BIT Numer Math (2009) 49: 21–53
DOI 10.1007/s10543-008-0206-8

BIT

Trust-region and other regularisations of linear least-squares problems

C. Cartis · N.I.M. Gould · P.L. Toint

Received: 3 February 2008 / Accepted: 14 October 2008 / Published online: 22 January 2009

Abstract We consider methods for regularising the least-squares solution of the linear system $Ax = b$. In particular, we propose iterative methods for solving large problems in which a trust-region bound $\|x\| \leq \Delta$ is imposed on the size of the solution, and in which the least value of linear combinations of $\|Ax - b\|_2^q$ and a regularisation term $\|x\|_2^p$ for various p and $q = 1, 2$ is sought. In each case, one or more "secular" equations are derived, and fast Newton-like solution procedures are suggested. The resulting algorithms are available as part of the GALAHAD optimization library.

Keywords Linear least-squares · Regularisation · Trust-region · Secular equation

Mathematics Subject Classification (2000) 65F22 · 65H05 · 65K05 · 90C25

Communicated by Lars Eldén.

This work was partially supported by EPSRC grants EP/E053351/1 and EP/F005369/1.

C. Cartis
School of Mathematics, The King's Buildings, University of Edinburgh, Edinburgh, EH9 3JZ, Scotland, UK
e-mail: coralia.cartis@ed.ac.uk

N.I.M. Gould (✉)
Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, OX11 0QX, England, UK
e-mail: nick.gould@stfc.ac.uk

P.L. Toint
Department of Mathematics, Facultés Universitaires ND de la Paix–University of Namur, 61, rue de Bruxelles, 5000 Namur, Belgium
e-mail: philippe.toint@fundp.ac.be

1 Introduction

1.1 Motivation

Let $A \in \Re^{m\times n}$ and $b \in \Re^m$ be given data, and let $\|\cdot\|$ denote the Euclidean ℓ_2 norm. We are interested in finding $x \in \Re^n$ so that both $\|Ax - b\|$ and $\|x\|$ are small. Traditionally this has been achieved by minimizing

$$\|Ax - b\|^2 + \lambda\|x\|^2$$

for some suitable positive regularisation parameter λ—this is often known as Tikhonov regularization or, in statistics, ridge regression. Many heuristics (for example, the discrepancy principle, generalised cross validation, the L-curve method, and the unbiased predictive risk estimator) [21, 34] have been proposed for selecting λ and, given λ, most methods use the observation that the problem may then be reformulated as the weighted least-squares problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \left\| \begin{pmatrix} A \\ \lambda^{\frac{1}{2}} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|, \tag{1.1}$$

where I is the appropriately-dimensioned identity matrix. In this paper, we consider both generalisations and alternatives to this form of regularisation.

While there are many real applications for (regularised) linear least-squares [3, 34], our main interests are in nonlinear problems for which linear least-squares problems arise as sub-problems. The best know example is nonlinear least-squares (fitting) in which the least value of the ℓ_2-norm $\|F(x)\|$ of a vector-valued function $F : \Re^n \to \Re^m$ is sought [8, Chap. 10]. Here $F(x_k + s)$ is often approximated locally about a current iterate x_k by $F(x_k) + J(x_k)s$, involving the Jacobian J of F. This leads to the Gauss-Newton method in which the correction s_k is chosen to minimize $\|F(x_k) + J(x_k)s\|$. In order to globalise such a scheme, Moré [30] proposed that the step be regularised to

$$\underset{s\in\mathbb{R}^n}{\text{minimize}} \|F(x_k) + J(x_k)s\| \quad \text{subject to} \quad \|s\| \le \Delta_k$$

for some dynamically adjusted radius $\Delta_k > 0$, making rigourous earlier heuristics by Levenberg, Morrison and Marquardt [26, 28, 31] in which the step was chosen to

$$\underset{s\in\mathbb{R}^n}{\text{minimize}} \frac{1}{2}\|F(x_k) + J(x_k)s\|^2 + \frac{1}{2}\sigma_k\|s\|^2$$

for some regularisation parameter $\sigma_k > 0$. This trust-region approach has been extended to the large-scale case by Lukšan [27]. More recently, Nesterov [32] suggested that choosing the step to

$$\underset{s\in\mathbb{R}^n}{\text{minimize}} \|F(x_k) + J(x_k)s\| + \frac{1}{2}\sigma_k\|s\|^2$$

leads to a good worst-case iteration complexity bound in some cases, while there are reasons to believe [5, 33] that similar results are possible for steps chosen to approximately

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2}\|F(x_k)+J(x_k)s\|^2+\frac{1}{3}\sigma_k\|s\|^3.$$

As a second example, in a number of current iterative methods for constrained optimization [1, 17, 25, 36], a so-called normal step s is computed to try to improve constraint infeasibility by approximately solving the subproblem

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ \|J(x_k)s+c(x_k)\| \quad \text{subject to} \quad \|s\|\le\Delta_k.$$

Here $J(x_k)s+c(x_k)$ is a linearization of the nonlinear constraints $c(x)=0$ about $x=x_k$, and the trust-region constraint $\|s\|\le\Delta_k$ for a given radius $\Delta_k>0$ is imposed to limit the size of the step [7, Sect. 15.4]. Such algorithms often compute Lagrange multiplier estimates y from the subproblem

$$\underset{y\in\mathbb{R}^m}{\text{minimize}}\ \|J^T(x_k)y-g(x_k)\| \quad \text{subject to} \quad \|y\|\le\eta_k,$$

where $g(x)$ is the gradient of the objective function and where η_k is chosen to preclude large multiplier estimates. Developing methods [18] replace the trust-region constraints in these subproblems by adding appropriate regularisation as above.

1.2 The problem

In this paper, we consider the generic linear least-squares trust-region problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \|Ax-b\| \quad \text{subject to} \quad \|x\|\le\Delta \tag{1.2}$$

for given $\Delta>0$, the regularised linear least-squares problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2}\|Ax-b\|^2+\frac{\sigma}{p}\|x\|^p \tag{1.3}$$

and the regularised linear least ℓ_2-norm problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \|Ax-b\|+\frac{\sigma}{p}\|x\|^p \tag{1.4}$$

for given $\sigma>0$ and $p\ge 2$; we shall be especially interested in methods appropriate when n is large. As the two examples in Sect. 1.1 indicate, we shall make no assumption concerning the size of m relative to n, and thus whether the un-regularised problem is under-, well- or over-determined.

1.3 Organisation

The paper is organised as follows. In Sects. 2–4 we propose iterative methods for finding approximate solutions to problems (1.2)–(1.4) respectively. Some details of software implementations of these ideas are reported in Sect. 5. We make further comments and draw conclusions in Sect. 6.

2 Solving the least-squares trust-region problem

We first consider the trust-region problem (1.2). There is a long history of work on this topic [6, 11, 13, 38, 39, 41, 42], which we will review as we proceed.

2.1 Solution characteristics

It is straightforward to derive [11, 41] usable optimality conditions for (1.2). Specifically, let $\lambda \geq 0$ and define $x(\lambda)$ so that

$$(A^T A + \lambda I)x(\lambda) = A^T b \tag{2.1}$$

or equivalently that $x(\lambda)$ solves the weighted least-squares problem (1.1). Then so long as $\|x(0)\| \leq \Delta$, $x(0)$ is the desired solution to (1.2). Otherwise the solution is $x(\lambda_*)$, where λ_* is the positive root of the so-called "secular" equation

$$\|x(\lambda)\| - \Delta = 0. \tag{2.2}$$

If it is feasible to factorize $A^T A + \lambda I$ (either explicitly using Cholesky/possibly-truncated SVD or implicitly by bi-diagonalising A, see e.g., [9]), a simple univariate root finding method may be used to determine the appropriate root of (2.2)—this might require the derivative of $\pi(\lambda) = \|x(\lambda)\|$, but it is easy to show that

$$\pi'(\lambda) = \frac{x^T(\lambda)x'(\lambda)}{\|x(\lambda)\|}, \quad \text{where } (A^T A + \lambda I)x'(\lambda) = -x(\lambda). \tag{2.3}$$

We give general details in Sect. 2.3.3. Our interest, however, is in the case for which a factorization of $A^T A + \lambda I$ is either impossible, through lack of memory, or too expensive to contemplate—applications such as three-dimensional PDE-constrained optimization [2] and those for which A has a significant number of non-sparse rows are good examples. We resort in this case to iterative methods. We note that although we describe an approach using LSQR, there is at least one alternative based on a parametric eigenvalue formulation [41, 42].

2.2 The unconstrained problem and LSQR

We now describe how we aim to solve (1.2). The basis of what we shall use is the LSQR method of Paige and Saunders [38, 39]. LSQR is designed to minimize the function

$$f(x) = \frac{1}{2}\|Ax - b\|^2$$

or its regularisation

$$f_\lambda(x) = \frac{1}{2}\|Ax - b\|^2 + \frac{1}{2}\lambda\|x\|^2$$

for some given $\lambda > 0$. It is to be preferred in practice to the theoretically-equivalent conjugate-gradient method in many cases since numerical properties are better for the former [39] and more accurately reflect the conditioning of the problem [3, Theorem 1.4.6 et. seq.].

We follow in the most part the notation in [39], and for completeness fill in some of the details of the slightly more terse aspects of Paige and Saunders' description.

2.2.1 Lower bi-diagonalisation of A

The iterative bi-diagonalisation algorithm due to Golub and Kahan [12] is a core component of LSQR. Sequences of unit vectors $\{u_k \in \Re^m\}$ and $\{v_k \in \Re^n\}$ are constructed as follows:

$$\begin{aligned}&\textbf{Initialization:}\quad \beta_1 u_1 = b \quad \text{and} \quad \alpha_1 v_1 = A^T u_1\\ &\textbf{Iteration:}\quad \beta_{k+1} u_{k+1} = A v_k - \alpha_k u_k \quad \text{and}\\ &\alpha_{k+1} v_{k+1} = A^T u_{k+1} - \beta_{k+1} v_k \quad \text{for } k \geq 1.\end{aligned} \tag{2.4}$$

This leads directly to the relationships

$$AV_k = U_{k+1} B_k \quad \text{and} \quad b = \beta_1 U_{k+1} e_1, \tag{2.5}$$

where $\beta_1 = \|b\|$, e_i denotes the ith column of the identity matrix, $U_k = (u_1\ u_2\ \ldots\ u_k)$, $U_k^T U_k = I$, $V_k = (v_1\ v_2\ \ldots\ v_k)$, $V_k^T V_k = I$ and

$$B_k = \begin{pmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \\ & & & \beta_{k+1} \end{pmatrix} \equiv \begin{pmatrix} B_{k-1} & \alpha_k e_k \\ 0 & \beta_{k+1} \end{pmatrix} \tag{2.6}$$

is $(k+1)$ by k and lower bi-diagonal. A further useful property is that

$$A^T U_{k+1} = V_k B_k^T + \alpha_{k+1} v_{k+1} e_{k+1}^T. \tag{2.7}$$

2.2.2 Reduction to upper bi-diagonal form

To approximately minimize $f(x)$, we find the sequence of minimizers of $f(V_k y)$ in the expanding subspace $x = V_k y$, $k = 1, 2, \ldots$. Thus we pick $x_k = V_k y_k$, where

$$y_k = \arg\min_{y \in \mathbb{R}^k} \|B_k y - \beta_1 e_1\|; \tag{2.8}$$

formally y_k satisfies the normal equations

$$B_k^T B_k y_k = \beta_1 B_k^T e_1. \tag{2.9}$$

To find y_k, B_k is reduced to upper triangular form by pre-multiplying it by a product of plane rotations $Q_k = Q_{k,k+1} \cdots Q_{1,2}$, where the plane rotation $Q_{j,j+1}$ operates solely on rows j and $j+1$ to eliminate the sub-diagonal entry in row j. This leads to

$$Q_k (B_k \quad \beta_1 e_1) = \begin{pmatrix} R_k & f_k \\ 0 & \bar{\phi}_{k+1} \end{pmatrix}, \tag{2.10}$$

where

$$R_k = \begin{pmatrix} \rho_1 & \theta_2 & & \\ & \ddots & \ddots & \\ & & \rho_{k-1} & \theta_k \\ & & & \rho_k \end{pmatrix} \equiv \begin{pmatrix} R_{k-1} & \theta_k e_{k-1} \\ 0 & \rho_k \end{pmatrix} \tag{2.11}$$

is k by k and upper bi-diagonal and

$$f_k = \begin{pmatrix} f_{k-1} \\ \phi_k \end{pmatrix} \in \Re^k. \tag{2.12}$$

To be specific, the nature of Q_k, together with (2.6) and (2.10) imply that

$$\begin{pmatrix} Q_{k-1} & 0 \\ 0 & 1 \end{pmatrix} (B_k \quad \beta_1 e_1) = \begin{pmatrix} Q_{k-1} B_{k-1} & Q_{k-1,k} \alpha_k e_k & Q_{k-1,k} \beta_1 e_1 \\ 0 & \beta_{k+1} & 0 \end{pmatrix}$$

$$= \begin{pmatrix} R_{k-1} & \theta_k e_{k-1} & f_{k-1} \\ 0 & \bar{\rho}_k & \bar{\phi}_k \\ 0 & \beta_{k+1} & 0 \end{pmatrix}.$$

