$$

whose optimality conditions may be written as the augmented system

$$\begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}, \tag{3.2}$$

or as the *normal equations*

$$\mathbf{A}^T\mathbf{M}^{-1}\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{M}^{-1}\mathbf{b}. \tag{3.3}$$

Note that (3.2) also represent the necessary and sufficient optimality conditions of the convex quadratic program

$$\underset{\mathbf{r}\in\mathbb{R}^m}{\text{minimize}} \ -\mathbf{b}^T\mathbf{r} + \tfrac{1}{2}\mathbf{r}^T\mathbf{M}\mathbf{r} \quad \text{subject to } \mathbf{A}^T\mathbf{r} = \mathbf{0},$$

and therefore the methods below equally apply to the problem stated in this form. Suppose that the context dictates that the appropriate norm to measure the variables $\mathbf{x}$ is the $\mathbf{N}$-norm, where $\mathbf{N}$ is a symmetric and positive-definite n-by-n matrix. Consider solving the central-preconditioned variant of (3.2)

$$\begin{bmatrix} \mathbf{M}^{-\frac{1}{2}} & \\ & \mathbf{N}^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{M}^{-\frac{1}{2}} & \\ & \mathbf{N}^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \mathbf{M}^{\frac{1}{2}}\mathbf{r} \\ \mathbf{N}^{\frac{1}{2}}\mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{M}^{-\frac{1}{2}}\mathbf{b} \\ \mathbf{0} \end{bmatrix}. \tag{3.4}$$

The coefficient matrix of (3.4) may be written

$$\bar{\mathbf{K}} := \begin{bmatrix} \mathbf{I} & \bar{\mathbf{A}} \\ \bar{\mathbf{A}}^T & \mathbf{0} \end{bmatrix}, \qquad \bar{\mathbf{A}} := \mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{N}^{-\frac{1}{2}}.$$

We follow Arioli and Orban (2013) and refer to the singular values of $\bar{\mathbf{A}}$ as the *elliptic singular values* of $\mathbf{A}$. By virtue of Proposition 2.1, those values govern the convergence of iterative methods for (2.1). Consider now the application of Algorithm 2.1 to the operator $\bar{\mathbf{A}}$ and right-hand side $\bar{\mathbf{b}} := \mathbf{M}^{-\frac{1}{2}}\mathbf{b}$. After the change of variable $\mathbf{u}_i := \mathbf{M}^{-\frac{1}{2}}\bar{\mathbf{u}}_i$ and $\mathbf{v}_i := \mathbf{N}^{-\frac{1}{2}}\bar{\mathbf{v}}_i$, the resulting process may be stated as Algorithm 3.1, in which the computation of each $\mathbf{u}_k$ requires the solution of a linear system with matrix $\mathbf{M}$ and that of each $\mathbf{v}_k$ requires the solution of a linear system with matrix $\mathbf{N}$. Algorithm 3.1 is given by Arioli (2013, Section 3) and Arioli and Orban (2013, Algorithm 4.2), and generalizes related processes of Benbow (1999).

Algorithm 3.1 Preconditioned Golub-Kahan Bidiagonalization Process

Require: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{M}^{-1}$, $\mathbf{N}^{-1}$
1: $\beta_1\mathbf{M}\mathbf{u}_1 = \mathbf{b}$ with $\beta_1 > 0$ so that $\|\mathbf{u}_1\|_{\mathbf{M}} = 1$
2: $\alpha_1\mathbf{N}\mathbf{v}_1 = \mathbf{A}^T\mathbf{u}_1$ with $\alpha_1 > 0$ so that $\|\mathbf{v}_1\|_{\mathbf{N}} = 1$
3: **for** $k = 1, 2, \ldots$ **do**
4: $\quad\beta_{k+1}\mathbf{M}\mathbf{u}_{k+1} = \mathbf{A}\mathbf{v}_k - \alpha_k\mathbf{M}\mathbf{u}_k$ with $\beta_{k+1} > 0$ so that $\|\mathbf{u}_{k+1}\|_{\mathbf{M}} = 1$
5: $\quad\alpha_{k+1}\mathbf{N}\mathbf{v}_{k+1} = \mathbf{A}^T\mathbf{u}_{k+1} - \beta_{k+1}\mathbf{N}\mathbf{v}_k$ with $\alpha_{k+1} > 0$ so that $\|\mathbf{v}_{k+1}\|_{\mathbf{N}} = 1$.

Algorithm 3.1 is precisely the Golub-Kahan process applied to the operator $\mathbf{A}$ and right-hand side $\mathbf{b}$ with preconditioners $\mathbf{M}$ and $\mathbf{N}$.

Defining as before $\mathbf{U}_k := \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_k\end{bmatrix}$ and $\mathbf{V}_k := \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_k\end{bmatrix}$, Algorithm 3.1 is characterized by the identities

$$\mathbf{A}\mathbf{V}_k = \mathbf{M}\mathbf{U}_k\mathbf{L}_k + \beta_{k+1}\mathbf{M}\mathbf{u}_{k+1}\mathbf{e}_k^T = \mathbf{M}\mathbf{U}_{k+1}\mathbf{B}_k, \tag{3.5a}$$

$$\mathbf{A}^T\mathbf{U}_{k+1} = \mathbf{N}\mathbf{V}_k\mathbf{B}_k^T + \alpha_{k+1}\mathbf{N}\mathbf{v}_{k+1}\mathbf{e}_{k+1}^T = \mathbf{N}\mathbf{V}_{k+1}\mathbf{L}_{k+1}^T, \tag{3.5b}$$

in place of (2.3). In addition, the identities

$$\mathbf{U}_k^T \mathbf{M} \mathbf{U}_k = \mathbf{I}_k \quad \text{and} \quad \mathbf{V}_k^T \mathbf{N} \mathbf{V}_k = \mathbf{I}_k$$

are satisfied in exact arithmetic, i.e., $\mathbf{M}^{\frac{1}{2}} \mathbf{U}_k$ and $\mathbf{N}^{\frac{1}{2}} \mathbf{V}_k$ are orthogonal matrices.

The rest of this section reviews several iterative methods for (3.1). The methods below are fully described by Arioli and Orban (2013) in the presence of regularization terms. The reason for presenting them again below is to highlight the differences with the case when no regularization term is present and to help establish connections between methods for (3.2) and methods for (3.3). Note that Arioli and Orban (2013) specify complete algorithms along with meaningful stopping criteria. We refer the interested reader to their paper for complete algorithmic details.

3.1 Preconditioned LSQR

By definition, the LSQR method of Paige and Saunders (1982) is equivalent to the conjugate gradient method (CG) applied to (3.3). Based on Algorithm 3.1, LSQR seeks an approximate solution as a combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, i.e., of the form $\mathbf{x}_k = \mathbf{V}_k x_k$. The vector x_k is chosen to minimize the least-squares residual. Using (3.5), we have

$$\mathbf{A}\mathbf{x}_k - \mathbf{b} = \mathbf{M}\mathbf{U}_{k+1}(\mathbf{B}_k x_k - \beta_1 \mathbf{e}_1)$$

and

$$\|\mathbf{A}\mathbf{x}_k - \mathbf{b}\|_{\mathbf{M}^{-1}} = \|\mathbf{M}^{\frac{1}{2}} \mathbf{U}_{k+1}(\mathbf{B}_k x_k - \beta_1 \mathbf{e}_1)\|_2 = \|\mathbf{B}_k x_k - \beta_1 \mathbf{e}_1\|_2,$$

where we used the fact that $\mathbf{M}^{\frac{1}{2}} \mathbf{U}_{k+1}$ is an orthogonal matrix in exact arithmetic. Finally, x_k is selected as the solution of the overdetermined bidiagonal linear least-squares problem

$$\underset{x \in \mathbb{R}^k}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{B}_k x - \beta_1 \mathbf{e}_1\|_2^2, \tag{3.6}$$

and therefore satisfies

$$\begin{bmatrix} \mathbf{I}_{k+1} & \mathbf{B}_k \\ \mathbf{B}_k^T & \end{bmatrix} \begin{bmatrix} r_k \\ x_k \end{bmatrix} = \begin{bmatrix} \beta_1 \mathbf{e}_1 \\ \mathbf{0} \end{bmatrix}, \tag{3.7}$$

where $r_k := \beta_1 \mathbf{e}_1 - \mathbf{B}_k x_k$.

The following result may be established exactly as (Arioli and Orban, 2013, Theorem 6.1).

