LINE-SEARCH AND TRUST-REGION EQUATIONS FOR A PRIMAL-DUAL INTERIOR METHOD FOR NONLINEAR OPTIMIZATION

Philip E. Gill* Vyacheslav Kungurtsev† Daniel P. Robinson‡

UCSD Center for Computational Mathematics
Technical Report CCoM-21-4
September 1, 2021

Abstract

The approximate Newton equations for a minimizing a shifted primal-dual penalty-barrier method are derived for a nonlinearly constrained problem in general form. These equations may be used in conjunction with either a line-search or trust-region method to force convergence from an arbitrary starting point. It is shown that under certain conditions, the approximate Newton equations are equivalent to a regularized form of the conventional primal-dual path-following equations.

Key words. Nonlinear programming, nonlinear constraints, shifted penalty-barrier methods, augmented Lagrangian methods, primal-dual interior methods, path-following methods, regularized methods.

AMS subject classifications. 49J20, 49J15, 49M37, 49D37, 65F05, 65K05, 90C30

*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112 (pgill@ucsd.edu). Research supported in part by National Science Foundation grants DMS-1318480 and DMS-1361421. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

†Agent Technology Center, Department of Computer Science, Faculty of Electrical Engineering, Czech Technical University in Prague. (vyacheslav.kungurtsev@fel.cvut.cz) Research supported by the OP VVV project CZ.02.1.01/0.0/0.0/16 019/0000765 "Research Center for Informatics".

‡Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015 (dpr219@lehigh.edu). Research supported in part by National Science Foundation grant DMS-1217153. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

1. Introduction

This note concerns that derivation of the line-search and trust-region equations for a shifted primal-dual penalty-barrier merit method for constrained optimization. These methods are intended for the minimization of a twice-continuously differentiable function subject to both equality and inequality constraints that may include a set of twice-continuously differentiable constraint functions. A description of the line-search and trust-region methods for a problem with nonlinear inequality constraints is given by Gill, Kungurtsev and Robinson [4] and Gill, Kungurtsev and Robinson [5]. The note concerns the formulation of the equations for problems written in the general form:

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad \begin{cases} c(x) - s = 0, & L_X s = h_X, \quad \ell^S \le L_L s, \quad L_U s \le u^S, \\ Ax - b = 0, & E_X x = b_X, \quad \ell^X \le E_L x, \quad E_U x \le u^X, \end{cases} \tag{NLP}$$

where A denotes a constant $m_A \times n$ matrix, and b, h_X, b_X, ℓ^S, u^S, ℓ^X and u^X are fixed vectors of dimension m_A, m_X, n_X, m_L, m_U, n_L and n_U, respectively. Similarly, L_X, L_L and L_U denote fixed matrices of dimension $m_X \times m$, $m_L \times m$ and $m_U \times m$, respectively, and E_X, E_L and E_U are fixed matrices of dimension $n_X \times n$, $n_L \times n$ and $n_U \times n$, respectively. Throughout the discussion, the functions $c : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ are assumed to be twice-continuously differentiable. The components of s may be interpreted as slack variables associated with the nonlinear constraints.

The quantity E_X denotes an $n_X \times n$ matrix formed from n_X independent rows of I_n, the identity matrix of order n. This implies that the equality constraints $E_X x = b_X$ fix n_X components of x at the corresponding values of b_X. Similarly, E_L and E_U denote $n_L \times n$ and $n_U \times n$ matrices formed from subsets of rows of I_n such that $E_X^T E_L = 0$, $E_X^T E_U = 0$, i.e., a variable is either fixed or free to move, possibly bounded by an upper or lower bound. Note that an x_j may be an unrestricted variable in the sense that it is neither fixed nor subject to an upper or lower bound, in which case e_j^T, the jth row of I_n, is not a row of E_X, E_L or E_U. Analogous definitions hold for L_X, L_L and L_U as subsets of rows of I_m. However, we impose the restriction that a given s_j must be either fixed or restricted by an upper or lower bound, i.e., there are no unrestricted slacks[1]. Let E_F denote the matrix of rows of I_n that are not rows of E_X, and let L_F denote the matrix of rows of I_m that are not rows of L_X. If $n_F = n - n_X$ and $m_F = m - m_X$, then E_F and L_F are $n_F \times n$ and $m_F \times m$ respectively. Note that $n_L + n_U$ may be less than n_F, but m_F must equal $m_L + m_U$. The matrices $\begin{pmatrix} E_X^T & E_F^T \end{pmatrix}$ and $\begin{pmatrix} L_X^T & L_F^T \end{pmatrix}$ are column permutations of I_n and I_m. Moreover, there are $n \times n$ and $m \times m$ permutation matrices P_x and P_s such that

$$P_x = \begin{pmatrix} E_F \\ E_X \end{pmatrix} \quad \text{and} \quad P_s = \begin{pmatrix} L_F \\ L_X \end{pmatrix}, \tag{1.1}$$

with $E_F E_F^T = I_F^x$, $E_X E_X^T = I_X^x$, and $E_F E_X^T = 0$, and $L_F L_F^T = I_F^s$, $L_X L_X^T = I_X^s$, and $L_F L_X^T = 0$.

All general inequality constraints are imposed indirectly using a shifted primal-dual barrier function. The general equality constraints $c(x) - s = 0$ and $Ax = b$ are enforced using an primal-dual augmented Lagrangian algorithm, which implies that the

[1]This is not a significant restriction because a "free" slack is equivalent to a unrestricted nonlinear constraint, which may be discarded from the problem. The shifted primal-dual penalty-barrier equations can be derived without this restriction, but the derivation is beyond the scope of this note.

equalities are satisfied in the limit. The exception to this is when the constraints $E_X x = b_X$, and $L_X s = h_X$ are used to fix a subset of the variables and slacks. These bounds are enforced at every iterate.

An equality constraint $c_i(x) = 0$ may be handled by introducing the slack variable s_i and writing the constraint as the two constraints $c_i(x) - s_i = 0$ and $s_i = 0$. In this case the ith coordinate vector e_i can be included as a row of L_X. Linear *inequality* constraints must be included as part of c. A linear equality constraint can be either included with the nonlinear equality constraints or the matrix A. The constraints involving A may be used to temporarily fix a subset of the variables at their bounds without altering the underlying structure of the approximate Newton equations. In this case, A and b have the form

$$A = \begin{pmatrix} A_L \\ -A_U \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} \ell_A \\ -u_A \end{pmatrix},$$

where A_L and A_U are rows of the identity matrix and ℓ_A and u_A are the associated vectors of temporarily fixed lower and upper bounds (see Gill, Kungurtsev and Robinson [4] for more details).