Thus if the plane rotation $Q_{k,k+1}$ has non-trivial elements c_k and s_k, we have

$$\begin{pmatrix} c_k & s_k \\ -s_k & c_k \end{pmatrix} \begin{pmatrix} \bar{\rho}_k & \bar{\phi}_k \\ \beta_{k+1} & 0 \end{pmatrix} = \begin{pmatrix} \rho_k & \phi_k \\ 0 & \bar{\phi}_{k+1} \end{pmatrix};$$

to prepare for the next step we also need $Q_{k,k+1} \alpha_{k+1} e_{k+1}$ for which the non-zero components are

$$\begin{pmatrix} c_k & s_k \\ -s_k & c_k \end{pmatrix} \begin{pmatrix} 0 \\ \alpha_{k+1} \end{pmatrix} = \begin{pmatrix} \theta_{k+1} \\ \bar{\rho}_{k+1} \end{pmatrix}.$$

Initial values $\bar{\rho}_1 = \alpha_1$ and $\bar{\phi}_1 = \beta_1$ are needed.

2.2.3 Solution of the problem in the subspace $V_k y$

It follows from (2.10) and $Q_k^T Q_k = I$ that the required solution to (2.8) satisfies

$$R_k y_k = f_k \tag{2.13}$$

and thus $x_k = V_k R_k^{-1} f_k = D_k f_k$, where

$$V_k R_k^{-1} = D_k = (d_1 \quad d_2 \quad \dots \quad d_k). \tag{2.14}$$

Hence

$$x_k = D_{k-1} f_{k-1} + d_k \phi_k = x_{k-1} + \phi_k d_k$$

with $x_0 = 0$. Fortunately the precise (upper bi-diagonal) form of R_k in (2.11) along with (2.14) imply that

$$\begin{aligned}(V_{k-1} \quad v_k) = V_k &= (D_{k-1} \quad d_k) \begin{pmatrix} R_{k-1} & \theta_k e_{k-1} \\ 0 & \rho_k \end{pmatrix} \\ &= (D_{k-1} R_{k-1} \quad \theta_k D_{k-1} e_{k-1} + \rho_k d_k) \\ &= (D_{k-1} R_{k-1} \quad \theta_k d_{k-1} + \rho_k d_k)\end{aligned}$$

and hence

$$d_k = (v_k - \theta_k d_{k-1})/\rho_k,$$

enabling us to recur d_k from d_{k-1} and v_k starting from $d_0 = 0$. A small saving can be made by using ρ_k from (2.11) and defining $w_k = \rho_k d_k$ in which case

$$\begin{aligned} x_k &= x_{k-1} + (\phi_k/\rho_k) w_k \quad \text{and} \\ w_{k+1} &= v_{k+1} - (\theta_{k+1}/\rho_k) w_k \end{aligned} \tag{2.15}$$

with $w_1 = v_1$.

2.2.4 Norms of required terms

It is important to monitor $\nabla_x f(x_k) = A^T(Ax_k - b)$ to decide when to stop the iteration. Fortunately, it follows directly from (2.5) and (2.7) that

$$\begin{aligned}\nabla_x f(x_k) &= A^T U_{k+1}(B_k y_k - \beta_1 e_1) \\ &= V_k^T B_k^T (B_k y_k - \beta_1 e_1) + \alpha_{k+1} v_{k+1} e_{k+1}^T (B_k y_k - \beta_1 e_1);\end{aligned} \tag{2.16}$$

the first term vanishes because of the normal equations (2.9), and thus

$$\nabla_x f(x_k) = \alpha_{k+1} v_{k+1} e_{k+1}^T (B_k y_k - \beta_1 e_1). \tag{2.17}$$

But (2.10), (2.13) and the precise form of Q_k together show that

$$e_{k+1}^T (B_k y_k - \beta_1 e_1) = e_{k+1}^T Q_k^T Q_k (B_k y_k - \beta_1 e_1) = \bar{\phi}_{k+1} e_{k+1}^T Q_k^T e_{k+1} = \bar{\phi}_{k+1} c_k,$$

and hence from (2.17) that

$$\|\nabla_x f(x_k)\| = |\bar{\phi}_{k+1} \alpha_{k+1} c_k|$$

using known quantities [39, Sect. 5.1]. Thus $\|\nabla_x f(x_k)\|$ is available without the expense of computing $\nabla_x f(x_k)$. It is also useful to monitor $\|Ax_k - b\|$ and again [39, Sect. 5.1] this is readily available since (2.5) and (2.10) give

$$\begin{aligned} Ax_k - b &= AV_k y_k - b = U_{k+1}(B_k y_k - \beta_1 e_1) \\ &= U_{k+1} Q_k^T \begin{pmatrix} R_k y_k - f_k \\ -\bar{\phi}_{k+1} \end{pmatrix} = -\bar{\phi}_{k+1} U_{k+1} Q_k^T e_{k+1} \end{aligned} \tag{2.18}$$

and hence

$$\|Ax_k - b\| = |\bar{\phi}_{k+1}|.$$

In what will follow, it is also vital to monitor $\|x_k\|$. This is not immediately available, but may be found with a modest amount of extra work [39, Sect. 5.2]. To be specific, since R_k is upper bi-diagonal, it may be reduced to lower bi-diagonal form by post-multiplying by a product of plane rotations $W_k = W_{1,2} \cdots W_{k-1,k}$. This produces

$$R_k W_k = \bar{L}_k = \begin{pmatrix} \lambda_1 & & & & \\ \gamma_2 & \lambda_2 & & & \\ & \ddots & \ddots & & \\ & & \gamma_{k-1} & \lambda_{k-1} & \\ & & & \gamma_k & \bar{\lambda}_k \end{pmatrix} \equiv \begin{pmatrix} L_{k-1} & \\ \gamma_k e_{k-1}^T & \bar{\lambda}_k \end{pmatrix} \tag{2.19}$$

which is k by k lower bi-diagonal. Note that the leading $(k-1)$ by $(k-1)$ sub-block L_{k-1} of $\bar{L}_k$ is not altered in subsequent iterations, but that the trailing diagonal entry $\bar{\lambda}_k$ of $\bar{L}_k$ will become λ_k on iteration $k+1$.

Now let z_k and $\bar{z}_k$ satisfy $L_k z_k = f_k$ and $\bar{L}_k \bar{z}_k = f_k$, respectively. Since L_k and $\bar{L}_k$ share the leading k by $(k-1)$ sub-block,

$$z_k \equiv \begin{pmatrix} z_{k-1} \\ \zeta_k \end{pmatrix} \quad \text{and} \quad \bar{z}_k \equiv \begin{pmatrix} z_{k-1} \\ \bar{\zeta}_k \end{pmatrix}, \quad \text{where } \bar{\zeta}_k = \frac{\lambda_k}{\bar{\lambda}_k} \zeta_k. \tag{2.20}$$

In this case

$$x_k = V_k R_k^{-1} f_k = V_k W_k \bar{L}_k^{-1} f_k = V_k W_k \bar{z}_k$$

and thus

$$\|x_k\| = \|\bar{z}_k\|$$

since W_k is orthogonal and $V_k^T V_k = I$. But (2.12)–(2.20) give that

$$\bar{L}_k \bar{z}_k = \begin{pmatrix} L_{k-1} & \\ \gamma_k e_{k-1}^T & \bar{\lambda}_k \end{pmatrix} \begin{pmatrix} z_{k-1} \\ \bar{\zeta}_k \end{pmatrix} = \begin{pmatrix} f_{k-1} \\ \phi_k \end{pmatrix} = f_k,$$

in which case

$$\bar{\zeta}_k = (\phi_k - \gamma_k \zeta_{k-1}) / \bar{\lambda}_k. \tag{2.21}$$

Thus

$$\|x_k\|^2 = \|\bar{z}_k\|^2 = \|z_{k-1}\|^2 + \bar{\zeta}_k^2 \quad \text{and} \quad \|z_k\|^2 = \|z_{k-1}\|^2 + \zeta_k^2$$

may be recurred as the iteration proceeds in terms of $\bar{\zeta}_k$ from (2.21) which needs $\zeta_{k-1} = \bar{\zeta}_{k-1}\bar{\lambda}_{k-1}/\lambda_{k-1}$ from (2.20). Moreover the decomposition (2.19) may be calculated step by step. For, given $\bar{L}_{k-1}$,

$$\begin{pmatrix} R_{k-1} & \theta_k e_{k-1} \\ & \rho_k \end{pmatrix} \begin{pmatrix} W_{k-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \bar{L}_{k-1} & \theta_k e_{k-1} \\ & \rho_k \end{pmatrix}$$

$$= \begin{pmatrix} L_{k-2} & & \\ \gamma_{k-1} e_{k-2}^T & \bar{\lambda}_{k-1} & \theta_k \\ & & \rho_k \end{pmatrix}.$$

Thus if the plane rotation $W_{k-1,k}$ operating on columns $k-1$ and k has non-trivial elements c_{k-1}^w and s_{k-1}^w, we have

$$\begin{pmatrix} \bar{\lambda}_{k-1} & \theta_k \\ 0 & \rho_k \end{pmatrix} \begin{pmatrix} c_{k-1}^w & -s_{k-1}^w \\ s_{k-1}^w & c_{k-1}^w \end{pmatrix} = \begin{pmatrix} \lambda_{k-1} & 0 \\ \gamma_k & \bar{\lambda}_k \end{pmatrix},$$

which gives λ_{k-1}, γ_k, $\bar{\lambda}_k$ and hence $\bar{L}_k$, The initial value $\bar{\lambda}_1 = \rho_1$ is needed.

2.3 Adding a trust region

It is well known [39, Sect. 7] that the iterates generated by LSQR are mathematically equivalent to those generated by applying the conjugate gradient method to minimize $f(x)$. Moreover the columns of the matrix V_k span precisely the Krylov space $\{(A^T A)^i A^T b\}_{i=1}^{k-1}$. This has the important consequence [43] that the norms $\|x_k\|$, $k = 0, 1, 2, \ldots$ are monotonically increasing (see also [27]). Thus if we apply LSQR to the problem (1.2) and we find

$$\|x_{k-1}\| \le \Delta < \|x_k\|, \tag{2.22}$$

immediately we may deduce that the solution to (1.2) lies on the boundary of the trust region.