Proposition 3.1 *The preconditioned* LSQR *iterates on* (3.1) *are the same as those generated by the standard conjugate gradient method on the positive semi-definite system* (3.3) *with preconditioner* $\mathbf{N}$. *The error* $(\mathbf{x}_k - \mathbf{x}_*)$ *decreases along the iterations in the* $(\mathbf{N}^{-\frac{1}{2}} \mathbf{A}^T \mathbf{M}^{-1} \mathbf{A} \mathbf{N}^{-\frac{1}{2}})$*-norm.*

Note that (3.3) with preconditioner $\mathbf{N}$ is none other than the system of normal equations

$$\bar{\mathbf{A}}^T \bar{\mathbf{A}} \bar{\mathbf{x}} = \bar{\mathbf{A}}^T \bar{\mathbf{b}}.$$

It is instructive to characterize the preconditioned LSQR method with its own Lanczos process. We have from (3.3) and (3.5) that

$$\mathbf{A}^T \mathbf{M}^{-1} \mathbf{A} \mathbf{V}_k = \mathbf{N}\mathbf{V}_k \mathbf{B}_k^T \mathbf{B}_k + \alpha_{k+1}\beta_{k+1} \mathbf{N}\mathbf{v}_{k+1} \mathbf{e}_k^T. \tag{3.8}$$

Note that $\mathbf{B}_k^T \mathbf{B}_k$ is indeed tridiagonal and symmetric, as in (2.6).

3.2 Preconditioned CRAIG

The method of Craig (1955), as described by Paige and Saunders (1982, §7.2), attempts to solve the saddle-point system (3.2) directly by seeking approximations $\mathbf{r}_k := \mathbf{U}_k r_k$ and $\mathbf{x}_k := \mathbf{V}_k x_k$. This amounts to seeking an approximation in the Krylov subspace spanned by the columns of $\text{blkdiag}(\mathbf{U}_k, \mathbf{V}_k)$, and yields the block equations

$$\mathbf{M}\mathbf{U}_k r_k + \mathbf{A}\mathbf{V}_k x_k = \mathbf{M}\mathbf{u}_{k+1}(\beta_1 \mathbf{e}_1) \quad \text{and} \quad \mathbf{A}^T\mathbf{U}_k r_k = \mathbf{0}.$$

Premultiplying the first equation with $\mathbf{U}_k^T$ and the second with $\mathbf{V}_k^T$ amounts to imposing a Galerkin condition on the residual of the saddle-point system. Combining with (3.5) and using the orthonormality of $\mathbf{M}^{\frac{1}{2}}\mathbf{U}_k$ and of $\mathbf{N}^{\frac{1}{2}}\mathbf{V}_k$, this yields the reduced saddle-point system

$$\begin{bmatrix} \mathbf{I}_k & \mathbf{L}_k \\ \mathbf{L}_k^T & \end{bmatrix} \begin{bmatrix} r_k \\ x_k \end{bmatrix} = \begin{bmatrix} \beta_1 \mathbf{e}_1 \\ \mathbf{0} \end{bmatrix}, \tag{3.9}$$

which represents the optimality conditions of the problem

$$\underset{x}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{L}_k x - \beta_1 \mathbf{e}_1\|_2^2.$$

Note the similarity with (3.6) and (3.7). It is clear that $\mathbf{L}_k$ being square and nonsingular, $x_k = \mathbf{L}_k^{-1}(\beta_1 \mathbf{e}_1)$ and $r_k = \mathbf{0}$. This method in itself is therefore of limited interest unless $\mathbf{r}$ indeed vanishes at a solution of (3.2), i.e., unless $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent. However, this method turns out to be useful in §4 to explain the behavior of SYMMLQ on (3.2).

3.3 Preconditioned LSMR

By definition, the preconditioned LSMR method of Fong and Saunders (2011) consists in applying MINRES (Paige and Saunders, 1975) to (3.3). Because Figure 2.1 indicates that the residual of (3.3) should be measured in the $\mathbf{N}^{-1}$-norm, this amounts to solving

$$\underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{A}^T\mathbf{M}^{-1}(\mathbf{A}\mathbf{x} - \mathbf{b})\|_{\mathbf{N}^{-1}}.$$

Seeking again an approximation of the form $\mathbf{x}_k = \mathbf{V}_k x_k$, we obtain the subproblem

$$\underset{x}{\text{minimize}} \ \left\| \begin{bmatrix} \mathbf{B}_k^T\mathbf{B}_k \\ \alpha_{k+1}\beta_{k+1}\mathbf{e}_k^T \end{bmatrix} x - \alpha_1\beta_1\mathbf{e}_1 \right\|_2,$$

and the Lanczos process (3.8). The details follow from the same reasoning as in previous sections and may be found in (Fong and Saunders, 2011) and (Arioli and Orban, 2013). The following result summarizes the defining properties of the preconditioned LSMR.

Proposition 3.2 *The preconditioned* LSMR *iterates on* (2.1) *are the same as those generated by the standard* MINRES *on the positive semi-definite system* (3.3) *with preconditioner* $\mathbf{N}$. *The residual* $\mathbf{A}^T\mathbf{M}^{-1}\mathbf{r}_k$ *decreases along the iterations in the* $\mathbf{N}^{-1}$*-norm, where* $\mathbf{r}_k := \mathbf{b} - \mathbf{A}\mathbf{x}_k$.

3.4 Preconditioned CRAIG-MR

By definition, the preconditioned CRAIG-MR method, introduced by Arioli and Orban (2013), consists in applying MINRES to

$$\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{b}. \tag{3.10}$$

Were $\mathbf{A}\mathbf{x} = \mathbf{b}$ consistent, this would be equivalent to solving the minimum-norm problem

$$\underset{\mathbf{x}}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{x}\|_{\mathbf{N}}^2 \quad \text{subject to} \ \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{3.11}$$

Because Figure 2.1 indicates that the residual of (3.10) should be measured in the $\mathbf{M}^{-1}$-norm, applying MINRES to (3.10) amounts to solving

$$\underset{\mathbf{y}}{\text{minimize}} \ \|(\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T)\mathbf{y} - \mathbf{b}\|_{\mathbf{M}^{-1}}.$$

Note that this approach is justified even if $\mathbf{A}\mathbf{x} = \mathbf{b}$ is not consistent, in which case the residual above is nonzero. Seeking an approximation of the form $\mathbf{y}_k = \mathbf{U}_k y_k$, (3.5) shows that the residual may be written

$$\begin{aligned}\mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T\mathbf{U}_k y_k - \mathbf{M}^{-\frac{1}{2}}\mathbf{b} &= \mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{V}_k\mathbf{L}_k^T y_k - \beta_1\mathbf{M}^{\frac{1}{2}}\mathbf{u}_1 \\ &= \mathbf{M}^{\frac{1}{2}}\mathbf{U}_{k+1}(\mathbf{B}_k\mathbf{L}_k^T y_k - \beta_1\mathbf{e}_1).\end{aligned}$$

Using the orthogonality of $\mathbf{M}^{\frac{1}{2}}\mathbf{U}_{k+1}$, we obtain the subproblem

$$\underset{y}{\text{minimize}} \ \left\| \begin{bmatrix} \mathbf{L}_k\mathbf{L}_k^T \\ \alpha_k\beta_{k+1}\mathbf{e}_k^T \end{bmatrix} y - \beta_1\mathbf{e}_1 \right\|_2,$$

and the Lanczos process

$$\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T\mathbf{U}_k = \mathbf{M}\mathbf{U}_k\mathbf{L}_k\mathbf{L}_k^T + \alpha_k\beta_{k+1}\mathbf{M}\mathbf{u}_{k+1}\mathbf{e}_k^T, \tag{3.12}$$

where $\mathbf{L}_k\mathbf{L}_k^T$ is symmetric and tridiagonal. This Lanczos process should be compared with (2.6). Arioli and Orban (2013) provide complete algorithmic details, including the case where regularization terms are present. The defining properties of CRAIG-MR are summarized in the next result.

Proposition 3.3 *The preconditioned* CRAIG-MR *iterates on* (2.1) *are the same as those generated by the standard* MINRES *on the positive semi-definite system* (3.10) *with preconditioner* $\mathbf{M}$. *The residual* $\mathbf{b} - (\mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T)\mathbf{y}_k$ *decreases along the iterations in the* $\mathbf{M}^{-1}$*-norm.*

In the next section, we establish connections between the methods of this section and Krylov methods applied to (3.2).