The optimality conditions for problem (NLP) are given in Section 2. The shifted path-following equations are formulated in Section 3. The shifted primal-dual penalty-barrier function associated with problem is discussed in Section 4. This function serves as a merit function for both the line-search and trust-region method. The equations for a line-search modified Newton method are formulated in Sections 5 and 6, and summarized in Section 7. The analogous equations for the trust-region method are derived in Section 8 and summarized in Section 9.

Notation. Given vectors x and y, the vector consisting of x augmented by y is denoted by (x, y). The subscript i is appended to vectors to denote the ith component of that vector. Given vectors a and b with the same dimension, the vector with ith component $a_i b_i$ is denoted by $a \cdot b$. Similarly, $\min(a, b)$ is a vector with components $\min(a_i, b_i)$. The vector e denotes the column vector of ones, and I denotes the identity matrix. The dimensions of e and I are defined by the context. The vector two-norm or its induced matrix norm are denoted by $\|\cdot\|$. For brevity, in some equations the vector $g(x)$ is used to denote $\nabla f(x)$, the gradient of $f(x)$. The matrix $J(x)$ denotes the $m \times n$ constraint Jacobian, which has ith row $\nabla c_i(x)^T$. Given a Lagrangian function $L(x, y) = f(x) - c(x)^T y$ with y a m-vector of dual variables, the Hessian of the Lagrangian with respect to x is denoted by $H(x, y) = \nabla^2 f(x) - \sum_{i=1}^m y_i \nabla^2 c_i(x)$. Both the line-search and trust-region equations utilize the Moore-Penrose pseudoinverse of a diagonal matrix. In particular, if $D = \text{diag}(d_1, d_2, \ldots, d_n)$, then the pseudoinverse $D^\dagger$ is diagonal with $D^\dagger_{ii} = 0$ for $d_i = 0$ and $D^\dagger_{ii} = 1/d_i$ for $d_i \neq 0$.

2. Optimality conditions

The first-order KKT conditions for problem (NLP) are

$$\left.\begin{array}{rrr}
\nabla f(x^*) - J(x^*)^T y^* - A^T v^* - E_X^T z_X^* - E_L^T z_1^* + E_U^T z_2^* = 0, & z_1^* \geq 0, & z_2^* \geq 0, \\
y^* - L_X^T w_X^* - L_L^T w_1^* + L_U^T w_2^* = 0, & w_1^* \geq 0, & w_2^* \geq 0, \\
c(x^*) - s^* = 0, & & L_X s^* - h_X = 0, \\
& Ax^* - b = 0, & E_X x^* - b_X = 0, \\
E_L x^* - \ell^X \geq 0, & u^X - E_U x^* \geq 0, & \\
L_L s^* - \ell^S \geq 0, & u^S - L_U s^* \geq 0, & \\
z_1^* \cdot (E_L x^* - \ell^X) = 0, & z_2^* \cdot (u^X - E_U x^*) = 0, & \\
w_1^* \cdot (L_L s^* - \ell^S) = 0, & w_2^* \cdot (u^S - L_U s^*) = 0, &
\end{array}\right\} \tag{2.1}$$

where y^*, w_X^*, and z_X^* are the multipliers for the equality constraints $c(x) - s = 0$, $L_X s^* = h_X$ and $E_X x^* = b_X$, and z_1^*, z_2^*, w_1^* and w_2^* may be interpreted as the Lagrange multipliers for the inequality constraints $E_L x - \ell^X \geq 0$, $u^X - E_U x \geq 0$, $L_L s - \ell^S \geq 0$ and $u^S - L_U s \geq 0$, respectively. The components of v^* are the multipliers for the linear equality constraints $Ax = b$.

The discussion that follows makes extensive use of the auxiliary quantities

$$x_1 = E_L x - \ell^X, \quad x_2 = u^X - E_U x, \quad s_1 = L_L s - \ell^S, \quad \text{and} \quad s_2 = u^S - L_U s. \tag{2.2}$$

In some cases x_1, x_2, s_1 and s_2 are used to simplify the appearance of certain equations, in others they are regarded as independent variables associated with the problem

$$\left.\begin{array}{lllll}
\underset{x,x_1,x_2,s,s_1,s_2}{\text{minimize}} & f(x) & & & \\
\text{subject to} & c(x) - s = 0, & Ax - b = 0, & & \\
& E_L x - x_1 = \ell^X, & L_L s - s_1 = \ell^S, & x_1 \geq 0, & s_1 \geq 0, \\
& E_U x + x_2 = u^X, & L_U s + s_2 = u^S, & x_2 \geq 0, & s_2 \geq 0, \\
& E_X x - b_X = 0, & L_X s - h_X = 0, & &
\end{array}\right\} \tag{NP}$$

which is equivalent to problem (NLP). In this case, the dual variables z_1^*, z_2^*, w_1^*, and w_2^* associated with the optimality conditions (2.1) are the Lagrange multipliers for the inequality constraints $x_1 \geq 0$, $x_2 \geq 0$, $s_1 \geq 0$, and $s_2 \geq 0$, respectively.

In the derivations that follow, the vectors z and w are defined as

$$z = E_X^T z_X + E_L^T z_1 - E_U^T z_2, \quad \text{and} \quad w = L_X^T w_X + L_L^T w_1 - L_U^T w_2. \tag{2.3}$$

3. The path-following equations

Penalty and barrier methods are closely related to path-following methods. These methods follow a continuous path that passes through a solution of (NLP). In the simplest case, the path is parameterized by a positive scalar parameter that serves as both a perturbation of the equality constraints and a perturbation of the complementarity conditions associated with the optimality conditions for problem (NLP). In Gill, Kungurtsev and Robinson [4], the perturbations involve estimates of the Lagrange multipliers for the equality and inequality constraints.