2.3.1 The Steihaug-Toint point

The Steihaug-Toint [43, 44] proposal is to generate iterates using the conjugate-gradient method—in our case, using LSQR—until an iterate for which (2.22) occurs, and then to replace x_k by the so-called Steihaug-Toint point $x_k^{\text{ST}} = x_{k-1} + \sigma w_k$, where σ is determined so that $\|x_{k-1} + \sigma w_k\| = \Delta$. This may be achieved by finding σ as the root of the quadratic equation

$$\|x_{k-1}\|^2 - \Delta^2 + 2x_{k-1}^T w_k \sigma + \|w_k\|^2 \sigma^2 = 0 \tag{2.23}$$

with the same sign as the too-large step-size ϕ_k/ρ_k in (2.15). Such a Steihaug-Toint approach was first proposed in the least-squares context, using LSQR, by

Lukšan [27]. While the required coefficients in (2.23) may be found directly as inner products, savings may be made by noting that $\|x_{k-1}\|$ is already being recurred. Furthermore (2.15) implies that

$$\begin{aligned}\|w_{k+1}\|^2 &= \|v_{k+1}\|^2 - 2(\theta_{k+1}/\rho_k)v_{k+1}^T w_k + (\theta_{k+1}/\rho_k)^2\|w_k\|^2 \\ &= 1 + (\theta_{k+1}/\rho_k)^2\|w_k\|^2 \end{aligned} \tag{2.24}$$

since v_k is a unit vector and $v_{k+1}^T w_k = \rho_k v_{k+1}^T d_k = \rho_k v_{k+1}^T V_k R_k^{-1} e_k = 0$ because v_{k+1} is orthogonal to V_k, and thus $\|w_k\|$ may also be cheaply recurred. Finally, since $\|x_{k+1}\|$ has been computed (and found to be too large), it follows immediately from (2.15) that

$$2x_{k-1}^T w_k = \frac{\|x_k\|^2 - \|x_{k-1}\|^2 - (\phi_k/\rho_k)^2\|w_k\|^2}{(\phi_k/\rho_k)}$$

using available data.

Given σ, it is also useful to know $\|Ax_k^{\text{ST}} - b\|$ without computing x_k^{ST}. It follows from (2.5), (2.10) and (2.14) that

$$\begin{aligned}Aw_k &= \rho_k Ad_k = \rho_k AV_k R_k^{-1} e_k = \rho_k U_{k+1} B_k R_k^{-1} e_k \\ &= \rho_k U_{k+1} Q_k^T \begin{pmatrix} I \\ 0 \end{pmatrix} e_k = \rho_k U_{k+1} Q_k^T e_k. \end{aligned} \tag{2.25}$$

But since

$$Q_k^T e_k = \begin{pmatrix} Q_{k-1}^T & 0 \\ 0 & 1 \end{pmatrix} Q_{k,k+1}^T e_k = \begin{pmatrix} Q_{k-1}^T & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_k e_k \\ s_k \end{pmatrix} = \begin{pmatrix} c_k Q_{k-1}^T e_k \\ s_k \end{pmatrix},$$

it immediately follows from (2.25) that

$$Aw_k = \rho_k c_k U_k Q_{k-1}^T e_k + \rho_k s_k u_{k+1}$$

and thus from (2.18)

$$A(x_{k-1} + \sigma w_k) - b = (\sigma\rho_k c_k - \bar{\phi}_k) U_k Q_{k-1}^T e_k + \sigma\rho_k s_k u_{k+1}.$$

As u_{k+1} and U_k are orthogonal, we then have the relationship

$$\|Ax_k^{\text{ST}} - b\|^2 = \|A(x_{k-1} + \sigma w_k) - b\|^2 = (\sigma\rho_k c_k - \bar{\phi}_k)^2 + (\sigma\rho_k s_k)^2$$

in terms of known (scalar) quantities.

There is an important result [45] concerning the application of the conjugate gradient method to minimize a strictly convex quadratic function within a spherical trust region, which has subsequently been extended [7, Theorem 7.5.9] to cover the convex case as is needed here. The result is that if x^{ST} is the Steihaug-Toint point and x_* is the solution of (1.2) then

$$\|b\|^2 - \|Ax_* - b\|^2 \leq 2(\|b\|^2 - \|Ax^{\text{ST}} - b\|^2).$$

In other words, the optimal decrease will be no more than twice that achieved at the Steihaug-Toint point. Thus it may become apparent at x^{ST} whether it is impossible to reduce $\|Ax - b\|$ to zero within the trust region since

$$\|Ax_* - b\|^2 \geq 2\|Ax^{\text{ST}} - b\|^2 - \|b\|^2,$$

which will be nonzero whenever $\|Ax^{\text{ST}} - b\| > \frac{1}{\sqrt{2}}\|b\|$. In view of this result, it is questionable whether it is really beneficial to try to improve upon the Steihaug-Toint point, but for completeness and for what follows in Sects. 3 and 4 we now show how this may be achieved.

2.3.2 Beyond the Steihaug-Toint point

Once it is known that the solution lies on the trust-region boundary, problem (1.2) is equivalent to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \|Ax - b\| \quad \text{subject to} \quad \|x\| = \Delta. \tag{2.26}$$

More particularly, from (2.8), the problem in the subspace $x = V_k y$ becomes

$$\underset{y\in\mathbb{R}^k}{\text{minimize}} \|B_k y - \beta_1 e_1\| \quad \text{subject to} \quad \|y\| = \Delta$$

or equivalently

$$\underset{y\in\mathbb{R}^k}{\text{minimize}} \frac{1}{2}\|B_k y - \beta_1 e_1\|^2 \quad \text{subject to} \quad \frac{1}{2}\|y\|^2 = \frac{1}{2}\Delta^2 \tag{2.27}$$

since $\|V_k y\| = \|y\|$ as V_k has orthogonal columns.

Necessary and sufficient conditions for y_k to solve (2.27) are that

$$B_k^T(B_k y_k - \beta_1 e_1) + \lambda_k y_k = 0 \quad \text{and} \quad \|y_k\| = \Delta \tag{2.28}$$

for some Lagrange multiplier $\lambda_k \geq 0$. A more useful interpretation is that given $\lambda = \lambda_k$, one could find $y_k = y_k(\lambda)$ from the equation

$$[B_k^T B_k + \lambda I]y_k(\lambda) - \beta_1 B_k^T e_1 = 0, \tag{2.29}$$

and the required λ satisfies the scalar secular equation

$$\|y_k(\lambda)\| - \Delta = 0. \tag{2.30}$$

Vitally, (2.29) is the stationarity condition for the convex function

$$\frac{1}{2}\|B_k y - \beta_1 e_1\|^2 + \frac{1}{2}\lambda\|y\|^2,$$

and as we observed in Sect. 1.1 we can thus find $y_k(\lambda)$ as the solution to the weighted linear least-squares problem

$$\underset{y\in\mathbb{R}^k}{\text{minimize}}\ \frac{1}{2}\left\| \begin{pmatrix} B_k \\ \lambda^{\frac{1}{2}} I \end{pmatrix} y - \begin{pmatrix} \beta e_1 \\ 0 \end{pmatrix} \right\|. \tag{2.31}$$

Thus we seek the positive root of the secular equation (2.29) where $y(\lambda)$ is defined implicitly as the solution of (2.31).

To solve (2.31), we simply use the method proposed by Paige and Saunders [38], but recognise that a new factorization will be required every time λ changes. To fill in the details, we proceed just as in (2.10) by reducing

$$\begin{pmatrix} B_k \\ \lambda^{\frac{1}{2}} I \end{pmatrix}$$

to upper bi-diagonal form using plane rotations. In particular, we apply the product[1] of plane rotations

$$Q_{2k}(\lambda) = Q_{k,k+1}(\lambda)Q_{k,2k+1}(\lambda)\cdots Q_{2,3}(\lambda)Q_{2,k+3}(\lambda)Q_{1,2}(\lambda)Q_{1,k+2}(\lambda)$$

to form

$$Q_{2k}(\lambda)\begin{pmatrix} B_k & \beta_1 e_1 \\ \lambda^{\frac{1}{2}} I & 0 \end{pmatrix} = \begin{pmatrix} R_k(\lambda) & f_k(\lambda) \\ 0 & \bar{\phi}_{k+1}(\lambda) \\ 0 & p_k(\lambda) \end{pmatrix}, \tag{2.32}$$

where $p_k(\lambda) \in \Re^k$. Once the upper bi-diagonal $R_k(\lambda)$ is known, the required solution $y_k(\lambda)$ to (2.31) may simply be recovered by back-substitution from

$$R_k(\lambda)y_k(\lambda) = f_k(\lambda). \tag{2.33}$$

Note that (2.32) shows that

$$B_k^T B_k + \lambda I = R_k^T(\lambda)R_k(\lambda) \tag{2.34}$$

since $Q_{2k}(\lambda)$ is orthogonal.

The seeds of this idea of expanding subspace minimization was first proposed, in the more general context of minimizing quadratic functions within spherical trust regions, by Gould, Lucidi, Roma and Toint [14], and forms the basis of the GLTR package within the GALAHAD optimization library [15]. In the least-squares case, Golub and von Matt [13] considered similar ideas for equality-constrained problems.

[1] As Paige and Saunders note, the rotations may be applied in other orders, but their experience suggests this order gives marginally more accurate results.

2.3.3 The secular equation and its solution

We now consider the secular equation (2.29)–(2.30) in a more general context. Namely, we aim to find the positive root, λ_*, of the secular equation

$$\phi(\lambda) \stackrel{\text{def}}{=} \|y(\lambda)\| - \Delta = 0, \tag{2.35}$$

where $y(\lambda)$ satisfies

$$[B^T B + \lambda I]y(\lambda) - B^T g = 0, \tag{2.36}$$

for a given (rectangular) matrix B, vector g and scalar $\Delta > 0$. We shall suppose that, as was the case in the previous section, (2.35)–(2.36) has a positive root—this need not be the case if Δ is too large. We shall also presume, as was the case in (2.34), that

$$H(\lambda) \stackrel{\text{def}}{=} B^T B + \lambda I = R^T(\lambda)R(\lambda) \tag{2.37}$$

for some upper-triangular (for (2.34), upper bi-diagonal) matrix $R(\lambda)$.

To find the required root it is vital to understand how $\|y(\lambda)\|$ behaves. To this end, here and later we shall use the following general result.

Lemma 2.1 *Given scalars β, a_i and b_i, $i = 1, \ldots, \ell$, with $b_i > 0$ and $\|a\| \neq 0$, let*

$$\chi(\lambda) \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{\ell} \left(\frac{a_i}{b_i + \lambda}\right)^2}$$

and

$$\psi(\lambda) \stackrel{\text{def}}{=} [\chi(\lambda)]^\beta.$$

Then $\psi(\lambda)$ is strictly decreasing and strictly convex on $[0, \infty)$ when $\beta > 0$, and strictly increasing and concave on $[0, \infty)$ when $\beta \in [-1, 0)$. The same is true of $\psi(\lambda)$ for $\lambda \in (0, \infty)$ if instead $b_i \geq 0$, $i = 1, \ldots, \ell$ and more generally for $\lambda \in (-\min_{1 \leq i \leq \ell} b_i, \infty)$ for any b_i, $i = 1, \ldots, \ell$.

Proof Differentiation gives

$$\psi'(\lambda) = \beta[\chi(\lambda)]^{\beta-1}\chi'(\lambda) \quad \text{and}$$

$$\psi''(\lambda) = \beta[\chi(\lambda)]^{\beta-2}\big[\chi(\lambda)\chi''(\lambda) + (\beta - 1)[\chi'(\lambda)]^2\big],$$

and since

$$[\chi(\lambda)]^2 = \sum_{i=1}^{\ell} \left(\frac{a_i}{b_i + \lambda}\right)^2$$

it follows that

$$\chi(\lambda)\chi'(\lambda) = -\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i + \lambda)^3} \quad \text{and} \quad [\chi'(\lambda)]^2 + \chi(\lambda)\chi''(\lambda) = 3\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i + \lambda)^4}.$$

Hence $\psi'(\lambda)$ has the opposite sign to β. Moreover, direct substitution and the Cauchy-Schwarz inequality gives

$$\begin{aligned}
&\chi(\lambda)\chi''(\lambda) + (\beta - 1)[\chi'(\lambda)]^2 \\
&\quad = \frac{3\Big(\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i+\lambda)^4}\Big)\Big(\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i+\lambda)^2}\Big) + (\beta - 2)\Big(\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i+\lambda)^3}\Big)^2}{\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i+\lambda)^2}} \\
&\quad \geq (\beta + 1)\frac{\Big(\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i+\lambda)^3}\Big)^2}{\sum_{i=1}^{\ell} \frac{a_i^2}{(b_i+\lambda)^2}}.
\end{aligned}$$

Thus if $\beta > 0$, $\psi''(\lambda) > 0$, while if $\beta \in [-1, 0]$, $\psi''(\lambda) \leq 0$ as required. □

Lemma 2.2 *Let*

$$\pi(\lambda) \stackrel{\text{def}}{=} \|y(\lambda)\| \equiv \big[y^T(\lambda)y(\lambda)\big]^{\frac{1}{2}}, \tag{2.38}$$

where $y(\lambda)$ satisfies (2.36). *Then $\pi(\lambda)$ is strictly convex on* $[0, \infty)$ *and decays monotonically to zero as λ increases from zero.*

Proof Briefly, suppose that B has the singular-value decomposition $B = PSY$, involving appropriately-dimensioned orthogonal matrices P and Y as well as the rectangular S, whose only nonzero entries are the "diagonals" $S_{ii} \equiv \sigma_i > 0, i = 1, \ldots, \ell$. Then $(S^T S + \lambda I)Yy(\lambda) = S^T P^T g$, and hence

$$[\pi(\lambda)]^2 \equiv \|y(\lambda)\|^2 \equiv \|Yy(\lambda)\|^2 = \sum_{i=1}^{\ell} \frac{\sigma_i^2 r_i^2}{(\sigma_i^2 + \lambda)^2}, \tag{2.39}$$

where $r = P^T g$. Here ℓ is no larger than the smaller of the row and column dimensions of B. Thus the result follows directly from Lemma 2.1 for the case when $\chi(\lambda) = \pi(\lambda)$ and $\beta = 1$. □

This has an immediate vital consequence.