4 Preconditioned Full-Space Methods

The relations (3.5) combine into

$$\begin{aligned}\begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix}\begin{bmatrix} \mathbf{U}_k & \\ & \mathbf{V}_k \end{bmatrix} &= \begin{bmatrix} \mathbf{M} & \\ & \mathbf{N} \end{bmatrix}\begin{bmatrix} \mathbf{U}_k & \\ & \mathbf{V}_k \end{bmatrix}\begin{bmatrix} \mathbf{I}_k & \mathbf{L}_k \\ \mathbf{L}_k^T & \end{bmatrix} + \begin{bmatrix} \beta_{k+1}\mathbf{M}\mathbf{u}_{k+1} \\ \mathbf{0} \end{bmatrix}\mathbf{e}_{2k}^T \\ \begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix}\begin{bmatrix} \mathbf{U}_{k+1} & \\ & \mathbf{V}_k \end{bmatrix} &= \begin{bmatrix} \mathbf{M} & \\ & \mathbf{N} \end{bmatrix}\begin{bmatrix} \mathbf{U}_{k+1} & \\ & \mathbf{V}_k \end{bmatrix}\begin{bmatrix} \mathbf{I}_{k+1} & \mathbf{B}_k \\ \mathbf{B}_k^T & \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \alpha_{k+1}\mathbf{N}\mathbf{v}_{k+1} \end{bmatrix}\mathbf{e}_{2k+1}^T.\end{aligned} \tag{4.1}$$

According to (2.6), (4.1) describe a Lanczos process applied to a symmetric saddle-point matrix in a metric defined by the block-diagonal matrix $\text{blkdiag}(\mathbf{M}, \mathbf{N})$. Indeed, as long as $\alpha_{k+1} \neq 0$ and $\beta_{k+1} \neq 0$, the symmetric saddle-point matrices in the right-hand side of (4.1) are nonsingular. Yet, those matrices may not have an LU factorization without pivoting and hence, the conjugate gradient method may not be well defined. The permutation $\mathbf{P}_k := \begin{bmatrix}\mathbf{e}_1 & \mathbf{e}_{k+1} & \mathbf{e}_2 & \mathbf{e}_{k+2} & \cdots & \mathbf{e}_k & \mathbf{e}_{2k}\end{bmatrix}$ pointed out by Paige (1974) restores the order in which the vectors $\mathbf{u}_k$ and $\mathbf{v}_k$ are generated by Algorithm 3.1, i.e.,

$$\mathbf{P}_k\begin{bmatrix} \mathbf{U}_k & \\ & \mathbf{V}_k \end{bmatrix}\mathbf{P}_k^T = \begin{bmatrix} \mathbf{u}_1 & \mathbf{0} & \cdots & \mathbf{u}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_1 & \cdots & \mathbf{0} & \mathbf{v}_k \end{bmatrix}.$$

The same permutation highlights the tridiagonal matrix generated after $2k$ iterations of the Lanczos process (4.1):

$$\mathbf{T}_{2k} = \mathbf{P}_k \begin{bmatrix} \mathbf{I}_k & \mathbf{L}_k \\ \mathbf{L}_k^T & \end{bmatrix} \mathbf{P}_k^T = \begin{bmatrix} 1 & \alpha_1 & & & \\ \alpha_1 & 0 & \beta_2 & & \\ & \beta_2 & 1 & \ddots & \\ & & \ddots & \ddots & \alpha_k \\ & & & \alpha_k & 0 \end{bmatrix},$$

$$\mathbf{T}_{2k+1} = \begin{bmatrix} \mathbf{T}_{2k} & \beta_{k+1}\mathbf{e}_{2k} \\ \beta_{k+1}\mathbf{e}_{2k}^T & 1 \end{bmatrix}.$$

Equipped with the above, we are in position to examine preconditioned full-space methods targeting (3.2) directly.

4.1 Preconditioned SYMMLQ

By definition, the SYMMLQ method of Paige and Saunders (1975) is the method referred to as D-Lanczos by Saad (2003, Algorithm 6.17) that solves the system $\mathbf{T}_k\mathbf{w}_k = \beta_1\mathbf{e}_1$ where $\mathbf{w} = (\mathbf{r}, \mathbf{x})$ by way of an LQ factorization of $\mathbf{T}_k$. For stability reasons, SYMMLQ generates iterates that differ from those $\mathbf{w}_k$, but that allow to recover the latter easily if desired and if they are well defined. In exact arithmetic, it is equivalent in principle to SYMMBK (Chandra, 1978), which solves the tridiagonal system by way of a symmetric indefinite factorization.

The Lanczos process (4.1) implies directly the following result.

Theorem 4.1 *The application of* SYMMLQ *to* (3.2) *with preconditioner* blkdiag$(\mathbf{M}, \mathbf{N})$ *amounts to performing an iteration of the preconditioned* CRAIG *method every even step and an iteration of the preconditioned* LSQR *method every odd step.*

Theorem 4.1 indicates that SYMMLQ may be viewed as a predictor-corrector process in that its iterations come in two kinds. The first sets a predicted solution estimate according to the CRAIG method, which would be appropriate were the residual to vanish at the solution. The second applies a correction given as an LSQR step.

Note that when applied to a symmetric positive definite operator, SYMMLQ is, in a sense, equivalent the conjugate gradient method (Saad, 2003, Section 6.7) because it may be used to generate the conjugate gradient iterates. In the presence of regularization, the permuted tridiagonal in (4.1) takes the form

$$\begin{bmatrix} \mathbf{I}_k & \mathbf{L}_k \\ \mathbf{L}_k^T & -\mathbf{I}_k \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{I}_k & \mathbf{B}_k \\ \mathbf{B}_k^T & -\mathbf{I}_k \end{bmatrix},$$

both of which are *symmetric and quasi definite* (Vanderbei, 1995). Arioli and Orban (2013) establish that SYMMLQ and CG are also equivalent in the presence of symmetric quasi-definite operators. However, in the absence of regularization, SYMMLQ and CG differ and in fact, CG may not be well defined. In this sense, Theorem 4.1 is not a special case of (Arioli and Orban, 2013, Theorem 8.1).

4.2 Preconditioned MINRES

Applying the MINRES method of Paige and Saunders (1975) to (3.2) amounts to directly minimizing the norm of the residual. As the latter should be measured in blkdiag$(\mathbf{M}^{-1}, \mathbf{N}^{-1})$-norm, the minimization problem may be stated as

$$\underset{\mathbf{r},\mathbf{x}}{\text{minimize}} \ \tfrac{1}{2} \left\| \begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \end{bmatrix} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_{\mathbf{H}^{-1}}, \tag{4.2}$$

where $\mathbf{H} := \text{blkdiag}(\mathbf{M}, \mathbf{N})$. Seeking approximations $\mathbf{r}_k = \mathbf{U}_k r_k$ and $\mathbf{x}_k = \mathbf{V}_k x_k$, the optimality conditions of (4.2) impose the Ritz-Galerkin condition

$$\begin{bmatrix} \mathbf{U}_k & \\ & \mathbf{V}_k \end{bmatrix}^T \begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix} \begin{bmatrix} \mathbf{M}^{-1} & \\ & \mathbf{N}^{-1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix} \begin{bmatrix} \mathbf{U}_k r_k \\ \mathbf{V}_k x_k \end{bmatrix} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right) = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}. \tag{4.3}$$

Using (3.5), this may be rewritten

$$\left(\begin{bmatrix} \mathbf{I}_k & \mathbf{L}_k \\ \mathbf{L}_k^T & \end{bmatrix}^2 + \beta_{k+1}^2 \mathbf{e}_{2k} \mathbf{e}_{2k}^T \right) \begin{bmatrix} r_k \\ x_k \end{bmatrix} = \begin{bmatrix} \beta_1 \mathbf{e}_1 \\ \alpha_1 \beta_1 \mathbf{e}_1 \end{bmatrix}.$$

This system represents the optimality conditions of

$$\underset{r,x}{\text{minimize}} \ \tfrac{1}{2} \left\| \begin{bmatrix} \mathbf{I}_{k+1} & \mathbf{B}_k \\ \mathbf{B}_k^T & \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} - \begin{bmatrix} \beta_1 \mathbf{e}_1 \\ \mathbf{0} \end{bmatrix} \right\|_2,$$

which is the subproblem solved at the k-th preconditioned MINRES iteration. Indeed,

$$\begin{bmatrix} \mathbf{I}_k & \mathbf{L}_k \\ \mathbf{L}_k^T & \end{bmatrix}^2 = \begin{bmatrix} \mathbf{I}_k + \mathbf{L}_k \mathbf{L}_k^T & \mathbf{L}_k \\ \mathbf{L}_k^T & \mathbf{L}_k^T \mathbf{L}_k \end{bmatrix}$$

and

$$\mathbf{B}_k^T \mathbf{B}_k = \mathbf{L}_k^T \mathbf{L}_k + \beta_{k+1}^2 \mathbf{e}_k \mathbf{e}_k^T.$$

Upon pasting the Lanczos processes (3.8) and (3.12) together, we obtain

$$\begin{aligned} \begin{bmatrix} \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T + \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{A}^T \mathbf{M}^{-1} \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{U}_k & \\ & \mathbf{V}_k \end{bmatrix} = & \\ \begin{bmatrix} \mathbf{M} & \\ & \mathbf{N} \end{bmatrix} \begin{bmatrix} \mathbf{U}_k & \\ & \mathbf{V}_k \end{bmatrix} \begin{bmatrix} \mathbf{I}_k + \mathbf{L}_k \mathbf{L}_k^T & \mathbf{L}_k \\ \mathbf{L}_k^T & \mathbf{B}_k^T \mathbf{B}_k \end{bmatrix} & \\ + \beta_{k+1} \begin{bmatrix} \mathbf{M} & \\ & \mathbf{N} \end{bmatrix} \begin{bmatrix} \alpha_k \mathbf{u}_{k+1} \mathbf{e}_k^T & \mathbf{u}_{k+1} \mathbf{e}_k^T \\ & \alpha_{k+1} \mathbf{v}_{k+1} \mathbf{e}_k^T \end{bmatrix}. \end{aligned} \tag{4.4}$$

The operator in the above Lanczos-type process is none other than

$$\begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix} \begin{bmatrix} \mathbf{M}^{-1} & \\ & \mathbf{N}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{N}^{-1}\mathbf{A}^T + \mathbf{M} & \mathbf{A} \\ \mathbf{A}^T & \mathbf{A}^T \mathbf{M}^{-1} \mathbf{A} \end{bmatrix}, \tag{4.5}$$

and is the operator appearing in (4.3). Thus, according to (2.6), (4.4) represents a Lanczos process on the operator (4.5) in the metric defined by $\mathbf{H}$. This Lanczos process characterizes the preconditioned MINRES method for (3.2).