Let z_1^E and z_2^E, w_1^E and w_2^E denote nonnegative estimates of z_1^* and z_2^*, w_1^* and w_2^*. Given small positive scalars μ^P, μ^A and μ^B, consider the perturbed optimality conditions

$$\left.\begin{array}{rrr}
\nabla f(x) - J(x)^{\mathrm{T}} y - A^{\mathrm{T}} v - E_X^{\mathrm{T}} z_X - E_L^{\mathrm{T}} z_1 + E_U^{\mathrm{T}} z_2 = 0, & z_1 \ge 0, & z_2 \ge 0,\\
y - L_X^{\mathrm{T}} w_X - L_L^{\mathrm{T}} w_1 + L_U^{\mathrm{T}} w_2 = 0, & w_1 \ge 0, & w_2 \ge 0,\\
c(x) - s = \mu^P (y^E - y), & E_X x - b_X = 0, & L_X s - h_X = 0,\\
Ax - b = \mu^A (v^E - v), & & \\
E_L x - \ell^X \ge 0, & u^X - E_U x \ge 0, & \\
L_L s - \ell^S \ge 0, & u^S - L_U s \ge 0, & \\
z_1 \cdot (E_L x - \ell^X) = \mu^B (z_1^E - z_1), & z_2 \cdot (u^X - E_U x) = \mu^B (z_2^E - z_2), & \\
w_1 \cdot (L_L s - \ell^S) = \mu^B (w_1^E - w_1), & w_2 \cdot (u^S - L_U s) = \mu^B (w_2^E - w_2). &
\end{array}\right\} \qquad (3.1)$$

Let v_P denote the vector of variables $v_P = (x,\, s,\, y,\, v,\, w_X,\, z_X,\, z_1,\, z_2,\, w_1,\, w_2)$. The primal-dual path-following equations are given by $F(v_P) = 0$, with

$$F(v_P) = \begin{pmatrix}
\nabla f(x) - J(x)^{\mathrm{T}} y - A^{\mathrm{T}} v - E_X^{\mathrm{T}} z_X - E_L^{\mathrm{T}} z_1 + E_U^{\mathrm{T}} z_2\\
y - L_X^{\mathrm{T}} w_X - L_L^{\mathrm{T}} w_1 + L_U^{\mathrm{T}} w_2\\
c(x) - s + \mu^P (y - y^E)\\
Ax - b + \mu^A (v - v^E)\\
E_X x - b_X\\
L_X s - h_X\\
z_1 \cdot (E_L x - \ell^X) + \mu^B (z_1 - z_1^E)\\
z_2 \cdot (u^X - E_U x) + \mu^B (z_2 - z_2^E)\\
w_1 \cdot (L_L s - \ell^S) + \mu^B (w_1 - w_1^E)\\
w_2 \cdot (u^S - L_U s) + \mu^B (w_2 - w_2^E)
\end{pmatrix} = \begin{pmatrix}
\nabla f(x) - J(x)^{\mathrm{T}} y - A^{\mathrm{T}} v - z\\
y - w\\
c(x) - s + \mu^P (y - y^E)\\
Ax - b + \mu^A (v - v^E)\\
E_X x - b_X\\
L_X s - h_X\\
z_1 \cdot (E_L x - \ell^X) + \mu^B (z_1 - z_1^E)\\
z_2 \cdot (u^X - E_U x) + \mu^B (z_2 - z_2^E)\\
w_1 \cdot (L_L s - \ell^S) + \mu^B (w_1 - w_1^E)\\
w_2 \cdot (u^S - L_U s) + \mu^B (w_2 - w_2^E)
\end{pmatrix}, \qquad (3.2)$$

where the first $n+m$ equations are written in terms of z and w such that $z = E_X^{\mathrm{T}} z_X + E_L^{\mathrm{T}} z_1 - E_U^{\mathrm{T}} z_2$ and $w = L_X^{\mathrm{T}} w_X + L_L^{\mathrm{T}} w_1 - L_U^{\mathrm{T}} w_2$. (To simplify the notation, the dependence of F on the parameters μ^A, μ^P, μ^B, y^E, v^E, z_1^E, z_2^E, w_1^E, w_2^E is omitted.) Any zero (x,

s, y, v, w_X, z_X, z_1, z_2, w_1, w_2) of F such that $\ell^X < E_L x$, $E_U x < u^X$, $\ell^S < L_L s$, $L_U s < u^S$, $z_1 > 0$, $z_2 > 0$, $w_1 > 0$, and $w_2 > 0$ approximates a point satisfying the optimality conditions (2.1), with the approximation becoming increasingly accurate as the terms $\mu^P(y - y^E)$, $\mu^A(v - v^E)$, $\mu^B(z_1 - z_1^E)$, $\mu^B(z_2 - z_2^E)$, $\mu^B(w_1 - w_1^E)$ and $\mu^B(w_2 - w_2^E)$ approach zero. For any sequence of z_1^E, z_2^E, w_1^E, w_2^E, v^E and y^E such that $z_1^E \to z_1^*$, $z_2^E \to z_2^*$, $w_1^E \to w_1^*$, $w_2^E \to w_2^*$, $v^E \to v^*$ and $y^E \to y^*$, and it must hold that solutions $(x, s, y, v, z_1, z_2, w_1, w_2)$ of (3.1) must satisfy $z_1 \cdot (x - \ell^X) \to 0$, $z_2 \cdot (u^X - x) \to 0$, $w_1 \cdot (s - \ell^S) \to 0$, and $w_2 \cdot (u^S - s) \to 0$, This implies that any solution $(x, s, y, v, w_X, z_X, z_1, z_2, w_1, w_2)$ of (3.1) will approximate a solution of (2.1) independently of the values of μ^P, μ^A and μ^B (i.e., it is not necessary that $\mu^P \to 0$, $\mu^A \to 0$ and $\mu^B \to 0$).

If $v_P = (x, s, y, v, w_X, z_X, z_1, z_2, w_1, w_2)$ is a given approximate zero of F such that $\ell^X - \mu^B e < E_L x$, $E_U x < u^X + \mu^B e$, $\ell^S - \mu^B e < L_L s$, $L_U s < u^S + \mu^B e$, $z_1 > 0$, $z_2 > 0$, $w_1 > 0$, and $w_2 > 0$, the Newton equations for the change in variables $\Delta v_P = (\Delta x, \Delta s, \Delta y, \Delta v, \Delta w_X, \Delta z_X, \Delta z_1, \Delta z_2, \Delta w_1, \Delta w_2)$ are given by $F'(v_P)\Delta v_P = -F(v_P)$, with

$$F'(v_P) = \begin{pmatrix} H(x,y) & 0 & -J(x)^T & -A^T & 0 & -E_X^T & -E_L^T & E_U^T & 0 & 0 \\ 0 & 0 & I_m & 0 & -L_X^T & 0 & 0 & 0 & -L_L^T & L_U^T \\ J(x) & -I_m & D_Y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A & 0 & 0 & D_A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Z_1 E_L & 0 & 0 & 0 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 E_U & 0 & 0 & 0 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 L_L & 0 & 0 & 0 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 L_U & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix}, \tag{3.3}$$

where $X_1^\mu = \mathrm{diag}(x_1 + \mu^B e)$, $X_2^\mu = \mathrm{diag}(x_2 + \mu^B e)$, $S_1^\mu = \mathrm{diag}(s_1 + \mu^B e)$, $S_2^\mu = \mathrm{diag}(s_2 + \mu^B e)$, $Z_1 = \mathrm{diag}(z_1)$, $Z_2 = \mathrm{diag}(z_2)$, $W_1 = \mathrm{diag}(w_1)$ and $W_2 = \mathrm{diag}(w_2)$, with x_1, x_2, s_1 and s_2 given by (2.2). Any s may be written as $s = L_F^T s_F + L_X^T s_X$, where s_F and s_X denote the components of s corresponding to the "free" and "fixed" components of s, respectively. Similarly, any x may be written as $x = E_F^T x_F + E_X^T x_X$, where x_F and x_X denote the free and fixed components of x.

