IMPLEMENTING A SMOOTH EXACT PENALTY FUNCTION FOR EQUALITY-CONSTRAINED NONLINEAR OPTIMIZATION SUPPLEMENTARY MATERIAL

RON ESTRIN*, MICHAEL P. FRIEDLANDER†, DOMINIQUE ORBAN‡, AND MICHAEL A. SAUNDERS§

1. Direct methods for solving the augmented system. We describe various direct methods for solving

$$\begin{bmatrix} I & A \\ A^T & -\delta^2 I \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} w \\ z \end{bmatrix}, \qquad \text{where} \qquad \mathcal{K} = \begin{bmatrix} I & A \\ A^T & -\delta^2 I \end{bmatrix}, \tag{1.1}$$

required to evaluate ϕ_σ and its derivatives. Recall that $A_\delta = \begin{bmatrix} A^T & \delta I \end{bmatrix}^T$ when $\delta > 0$; otherwise $A_\delta = A$. For this section, given a matrix R and vector b, the shorthand notation $x \leftarrow R \backslash b$ means that x solves the system $Rx = b$ (typically via forward or backward substitution).

1.1. QR factorization. Algorithm 1 computes (p, q) using the thin QR factorization of $A_\delta = QR$, with Q orthogonal and $R \in \mathbb{R}^{m \times m}$.

Algorithm 1 Solving (1.1) using the QR factorization.

1: $Q,\ R \leftarrow \mathrm{qr}(A_\delta)$
2: $\bar{w} \leftarrow Q_{1:n,:}^T w$
3: $\bar{z} \leftarrow R^T \backslash z$
4: $p \leftarrow w - Q_{1:n,:}(\bar{w} - \bar{z})$
5: $q \leftarrow R \backslash (\bar{w} - \bar{z})$
6: **return** (p, q)

The advantage is that A_δ is factorized instead of $\mathcal{K}$, and this method is backward stable for both p and q (Golub and Van Loan, 2013, §5.3.6). If A_δ is sparse, R is likely to be sparse (for some column permutation of A_δ) but unfortunately Q is not. For large problems, it may not be practical to store Q in order to solve (1.1).

1.2. Corrected semi-normal equations. The R factor from $A_\delta = QR$ can be computed without storing Q. We can then solve the semi-normal equations $R^T Rq = A^T w - z$ and set $p = w - Aq$. Björck and Paige (1994) show that this is not acceptable-error stable for p, possibly giving large error in p, particularly when $\|p\| \ll \|w\|$. Note that $p = g_\sigma$ when solving for the multiplier estimate means we may

*Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA. (E-mail: restrin@stanford.edu.)

†Department of Computer Science, University of British Columbia, Vancouver V6T 1Z4, BC, Canada (E-mail: mpf@cs.ubc.ca). Research partially supported by the Office of Naval Research [award N00014-17-1-2009].

‡GERAD and Department of Mathematics and Industrial Engineering, École Polytechnique, Montréal, QC, Canada (E-mail: dominique.orban@gerad.ca). Research supported by supported by NSERC Discovery Grant 299010-04.

§Systems Optimization Laboratory, Department of Management Science and Engineering, Stanford University, Stanford, CA. E-mail: E-mail: saunders@stanford.edu. Research partially supported by the National Institute of General Medical Sciences of the National Institutes of Health [award U01GM102098].

This manuscript is for review purposes only.

obtain large errors in the gradient near the solution if care is not taken. Fortunately, Björck and Paige (1994) show that one step of iterative refinement ensures p is acceptable-error stable; see Algorithm 2.

Algorithm 2 Solving (1.1) using the semi-normal equations.

```
R ← qr(A_δ)
q ← (R^T R)\(A^T w − z)        p ← w − Aq
Δq ← (R^T R)\(A^T p − δ^2 q − z)                ▷ Iterative refinement
q ← q + Δq        p ← p − AΔq
return (p, q)
```

1.3. LDL and Bunch-Kaufman factorization. When it is not practical to store Q from the QR factors of A, or the semi-normal equations do not provide sufficient accuracy, it may be possible to compute the LDL or Bunch-Kaufman factorization of $\mathcal{K}$ directly. Although an $(n+m) \times (n+m)$ matrix is factorized (rather than an $n \times m$ matrix), the entire factorization is likely to be sparse, and the solution is typically more accurate than with the semi-normal equations.

Björck (1967) and Saunders (1995) discuss scaling of the $(1, 1)$ identity block to improve the condition number of $\mathcal{K}$. Saunders (1995) also considers the case where $\mathcal{K}$ is regularized with $-\delta^2 I$ in the $(2, 2)$ block.

References.

A. Björck. Iterative refinement of linear least squares solutions. I. *Nordisk Tidskr. Informations-Behandling (BIT)*, 7:257–278, 1967.

A. Björck and C. C. Paige. Solution of augmented linear systems using orthogonal factorizations. *BIT*, 34(1):1–24, 1994. DOI: 10.1007/BF01935013.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

M. A. Saunders. Solution of sparse rectangular systems using LSQR and Craig. *BIT*, 35(4):588–604, 1995. DOI: 10.1007/BF01739829.

This manuscript is for review purposes only.