We have established the following result, which parallels (Arioli and Orban, 2013, Theorem 8.3).

Theorem 4.2 *The application of* MINRES *to* (3.2) *with preconditioner* $\text{blkdiag}(\mathbf{M}, \mathbf{N})$ *amounts to performing an iteration of the preconditioned* CRAIG-MR *method every even step and an iteration of the preconditioned* LSMR *method every odd step.*

5 The Projected Golub-Kahan Bidiagonalization Process

Let us now return our attention to the nullspace-constrained problem (1.1). Let the columns of the n-by-q matrix $\mathbf{Z}$ form a basis for $\text{Null}(\mathbf{E})$ and let $\mathbf{x} = \mathbf{Z}x$, where $x \in \mathbb{R}^q$. Then (1.1) may be equivalently written

$$\underset{x \in \mathbb{R}^q}{\text{minimize}} \ \tfrac{1}{2} \|\mathbf{A}\mathbf{Z}x - \mathbf{b}\|_{\mathbf{M}^{-1}}^2. \tag{5.1}$$

The common optimality conditions of (1.1) and (5.1) now simplify to

$$\begin{bmatrix} \mathbf{M} & \mathbf{AZ} \\ \mathbf{Z}^T\mathbf{A}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ x \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}. \tag{5.2}$$

Numerical methods for (5.1) typically attempt to solve the normal equations

$$\mathbf{Z}^T\mathbf{A}^T\mathbf{M}^{-1}\mathbf{AZ}x = \mathbf{Z}^T\mathbf{A}^T\mathbf{M}^{-1}\mathbf{b}, \tag{5.3}$$

for instance using a suitable implementation of the conjugate gradient or minimum residual method. In this context, it is natural to select a preconditioner of the form $\mathbf{N} := \mathbf{Z}^T\mathbf{GZ}$ for some $\mathbf{G}$ that is symmetric and positive definite on Null($\mathbf{E}$), i.e., such that $\mathbf{Z}^T\mathbf{GZ}$ is symmetric and positive definite. Intuitively, it should be beneficial to select $\mathbf{G}$ as a suitable approximation to $\mathbf{A}^T\mathbf{M}^{-1}\mathbf{A}$. This intuition is confirmed by, e.g., (Gould et al., 2013, Theorem 4.1).

Equipped with the above observations, we apply Algorithm 3.1 to the operator $\mathbf{AZ}$ with right-hand side $\mathbf{b}$ and preconditioner $\mathbf{N} := \mathbf{Z}^T\mathbf{GZ}$. After the change of variable $\mathbf{v}_i \leftarrow \mathbf{Z}\mathbf{v}_i$, we obtain Algorithm 5.1, where

$$\mathbf{P}_\mathbf{G} := \mathbf{Z}(\mathbf{Z}^T\mathbf{GZ})^{-1}\mathbf{Z}^T, \tag{5.4}$$

and where we define the seminorm

$$\|\mathbf{w}\|^2_{[\mathbf{G}]} := \mathbf{w}^T\mathbf{P}_\mathbf{G}\mathbf{w}, \tag{5.5}$$

and its dual seminorm, defined on Null($\mathbf{E}$),

$$\|\mathbf{v}\|^2_{\{\mathbf{G}\}} = \mathbf{v}^T\mathbf{G}\mathbf{v}. \tag{5.6}$$

If $\mathbf{v} \in$ Null($\mathbf{E}$) and $\mathbf{v} = \mathbf{Z}\bar{\mathbf{v}}$, then $\|\mathbf{v}\|_{\{\mathbf{G}\}} = \|\bar{\mathbf{v}}\|_{\mathbf{Z}^T\mathbf{GZ}}$. It is easy to establish by a recursion argument that $\mathbf{v}_k \in$ Null($\mathbf{E}$) for all k.

Algorithm 5.1 Projected Golub-Kahan Bidiagonalization Process for (5.1)

Require: $\mathbf{A}$, $\mathbf{b}$, $\mathbf{M}^{-1}$, $\mathbf{G}$
1: $\beta_1\mathbf{M}\mathbf{u}_1 = \mathbf{b}$ with $\beta_1 > 0$ so that $\|\mathbf{u}_1\|_\mathbf{M} = 1$
2: $\alpha_1\mathbf{v}_1 = \mathbf{P}_\mathbf{G}\mathbf{A}^T\mathbf{u}_1$ with $\alpha_1 > 0$ so that $\|\mathbf{v}_1\|_{\{\mathbf{G}\}} = 1$
3: **for** $k = 1, 2, \ldots$ **do**
4: $\quad\beta_{k+1}\mathbf{M}\mathbf{u}_{k+1} = \mathbf{A}\mathbf{v}_k - \alpha_k\mathbf{M}\mathbf{u}_k$ with $\beta_{k+1} > 0$ so that $\|\mathbf{u}_{k+1}\|_\mathbf{M} = 1$
5: $\quad\alpha_{k+1}\mathbf{v}_{k+1} = \mathbf{P}_\mathbf{G}\mathbf{A}^T\mathbf{u}_{k+1} - \beta_{k+1}\mathbf{v}_k$ with $\alpha_{k+1} > 0$ so that $\|\mathbf{v}_{k+1}\|_{\{\mathbf{G}\}} = 1$.

Algorithm 5.1 may be summarized with the identities

$$\mathbf{A}\mathbf{V}_k = \mathbf{M}\mathbf{U}_k\mathbf{L}_k + \beta_{k+1}\mathbf{M}\mathbf{u}_{k+1}\mathbf{e}_k^T = \mathbf{M}\mathbf{U}_{k+1}\mathbf{B}_k, \tag{5.7a}$$

$$\mathbf{P}_\mathbf{G}\mathbf{A}^T\mathbf{U}_{k+1} = \mathbf{V}_k\mathbf{B}_k^T + \alpha_{k+1}\mathbf{v}_{k+1}\mathbf{e}_{k+1}^T \quad = \mathbf{V}_{k+1}\mathbf{L}_{k+1}^T, \tag{5.7b}$$

which replace (3.5). In addition, the identities

$$\mathbf{U}_k^T\mathbf{M}\mathbf{U}_k = \mathbf{I}_k \quad \text{and} \quad \mathbf{V}_k^T\mathbf{G}\mathbf{V}_k = \mathbf{I}_k$$

are satisfied in exact arithmetic. As before, this implies that $\mathbf{M}^{\frac{1}{2}}\mathbf{U}_k$ is orthogonal but care must be exercised because $\mathbf{G}$ may be singular.

We refer to Algorithm 5.1 as the *projected* variant of the Golub-Kahan process because it computes basis vectors $\mathbf{v}_k$ of the intersection of a Krylov space associated to $\mathbf{A}^T\mathbf{M}^{-1}\mathbf{A}$ with Null($\mathbf{E}$) by repeated projection. This is a slight abuse of terminology since the operator $\mathbf{P}_\mathbf{G}$ is not a projector but rather shares similarities with a projector into Null($\mathbf{E}$). More precisely, the operator $\mathbf{P}_\mathbf{G}\mathbf{G}$ is an oblique projector into Null($\mathbf{E}$).

Implementing projected variants of classic methods for linear least-squares problems such as LSQR, CRAIG, LSMR and CRAIG-MR simply consists in using Algorithm 5.1 in place of Algorithms 2.1 or 3.1. The computation of $\mathbf{v} := \mathbf{P}_\mathbf{G}\mathbf{w}$ may be performed by solving the symmetric saddle-point system

$$\begin{bmatrix} \mathbf{G} & \mathbf{E}^T \\ \mathbf{E} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} \mathbf{w} \\ \mathbf{0} \end{bmatrix} \tag{5.8}$$

by way of a factorization or an iterative method that ensures that $\mathbf{E}\mathbf{v} = \mathbf{0}$ holds accurately. Therefore, it is possible to view the operator $\mathbf{P}_\mathbf{G}$ as the composition

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{G} & \mathbf{E}^T \\ \mathbf{E} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}, \tag{5.9}$$

which shows that $\mathbf{P}_\mathbf{G}$ is in fact independent of the choice of $\mathbf{Z}$. For this reason, we intentionally omitted $\mathbf{Z}$ from the requirements of Algorithm 5.1.