The partition of x into free and fixed variables induces a partition of $H(x, y)$, A, $J(x)$, E_L and E_U. We use H_F to denote the $n_F \times n_F$ symmetric matrix of rows and columns of H associated with the free variables and A_F, A_X, J_F, J_X to denote the free and fixed columns of A and $J(x)$. In particular,

$$H_F = E_F H(x,y) E_F^T, \qquad A_F = A E_F^T, \qquad A_X = A E_X^T, \qquad J_F = J(x) E_F^T, \quad \text{and} \quad J_X = J(x) E_X^T,$$

Similarly, the $n_L \times n_F$ matrix E_{LF} and $n_U \times n_F$ matrix E_{UF} comprise the free columns of E_L and E_U, with

$$E_{LF} = E_L E_F^T, \quad \text{and} \quad E_{UF} = E_U E_F^T.$$

It follows that the components of $E_{LF} x_F$ are the values of the free variables that are subject to lower bounds. A similar interpretation applied for $E_{UF} x_F$. Analogous definitions apply for the $m_L \times m_F$ matrix L_{LF} and $m_U \times m_F$ matrix L_{UF}.

The next step is to transform the path-following equations to reflect the structure of free and fixed variables. Consider the block-diagonal orthogonal matrix $Q = \operatorname{diag}(P_{\scriptscriptstyle X},\, P_{\scriptscriptstyle S},\, I_m,\, I_{\scriptscriptstyle A},\, I_X^s,\, I_X^x,\, I_L^x,\, I_U^x,\, I_L^s,\, I_U^s)$, where $P_{\scriptscriptstyle X}$ and $P_{\scriptscriptstyle S}$ are defined in (1.1). Given the identities

$$\begin{pmatrix} \Delta x_F \\ \Delta x_X \end{pmatrix} = P_x \Delta x = \begin{pmatrix} E_F \Delta x \\ E_X \Delta x \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \Delta s_F \\ \Delta s_X \end{pmatrix} = P_S \Delta s = \begin{pmatrix} L_F \Delta s \\ L_X \Delta s \end{pmatrix},$$

and $QF'(v_P)Q^{\mathrm{T}}Q\Delta v_P = -QF(v_P)$, we obtain the transformed equations

$$\begin{pmatrix}
H_F & H_O^{\mathrm{T}} & 0 & 0 & -J_F^{\mathrm{T}} & -A_F^{\mathrm{T}} & 0 & 0 & -E_{LF}^{\mathrm{T}} & E_{UF}^{\mathrm{T}} & 0 & 0 \\
H_O & H_X & 0 & 0 & -J_X^{\mathrm{T}} & -A_X^{\mathrm{T}} & 0 & -I_X^x & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & L_F & 0 & 0 & 0 & 0 & 0 & -L_{LF}^{\mathrm{T}} & L_{UF}^{\mathrm{T}} \\
0 & 0 & 0 & 0 & L_X & 0 & -I_X^s & 0 & 0 & 0 & 0 & 0 \\
J_F & J_X & -L_F^{\mathrm{T}} & -L_X^{\mathrm{T}} & D_Y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
A_F & A_X & 0 & 0 & 0 & D_A & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & I_X^s & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & I_X^x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
Z_1 E_{LF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_1^{\mu} & 0 & 0 & 0 \\
-Z_2 E_{UF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_2^{\mu} & 0 & 0 \\
0 & 0 & W_1 L_{LF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_1^{\mu} & 0 \\
0 & 0 & -W_2 L_{UF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_2^{\mu}
\end{pmatrix}
\begin{pmatrix}
\Delta x_F \\ \Delta x_X \\ \Delta s_F \\ \Delta s_X \\ \Delta y \\ \Delta v \\ \Delta w_X \\ \Delta z_X \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2
\end{pmatrix}
= -\begin{pmatrix}
g_F - J_F^{\mathrm{T}} y - A_F^{\mathrm{T}} v - z_F \\
g_X - J_X^{\mathrm{T}} y - A_X^{\mathrm{T}} v - z_X \\
y_F - w_F \\
y_X - w_X \\
c(x) - s + \mu^P (y - y^E) \\
Ax - b + \mu^A (v - v^E) \\
E_X x - b_X \\
L_X s - h_X \\
z_1 \cdot (E_L x - \ell^X) + \mu^B (z_1 - z_1^E) \\
z_2 \cdot (u^X - E_U x) + \mu^B (z_2 - z_2^E) \\
w_1 \cdot (L_L s - \ell^S) + \mu^B (w_1 - w_1^E) \\
w_2 \cdot (u^S - L_U s) + \mu^B (w_2 - w_2^E)
\end{pmatrix},$$

where $H_O = E_F H(x,y) E_X^{\mathrm{T}}$, $H_X = E_X H(x,y) E_X^{\mathrm{T}}$, $g_F = E_F g$, $z_F = E_F z$ and $y_F = L_F y$.