Theorem 2.3 *Newton's method applied to* (2.35) *will converge monotonically, globally Q-linearly and ultimately Q-superlinearly to the positive root λ_* of* (2.35) *for any initial estimate $\lambda_0 \in (0, \lambda_*]$. The same is true for the secant method for initial estimates $0 \leq \lambda_0 < \lambda_1 \leq \lambda_*$.*

Proof This follows directly from Lemma 2.2 because of the known convergence properties of Newton-like methods applied to univariate convex functions. See Lemma A.1 for details. □

We return to this, albeit in more generality, shortly. We comment that although in many cases $\lambda_0 = 0$ might also be permitted, we avoid this here and hereafter since, at least in the under-determined case, the derivatives of $\pi(\lambda)$ at 0 may be infinite.

In practice, instead of seeking the positive root of (2.35), one might equally seek the same root of

$$\psi(\lambda) \stackrel{\text{def}}{=} \Psi(\|y(\lambda)\|) - \Psi(\Delta) = 0 \tag{2.40}$$

for some "suitable" differentiable function Ψ; the choice $\Psi(t) = 1/t$ has strong advantages since this removes the poles present in (2.36) and produces a virtually linear function within a large neighbourhood of the required root [6, 22, 40].

In the special case in which

$$\Psi(t) = t^\alpha \tag{2.41}$$

for a given scalar α, we may generalise Lemma 2.2.

Lemma 2.4 *For given real α, let*

$$\psi(\lambda; \alpha) \stackrel{\text{def}}{=} [\pi(\lambda)]^\alpha,$$

where $\pi(\lambda)$ satisfies (2.38), and suppose that $\lambda \geq 0$. Then $\psi(\lambda; \alpha)$ is strictly convex and decreasing for all $\alpha > 0$ and concave and increasing for all $\alpha \in [-1, 0)$.

Proof The result follows directly from (2.39) and Lemma 2.1 with $\chi(\lambda) = \pi(\lambda)$. □

The situation when $\alpha < -1$ is less clear, although the identity

$$\psi''(\lambda; \alpha) = \alpha[\pi(\lambda)]^{\alpha-4}(3\|y(\lambda)\|^2\|y'(\lambda)\|^2 - (2-\alpha)[y^T(\lambda)y'(\lambda)]^2), \tag{2.42}$$

which follows by differentiating (2.36) and (2.38) twice, may be rewritten as

$$\psi''(\lambda; \alpha) = \alpha[\pi(\lambda)]^{\alpha-2}\|y'(\lambda)\|^2 \left(3 - (2-\alpha)\frac{y'^T(\lambda)H(\lambda)y'(\lambda)}{\|y'(\lambda)\|^2}\frac{y^T(\lambda)H^{-1}(\lambda)y(\lambda)}{\|y(\lambda)\|^2}\right).$$

It is then straightforward to deduce that $\psi(\lambda; \alpha)$ is convex if $\alpha < 2 - 3/\kappa(H(\lambda))$, where $\kappa(H(\lambda))$ is the spectral condition number $(\lambda + \sigma_{\max}^2)/(\lambda + \sigma_{\min}^2)$. In particular, if $\alpha < \alpha_c \stackrel{\text{def}}{=} 2 - 3\sigma_{\max}^2/\sigma_{\min}^2$, $\psi(\lambda; \alpha)$ is convex for all $\lambda \geq 0$. For $\alpha \in (\alpha_c, -1)$, $\psi(\lambda; \alpha)$ may not be unimodal for all $\lambda \geq 0$, but appears often to be so over an (unfortunately unknown) interval surrounding the required root.

As before, this has immediate consequences.

Theorem 2.5 *Newton's method applied to (2.40) in the case $\Psi(t) = t^\alpha$ for any nonzero $\alpha \geq -1$ will converge monotonically, globally Q-linearly and ultimately Q-superlinearly to its positive root λ_* of (2.35) for any initial estimate $\lambda_0 \in (0, \lambda_*]$. The same is true for the secant method for initial estimates $0 \leq \lambda_0 < \lambda_1 \leq \lambda_*$.*

Proof This again follows directly from Lemma 2.4 because of the known convergence properties of Newton-like methods applied to univariate convex function. See Lemma A.1 for details. □

While one might apply the secant method to solve (2.40) without needing derivatives [6], most effective methods require at least first derivatives. Presuming that $\Psi(t)$ and its derivatives are known analytically, the only remaining obstacle is then the need to find the derivative of $\pi(\lambda)$. Direct differentiation of (2.38) immediately gives

$$\pi'(\lambda) = \frac{y^T(\lambda)y'(\lambda)}{\|y(\lambda)\|},$$

while that of (2.36) yields

$$H(\lambda)y'(\lambda) + y(\lambda) = 0.$$

Thus, using (2.34),

$$y^T(\lambda)y'(\lambda) = -y(\lambda)^T H^{-1}(\lambda)y(\lambda) = -h^T(\lambda)h(\lambda),$$

and hence

$$\pi'(\lambda) = -\frac{\|h(\lambda)\|^2}{\|y(\lambda)\|}, \quad \text{where } R^T(\lambda)h(\lambda) = y(\lambda).$$

So the first derivative of $\pi(\lambda)$ is available by forward substitution from $y(\lambda)$ using the lower triangular—for (2.34), lower bi-diagonal—matrix $R^T(\lambda)$. If higher-order derivatives are required, they may be computed successively, each at the cost of a further forward or back substitution [9].

We thus conclude that given λ_0 in $[0, \lambda_*)$, the Newton iterates for (2.40) are generated as

$$\lambda_{j+1} = \lambda_j + \frac{\|y(\lambda_j)\|}{\|h(\lambda_j)\|^2}\frac{[\Psi(\|y(\lambda_j)\|) - \Psi(\Delta)]}{\Psi'(\|y(\lambda_j)\|)} \quad \text{for } j \geq 0 \tag{2.43}$$

and when $\Psi(t) = t^\alpha$ for $\alpha \geq -1$ the iterates converge to λ_*; any starting value $\lambda_0 > 0$ for which $[\Psi(\|y(\lambda_0)\|) - \Psi(\|\Delta\|)]/\Psi'(\|y(\lambda_0)\|) > 0$ is allowed, and the simple expedient of choosing λ_0 to be a tiny positive number almost always suffices. We note that it is possible to compute better starting values [6, 13], but since the above Newton iteration has proved to be so effective in practice, we have not done so.

Since (2.43) with $\Psi(t) = t^\alpha$ converges monotonically to λ_* from the left for all $\alpha \geq -1$, this leads to the interesting opportunity to choose α at each iteration to give the best possible next iterate. Specifically, the Newton correction for a particular α is

$$\Delta\lambda_j(\alpha) = \frac{\|y(\lambda_j)\|^2}{\|h(\lambda_j)\|^2}\frac{(1-\mu_j^\alpha)}{\alpha}, \quad \text{where } \mu_j = \frac{\Delta}{\|y(\lambda_j)\|} \leq 1.$$

But

$$\xi(\alpha) \stackrel{\text{def}}{=} \frac{1-\mu^\alpha}{\alpha}$$

decreases monotonically on $\Re$, since

$$\xi'(\alpha) = \frac{e^{\alpha \ln \mu}}{\alpha^2}[1 - \alpha \ln \mu - e^{-\alpha \ln \mu}] \leq 0$$

which follows because $1-t \le e^{-t}$ for all t, and thus $\xi(\alpha)$ attains its maximum in the region of interest when $\alpha = -1$. Thus, there are good theoretical grounds to support the popular transformation $\Psi(t) = 1/t$. In our experience it is rare to require more than five Newton steps to attain full working accuracy, and frequently one or two iterations are enough (see Sect. 5.1).

We note in passing that an alternative way of transforming the original secular equation (2.35) into one which may be more easily solved, using a nonlinear transformation of the independent variable, has been proposed by Melman [29]. We have not explored this possibility here.

2.3.4 Recovering the solution

Once the boundary has been attained, we stop the iteration as soon as $A^T(Ax_k - b) + \lambda_k x_k$ is sufficiently small. Since (2.16) gives that

$$A^T(Ax_k - b) + \lambda_k x_k = V_k^T[B_k^T B_k y_k + \lambda_k y_k - \beta_1 B_k^T e_1] + \alpha_{k+1} v_{k+1} e_{k+1}^T (B_k y_k - \beta_1 e_1)$$

and as (2.28) implies that the first term vanishes, we have

$$A^T(Ax_k - b) + \lambda_k x_k = \alpha_{k+1} v_{k+1} e_{k+1}^T B_k y_k = \alpha_{k+1} v_{k+1} \beta_{k+1} e_k^T y_k.$$

Hence

$$\|A^T(Ax_k - b) + \lambda_k x_k\| = |\alpha_{k+1} v_{k+1} \beta_{k+1} e_k^T y_k|$$

may be computed trivially from available data.

As soon as the required y_ℓ is known, the estimate $x_\ell = V_\ell y_\ell$ may be recovered by regenerating the vectors v_k, $1 \le l \le \ell$ as needed, or by recovering them from memory or backing store. We have found it advantageous to store a small number t (say $t = 10$) of the first v_k, $1 \le k \le t$ along with u_t to avoid the expense of regenerating these early vectors, and to start the second pass iteration to determine x_ℓ from $k = t$ if necessary. We also take the precaution of recording all previous residuals $\|Ax_k - b\|$, and picking ℓ to give a specified fraction of the best reduction found in the first pass. To do this requires that we know $\|Ax_k - b\|$. Fortunately, again this is easy to compute from available data. For it follows by forming the product of (2.32) with the vector $(y_k^T \; -1)^T$ and from (2.33) that

$$\begin{aligned}\|Ax_k - b\|^2 + \lambda_k \Delta^2 &= \|B_k y_k - \beta_1 e_1\|^2 + \lambda_k \|y_k\|^2 \\ &= \|R_k(\lambda_k) y_k - f_k(\lambda_k)\|^2 + \bar{\phi}_{k+1}^2(\lambda_k) + \|p_k(\lambda_k)\|^2 \\ &= \bar{\phi}_{k+1}^2(\lambda_k) + \|p_k(\lambda_k)\|^2,\end{aligned}$$

and thus

$$\|Ax_k - b\| = \sqrt{\bar{\phi}_{k+1}^2(\lambda_k) + \|p_k(\lambda_k)\|^2 - \lambda_k \Delta^2}.$$

3 Solving the regularised least-squares problem

We next turn to our second, regularised linear least-squares problem (1.3).

3.1 Solution characteristics

As in Sect. 2.1, computationally viable optimality conditions are available. Indeed, the required solution is given by $x(\lambda_*)$ satisfying (2.1), where λ_* is the positive root of a different secular equation

$$\sigma \|x(\lambda)\|^{p-2} - \lambda = 0; \tag{3.1}$$

this latter condition simply arises from the first-order optimality conditions for problem (1.3). Again, if it is feasible to factorize $A^T A + \lambda I$, a simple univariate root finding method—perhaps using the derivative (2.3) of $\|x(\lambda)\|$—may be used to determine the appropriate root of (3.1), while otherwise we must resort to iteration.