The computation of the scaling factor $\alpha_{k+1} > 0$ such that $\|\mathbf{v}_{k+1}\|_{\{\mathbf{G}\}} = 1$ in Algorithm 5.1 is performed as follows. First, compute $\mathbf{P}_\mathbf{G}\mathbf{A}^T\mathbf{u}_{k+1}$ by solving (5.8) with $\mathbf{w} = \mathbf{A}^T\mathbf{u}_{k+1}$ and denote the top segment of the solution vector as $\mathbf{v}_{k+1}$. Because $\mathbf{G}\mathbf{v}_{k+1} + \mathbf{E}^T\mathbf{z} = \mathbf{A}^T\mathbf{u}_{k+1}$ and $\mathbf{E}\mathbf{v}_{k+1} = \mathbf{0}$, taking the inner product of the first block equation with $\mathbf{v}_{k+1}$ yields $\mathbf{v}_{k+1}^T\mathbf{G}\mathbf{v}_{k+1} = \mathbf{v}_{k+1}\mathbf{A}^T\mathbf{u}_{k+1}$. By orthogonality, $\mathbf{v}_{k+1}^T\mathbf{G}\mathbf{v}_k = 0$. In other words, $\|\mathbf{v}_{k+1}\|_{\{\mathbf{G}\}} = \|\mathbf{A}^T\mathbf{u}_{k+1}\|_{[\mathbf{G}]}$.

Note that in the case where $\mathbf{G}$ is symmetric and positive definite, the explicit form of the inverse

$$\begin{bmatrix} \mathbf{G} & \mathbf{E}^T \\ \mathbf{E} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{G}^{-1} - \mathbf{G}^{-1}\mathbf{E}^T\mathbf{S}^{-1}\mathbf{E}\mathbf{G}^{-1} & \mathbf{G}^{-1}\mathbf{E}^T\mathbf{S}^{-1} \\ \mathbf{S}^{-1}\mathbf{E}\mathbf{G}^{-1} & \mathbf{S}^{-1} \end{bmatrix},$$

where $\mathbf{S} := \mathbf{E}\mathbf{G}^{-1}\mathbf{E}^T$ is the (negative) Schur complement, shows that the leading block of the inverse is indeed a projector. In this case, $\mathbf{w}$ may equivalently be computed as the (preconditioned) residual of

$$\underset{\mathbf{z}}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{E}^T\mathbf{z} - \mathbf{w}\|^2_{\mathbf{G}^{-1}},$$

and any appropriate direct or iterative method may be employed provided it ensures that $\mathbf{E}\mathbf{v} = \mathbf{0}$ holds accurately.

In the central-preconditioned variant of (5.2), the coefficient matrix corresponding to (3.4) may now be written

$$\bar{\mathbf{K}} := \begin{bmatrix} \mathbf{I} & \bar{\mathbf{A}} \\ \bar{\mathbf{A}}^T & \mathbf{0} \end{bmatrix}, \qquad \bar{\mathbf{A}} := \mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{Z}\mathbf{N}^{-\frac{1}{2}}.$$

We extend the nomenclature of Arioli and Orban (2013) and refer to the singular values of $\bar{\mathbf{A}}$ as the *constrained elliptical values* of $\mathbf{A}$. By virtue of Proposition 2.1, those values govern the convergence of iterative methods for (1.1). Algorithm 5.1 is precisely the Golub-Kahan process applied to the operator $\bar{\mathbf{A}}$ and right-hand side $\bar{\mathbf{b}} = \mathbf{M}^{-\frac{1}{2}}\mathbf{b}$.

A direct consequence of (Paige, 1974, Section 2) and (Arioli and Orban, 2013, Section 4) is that Algorithm 5.1 terminates with $\beta_{k+1} = 0$ in exact arithmetic if and only if $\mathbf{b} \in \text{Range}(\mathbf{A}\mathbf{Z})$, i.e., in the range of $\mathbf{A}$ restricted to $\text{Null}(\mathbf{E})$. In the case of nonzero residual problems, Algorithm 5.1 must terminate with $\alpha_{k+1} = 0$, i.e., $\mathbf{v}_{k+1} = \mathbf{0}$. Geometrically, this means that the next basis vector would have been orthogonal to $\text{Null}(\mathbf{E})$ in the $\mathbf{G}$-norm.

We are now in position to describe numerical methods for the reduced linear least-squares problem (5.1). Because $\mathbf{Z}$ is an artefact of the analysis and never need be computed, the methods below also solve (1.1). Conceptually, they may be seen as solving (5.2), or equivalently, (5.3), and in this sense, they correspond to the nullspace method for saddle-point systems (Benzi, Golub, and Liesen, 2005, §6), with the difference that in the present case, the reduced problem is itself stated as a symmetric saddle-point system.

Gould, Orban, and Rees (2013) describe projected Krylov methods for saddle-point systems. Such methods applied to (1.3) are formally equivalent to applying a Krylov method to the corresponding nullspace system, which happens to be precisely (5.3). This observation allows us to establish connections between numerical methods for (1.1) and projected Krylov methods for (1.3). The same authors establish that projected Krylov methods are also equivalent to *constraint-preconditioned* Krylov methods (Keller, Gould, and Wathen, 2000), i.e., Krylov methods applied to (1.3) with an initial guess $(\mathbf{x}_0, \mathbf{y}_0)$ satisfying $\mathbf{E}\mathbf{x}_0 = \mathbf{0}$ and with the preconditioner (5.8)—note that this is quite different from projected methods, which use the

preconditioner (5.9). In this case, they establish that any constraint-preconditioned Krylov method guarantees that $\mathbf{E}\mathbf{x}_k = \mathbf{0}$ remains satisfied throughout the iterations. Note that the preconditioner (5.8) is indefinite but constraint-preconditioned Krylov methods are a special case where applying methods such as CG, SYMMLQ and MINRES with an indefinite preconditioner is permitted.

We term the methods below *constrained* to emphasize the fact that they solve the nullspace-constrained problem (1.1). For each method, the subproblem solved at iteration k is identical to that in the corresponding method of §3.

5.1 Constrained LSQR

Replacing Algorithm 3.1 with Algorithm 5.1, the constrained LSQR method seeks an approximation in Null($\mathbf{E}$) as a combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, i.e., of the form $\mathbf{x}_k = \mathbf{V}_k x_k \in \text{Null}(\mathbf{E})$. The vector x_k is chosen to minimize the least-squares residual (5.1). The next result follows directly from Proposition 3.1.

Corollary 5.1 *The constrained* LSQR *iterates on* (1.1) *are the same as those generated by the standard* LSQR *on* (5.1). *They also coincide with those generated by the standard* CG *on the positive definite system* (5.3) *with preconditioner* $\mathbf{N} = \mathbf{Z}^T\mathbf{G}\mathbf{Z}$.

By definition of LSQR, we have also established a link with the projected CG method described by Gould et al. (2001) and the projected and constraint-preconditioned CG methods of Gould et al. (2013).

Corollary 5.2 *The constrained* LSQR *method is equivalent to the projected* CG *method applied to* (1.1) *or, equivalently, to* (1.3). *In turn, this is equivalent to the constraint-preconditioned* CG, *i.e., the standard* CG *method applied to the symmetric indefinite system* (1.3) *with preconditioner* (5.8).

It follows from (5.7) that the projected LSQR method is characterized by the Lanczos-like process

$$\mathbf{P}_{\mathbf{G}}\mathbf{A}^T\mathbf{M}^{-1}\mathbf{A}\mathbf{V}_k = \mathbf{V}_k\mathbf{B}_k^T\mathbf{B}_k + \alpha_{k+1}\beta_{k+1}\mathbf{v}_{k+1}\mathbf{e}_{k+1}^T. \tag{5.10}$$

Note that (5.10), due to a projection operation, does not exactly have the form (2.6) because the operator in the left-hand side is not symmetric.

We may expect that the benefits of LSQR over the conjugate gradient method in terms of stability and accuracy carry over to the present constrained context.

5.2 Constrained CRAIG

This method is identical in principle to that of §3.2. We first rewrite (5.2) as

$$\begin{bmatrix} \mathbf{M} & \mathbf{A} \\ \mathbf{Z}^T\mathbf{A}^T & \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}$$

and seek approximations $\mathbf{r} \approx \mathbf{U}_k r_k$ and $\mathbf{x} \approx \mathbf{V}_k x_k$. The first block equation becomes $\mathbf{M}\mathbf{U}_k r_k + \mathbf{A}\mathbf{V}_k x_k = \mathbf{b}$. After substituting $\mathbf{b} = \beta_1\mathbf{M}\mathbf{u}_1$ and premultiplying with $\mathbf{U}_k^T$, we obtain $r_k + \mathbf{L}_k s_k = \beta_1\mathbf{e}_1$. Premultiplying the second block equation with $\mathbf{Z}(\mathbf{Z}^T\mathbf{G}\mathbf{Z})^{-1}$ yields $\mathbf{0} = \mathbf{P}_{\mathbf{G}}\mathbf{A}^T\mathbf{U}_k r_k = \mathbf{V}_k\mathbf{L}_k^T r_k$. Premultiplying the result with $\mathbf{V}_k^T\mathbf{G}$ finally yields $\mathbf{L}_k^T r_k = \mathbf{0}$. In summary, we have again obtained the reduced system (3.9). Although admittedly of limited interest, the following result summarizes the above.