As the constraints $L_X s - h_X = 0$ and $E_X x - b_X = 0$ are enforced throughout, it follows that $\Delta s_X = 0$ and $\Delta x_X = 0$, in which case Δs and Δx satisfy

$$\Delta s = L_F^T \Delta s_F + L_X^T \Delta s_X = L_F^T \Delta s_F \quad \text{and} \quad \Delta x = E_F^T \Delta x_F + E_X^T \Delta x_X = E_F^T \Delta x_F.$$

After scaling the last four blocks of equations by (respectively) Z_1^{-1}, Z_2^{-1}, W_1^{-1} and W_2^{-1}, collecting terms and reordering the equations and unknowns, we obtain

$$\begin{pmatrix} H_F & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - z_F \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}, \tag{3.4}$$

where

$$\left. \begin{aligned} D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c - s), & D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax - b), \\ D_1^W &= S_1^\mu W_1^{-1}, & \pi_1^W &= \mu^B (S_1^\mu)^{-1} w_1^E, & D_1^Z &= X_1^\mu Z_1^{-1}, & \pi_1^Z &= \mu^B (X_1^\mu)^{-1} z_1^E, \\ D_2^W &= S_2^\mu W_2^{-1}, & \pi_2^W &= \mu^B (S_2^\mu)^{-1} w_2^E, & D_2^Z &= X_2^\mu Z_2^{-1}, & \pi_2^Z &= \mu^B (X_2^\mu)^{-1} z_2^E. \end{aligned} \right\} \tag{3.5}$$

Given the definitions (2.3), the vectors Δs and Δw_X are recovered as $\Delta s = L_F^T \Delta s_F$ and $\Delta w_X = [\, y + \Delta y - w\,]_X$. Similarly, Δx and Δz_X are recovered as $\Delta x = E_F^T \Delta x_F$ and $\Delta z_X = [\, \nabla f(x) + H(x,y)\Delta x - J(x)^T(y + \Delta y) - A^T(v + \Delta v) - z\,]_X$.

4. A shifted primal-dual penalty-barrier function

Consider the shifted primal-dual penalty-barrier problem applied to (NP):

$$\left. \begin{aligned} \underset{\substack{x,x_1,x_2,s,s_1,s_2,\\ y,v,z_1,z_2,w_1,w_2}}{\text{minimize}} \quad & M(x, x_1, x_2, s, s_1, s_2, y, v, w_1, w_2\,; \mu^P, \mu^B, y^E, v^E, w_1^E, w_2^E) \\ \text{subject to} \quad & E_L x - x_1 = \ell^X, \quad L_L s - s_1 = \ell^S, \;\; x_1 + \mu^B e > 0, \;\; z_1 > 0, \;\; s_1 + \mu^B e > 0, \;\; w_1 > 0, \\ & E_U x + x_2 = u^X, \quad L_U s + s_2 = u^S, \;\; x_2 + \mu^B e > 0, \;\; z_2 > 0, \;\; s_2 + \mu^B e > 0, \;\; w_2 > 0, \\ & E_X x - b_X = 0, \quad L_X s - h_X = 0, \end{aligned} \right\} \tag{4.1}$$

where $M(x,\, x_1,\, x_2,\, s,\, s_1,\, s_2,\, y,\, v,\, z_1,\, z_2,\, w_1,\, w_2\,; \mu^P,\, \mu^B,\, y^E,\, v^E,\, z_1^E,\, z_2^E,\, w_1^E,\, w_2^E)$ is the shifted primal-dual penalty-barrier function

$$
\begin{aligned}
f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\
- (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A(v - v^E)\|^2 \\
- \sum_{j=1}^{n_L}\Big\{\mu^B[\, z_1^E\,]_j \ln\big([\, z_1\,]_j[\, x_1 + \mu^B e\,]_j^2\big) - [\, z_1 \cdot (x_1 + \mu^B e)\,]_j\Big\} \\
- \sum_{j=1}^{n_U}\Big\{\mu^B[\, z_2^E\,]_j \ln\big([\, z_2\,]_j[\, x_2 + \mu^B e\,]_j^2\big) - [\, z_2 \cdot (x_2 + \mu^B e)\,]_j\Big\} \\
- \sum_{i=1}^{m_L}\Big\{\mu^B[\, w_1^E\,]_i \ln\big([\, w_1\,]_i[\, s_1 + \mu^B e\,]_i^2\big) - [\, w_1 \cdot (s_1 + \mu^B e)\,]_i\Big\} \\
- \sum_{i=1}^{m_U}\Big\{\mu^B[\, w_2^E\,]_i \ln\big([\, w_2\,]_i[\, s_2 + \mu^B e\,]_i^2\big) - [\, w_2 \cdot (s_2 + \mu^B e)\,]_i\Big\}. \qquad (4.2)
\end{aligned}
$$

The gradient $\nabla M(x,\, x_1,\, x_2,\, s,\, s_1,\, s_2,\, y,\, v,\, z_1,\, z_2,\, w_1,\, w_2)$ may be defined in terms of the quantities $X_1^\mu = \operatorname{diag}(E_L x - \ell^X + \mu^B e)$, $X_2^\mu = \operatorname{diag}(u^X - E_U x + \mu^B e)$, $Z_1 = \operatorname{diag}(z_1)$, $Z_2 = \operatorname{diag}(z_2)$, $W_1 = \operatorname{diag}(w_1)$, $W_2 = \operatorname{diag}(w_2)$, $S_1^\mu = \operatorname{diag}(L_L s - \ell^S + \mu^B e)$ and

$S_2^\mu = \operatorname{diag}(u^{\scriptscriptstyle S} - L_{\scriptscriptstyle U} s + \mu^{\scriptscriptstyle B} e)$, in particular

$$\nabla M = \begin{pmatrix} g - A^{\mathrm{T}}\big(2(v^{\scriptscriptstyle E} + \frac{1}{\mu^A}(Ax-b)) - v\big) - J^{\mathrm{T}}\big(2(y^{\scriptscriptstyle E} - \frac{1}{\mu^P}(c-s)) - y\big) \\ z_1 - 2\mu^{\scriptscriptstyle B}(X_1^\mu)^{-1} z_1^{\scriptscriptstyle E} \\ z_2 - 2\mu^{\scriptscriptstyle B}(X_2^\mu)^{-1} z_2^{\scriptscriptstyle E} \\ 2\big(y^{\scriptscriptstyle E} - \frac{1}{\mu^P}(c-s)\big) - y \\ w_1 - 2\mu^{\scriptscriptstyle B}(S_1^\mu)^{-1} w_1^{\scriptscriptstyle E} \\ w_2 - 2\mu^{\scriptscriptstyle B}(S_2^\mu)^{-1} w_2^{\scriptscriptstyle E} \\ c - s + \mu^{\scriptscriptstyle P}(y - y^{\scriptscriptstyle E}) \\ Ax - b + \mu^{\scriptscriptstyle A}(v - v^{\scriptscriptstyle E}) \\ x_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_1^{-1} z_1^{\scriptscriptstyle E} \\ x_2 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_2^{-1} z_2^{\scriptscriptstyle E} \\ s_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} W_1^{-1} w_1^{\scriptscriptstyle E} \\ s_2 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} W_2^{-1} w_2^{\scriptscriptstyle E} \end{pmatrix}$$