3.2 Iterative solution

As before, we shall seek an approximate solution in a sequence of expanding subspaces, and once again we shall use the Golub–Kahan bi-diagonalisation algorithm as our core ingredient. Thus we seek the solution to (1.3) when $x = V_k y$, where V_k satisfies (2.5). This solution is thus $x_k = V_k y_k$, where

$$y_k = \arg\min_{y \in \mathbb{R}^k} \frac{1}{2}\|B_k y - \beta_1 e_1\|^2 + \frac{\sigma}{p}\|y\|^p. \tag{3.2}$$

Thus

$$B_k^T(B_k y_k - \beta_1 e_1) + \sigma \|y_k\|^{p-2} y_k = 0$$

or alternatively

$$B_k^T(B_k y_k - \beta_1 e_1) + \lambda_k y_k = 0 \quad \text{where } \lambda_k = \sigma \|y_k\|^{p-2}.$$

Hence we must find the (positive) root $\lambda = \lambda_k$ of the secular equation

$$\sigma \|y_k(\lambda)\|^{p-2} - \lambda = 0, \tag{3.3}$$

where just as in (2.30)

$$[B_k^T B_k + \lambda I] y_k(\lambda) - \beta_1 B_k^T e_1 = 0. \tag{3.4}$$

We may solve (3.4) exactly as we did in Sect. 2.3.2, and thus it remains to consider the secular equation (3.3). For $p = 2$, this is just the problem considered in detail by Paige and Saunders [38]; in this case $\lambda_k = \sigma$ throughout and the solution can be obtained in a single pass. Thus, in what follows, we shall assume that $p > 2$.

3.3 The secular equation and its solution

Once again, rather than considering (3.3)–(3.4), we prefer the generic case of finding the positive root of

$$\phi_\sigma(\lambda) \stackrel{\text{def}}{=} \sigma \|y(\lambda)\|^{p-2} - \lambda = 0, \tag{3.5}$$

where $y(\lambda)$ satisfies (2.36). But as before, there are advantages in seeking instead the same root of

$$\psi_\sigma(\lambda) \stackrel{\text{def}}{=} \Psi(\sigma \|y(\lambda)\|^{p-2}) - \Psi(\lambda) = 0 \tag{3.6}$$

for some "suitable" differentiable function Ψ. The choices $\Psi_\sigma(t) = (t/\sigma)^\beta$ for some real β, giving rise to the secular equation

$$\|y(\lambda)\|^{\beta(p-2)} - (\lambda/\sigma)^\beta = 0 \tag{3.7}$$

(particularly with $\beta = -1$), or $\Psi_\sigma(t, \lambda) = (\lambda\sigma/t)^\beta$, yielding the secular equation

$$\frac{\lambda^\beta}{\|y(\lambda)\|^{\beta(p-2)}} - \sigma^\beta = 0, \tag{3.8}$$

have both been proposed for the special case $p = 3$ [5].

For the secular equation (3.7), we have the following result.

Lemma 3.1 *For given real β and $p > 2$, let*

$$\theta(\lambda; \beta) \stackrel{\text{def}}{=} \|y(\lambda)\|^{\beta(p-2)} - (\lambda/\sigma)^\beta ,$$

where $y(\lambda)$ satisfies (2.36), and suppose that $\lambda \geq 0$. Then $\theta(\lambda; \beta)$ is strictly convex and decreasing for all $\beta \in (0, 1]$ and concave and increasing for all $-1/(p-2) \leq \beta < 0$.

Proof Since $\zeta(\lambda) \stackrel{\text{def}}{=} -\lambda^\gamma$ is strictly convex and decreasing when $\lambda \geq 0$ for $\gamma \in (0, 1]$, it follows from Lemma 2.4 that the same is true for $\theta(\lambda; \beta)$ for $\beta \in (0, 1]$. Likewise, as $\zeta(\lambda)$ is strictly concave and increasing when $\lambda \geq 0$ for $\gamma < 0$, Lemma 2.4 shows that the same is true for $\theta(\lambda; \beta)$ for $-1/(p-2) \leq \beta < 0$. □

Thus, as in the trust-region case, appropriately initialized secant and Newton's methods applied to (3.7) possess powerful convergence properties.

Theorem 3.2 *Newton's method applied to (3.7) for nonzero $\beta \in [-1/(p-2), 1]$ will converge monotonically, globally Q-linearly and ultimately Q-superlinearly to its positive root λ_* of (3.5) for any initial estimate $\lambda_0 \in (0, \lambda_*]$. The same is true for the secant method for initial estimates $0 \leq \lambda_0 < \lambda_1 \leq \lambda_*$.*

Proof As before, this follows directly from Lemma 3.1 because of the known convergence properties of Newton-like methods applied to univariate convex function. See Lemma A.1 for details. □

By contrast, it is easy to find examples for which the curvature for the function in (3.8) changes sign, and thus we are unable to conclude in general that Newton-like methods for this secular equation will converge globally in $[0, \lambda_*]$.

The Newton iterates for (3.7) satisfy

$$\lambda_{j+1} = \lambda_j + \frac{\|y(\lambda_j)\|^{\beta(p-2)} - (\lambda_j/\sigma)^\beta}{\beta[(p-2)\|y(\lambda_j)\|^{\beta(p-2)-2}\|h(\lambda_j)\|^2 + \lambda_j^{\beta-1}/\sigma^\beta]}$$

and thus for given β, the Newton correction is

$$\Delta\lambda_j(\beta) = \frac{\|y(\lambda_j)\|^2}{(p-2)\|h(\lambda_j)\|^2} \frac{(1-\mu_j^\beta)}{\beta(1+\tau_j\mu_j^\beta)},$$

where, if $\lambda_0 \in [0, \lambda_*]$ and $\beta \in [-1/(p-2), 1]$,

$$\tau_j = \frac{\|y(\lambda_j)\|^2}{(p-2)\lambda_j\|h(\lambda_j)\|^2} \quad \text{and} \quad \mu_j = \frac{\lambda_j}{\sigma_j\|y(\lambda_j)\|^{p-2}} \le 1.$$

This again gives us the opportunity to pick β to give the best (largest) Newton correction. Unfortunately, unlike in the trust-region case, the correction may be multi-modal in the region of interest, and thus the best step may have to be picked by iteration to maximize

$$\eta_j(\beta) \stackrel{\text{def}}{=} \frac{1-\mu_j^\beta}{\beta(1+\tau_j\mu_j^\beta)}$$

for the given data μ_j and τ_j.

When $2 < p \le 3$, another acceleration is possible by choosing $\beta = -1$ in (3.7). This gives

$$\|y(\lambda)\|^{2-p} - \sigma/\lambda = 0. \tag{3.9}$$

Rather than applying Newton's method to (3.9), it then pays instead to linearize the term $\omega(\lambda) \stackrel{\text{def}}{=} \|y(\lambda)\|^{2-p}$, but not the remaining term σ/λ, when computing a correction $\Delta\lambda_j^{\text{C}}$ to the estimate λ_j of the required root of (3.9). The resulting correction thus satisfies the equation

$$\omega(\lambda_j) + \omega'(\lambda_j)\Delta\lambda_j^{\text{C}} \equiv \frac{1}{\|y(\lambda_j)\|^{p-2}} + (p-2)\frac{\|h(\lambda_j)\|^2}{\|y(\lambda_j)\|^p}\Delta\lambda_j^{\text{C}} = \frac{\sigma}{\lambda_j + \Delta\lambda_j^{\text{C}}}, \tag{3.10}$$

which may be rewritten as a quadratic equation for $\Delta\lambda_j^{\text{C}}$.

Before we analyse the correction given by (3.10), we have the following general result.

Lemma 3.3 *Let the interval* $\mathcal{I} \subseteq \Re^+ \equiv [0, \infty)$ *and* $\sigma > 0$. *Suppose that* $\phi : \mathcal{I} \to \Re^+$ *is concave, strictly increasing and continuously differentiable, and that* $\theta(\lambda) \stackrel{\text{def}}{=} \phi(\lambda) - \sigma/\lambda$ *has a (unique) zero* $\lambda_* \in \mathcal{I}$. *Let* $\lambda_e \in \mathcal{I}$ *be such that* $\theta(\lambda_e) < 0$.

Then both the Newton iterate $\lambda_e + \Delta\lambda_e^{\mathrm{N}}$ *for the equation* $\theta(\lambda) = 0$ *and the approximation* $\lambda_e + \Delta\lambda_e^{\mathrm{C}}$, *where* $\Delta\lambda_e^{\mathrm{C}}$ *is the largest root of*

$$\phi(\lambda_e) + \phi'(\lambda_e)\Delta\lambda_e^{\mathrm{C}} = \frac{\sigma}{\lambda_e + \Delta\lambda_e^{\mathrm{C}}}, \tag{3.11}$$

also lie in $\mathcal{I}$ *and yield negative values of* θ, *and* (*if repeated*) *converge monotonically towards* λ_*. *The convergence is globally Q-linear with factor at least* $1 - \theta'(\lambda_*)/\theta'(\lambda_e) < 1$ *and is ultimately Q-superlinear. Moreover* $\lambda_e + \Delta\lambda_e^{\mathrm{N}} \le \lambda_e + \Delta\lambda_e^{\mathrm{C}} \le \lambda_*$.

Proof Since $\upsilon(\lambda) \stackrel{\text{def}}{=} -\sigma/\lambda$ is concave, it is strictly increasing and continuously differentiable on $\mathcal{I}$, the same is true of $\theta(\lambda)$ by assumption on ϕ. Thus it follows from Lemma A.1 that the Newton iterates remain in $[\lambda_e, \lambda_*]$ and convergence occurs as described.

Since $\phi(\lambda_e)$ is a concave function of λ, (3.9) and (3.10) give that

$$\begin{aligned}\theta(\lambda_e + \Delta\lambda_e^{\mathrm{C}}) &= \phi(\lambda_e + \Delta\lambda_e^{\mathrm{C}}) - \frac{\sigma}{\lambda_e + \Delta\lambda_e^{\mathrm{C}}} \\ &\le \phi(\lambda_e) + \phi'(\lambda_e)\Delta\lambda_e^{\mathrm{C}} - \frac{\sigma}{\lambda_e + \Delta\lambda_e^{\mathrm{C}}} = 0.\end{aligned}$$

The Newton correction satisfies the linearized equation

$$\phi(\lambda_e) + \phi'(\lambda_e)\Delta\lambda_e^{\mathrm{N}} = \frac{\sigma}{\lambda_e} - \frac{\sigma}{\lambda_e^2}\Delta\lambda_e^{\mathrm{N}}. \tag{3.12}$$

But, as $\sigma/\lambda \equiv -\upsilon(\lambda)$ is a convex function of λ,

$$\frac{\sigma}{\lambda_e + \Delta\lambda_e^{\mathrm{C}}} \ge \frac{\sigma}{\lambda_e} - \frac{\sigma}{\lambda_e^2}\Delta\lambda_e^{\mathrm{C}},$$

and hence

$$\phi(\lambda_e) + \phi'(\lambda_e)\Delta\lambda_e^{\mathrm{C}} \ge \frac{\sigma}{\lambda_e} - \frac{\sigma}{\lambda_e^2}\Delta\lambda_e^{\mathrm{C}},$$

from (3.11). Combining this with (3.12), we obtain

$$\theta'(\lambda_e)(\Delta\lambda_e^{\mathrm{C}} - \Delta\lambda_e^{\mathrm{N}}) = \left(\phi'(\lambda_e) + \frac{\sigma}{\lambda_e^2}\right)(\Delta\lambda_e^{\mathrm{C}} - \Delta\lambda_e^{\mathrm{N}}) \ge 0$$

and hence $\Delta\lambda_e^{\mathrm{C}} \ge \Delta\lambda_e^{\mathrm{N}} > 0$ since $\theta'(\lambda_e) > 0$. Thus the alternative iterates improves on the Newton one, and the remaining results follow immediately. □

Applying Lemma 3.3 to the largest root of (3.10) then gives the following improvement on Newton's method.