Corollary 5.3 *The constrained* CRAIG *iterates on* (1.1) *are the same as those generated by the standard* CRAIG *method on* (5.1).

5.3 Constrained LSMR

The constrained LSMR method consists in applying MINRES to (5.3), i.e., to solve

$$\underset{x}{\text{minimize}} \ \|\mathbf{Z}^T\mathbf{A}^T\mathbf{M}^{-1}(\mathbf{A}\mathbf{Z}x - \mathbf{b})\|_{(\mathbf{Z}^T\mathbf{G}\mathbf{Z})^{-1}}.$$

This relationship is formalized in the next corollary.

Corollary 5.4 *The constrained* LSMR *iterates on* (1.1) *are the same as those generated by the standard* LSMR *on* (5.1). *They also coincide with those generated by the standard* MINRES *on the positive definite system* (5.3) *with preconditioner* $\mathbf{N} = \mathbf{Z}^T\mathbf{G}\mathbf{Z}$.

The next result states a connection between the constrained LSMR method and the projected and constraint-preconditioned MINRES methods of Gould et al. (2013). Indeed by construction, the projected MINRES is equivalent to applying MINRES to (5.3).

Corollary 5.5 *The constrained* LSMR *method is equivalent to the projected* MINRES *method applied to* (1.1) *or, equivalently, to* (1.3). *In turn, this is equivalent to the constraint-preconditioned* MINRES, *i.e., the standard* MINRES *method applied to the symmetric indefinite system* (1.3) *with preconditioner* (5.8).

The Lanczos process associated to the projected LSMR method is identical to (5.10).

5.4 Constrained CRAIG-MR

When $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent, the constrained minimum-norm problem corresponding to (3.11) reads

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{x}\|^2_{[\mathbf{G}]} \quad \text{subject to} \ \ \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{E}\mathbf{x} = \mathbf{0}, \tag{5.11}$$

where $\|\mathbf{x}\|_{[\mathbf{G}]}$ denotes the seminorm (5.5) induced by $\mathbf{G}$ in Null($\mathbf{E}$). The optimality conditions of (5.11) may be expressed as

$$\mathbf{A}\mathbf{Z}(\mathbf{Z}^T\mathbf{G}\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{A}^T\mathbf{y} = \mathbf{b}, \tag{5.12}$$

where $\mathbf{y}$ are Lagrange multipliers associated to the constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$. Note that (3.10) may also be written $\mathbf{A}\mathbf{P}_{\mathbf{G}}\mathbf{A}^T\mathbf{y} = \mathbf{b}$.

The constrained CRAIG-MR method consists in applying MINRES to (5.12), i.e., to solve

$$\underset{\mathbf{y}}{\text{minimize}} \ \|\mathbf{A}\mathbf{P}_{\mathbf{G}}\mathbf{A}^T\mathbf{y} - \mathbf{b}\|_{\mathbf{M}^{-1}}. \tag{5.13}$$

As before, CRAIG-MR is well defined even if $\mathbf{A}\mathbf{x} = \mathbf{b}$ is not consistent. The following result summarizes the relationship between the constrained and standard CRAIG-MR, and their relation to MINRES.

Corollary 5.6 *The constrained* CRAIG-MR *iterates on* (1.1) *are the same as those generated by the standard* CRAIG-MR *on* (5.1). *They also coincide with those generated by the standard* MINRES *on the positive semi-definite system* (5.12) *with preconditioner* $\mathbf{M}$.

Seeking an approximation $\mathbf{y} \approx \mathbf{y}_k = \mathbf{U}_k y_k$ in (5.13), it follows from (5.7) that the constrained CRAIG-MR method is characterized by the Lanczos process

$$\mathbf{A}\mathbf{P}_{\mathbf{G}}\mathbf{A}^T\mathbf{U}_k = \mathbf{M}\mathbf{U}_k\mathbf{L}_k\mathbf{L}_k^T + \alpha_k\beta_{k+1}\mathbf{M}\mathbf{u}_{k+1}\mathbf{e}_k^T. \tag{5.14}$$

6 Constrained Full-Space Methods and Full-Space Methods

This section summarizes connections between the constrained methods of the previous section and on the one hand, *constrained full-space methods*, i.e., methods applied to (5.2), and on the other hand, with *full-space methods*, i.e., methods applied directly to (1.4). We use the notation $\mathbf{Q}$, $\mathbf{J}$ and $\mathbf{f}$ introduced in (1.5).

The projected Krylov methods described by Gould et al. (2013) are full-space formulations of nullspace methods. For (1.4), this is equivalent to applying a Krylov method to the restriction of $\mathbf{Q}$ to the nullspace of $\mathbf{J}$. A basis for the nullspace of $\mathbf{J}$ is given by the columns of

$$\tilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times q} \\ \mathbf{0}_{n\times m} & \mathbf{Z} \end{bmatrix},$$

where, as before, the columns of $\mathbf{Z}$ form a basis for Null($\mathbf{E}$) and where we indicated the block dimensions for clarity. The reduced operator $\tilde{\mathbf{Z}}^T\mathbf{Q}\tilde{\mathbf{Z}}$ and right-hand side $\tilde{\mathbf{Z}}\mathbf{f}$ are precisely the operator and right-hand side of (5.2). Projected Krylov methods rely on a preconditioner $\mathbf{G} = \mathbf{G}^T \approx \mathbf{Q}$ such that $\tilde{\mathbf{Z}}^T\mathbf{G}\tilde{\mathbf{Z}}$ is positive definite on Null($\mathbf{J}$). Using the same block decomposition as (1.5), we may write

$$\begin{aligned} \tilde{\mathbf{Z}}^T\mathbf{G}\tilde{\mathbf{Z}} &= \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times q} \\ \mathbf{0}_{n\times m} & \mathbf{Z} \end{bmatrix}^T \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{21}^T \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m\times q} \\ \mathbf{0}_{n\times m} & \mathbf{Z} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{21}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{G}_{21} & \mathbf{Z}^T\mathbf{G}_{22}\mathbf{Z} \end{bmatrix}. \end{aligned} \tag{6.1}$$

A natural choice is to pick $\mathbf{G}_{11}$ positive definite—and possibly $\mathbf{G}_{11} = \mathbf{M}$—$\mathbf{G}_{21} = \mathbf{0}$ and $\mathbf{G}_{22}$ to be positive definite over Null($\mathbf{E}$).

Constraint-preconditioned Krylov methods are Krylov methods applied directly to (1.4) and correspondingly rely on the preconditioner

$$\begin{bmatrix} \mathbf{G} & \mathbf{J}^T \\ \mathbf{J} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{21}^T & \\ \mathbf{G}_{21} & \mathbf{G}_{22} & \mathbf{E}^T \\ & \mathbf{E} & \end{bmatrix}. \tag{6.2}$$

As in (5.9), the preconditioner (6.1) is related to (6.2) according to

$$\tilde{\mathbf{Z}}^T\mathbf{G}\tilde{\mathbf{Z}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{G} & \mathbf{J}^T \\ \mathbf{J} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}. \tag{6.3}$$

6.1 SYMMLQ

The discussion of the previous section and (Gould et al., 2013, Theorem 2.3) establish the following result.

Theorem 6.1 *The iterates generated by the projected* SYMMLQ *method applied to* (1.4) *are the same as those generated by the constraint-preconditioned* SYMMLQ *method applied to* (1.4)*, i.e., with preconditioner* (6.2)*. They also coincide with those generated by* SYMMLQ *applied to* (5.2) *with preconditioner* (6.1)*.*

We may now use the parallel between, on the one hand, (1.4) and (5.2), and on the other hand, (1.3) and (5.3) to establish a result corresponding to Theorem 4.1.

Theorem 6.2 *The application of* SYMMLQ *to* (1.4) *with preconditioner* (6.2)*, or of any equivalent form of* SYMMLQ *described in Theorem 6.1, amounts to performing an iteration of the constrained* CRAIG *method every even step and an iteration of the constrained* LSQR *method every odd step.*

6.2 MINRES

The results of the previous section carry over to MINRES using the same principles. We state them for completeness.