$$= \begin{pmatrix} g - A^{\mathrm{T}}\big(2(v^{\scriptscriptstyle E} + \frac{1}{\mu^A}(Ax-b)) - v\big) - J^{\mathrm{T}}\big(2(y^{\scriptscriptstyle E} - \frac{1}{\mu^P}(c-s)) - y\big) \\ (X_1^\mu)^{-1}\big(z_1 \cdot x_1 + \mu^{\scriptscriptstyle B} z_1^{\scriptscriptstyle E} + \mu^{\scriptscriptstyle B}(z_1 - z_1^{\scriptscriptstyle E})\big) \\ (X_2^\mu)^{-1}\big(z_2 \cdot x_2 + \mu^{\scriptscriptstyle B} z_2^{\scriptscriptstyle E} + \mu^{\scriptscriptstyle B}(z_2 - z_2^{\scriptscriptstyle E})\big) \\ 2\big(y^{\scriptscriptstyle E} - \frac{1}{\mu^P}(c-s)\big) - y \\ (S_1^\mu)^{-1}\big(w_1 \cdot s_1 + \mu^{\scriptscriptstyle B} w_1^{\scriptscriptstyle E} + \mu^{\scriptscriptstyle B}(w_1 - w_1^{\scriptscriptstyle E})\big) \\ (S_2^\mu)^{-1}\big(w_2 \cdot s_2 + \mu^{\scriptscriptstyle B} w_2^{\scriptscriptstyle E} + \mu^{\scriptscriptstyle B}(w_2 - w_2^{\scriptscriptstyle E})\big) \\ c - s + \mu^{\scriptscriptstyle P}(y - y^{\scriptscriptstyle E}) \\ Ax - b + \mu^{\scriptscriptstyle A}(v - v^{\scriptscriptstyle E}) \\ Z_1^{-1}\big(z_1 \cdot x_1 + \mu^{\scriptscriptstyle B}(z_1 - z_1^{\scriptscriptstyle E})\big) \\ Z_2^{-1}\big(z_2 \cdot x_2 + \mu^{\scriptscriptstyle B}(z_2 - z_2^{\scriptscriptstyle E})\big) \\ W_1^{-1}\big(w_1 \cdot s_1 + \mu^{\scriptscriptstyle B}(w_1 - w_1^{\scriptscriptstyle E})\big) \\ W_2^{-1}\big(w_2 \cdot s_2 + \mu^{\scriptscriptstyle B}(w_2 - w_2^{\scriptscriptstyle E})\big) \end{pmatrix} = \begin{pmatrix} g - A^{\mathrm{T}}\big(\pi^{\scriptscriptstyle V} + (\pi^{\scriptscriptstyle V} - v)\big) - J^{\mathrm{T}}\big(\pi^{\scriptscriptstyle Y} + (\pi^{\scriptscriptstyle Y} - y)\big) \\ -\big(\pi_1^{\scriptscriptstyle Z} + (\pi_1^{\scriptscriptstyle Z} - z_1)\big) \\ -\big(\pi_2^{\scriptscriptstyle Z} + (\pi_2^{\scriptscriptstyle Z} - z_2)\big) \\ \pi^{\scriptscriptstyle Y} + (\pi^{\scriptscriptstyle Y} - y) \\ -\big(\pi_1^{\scriptscriptstyle W} + (\pi_1^{\scriptscriptstyle W} - w_1)\big) \\ -\big(\pi_2^{\scriptscriptstyle W} + (\pi_2^{\scriptscriptstyle W} - w_2)\big) \\ -D_{\scriptscriptstyle Y}(\pi^{\scriptscriptstyle Y} - y) \\ -D_{\scriptscriptstyle A}(\pi^{\scriptscriptstyle V} - v) \\ -D_1^{\scriptscriptstyle Z}(\pi_1^{\scriptscriptstyle Z} - z_1) \\ -D_2^{\scriptscriptstyle Z}(\pi_2^{\scriptscriptstyle Z} - z_2) \\ -D_1^{\scriptscriptstyle W}(\pi_1^{\scriptscriptstyle W} - w_1) \\ -D_2^{\scriptscriptstyle W}(\pi_2^{\scriptscriptstyle W} - w_2) \end{pmatrix},$$

where the quantities $D_{\scriptscriptstyle Y}$, $\pi^{\scriptscriptstyle Y}$, $D_{\scriptscriptstyle A}$, $\pi^{\scriptscriptstyle V}$, $D_1^{\scriptscriptstyle W}$, $D_2^{\scriptscriptstyle W}$, $\pi_1^{\scriptscriptstyle W}$, $\pi_2^{\scriptscriptstyle W}$, $D_1^{\scriptscriptstyle Z}$, $D_2^{\scriptscriptstyle Z}$, $\pi_1^{\scriptscriptstyle Z}$, and $\pi_2^{\scriptscriptstyle Z}$ are defined in (3.5).

The Hessian $\nabla^2 M(x,\, x_1,\, x_2,\, s,\, s_1,\, s_2,\, y,\, v,\, z_1,\, z_2,\, w_1,\, w_2)$ is given by

$$\begin{pmatrix}
H_1 & 0 & 0 & -2J^T D_Y^{-1} & 0 & 0 & J^T & A^T & 0 & 0 & 0 & 0 \\
0 & 2\Pi_1^Z (X_1^\mu)^{-1} & 0 & 0 & 0 & 0 & -I_m & 0 & I_L^x & 0 & 0 & 0 \\
0 & 0 & 2\Pi_2^Z (X_2^\mu)^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & I_U^x & 0 & 0 \\
-2D_Y^{-1} J & 0 & 0 & 2D_Y^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2\Pi_1^W (S_1^\mu)^{-1} & 0 & 0 & 0 & 0 & 0 & I_L^s & 0 \\
0 & 0 & 0 & 0 & 0 & 2\Pi_2^W (S_2^\mu)^{-1} & 0 & 0 & 0 & 0 & 0 & I_U^s \\
J & 0 & 0 & -I_m & 0 & 0 & D_Y & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 0 & 0 & 0 & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\
0 & I_L^x & 0 & 0 & 0 & 0 & 0 & 0 & X_1^\mu Z_1^{-2} \Pi_1^Z & 0 & 0 & 0 \\
0 & 0 & I_U^x & 0 & 0 & 0 & 0 & 0 & 0 & X_1^\mu Z_2^{-2} \Pi_2^Z & 0 & 0 \\
0 & 0 & 0 & 0 & I_L^s & 0 & 0 & 0 & 0 & 0 & S_1^\mu W_1^{-2} \Pi_1^W & 0 \\
0 & 0 & 0 & 0 & 0 & I_U^s & 0 & 0 & 0 & 0 & 0 & S_2^\mu W_2^{-2} \Pi_2^W
\end{pmatrix},$$

where

$$H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^A} A^T A + \frac{2}{\mu^P} J(x)^T J(x) = H(x, 2\pi^Y - y) + 2A^T D_A^{-1} A + 2J(x)^T D_Y^{-1} J(x),$$

and I_L^x, I_U^x, I_L^s, and I_U^s denote identity matrices of dimension n_L, n_U, m_L and m_U respectively. The usual convention regarding diagonal matrices formed from vectors applies, with $\Pi_1^Z = \operatorname{diag}(\pi_1^Z)$, $\Pi_2^Z = \operatorname{diag}(\pi_2^Z)$, $\Pi_1^W = \operatorname{diag}(\pi_1^W)$, and $\Pi_2^W = \operatorname{diag}(\pi_2^W)$.