Corollary 3.4 *Suppose that* $2 < p \le 3$. *Then the sequence* $\{\lambda_j\}$, $j \ge 0$, *where* $\lambda_{j+1} = \lambda_j + \Delta\lambda_j^{\mathrm{C}}$ *and* $\Delta\lambda_j^{\mathrm{C}}$ *is the largest root of* (3.10), *will converge monotonically, globally*

Q-linearly (with factor at least $1-\theta'(\lambda_*)/\theta'(\lambda_0) < 1$) *and ultimately Q-superlinearly to its positive root* λ_* *of* (3.5) *for any initial estimate* $\lambda_0 \in (0, \lambda_*]$. *Moreover,* $\lambda_j + \Delta\lambda_j^{\mathrm{N}} \le \lambda_{j+1} \le \lambda_*$, *where* $\Delta\lambda_j^{\mathrm{N}}$ *is the Newton correction for the equation* $\theta(\lambda) = 0$ *at* $\lambda = \lambda_j$.

Proof The function ω in (3.10) satisfies the assumptions required by ϕ in Lemma 3.3 because of Lemma 2.4. The result then follows immediately from Lemma 3.3. □

In practice, the improvements from using $\Delta\lambda_j^{\mathrm{C}}$ from (3.10) rather than the Newton correction are sometimes dramatic, particularly when λ is small since then linearization of σ/λ gives a poor approximation. Similar accelerations, appropriate when the coefficients σ_i and r_i in (2.39) are known explicitly, are given by Bunch, Nielsen and Sorensen [24] and Melman [29].

4 Solving the regularised least-ℓ_2-norm problem

Our final topic is the solution of the regularised linear least ℓ_2-norm problem (1.4). We note in passing that (1.4) is an exact penalty function [35, Sect. 15.1] for the problem of minimizing $\|x\|$ subject to $Ax = b$, and thus if the latter is compatible we will expect these equations to be satisfied for all sufficiently small σ. By contrast (1.3) is the quadratic penalty function [35, Sect. 15.1] for the same problem and thus there is no expectation that $Ax = b$ will be satisfied even if it is compatible.

4.1 Solution characteristics

Let $\nu = \|Ax - b\|$. In this case (1.4) is equivalent to the differentiable constrained problem

$$\underset{x\in\mathbb{R}^n,\nu\in\mathbb{R}}{\text{minimize}}\ \nu + \frac{\sigma}{p}\|x\|^p \quad \text{subject to} \quad \frac{1}{2}\|Ax-b\|^2 = \frac{1}{2}\nu^2. \tag{4.1}$$

So long as $\nu > 0$, first-order optimality conditions for (4.1) require that

$$\begin{pmatrix} \sigma x\|x\|^{p-2} \\ 1 \end{pmatrix} = \mu \begin{pmatrix} A^T(Ax-b) \\ -\nu \end{pmatrix} \tag{4.2}$$

for some Lagrange multiplier μ. Letting $\lambda = \sigma\nu\|x\|^{p-2}$, (4.2) implies that the required solution is $x(\lambda_*)$, where $x(\lambda)$ is given by (2.1) and λ_* satisfies yet another secular equation

$$\|Ax(\lambda) - b\| - \frac{\lambda}{\sigma\|x(\lambda)\|^{p-2}} = 0. \tag{4.3}$$

Once again, if factorizing $A^TA + \lambda I$ is feasible, a simple univariate root finding method might be used to determine the appropriate root of (4.3)—this might require the derivatives (2.3) of $\|x(\lambda)\|$ and

$$\nu'(\lambda) = \frac{(Ax(\lambda)-b)^TAx'(\lambda)}{\nu(\lambda)} = -\lambda\frac{x^T(\lambda)x'(\lambda)}{\nu(\lambda)}$$

of $\nu(\lambda) = \|Ax(\lambda) - b\|$—but otherwise we shall resort to an iterative method.

If $\nu = 0$, as we have mentioned, we simply require the minimum ℓ_2-norm solution to $Ax = b$. Notice that in this case, the required solution satisfies

$$\begin{pmatrix} I & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ r \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}, \tag{4.4}$$

while the solution $x(\lambda)$ to (2.1) when $\nu > 0$ satisfies

$$\begin{pmatrix} I & A^T \\ A & -\lambda I \end{pmatrix} \begin{pmatrix} x(\lambda) \\ r(\lambda) \end{pmatrix} = \begin{pmatrix} 0 \\ b \end{pmatrix}. \tag{4.5}$$

Hence if the secular equation (4.3) has no positive root, the required solution may be recovered either directly from (4.4) or indirectly as $\lim_{\lambda \to 0_+} x(\lambda)$ from (2.1) or (4.5).

4.2 Iterative solution

Unsurprisingly, we seek an approximate solution in a sequence of expanding subspaces based on Golub–Kahan bi-diagonalisation. Thus we seek the solution to (1.4) when $x = V_k y$, where V_k satisfies (2.5). This solution is thus $x_k = V_k y_k$, where

$$y_k = \arg\min_{y \in \mathbb{R}^k} \frac{1}{2}\|B_k y - \beta_1 e_1\| + \frac{\sigma}{p}\|y\|^p. \tag{4.6}$$

Crucially, the norm of the residuals $\nu_k = \|B_k y_k(\lambda) - \beta_1 e_1\|$ is necessarily nonzero so long as the bi-diagonalisation continues, and thus any potential difficulties that might occur if ν_k were zero cannot arise. For otherwise, if $B_k y_k = \beta_1 e_1$ and each α_i and β_i is nonzero, the first k equations imply that every component of y_k must also be nonzero. But then the residual of the $k+1$st equation cannot be zero. Thus, as in Sect. 3.2, we seek $y_k = y_k(\lambda_k)$ where $y_k(\lambda)$ satisfies (3.4) and λ_k is the positive root of the secular equation

$$\|B_k y_k(\lambda) - \beta_1 e_1\| - \frac{\lambda}{\sigma \|y_k(\lambda)\|^{p-2}} = 0. \tag{4.7}$$

It remains to examine the secular equation (4.7).

4.3 The secular equation and its solution

Once again, rather than considering specifically (3.4) and (4.7), we investigate the generic problem of finding the positive root of

$$\rho(\lambda) \stackrel{\text{def}}{=} \sigma \frac{\|By(\lambda) - g\|}{\lambda} - \frac{1}{\|y(\lambda)\|^{p-2}} = 0, \tag{4.8}$$

where $y(\lambda)$ satisfies (2.36); as we shall see, there is a good reason for dividing both sides of the original equation by λ. But more generally, we may prefer

$$\sigma^\beta \left(\frac{\|By(\lambda) - g\|}{\lambda} \right)^\beta - \frac{1}{\|y(\lambda)\|^{\beta(p-2)}} = 0 \tag{4.9}$$

or

$$\left(\frac{\|By(\lambda)-g\|}{\lambda}\right)^{\beta}\|y(\lambda)\|^{\beta(p-2)}-\frac{1}{\sigma^{\beta}}=0 \tag{4.10}$$

for some real β. To this end, we have the following result.

Lemma 4.1 *Let*

$$\tau(\lambda)\stackrel{\text{def}}{=}\frac{\|By(\lambda)-g\|}{\lambda}$$

and suppose that $\lambda>0$. *Then* $[\tau(\lambda)])^{\beta}$ *is strictly convex and strictly decreasing for all* $\beta>0$ *and concave and strictly increasing for all* $\beta\in[-1,0)$.

Proof Using the notation introduced in the proof of Lemma 2.2 and supposing B has m rows, we have that $By(\lambda)-g=P(S(S^TS+\lambda I)^{-1}S^Tr-r)$, and hence

$$[\tau(\lambda)]^2=\frac{\|By(\lambda)-g\|^2}{\lambda^2}=\sum_{i=1}^{\ell}\frac{r_i^2}{(\sigma_i^2+\lambda)^2}+\sum_{i=\ell+1}^{m}\frac{r_i^2}{\lambda^2}. \tag{4.11}$$

The result then follows directly by applying Lemma 2.1 with $\chi(\lambda)=\tau(\lambda)$. □

Consider first the secular equation (4.9). If $\beta>0$, the leading term is strictly convex and decreasing (Lemma 4.1) while the second term is convex and decreasing for $\beta\le 1/(p-2)$ (Lemma 2.4) and hence so is their sum. Similarly, if $\beta<0$, the leading term is concave and increasing for $\beta\ge -1$ (Lemma 4.1) while the remaining term is strictly concave (just concave if $p=2$) and increasing (Lemma 2.4) as is the sum of the two terms. Thus we have the following convergence result.

Theorem 4.2 *Newton's method applied to* (4.9) *for nonzero* $\beta\in[-1,1/(p-2)]$ *will converge monotonically, globally Q-linearly and ultimately Q-superlinearly to its positive root* λ_* *of* (4.7) *for any initial estimate* $\lambda_0>0$ *for which* $\rho(\lambda_0)\ge 0$. *The same is true for the secant method for initial estimates* $0\le\lambda_0<\lambda_1$ *when* $\rho(\lambda_1)\ge 0$.

Proof The fact that $\rho(\lambda)$ is decreasing on $[0,\infty)$ together with the stated requirements on λ_0 and, if necessary, λ_1 imply that the starting values are to the left of the required root. The result then follows directly from the above discussion since the function in (4.9) is convex and decreasing ($0<\beta\le 1/(p-2)$) or concave and increasing ($-1<\beta<0$), and because of the known convergence properties of Newton-like methods applied to such functions. See Lemma A.1 for details. □

While Theorem 4.2 appears encouraging, the convergence may initially be slow when $p>2$ since both $\|y(\lambda)\|$ and $\tau(\lambda)$ may be large (and have large derivatives) when λ is close to zero. This defect might in principle be avoided by considering secular equations involving their reciprocals, such as (4.10) when $\beta<0$. If $\beta>0$, the leading term in (4.10) is the product of two decreasing, convex, positive functions (Lemmas 2.4 and 4.1) and thus decreasing, convex and positive [4, Exer. 3.32]. Thus

Newton-like methods for (4.10) will converge as above in this case. However, for negative β it is not clear when the leading term

$$\xi(\lambda) \stackrel{\text{def}}{=} \left(\frac{\|By(\lambda) - g\|}{\lambda} \|y(\lambda)\|^{p-2} \right)^{\beta} \tag{4.12}$$

in (4.10) will be concave; it is the product of increasing, concave terms when $\max(-1, 1/(2-p)) \leq \beta < 0$ (Lemmas 2.4 and 4.1), but this is insufficient to ensure concavity. Plots of (4.12) for various examples suggest that the term in question may be concave for sufficiently small negative β, and indeed it can be shown that $\xi(\lambda)$ is bounded below and above by known concave functions[2] when $\beta \in [-\frac{1}{2}, 0)$ and $p \leq 3$.

In practice, we have found that Newton steps for (4.10) with $\beta = -1/(p-1)$ always seem to outperform those for (4.9) with β in the range allowed by Theorem 4.2. We thus use such steps by default, but with the safeguard that if $\rho(\lambda)$ in (4.8) following the step becomes negative, we revert to the Newton step for (4.9) with $\beta = -1/(p-2)$. To date this safeguard has not been needed, and between two and six Newton steps appear to be necessary to achieve full working accuracy when $p = 2$; this increases slightly for larger p (see Sect. 5.1).

The special case $p = 2$ is not affected by these deliberations since then (4.10) becomes

$$\left(\frac{\|By(\lambda) - g\|}{\lambda} \right)^{\beta} - \frac{1}{\sigma^{\beta}} = 0, \tag{4.13}$$

for which the leading term is concave and increasing for all $\beta \in [-1, 0)$. Thus, for this case, Newton-like methods for (4.13) will converge as in Theorem 4.2, and the choice $\beta = -1$ gives the best behaviour for the same reasons as those discussed at the end of Sect. 2.3.3.