Theorem 6.3 *The iterates generated by the projected* MINRES *method applied to* (1.4) *are the same as those generated by the constraint-preconditioned* MINRES *method applied to* (1.4)*, i.e., with preconditioner* (6.2)*. They also coincide with those generated by* MINRES *applied to* (5.2) *with preconditioner* (6.1)*.*

Theorem 6.4 *The application of* MINRES *to* (1.4) *with preconditioner* (6.2)*, or of any equivalent form of* MINRES *described in Theorem 6.3, amounts to performing an iteration of the constrained* CRAIG-MR *method every even step and an iteration of the constrained* LSMR *method every odd step.*

7 Application to Optimization

In constrained optimization, it is often beneficial to treat linear constraints explicitly. At iteration k of a sequential linear or quadratic programming scheme to solve (1.2), a step is computed that must improve two conflicting objectives: the first is the objective function value, and the second is constraint satisfaction. For this reason, the step is often decomposed as the sum of a normal and a tangential step (Nocedal and Wright, 2006). The role of the normal step, denoted $\mathbf{s}$ for simplicity, is to improve constraint satisfaction without regard to the objective function. Assuming each $\mathbf{x}_k$ is computed so as to satisfy the linear equality constraints, the normal subproblem is typically formulated as (Conn, Gould, and Toint, 2000, Section 15.4)

$$\underset{\mathbf{s}}{\text{minimize}} \ \tfrac{1}{2}\|\mathbf{A}_k\mathbf{s} + \mathbf{c}_k\|^2_{\mathbf{M}^{-1}} \quad \text{subject to } \mathbf{E}\mathbf{s} = \mathbf{0}, \ \|\mathbf{s}\|_{\mathbf{N}} \leq \Delta, \tag{7.1}$$

where $\mathbf{c}_k := c(\mathbf{x}_k)$, the rows of $\mathbf{A}_k := \nabla c(\mathbf{x}_k)^T$ are the constraint gradients $\nabla c_i(\mathbf{x}_k)^T$, $\Delta > 0$ is a trust-region radius, and $\mathbf{M}$ and $\mathbf{N}$ are symmetric and positive-definite matrices. Note that in this case, the least-squares problem is underdetermined. A sophisticated user might include some linear constraints in $\mathbf{A}_k$ as appropriate so as to keep the factorization of (5.8) efficient. Standard constraint qualification conditions for (1.2) impose that the constraint gradients be linearly independent at a solution $\mathbf{x}^*$. In particular, this implies that $\mathbf{E}$ has full row rank. In the notation of (1.5), the matrix of (1.4) is nonsingular if $\mathbf{Q}$ is positive definite on the nullspace of $\mathbf{J}$. Let the columns of $\mathbf{Z}$ form an orthonormal basis for the nullspace of $\mathbf{E}$ and complete it with $\mathbf{Y}$ such that $\mathbf{W}$ defined via $\mathbf{W}^T := \begin{bmatrix}\mathbf{Y}^T & \mathbf{Z}^T\end{bmatrix}$ is orthogonal. The columns of $\mathbf{W}$ form an orthonormal basis of the nullspace of $\mathbf{J}$. Then

$$\mathbf{W}^T\mathbf{Q}\mathbf{W} = \mathbf{Y}^T\mathbf{M}\mathbf{Y} + \mathbf{Y}^T\mathbf{A}_k\mathbf{Z} + (\mathbf{Y}^T\mathbf{A}_k\mathbf{Z})^T$$

is positive definite provided $\mathbf{A}_k$ has been scaled so that $\|\mathbf{A}_k\|_2 < \lambda_{\min}(\mathbf{M})$.

Practical trust-region methods do not require that (7.1) be solved to optimality. It is sufficient to compute a step $\mathbf{s}$ that results in sufficient decrease of the least-squares objective (Conn et al., 2000, Assumption AA.1j, p. 696). The sufficient decrease condition imposes that the step $\mathbf{s}$ result in decrease of the least-squares objective that is at least a fixed fraction of the decrease achieved along the intersection of the projection of the steepest descent direction into the nullspace of $\mathbf{E}$ and the trust region. This condition is met by the projected LSQR method initialized with $\mathbf{s}_0 := \mathbf{0}$ because the first direction explored is precisely the steepest descent direction. Subsequent iterations of the projected LSQR method only result in further decrease. It follows from (Fong, 2011, Theorem 3.3.11) that the projected LSMR also decreases the least-squares residual monotonically, so that it may also be employed to solve (7.1) and yields a step that satisfies the sufficient decrease condition. An additional property of both LSQR and LSMR that makes them suitable for trust-region subproblems is the guarantee that they generate successive estimates $\mathbf{s}_j$ such that $\|\mathbf{s}_j\|_{\mathbf{N}} > \|\mathbf{s}_{j-1}\|_{\mathbf{N}}$—see (Steihaug, 1983, Theorem 2.1) and (Fong, 2011, Theorems 2.1.6, 3.3.1 and 3.3.6). Consequently, if the iterates ever leave the trust region, they will never return. Finally, in the case of LSQR, the decrease achieved is known to be at least

half of that achieved at the global minimizer of (7.1) (Conn et al., 2000, Theorem 7.5.9), independently of the elliptic singular values of $\mathbf{A}_k$.

Once a normal step $\mathbf{s}$ has been computed, SQP methods compute the total step as the sum of $\mathbf{s}$ and a tangential step $\mathbf{p}$. The role of the latter is to decrease the objective value while maintaining the feasibility improvements achieved by the normal step. The tangential subproblem can be stated as

$$\underset{\mathbf{p}}{\text{minimize}} \;\; q_k(\mathbf{p}) \quad \text{subject to} \quad \begin{bmatrix} \mathbf{A}_k \\ \mathbf{E} \end{bmatrix} \mathbf{p} = \mathbf{0}, \;\; \|\mathbf{p}\|_{\mathbf{N}} \leq \bar{\Delta}, \tag{7.2}$$

where $q_k(\mathbf{p}) := \mathbf{g}_k^{\mathsf{T}}\mathbf{p} + \frac{1}{2}\mathbf{p}^{\mathsf{T}}\mathbf{H}_k\mathbf{p}$ is a quadratic approximation of the Lagrangian of (1.2), and $\bar{\Delta}$ is a trust-region radius. Interestingly, the methods proposed in the previous sections may be applied to solve (7.2) provided $\mathbf{H}_k$ is symmetric and positive definite. Indeed, in this case, it is possible to write $\mathbf{H}_k = \mathbf{R}_k\mathbf{R}_k^{\mathsf{T}}$ for some nonsingular matrix $\mathbf{R}_k$—the Cholesky factor being but one example. If we set $\mathbf{b}_k := -\mathbf{R}_k^{-1}\mathbf{g}_k$, we see that (7.2) is equivalent to

$$\underset{\mathbf{p}}{\text{minimize}} \;\; \tfrac{1}{2}\|\mathbf{R}_k^{\mathsf{T}}\mathbf{p} - \mathbf{b}_k\|^2 \quad \text{subject to} \quad \begin{bmatrix} \mathbf{A}_k \\ \mathbf{E} \end{bmatrix} \mathbf{p} = \mathbf{0}, \;\; \|\mathbf{p}\|_{\mathbf{N}} \leq \bar{\Delta}, \tag{7.3}$$

which has the form (1.1) with an additional trust-region constraint. The formulation (7.3) is practical in situations where $\mathbf{R}_k$ is readily available, as in quasi-Newton methods in factored form. That we may apply the projected conjugate gradient method to (7.2) or, equivalently, the projected LSQR method to (7.3) is no surprise. But it is possible that we may also apply the projected LSMR method to (7.3), and that would be equivalent to applying the projected MINRES method to (7.2). The defining property of LSMR implies that $q_k(\mathbf{p})$ will be monotonically decreasing along the MINRES iterations. This is an alternative proof of (Fong and Saunders, 2012, Theorem 2.5). We restate this observation below in this alternative form as it is important in its own right for optimization purposes.

Theorem 7.1 (Fong and Saunders 2012, Theorem 2.5) *Let $\mathbf{H} = \mathbf{H}^{\mathsf{T}}$ be positive definite. The quadratic $q(\mathbf{p}) := \mathbf{g}^{\mathsf{T}}\mathbf{p} + \frac{1}{2}\mathbf{p}^{\mathsf{T}}\mathbf{H}\mathbf{p}$ is monotonically decreasing along the* MINRES *iterations applied to the system $\mathbf{H}\mathbf{p} = -\mathbf{g}$.*

MINRES also ensures that its approximations satisfy $\|\mathbf{p}_j\|_{\mathbf{N}} > \|\mathbf{p}_{j-1}\|_{\mathbf{N}}$ so that the trust-region constraint may safely be treated by stopping the iterations as soon as the boundary of the trust region is crossed. The complete procedure for identifying an approximate minimizer of a convex quadratic inside a trust region by way of MINRES is as follows. Initialize MINRES to solve $\mathbf{H}\mathbf{p} = -\mathbf{g}$ with $\mathbf{p}_0 := \mathbf{0}$. Let k_0 be the smallest index such that $\|\mathbf{p}_{k_0}\| > \Delta$. If no such k_0 exists, MINRES identifies a solution. Otherwise, the approximate solution is defined as $\tilde{\mathbf{p}} := (1-\alpha)\mathbf{p}_{k_0-1} + \alpha\mathbf{p}_{k_0}$ where $\alpha \in [0,1)$ is chosen so that $\|\tilde{\mathbf{p}}\| = \Delta$.