5. Derivation of the primal-dual line-search direction

The primal-dual penalty-barrier problem (4.1) may be written in the form

$$\underset{p\in\mathcal{I}}{\text{minimize}} \;\; M(p) \quad \text{subject to} \quad Cp = b_C,$$

where

$$\mathcal{I} = \big\{p : p = (x, x_1, x_2, s, s_1, s_2, y, v, z_1, z_2, w_1, w_2), \;\; \text{with } x_i + \mu^B e > 0,\ s_i + \mu^B e > 0,\ z_i > 0,\ w_i > 0 \text{ for } i = 1, 2\big\},$$

and

$$C = \begin{pmatrix}
E_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
E_L & -I_L^x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
E_U & 0 & I_U^x & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & L_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & L_L & -I_L^s & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & L_U & 0 & I_U^s & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}, \quad \text{with} \quad b_C = \begin{pmatrix} b_X \\ \ell^X \\ u^X \\ h_X \\ \ell^S \\ u^S \end{pmatrix}. \tag{5.1}$$

Let p be any vector in $\mathcal{I}$ such that $Cp = b_C$. The Newton direction Δp is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^{\mathrm{T}}\Delta p + \tfrac{1}{2}\Delta p^{\mathrm{T}}\nabla^2 M(p)\Delta p \quad \text{subject to} \quad C\Delta p = b_C - Cp = 0. \tag{5.2}$$

Let N denote a matrix whose columns form a basis for null(C), i.e., the columns of N are linearly independent and $CN = 0$. Every feasible direction Δp may be written in the form $\Delta p = Nd$. This implies that d satisfies the reduced equations $N^{\mathrm{T}}\nabla^2 M(p)Nd = -N^{\mathrm{T}}\nabla M(p)$. However, instead of solving (5.2), we formulate a linearly constrained *approximate* Newton method by approximating the Hessian $\nabla^2 M$ by a matrix B such that $N^{\mathrm{T}}B(p)N$ is positive definite with $N^{\mathrm{T}}B(p)N \approx N^{\mathrm{T}}\nabla^2 M(p)N$. Consider the matrix defined by replacing $\pi^{\scriptscriptstyle Y}$ by y, $\pi_1^{\scriptscriptstyle Z}$ by z_1, $\pi_2^{\scriptscriptstyle Z}$ by z_2, $\pi_1^{\scriptscriptstyle W}$ by w_1, $\pi_2^{\scriptscriptstyle W}$ by w_2 in the matrix $\nabla^2 M(x,\, x_1,\, x_2,\, s,\, s_1,\, s_2,\, y,\, v,\, z_1,\, z_2,\, w_1,\, w_2)$. This gives an approximate Hessian $B(x,\, x_1,\, x_2,\, s,\, s_1,\, s_2,\, y,\, v,\, z_1,\, z_2,\, w_1,\, w_2)$ of the form

$$\begin{pmatrix}
H^{\scriptscriptstyle B} + 2A^{\mathrm{T}}D_{\scriptscriptstyle A}^{-1}A + 2J^{\mathrm{T}}D_{\scriptscriptstyle Y}^{-1}J & 0 & 0 & -2J^{\mathrm{T}}D_{\scriptscriptstyle Y}^{-1} & 0 & 0 & J^{\mathrm{T}} & A^{\mathrm{T}} & 0 & 0 & 0 & 0 \\
0 & 2(D_1^{\scriptscriptstyle Z})^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle L}^x & 0 & 0 & 0 \\
0 & 0 & 2(D_2^{\scriptscriptstyle Z})^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle U}^x & 0 & 0 \\
-2D_{\scriptscriptstyle Y}^{-1}J & 0 & 0 & 2D_{\scriptscriptstyle Y}^{-1} & 0 & 0 & -I_m & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2(D_1^{\scriptscriptstyle W})^{-1} & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle L}^s & 0 \\
0 & 0 & 0 & 0 & 0 & 2(D_2^{\scriptscriptstyle W})^{-1} & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle U}^s \\
J & 0 & 0 & -I_m & 0 & 0 & D_{\scriptscriptstyle Y} & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 0 & 0 & 0 & 0 & 0 & D_{\scriptscriptstyle A} & 0 & 0 & 0 & 0 \\
0 & I_{\scriptscriptstyle L}^x & 0 & 0 & 0 & 0 & 0 & 0 & D_1^{\scriptscriptstyle Z} & 0 & 0 & 0 \\
0 & 0 & I_{\scriptscriptstyle U}^x & 0 & 0 & 0 & 0 & 0 & 0 & D_2^{\scriptscriptstyle Z} & 0 & 0 \\
0 & 0 & 0 & 0 & I_{\scriptscriptstyle L}^s & 0 & 0 & 0 & 0 & 0 & D_1^{\scriptscriptstyle W} & 0 \\
0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle U}^s & 0 & 0 & 0 & 0 & 0 & D_2^{\scriptscriptstyle W}
\end{pmatrix},$$

where $H^{\scriptscriptstyle B} \approx H(x,y)$ is chosen so that the approximate reduced Hessian $N^{\mathrm{T}}B(p)N$ is positive definite (see Section 7). Given $B(p)$, the approximate Newton direction is given by the solution of the QP subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^{\mathrm{T}}\Delta p + \tfrac{1}{2}\Delta p^{\mathrm{T}}B(p)\Delta p \quad \text{subject to} \quad C\Delta p = 0.$$