5 Software

The ideas developed in this paper have been implemented as three thread-safe Fortran 95 packages—respectively LSTR, LSRT and L2RT for problems (1.2)–(1.4)—as part of version 2.1 of the GALAHAD optimization library [15]. All use reverse communication to obtain the matrix-vector products

$$u := u + Av \quad \text{and} \quad v := v + A^T u,$$

[2] Specifically, given (2.39) and (4.11), it can be shown that if $\alpha \in (0, 1]$

$$\kappa_1 [\pi(\lambda)]^2 \min(1, [\pi(\lambda)]^2) \leq \left([\pi(\lambda)]^{\alpha} \tau(\lambda)\right)^2 \leq \kappa_2 [\pi(\lambda)]^2 \max(1, [\pi(\lambda)]^2)$$

for some constants κ_1 and κ_2. In this case

$$\kappa_1^{\beta} \min([\pi(\lambda)]^{\beta}, [\pi(\lambda)]^{2\beta}) \leq \left([\pi(\lambda)]^{\alpha} \tau(\lambda)\right)^{\beta} \leq \kappa_2^{\beta} \max([\pi(\lambda)]^{\beta}, [\pi(\lambda)]^{2\beta})$$

for which the bounding functions are concave by Lemma 2.4 when $\beta \in [-\frac{1}{2}, 0)$.

as required, and offer a variety of options. In particular, for the trust-region problem, the user can decide whether to stop at the Steihaug-Toint point if encountered (Sect. 2.3.1), or to continue around the trust-region boundary (Sect. 2.3.2). For all three problems, as we have mentioned in Sect. 2.3.4, the second-phase may be accelerated if needed by storing the first t (say) vectors v_i, $i = 1, \dots, t$, along with u_t as calculated in the first pass so that the bi-diagonalisation (2.4) may be restarted at iteration $k = t$. Moreover (Sect. 2.3.4), as the second pass may be an additional expense, a record is kept of the optimal objective function values for each value of k, and the second pass is only performed so far as to ensure a given fraction of the final optimal objective value. Large savings may be made in the second pass by choosing the required fraction to be significantly smaller than one.

The software may also be used to solve weighted least-squares problems involving the objective $\|W(Ax - b)\|$ and a scaled trust region $\|Sx\| \leq \Delta$ simply by solving instead the problem

$$\operatorname*{minimize}_{\bar{x} \in \mathbb{R}^n} \|\bar{A}\bar{x} - \bar{b}\| \quad \text{subject to} \quad \|\bar{x}\| \leq \Delta,$$

where $\bar{A} = WAS^{-1}$ and $\bar{b} = Wb$ and then recovering $x = S^{-1}\bar{x}$. Note the implication here that S must be non-singular. Similarly the weighted regularised problems

$$\operatorname*{minimize}_{x \in \mathbb{R}^n} \frac{1}{q} \|W(Ax - b)\|^q + \frac{1}{p} \sigma \|Sx\|^p$$

($q = 1, 2$) may be solved instead as

$$\operatorname*{minimize}_{\bar{x} \in \mathbb{R}^n} \frac{1}{q} \|\bar{A}\bar{x} - \bar{b}\|^q + \frac{1}{p} \sigma \|\bar{x}\|^p.$$

Note that the choice of W and S will affect the convergence of the method, and thus good choices may be used to accelerate its convergence. This is often known as preconditioning, but be aware that preconditioning changes the norms that define the problem. Good preconditioners will cluster the singular values of $\bar{A}$ around a few distinct values, and ideally (but usually unrealistically) all the singular values will be mapped to 1.

As we indicated in Sect. 1.1, our intention has always been to use these packages to solve problems arising in nonlinear fitting and constrained optimization. We shall delay numerical comparisons until we have done so. However at least one comment is in order here. We mentioned in Sect. 2.3.1 that the improvement possible if we solve the trust-region problem (1.2) accurately is no more than twice that derived from the Steihaug-Toint point. In practice, our experience has been far less optimistic, and often less than a ten percent—and sometimes less than one percent—improvement has been observed. Thus in the case of (1.2), we do not recommend going beyond the Steihaug-Toint point, since to do so will incur the cost of a second pass to recover x_k from y_k. This is by contrast to the problem of minimizing general quadratic functions within an ℓ_2 trust-region where the Steihaug-Toint point can be a very poor predictor of the possible reduction. This issue is not relevant for our other two, regularised, problems (1.3) and (1.4).

5.1 Numerical experience

We now supply evidence supporting the claims made in Sects. 2.3.3, 3.3 and 4.3 that Newton/Newton-like methods for the relevant secular equations are efficient in practice. To illustrate this, we construct a problem in which we may control the conditioning, and report the performance of the Newton-like iteration on a sequence of examples generated by restricting the given data to the subspace $x = V_k y$ (see Sect. 2.2.2) for each of the problems (1.2)–(1.4).

For given $w \in \Re^m$, $z \in \Re^n$, and "diagonal" $D \in \Re^{m\times n}$ for which $d_{i,j} = 0$ $(i \neq j)$, we construct A using Householder reflections and the singular-value decomposition

$$A = \left(I - \frac{2ww^T}{w^T w}\right) D \left(I - \frac{2zz^T}{z^T z}\right).$$

In particular, we chose $w_i = 1$ for all i, $z_i = 1$ for odd i and -1 for even i, and the diagonals $d_{i,i}$ to decrease linearly from 1 to some given $\rho < 1$; thus the spectral condition number is $1/\rho$. The components of b are all chosen to be 1. We considered three configurations $(m, n) = (1000, 5000)$, $(5000, 1000)$ and $(5000, 5000)$, two values of $\rho = 0.01$ and 0.0001, and three trust-region radii $\Delta = 1, 100$ and 10000. In Table 1 we report statistics on the number of required Newton steps (minimum, mean and maximum over all k) for the sequence of problems (1.2) restricted to $V_k y$ as implemented in LSTR.

Notice that the solution to some of these problems lies interior to the trust-region, and thus no Newton root-finding is needed. For those that do, between one and six Newton steps are required for each k, with a mean between two and four. This behaviour agrees with our observations on other "realistic" examples from the CUTEr [16] test set. It is interesting to observe that the number of iterations required grows (modestly) with the size of the radius, but this might be expected as this corresponds to required smaller values of λ for which ϕ in (2.35) and its variants change more rapidly.

Similar observations may be made for the iterations we have proposed for problems (1.3) and (1.4) using LSRT and L2RT. With the same data, but now with five different values of the regularisation weight $\sigma = 0.0001, 0.01, 1, 100$ and 10000, we report in Tables 2 and 3 the performance for the most-likely practical values of p (3 for problem (1.3) and 2 for (1.4)).

Table 1 Newton iterations for (2.40) in the case $\Psi(t) = 1/t$

Δ	Condition number	$m = 1000, n = 5000$ min	mean	max	$m = 5000, n = 1000$ min	mean	max	$m = 5000, n = 5000$ min	mean	max
1	100	1	2.0	3	1	2.0	3	1	2.0	3
1	10000	1	2.0	3	1	2.0	3	1	2.0	3
100	100	1	2.7	5	2	2.7	4	1	2.7	5
100	10000	1	2.6	5	2	2.7	4	1	2.7	5
10000	100		Interior			Interior			Interior	
10000	10000	2	2.7	5		Interior		3	3.8	6

Table 2 Newton-like iterations for (3.5) as summarised in Corollary 3.4 ($p = 3$)

σ	Condition number	$m = 1000, n = 5000$ min	mean	max	$m = 5000, n = 1000$ min	mean	max	$m = 5000, n = 5000$ min	mean	max
.0001	100	1	2.6	4	2	2.6	4	1	2.6	4
.0001	10000	1	2.6	4	2	2.6	4	1	2.6	4
.01	100	1	2.4	4	1	2.4	4	1	2.4	4
.01	10000	1	2.4	4	1	2.4	4	1	2.4	4
1	100	1	2.1	3	1	2.0	3	1	2.1	3
1	10000	1	2.1	3	1	2.0	3	1	2.1	3
100	100	1	1.8	2	1	1.8	2	1	1.8	2
100	10000	1	1.8	2	1	1.8	2	1	1.8	2
10000	100	1	1.7	2	1	1.7	2	1	1.7	2
10000	10000	1	1.7	2	1	1.7	2	1	1.7	2

Table 3 Newton iterations for (4.10) with $\beta = -1$

σ	Condition number	$m = 1000, n = 5000$ min	mean	max	$m = 5000, n = 1000$ min	mean	max	$m = 5000, n = 5000$ min	mean	max
.0001	100	1	2.7	4	1	2.0	3	1	2.6	4
.0001	10000	1	2.5	4	1	2.0	3	1	2.5	4
.01	100	1	2.5	5	1	2.2	4	1	2.5	5
.01	10000	1	2.5	5	1	2.2	4	1	2.5	5
1	100	1	2.2	4	1	2.0	4	1	2.2	4
1	10000	1	2.2	4	1	2.0	4	1	2.2	4
100	100	2	3.0	4	1	2.5	4	1	2.5	4
100	10000	2	3.0	4	1	2.5	4	1	2.5	4
10000	100	1	2.0	3	1	2.0	3	1	2.0	3
10000	10000	1	2.0	3	1	2.0	3	1	2.0	3

Now we observe between one and five Newton steps are required for each k, with a mean between two and three. Again slightly more effort is required when σ and hence the required λ are small.

Finally, when a non-standard value of p is used, we see in Table 4 that the number of iterations rises; for $p = 3$, now between one and nine iterations are needed, with the mean between two and four. Small values of the regularisation lead to small values of λ for which $\rho(\lambda)$ in (4.8) and its variants change rapidly.

6 Comments and conclusions

We have proposed a framework for solving a variety of (implicitly or explicitly) regularised linear-least squares problems. All proceed by approximating the solution to the given problem in an increasing set of convenient subspaces. Each leads to its

Table 4 Newton iterations (4.10) with $\beta = -0.5$ ($p = 3$)

σ	Condition number	$m = 1000, n = 5000$ min	mean	max	$m = 5000, n = 1000$ min	mean	max	$m = 5000, n = 5000$ min	mean	max
.0001	100	1	2.6	9	1	2.6	9	1	2.7	9
.0001	10000	1	2.7	9	1	2.6	8	1	2.6	8
.01	100	1	2.8	7	1	2.8	7	1	2.8	7
.01	10000	1	2.8	7	1	2.9	7	1	2.9	7
1	100	1	2.8	6	1	2.8	6	2	3.2	6
1	10000	2	3.2	6	1	2.8	6	1	2.8	6
100	100	2	3.0	5	2	3.0	5	2	3.0	5
100	10000	2	3.0	5	1	2.7	5	1	2.7	5
10000	100	1	2.7	5	1	2.7	5	2	3.5	5
10000	10000	2	3.5	5	2	3.5	5	2	3.5	5

own secular equation—a root-finding problem—for which Newton-like and other approaches are most effective. Software for each of the problems is available as part of GALAHAD. The methods considered may easily be extended to the more general regularisation

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \frac{1}{q}\|Ax - b\|^q + \frac{1}{p}\sigma\|x\|^p$$

for $p, q \geq 1$ but we do not give details here.

One alternative we have not yet considered is to apply the ideas first proposed by Hager and Park [19, 20], and subsequently refined by Erway, Gill, and Griffin [10], for the problem of minimizing a general quadratic function $q(x)$ within a spherical trust-region. These recognise that a possible disadvantage of the earlier GLTR approach [14] to the same problem—and by implication for the methods we have considered here—is the need for a second pass to recover the solution $x_k = V_k y_k$ once a suitable y_k has been determined. The idea is simply that once it has been established that the solution lies on the trust-region boundary, a sequence of points $\{x_k\}$ are generated by choosing x_{k+1} to solve the given problem over a low-dimensional subspace $\mathcal{S}_k$ containing at least x_k and a mixture of $\nabla_x q(x_k)$, a crude Newton-based approximation to the solution $x(\lambda)$ to the relevant secular equation and an approximation to the eigenvector corresponding to the left-most eigenvalue of $\nabla_{xx} q(x_k)$; since in our cases the objective is convex, the latter would not be needed. It has been established [20] that such an iteration converges to the solution to the problem, although it is unclear quite how this compares in cost with that of the second pass in the GLTR approach. This general approach can clearly be adapted—in the case of problem (1.2)—or generalised to the regularised problems (1.3) and (1.4). It remains to see how effective this is in comparison to the methods we have given in all of these cases.

Acknowledgements This paper is dedicated to the memory of Gene Golub. The second author enjoyed a number of interesting discussions with Gene on the subject of secular equations during the summer of 2007. The paper also benefited from fruitful discussions with Michael Saunders concerning LSQR, and those with Stefania Bellavia, Benedetta Morini and Margherita Porcelli on crucial details in Sect. 4. Useful comments from a pair of anonymous referees are also gratefully acknowledged.

Appendix A

The following result is stated, in part, in other sources, e.g., [23, Theorem 4.8]. For completeness, here we state and prove the version we require.