In order to conclude that MINRES may be used to solve trust-region subproblems in the strictly convex case, it remains to etablish that the solution that it returns satisfies the sufficient decrease condition. This is the subject of ongoing research and experiments.

8 Discussion

It has been observed several times in the literature that certain Krylov methods, and in particular MINRES, applied to symmetric saddle-point systems appear to perform redundant work in the sense that, roughly speaking, every other iteration only yields neglibigle improvement—see, e.g., (Fischer, Ramage, Silvester, and Wathen, 1998). In the present paper, we show that the iterations of such Krylov methods are combinations of iterations of related methods for least-squares or least-norm problems. In the context of (1.1), the constrained LSQR is the appropriate implementation of CG and should be expected to terminate in about half the number of iterations as would be required by SYMMLQ applied to (1.3) or (1.4). Similarly, LSMR and CRAIG-MR are the appropriate implementation of MINRES and should be expected to terminate in about half the number of

iterations as would be required by MINRES applied to (1.3) or (1.4). Fischer et al. (1998, §4) allude to this relationship.

In the presence of regularization, the method of Craig (1955) in its extended form is described by Saunders (1995), and its generalized form is described by Arioli (2013) and Arioli and Orban (2013). In this situation, symmetric saddle-point systems become symmetric quasi-definite systems (Vanderbei, 1995). Arioli and Orban (2013) establish that the conjugate gradient method is well defined, may be used in place of SYMMLQ, and alternates between generalized CRAIG and LSQR steps.

An extension of our work to linear inequality constraints is possible using either active-set or interior-point methods. Both types of methods are extensively documented in the literature. Recently, Dehghani and Orban (2013) proposed an interior-point method for the more general problem

$$\underset{\mathbf{x}\in\mathbb{R}^n}{\text{minimize}}\ \mathbf{c}^T\mathbf{x} + \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{M}^{-1}} \quad \text{subject to } \mathbf{E}\mathbf{x}=\mathbf{d},\ \mathbf{x}\geq 0.$$

At each iteration, a search direction is computed as the solution to a symmetric and quasi-definite system. Equivalently, this may be viewed as an unconstrained regularized linear least-squares problem in iteration-dependent metrics. Notably, the regularization operator is diagonal.

To see that symmetric and quasi-definite systems may be viewed as the optimality conditions of unconstrained linear least-squares problems, consider the system

$$\begin{bmatrix} \mathbf{Q} & \mathbf{C} \\ \mathbf{C}^T & -\mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}$$

and reformulate it as

$$\begin{bmatrix} \mathbf{Q} & & \mathbf{C} \\ & \mathbf{D}^{-1} & \mathbf{E} \\ \mathbf{C}^T & \mathbf{E}^T & \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{w} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix},$$

where we used the factorization $\mathbf{S} = \mathbf{E}^T\mathbf{D}\mathbf{E}$ where $\mathbf{D}$ is positive definite. We may choose, e.g., $\mathbf{D} = \mathbf{S}$ and $\mathbf{E} = \mathbf{I}$, or $\mathbf{D} = \mathbf{I}$ and $\mathbf{E} = \mathbf{S}^{\frac{1}{2}}$. The latter system gives the optimality conditions of

$$\underset{\mathbf{y}}{\text{minimize}}\ \left\| \begin{bmatrix} \mathbf{C} \\ \mathbf{E} \end{bmatrix} \mathbf{y} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_{\mathbf{M}^{-1}},$$

where $\mathbf{M} = \text{blkdiag}(\mathbf{Q}, \mathbf{D}^{-1})$. By contrast, the optimality conditions of the regularized constrained problem

$$\underset{\mathbf{x}}{\text{minimize}}\ \tfrac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{M}^{-1}} + \tfrac{1}{2}\|\mathbf{x}\|^2_{\mathbf{N}} \quad \text{subject to } \mathbf{E}\mathbf{x}=\mathbf{0},$$

may be stated as the symmetric saddle-point system

$$\begin{bmatrix} \mathbf{M} & \mathbf{A} & \\ \mathbf{A}^T & -\mathbf{N} & \mathbf{E}^T \\ & \mathbf{E} & \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix},$$

where $\mathbf{y}$ are the Lagrange multipliers. The methods presented in the present paper are directly applicable by substituting each of LSQR, CRAIG, LSMR and CRAIG-MR with their generalized version and setting the regularization parameter to 1—see Arioli and Orban (2013) for details.

References

M. Arioli. Generalized Golub-Kahan bidiagonalization and stopping criteria. *SIAM Journal on Matrix Analysis and Applications*, 34(2):571–592, 2013. DOI: 10.1137/120866543.

M. Arioli and D. Orban. Iterative methods for symmetric quasi-definite linear systems—Part I: Theory. *Les Cahiers du GERAD* G–2013–32, HEC Montréal, Canada, 2013.

S. J. Benbow. Solving generalized least-squares problems with LSQR. *SIAM Journal on Matrix Analysis and Applications*, 21(1):166–177, 1999. DOI: 10.1137/S0895479897321830.

M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005. DOI: 10.1017/S0962492904000212.

Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, 3rd edition, 1996. DOI: 10.1137/1.9781611971484.

R. Chandra. *Conjugate gradient methods for partial differential equations*. Ph.D. Thesis, Department of Computer Science, Yale University, New Haven CT, USA, 1978.

A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, USA, 2000.

J. E. Craig. The n-step iteration procedures. *Journal of Mathematics and Physics*, 34(1):64–73, 1955.

M. Dehghani and D. Orban. A regularized primal-dual interior-point method for constrained linear least squares. *Les Cahiers du GERAD*, HEC Montréal, Canada, 2013. In preparation.

L. El Ghaoui. Optimization models and applications. Online, December 2013. URL `https://inst.eecs.berkeley.edu/~ee127a/book/login`. Work in progress.

B. Fischer, A. Ramage, D. Silvester, and A. Wathen. Minimum residual methods for augmented systems. *BIT*, 38(3): 527–543, 1998.

D. C.-L. Fong. *Minimum Residual Methods for Sparse Least-Squares Using Golub-Kahan Bidiagonalization*. PhD thesis, Stanford University, Stanford, CA, December 2011.

D. C.-L. Fong and M. A. Saunders. LSMR: An iterative algorithm for sparse least-squares problems. *SIAM Journal on Scientific Computing*, 33(5):2950–2971, 2011.

D. C.-L. Fong and M. A. Saunders. CG versus MINRES: An empirical comparison. *SQU Journal for Science*, 17(1): 44–62, 2012.

G. H. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *SIAM Journal on Numerical Analysis*, 2(2):205–224, 1965. DOI: 10.1137/0702016.

N. I. M. Gould, M. E. Hribar, and J. Nocedal. On the solution of equality constrained quadratic problems arising in optimization. *SIAM Journal on Scientific Computing*, 23(4):1375–1394, 2001. DOI: 10.1137/S1064827598345667.

N. I. M. Gould, D. Orban, and T. Rees. Projected Krylov methods for saddle-point systems. *Les Cahiers du GERAD* G–2013–23, HEC Montréal, Canada, 2013.

M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.

C. Keller, N. I. M. Gould, and A. J. Wathen. Constraint preconditioning for indefinite linear systems. *SIAM Journal on Matrix Analysis and Applications*, 21(4):1300–1317, 2000. DOI: 10.1137/S0895479899351805.

C. Lanczos. Solution of systems of linear equations by minimized iterations. *Journal of Research of the National Bureau of Standards*, 49(1):33–53, 1952.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer series in Operations Research and Financial Engineering. Springer-Verlag New York, Inc., New York, NY, USA, second edition, 2006.

C. C. Paige. Bidiagonalization of matrices and solution of linear equations. *SIAM Journal on Numerical Analysis*, 11 (1):197–209, 1974. DOI: 10.1137/0711019.

C. C. Paige and M. A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM Journal on Numerical Analysis*, 12(4):617–629, 1975. DOI: 10.1137/0712047.

C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, 1982. DOI: 10.1145/355984.355989.

Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, 2nd edition, 2003.

M. A. Saunders. Solution of sparse rectangular systems using LSQR and CRAIG. *BIT*, 35:588–604, 1995. DOI: 10.1007/BF01739829.

T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM Journal on Numerical Analysis*, 20(3):626–637, 1983. DOI: 10.1137/0720042.

C. Van Loan. On the method of weighting for equality-constrained least-squares problems. *SIAM Journal on Numerical Analysis*, 22(5):851–864, 1985. DOI: 10.1137/0722051.

R. J. Vanderbei. Symmetric quasi-definite matrices. *SIAM Journal on Optimization*, 5(1):100–113, 1995. DOI: 10.1137/0805005.