Lemma A.1 *Suppose that $\theta : \mathcal{I} \rightarrow \Re$ is convex (resp. concave), strictly decreasing (resp. strictly increasing) and continuously differentiable on some interval $\mathcal{I} = [\lambda_{\min}, \lambda_{max}] \subseteq \Re$, and suppose further that there is a $\lambda_* \in \mathcal{I}$ for which $\theta(\lambda_*) = 0$.*

(i) *Now suppose that $\theta(\lambda_0) > 0$ for some given $\lambda_0 \in \mathcal{I}$. Then the Newton iterates $\{\lambda_j\}$, where*

$$\lambda_{j+1} = \lambda_j - \frac{\theta(\lambda_j)}{\theta'(\lambda_j)}, \tag{A.1}$$

for $j \geq 0$, all lie in $[\lambda_0, \lambda_*]$ and increase monotonically to λ_*. The convergence is globally Q-linear with factor at least

$$\gamma^{N} \stackrel{\text{def}}{=} 1 - \frac{\theta'(\lambda_*)}{\theta'(\lambda_0)} < 1$$

and is ultimately Q-superlinear (Q-quadratic if additionally θ' is Lipschitz continuous around λ_*).

(ii) *Suppose that $\theta(\lambda_0)$ and $\theta(\lambda_1) > 0$ for some given $\lambda_0 < \lambda_1 \in \mathcal{I}$. Then the secant iterates $\{\lambda_j\}$, where*

$$\lambda_{j+1} = \lambda_j - \frac{(\lambda_j - \lambda_{j-1})\theta(\lambda_j)}{\theta(\lambda_j) - \theta(\lambda_{j-1})}, \tag{A.2}$$

for $j \geq 1$, all lie in $[\lambda_0, \lambda_*]$ and increase monotonically to λ_*. The convergence is globally Q-linear with factor at least γ^{N}, and is ultimately Q-superlinear.

Proof We consider the convex case; the concave case then follows directly by considering $-\theta$. The assumptions are such that $\lambda \in \mathcal{I}$ with $\lambda < \lambda_*$ if and only if $\theta(\lambda) > 0$.

(i) By induction, suppose that $\theta(\lambda_j) > 0$. Since by assumption $\theta'(\lambda_j) < 0$, (A.1) shows that $\lambda_{j+1} > \lambda_j$. Additionally, the convexity of θ and (A.1) imply that

$$\theta(\lambda_{j+1}) \geq \theta(\lambda_j) + \theta'(\lambda_j)(\lambda_{j+1} - \lambda_j) = 0,$$

and thus $\theta(\lambda_{j+1}) > 0$. Convexity also implies that

$$\theta'(\lambda_*)(\lambda_j - \lambda_*) = \theta(\lambda_*) + \theta'(\lambda_*)(\lambda_j - \lambda_*) \geq \theta(\lambda_j), \tag{A.3}$$

in which case

$$\lambda_* - \lambda_{j+1} = \lambda_* - \lambda_j + \frac{\theta(\lambda_j)}{\theta'(\lambda_j)} \leq (\lambda_* - \lambda_j)\left(1 - \frac{\theta'(\lambda_*)}{\theta'(\lambda_j)}\right) \leq \gamma^{\text{N}}(\lambda_* - \lambda_j), \tag{A.4}$$

which establishes both that $\{\lambda_j\}$ converges to λ_* and that the convergence is at least linear. Ultimate superlinear convergence follows from (A.4) since $\theta'(\lambda_j) \to \theta'(\lambda_*)$, while quadratic convergence for Lipschitz continuous θ' follows since $\theta'(\lambda_*) < 0$ [37, Theorem 10.2.2].

(ii) By induction, suppose that $\lambda_{j-1} < \lambda_j$ and $\theta(\lambda_j) > 0$ (in which case $\theta(\lambda_{j-1}) > \theta(\lambda_j)$). Then it follows directly from (A.2) that $\lambda_{j+1} > \lambda_j$. Thus, the convexity of θ and (A.2) imply that

$$\theta(\lambda_{j+1}) \geq \theta(\lambda_j) + \frac{\lambda_{j+1} - \lambda_j}{\lambda_{j-1} - \lambda_j}(\theta(\lambda_{j-1}) - \theta(\lambda_j)) = 0.$$

Furthermore, the mean-value theorem implies that $\theta(\lambda_j) - \theta(\lambda_{j-1}) = \theta'(\xi_j)(\lambda_j - \lambda_{j-1})$ for some $\xi_j \in (\lambda_{j-1}, \lambda_j)$, and thus from (A.2)

$$\lambda_{j+1} = \lambda_j - \frac{\theta(\lambda_j)}{\theta'(\xi_j)}. \tag{A.5}$$

Thus, using (A.3) and (A.5),

$$\lambda_* - \lambda_{j+1} = \lambda_* - \lambda_j + \frac{\theta(\lambda_j)}{\theta'(\xi_j)} \leq (\lambda_* - \lambda_j)\left(1 - \frac{\theta'(\lambda_*)}{\theta'(\xi_j)}\right) \leq \gamma^{\text{N}}(\lambda_* - \lambda_j), \tag{A.6}$$

once again establishing both that $\{\lambda_j\}$ converges to λ_* and that the convergence is at least linear. Ultimate superlinear convergence follows from (A.6) since $\theta'(\xi_j) \to \theta'(\lambda_*)$; a more precise estimate of the Q-rate may be established if θ' is Lipschitz continuous [37, Theorem 11.2.8]. □

References

1. Arora, N., Biegler, L.T.: A trust-region SQP algorithm for equality constrained parameter estimation with simple parameter bounds. Comput. Optim. Appl. **28**(1), 51–86 (2004)
2. Biegler, L.T., Ghattas, O., Heinkenschloss, M., van Bloemen Waanders, B. (eds.): Large-Scale PDE-Constrained Optimization. Lecture Notes in Computational Science and Engineering, vol. 30. Springer, Berlin (2003)
3. Björck, Å.: Numerical Methods for Least Squares Problems. SIAM, Philadelphia (1996)
4. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
5. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Adaptive cubic overestimation methods for unconstrained optimization. Technical Report RAL-TR-2007-007, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England (2007)
6. Chan, T.F., Olkin, J.A., Cooley, D.W.: Solving quadratically constrained least squares using black box solvers. BIT **32**(3), 481–495 (1992)
7. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. SIAM, Philadelphia (2000)
8. Dennis, J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs (1983). Reprinted as Classics in Applied Mathematics 16, SIAM, Philadelphia, 1996
9. Eldén, L.: Algorithms for the regularization of ill-conditioned least squares problems. BIT **17**(2), 134–145 (1977)
10. Erway, J.B., Gill, P.E., Griffin, J.D.: Iterative methods for finding a trust-region step. Technical Report NA 07-2, Department of Mathematics, University of California, San Diego (2007)
11. Gander, W.: Least squares with a quadratic constraint. Numer. Math. **36**(3), 291–307 (1981)

12. Golub, G.H., Kahan, W.: Calculating the singular values and pseudo-inverse of a matrix. SIAM J. Numer. Anal. **2**(2), 205–224 (1965)
13. Golub, G.H., von Matt, U.: Quadratically constrained least squares and quadratic problems. Numer. Math. **59**(1), 561–580 (1991)
14. Gould, N.I.M., Lucidi, S., Roma, M., Toint, Ph.L.: Solving the trust-region subproblem using the Lanczos method. SIAM J. Optim. **9**(2), 504–525 (1999)
15. Gould, N.I.M., Orban, D., Toint, Ph.L.: GALAHAD—a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. ACM Trans. Math. Softw. **29**(4), 353–372 (2003)
16. Gould, N.I.M., Orban, D., Toint, Ph.L.: CUTEr (and SifDec), a constrained and unconstrained testing environment, revisited. ACM Trans. Math. Softw. **29**(4), 373–394 (2003)
17. Gould, N.I.M., Toint, Ph.L.: Nonlinear programming without a penalty function or a filter. Technical Report RAL-TR-2007-016, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England (2007)
18. Gould, N.I.M., Toint, Ph.L.: Regularised-funnel methods for nonlinear programming. Technical Report, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England (2009, in preparation)
19. Hager, W.W.: Minimizing a quadratic over a sphere. SIAM J. Optim. **12**(1), 188–208 (2001)
20. Hager, W.W., Park, S.: Global convergence of SSM for minimizing a quadratic over a sphere. Math. Comput. **74**(251), 1413–1423 (2004)
21. Hansen, P.C.: Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion. Monographs on Mathematical Modeling and Computation, vol. 4. SIAM, Philadelphia (1997)
22. Hebden, M.D.: An algorithm for minimization using exact second derivatives. Technical Report T.P. 515, AERE Harwell Laboratory, Harwell, Oxfordshire, England (1973)
23. Henrici, P.: Elements of Numerical Analysis. Wiley, New York (1964)
24. Nielsen J, C.P., Bunch, R., Sorensen, D.C.: Rank-one modification of the symmetric eigenproblem. Numer. Math. **31**(1), 31–48 (1978)
25. Lalee, M., Nocedal, J., Plantenga, T.: On the implementation of an algorithm for large-scale equality constrained optimization. SIAM J. Optim. **8**(3), 682–706 (1998)
26. Levenberg, K.: A method for the solution of certain problems in least squares. Q. Appl. Math. **2**(2), 164–168 (1944)
27. Lukšan, L.: Inexact trust region method for large sparse nonlinear least-squares. Kybernetica **29**(4), 305–324 (1993)
28. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11**(2), 431–441 (1963)
29. Melman, A.: Numerical solution of a secular equation. Numer. Math. **69**(4), 483–493 (1995)
30. Moré, J.J.: The Levenberg-Marquardt algorithm: implementation and theory. In: Watson, G.A. (ed.) Numerical Analysis, Dundee 1977. Lecture Notes in Mathematics, vol. 630, pp. 105–116. Springer, Berlin (1978)
31. Morrison, D.D.: Methods for nonlinear least squares problems and convergence proofs. In: Lorell, J., Yagi, F. (eds.) Proceedings of the Seminar on Tracking Programs and Orbit Determination, pp. 1–9. Jet Propulsion Laboratory, Pasadena (1960)
32. Nesterov, Yu.: Modified Gauss-Newton scheme with worst-case guarantees for global performance. Optim. Methods Softw. **22**(3), 469–483 (2007)
33. Nesterov, Yu., Polyak, B.T.: Cubic regularization of Newton method and its global performance. Math. Program. **108**(1), 177–205 (2006)
34. Neumaier, A.: Solving ill-conditioned and singular linear systems: a tutorial on regularization. SIAM Rev. **40**(3), 636–666 (1998)
35. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Series in Operations Research. Springer, Heidelberg (2006)
36. Omojokun, E.O.: Trust region algorithms for optimization with nonlinear equality and inequality constraints. Ph.D. Thesis, University of Colorado, Boulder, Colorado, USA (1989)
37. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, San Diego (1970)
38. Paige, C.C., Saunders, M.A.: ALGORITHM 583: LSQR: an algorithm for sparse linear equations and sparse least squares. ACM Trans. Math. Softw. **8**(2), 195–209 (1982)
39. Paige, C.C., Saunders, M.A.: LSQR: an algorithm for sparse linear equations and sparse least squares. ACM Trans. Math. Softw. **8**(1), 43–71 (1982)
40. Reinsch, C.: Smoothing by spline functions II. Numer. Math. **16**(5), 451–454 (1971)
41. Rojas, M.: A large-scale trust-region approach to the regularization of discrete ill-posed problems. Ph.D. Thesis, Rice University, Houston, Texas, USA (1998)

42. Rojas, M., Sorensen, D.C.: A trust-region approach to the regularization of large-scale discrete forms of ill-posed problems. SIAM J. Sci. Stat. Comput. **23**(6), 1842–1860 (2002)
43. Steihaug, T.: The conjugate gradient method and trust regions in large scale optimization. SIAM J. Numer. Anal. **20**(3), 626–637 (1983)
44. Toint, Ph.L.: Towards an efficient sparsity exploiting Newton method for minimization. In: Duff, I.S. (ed.) Sparse Matrices and Their Uses, pp. 57–88.1 Academic Press, London (1981)
45. Yuan, Y.: On the truncated conjugate-gradient method. Math. Program. Ser. A **87**(3), 561–573 (2000)