Consider the null-space basis defined from the columns of the matrix

$$N = \begin{pmatrix} E_F^T & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_F^T & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_A & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_L^x & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_U^x & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_L^s & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & I_U^s \end{pmatrix}, \tag{5.3}$$

where $E_{LF} = E_L E_F^T$, $E_{UF} = E_U E_F^T$, $L_{LF} = L_L L_F^T$ and $L_{UF} = L_U L_F^T$. The definition of N of (5.3) gives the reduced approximate Hessian $N^T B(p) N$ such that

$$\begin{pmatrix} \widehat{H}_F & -2J_F^T D_Y^{-1} L_F^T & J_F^T & A_F^T & E_{LF}^T & -E_{UF}^T & 0 & 0 \\ -2L_F D_Y^{-1} J_F & 2L_F\big(D_Y^{-1} + D_W^{\dagger}\big)L_F^T & -L_F & 0 & 0 & 0 & L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^w & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^w \end{pmatrix},$$

where $J_F = J(x)E_F^T$, $A_F = AE_F^T$, $\widehat{H}_F = E_F\big(H^B + 2A^T D_A^{-1} A + 2J(x)^T D_Y^{-1} J(x) + 2D_Z^{\dagger}\big)E_F^T$, with

$$D_Z^{\dagger} = E_L^T (D_1^z)^{-1} E_L + E_U^T (D_2^z)^{-1} E_U, \quad \text{and} \quad D_W^{\dagger} = L_L^T \big(D_1^w\big)^{-1} L_L + L_U^T \big(D_2^w\big)^{-1} L_U,$$

Similarly, the reduced gradient $N^{\mathrm{T}}\nabla M(p)$ is given by

$$\begin{pmatrix} g_F - A_F^{\mathrm{T}}\big(2\pi^V - v\big) - J_F^{\mathrm{T}}\big(2\pi^Y - y\big) - E_{LF}^{\mathrm{T}}(2\pi_1^Z - z_1) + E_{UF}^{\mathrm{T}}(2\pi_2^Z - z_2) \\ 2\pi_F^Y - y_F - L_{LF}^{\mathrm{T}}(2\pi_1^W - w_1) + L_{UF}^{\mathrm{T}}(2\pi_2^W - w_2) \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix},$$

where $g_F = E_F\nabla f(x)$, $\pi_F^Y = L_F\pi^Y$ and $y_F = L_F y$. The reduced approximate Newton equations $N^{\mathrm{T}}B(p)Nd = -N^{\mathrm{T}}\nabla M(p)$ are then

$$\begin{pmatrix} \widehat{H}_F & -2J_F^{\mathrm{T}}D_Y^{-1}L_F^{\mathrm{T}} & J_F^{\mathrm{T}} & A_F^{\mathrm{T}} & E_{LF}^{\mathrm{T}} & -E_{UF}^{\mathrm{T}} & 0 & 0 \\ -2L_FD_Y^{-1}J_F & 2L_F(D_Y^{-1} + D_W^{\dagger})L_F^{\mathrm{T}} & -L_F & 0 & 0 & 0 & L_{LF}^{\mathrm{T}} & L_{UF}^{\mathrm{T}} \\ J_F & -L_F^{\mathrm{T}} & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix}$$

$$= -\begin{pmatrix} g_F - A_F^{\mathrm{T}}\big(2\pi^V - v\big) - J_F^{\mathrm{T}}\big(2\pi^Y - y\big) - E_{LF}^{\mathrm{T}}(2\pi_1^Z - z_1) + E_{UF}^{\mathrm{T}}(2\pi_2^Z - z_2) \\ 2\pi_F^Y - y_F - L_{LF}^{\mathrm{T}}(2\pi_1^W - w_1) + L_{UF}^{\mathrm{T}}(2\pi_2^W - w_2) \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}, \quad (5.4)$$

Given any nonsingular matrix R, the direction d satisfies $RN^{\mathrm{T}}B(p)Nd = -RN^{\mathrm{T}}\nabla M(p)$. In particular, if R is the block upper-triangular matrix R such that

$$R = \begin{pmatrix} I_F^x & 0 & -2J_F^{\mathrm{T}}D_Y^{-1} & -2A_F^{\mathrm{T}}D_A^{-1} & -2E_{LF}^{\mathrm{T}}(D_1^z)^{-1} & 2E_{UF}^{\mathrm{T}}(D_2^z)^{-1} & 0 & 0 \\ & I_F^s & 2L_F D_Y^{-1} & 0 & 0 & 0 & -2L_{LF}^{\mathrm{T}}(D_1^w)^{-1} & 2L_{UF}^{\mathrm{T}}(D_2^w)^{-1} \\ & & I_m & 0 & 0 & 0 & 0 & 0 \\ & & & I_A & 0 & 0 & 0 & 0 \\ & & & & I_L^x & 0 & 0 & 0 \\ & & & & & I_U^x & 0 & 0 \\ & & & & & & I_L^s & 0 \\ & & & & & & & I_U^s \end{pmatrix},$$

where I_L^x, I_U^x, I_L^s, I_U^s are identity matrices of size n_L, n_U, m_L, and m_U respectively, then R is nonsingular with

$$RN^{\mathrm{T}}B(p)N = \begin{pmatrix} H_F^B & 0 & -J_F^{\mathrm{T}} & -A_F^{\mathrm{T}} & -E_{LF}^{\mathrm{T}} & E_{UF}^{\mathrm{T}} & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^{\mathrm{T}} & L_{UF}^{\mathrm{T}} \\ J_F & -L_F^{\mathrm{T}} & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^w & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^w \end{pmatrix},$$

and

$$RN^{\mathrm{T}}\nabla M(p) = \begin{pmatrix} g_F - J_F^{\mathrm{T}}y - A_F^{\mathrm{T}}v - z_F \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \\ -D_1^w(\pi_1^w - w_1) \\ -D_2^w(\pi_2^w - w_2) \end{pmatrix} = \begin{pmatrix} E_F\big(g - J^{\mathrm{T}}y - A^{\mathrm{T}}v - z\big) \\ L_F\big(y - w\big) \\ c(x) - s + \mu^P(y - y^E) \\ Ax - b + \mu^A(v - v^E) \\ Z_1^{-1}\big(z_1 \cdot (E_Lx - \ell^X) + \mu^B(z_1 - z_1^E)\big) \\ Z_2^{-1}\big(z_2 \cdot (u^X - E_Ux) + \mu^B(z_2 - z_2^E)\big) \\ W_1^{-1}\big(w_1 \cdot (L_Ls - \ell^S) + \mu^B(w_1 - w_1^E)\big) \\ W_2^{-1}\big(w_2 \cdot (u^S - L_Us) + \mu^B(w_2 - w_2^E)\big) \end{pmatrix},$$

with $H_F^B = E_F H^B E_F^T$. This implies that we may solve the following (unsymmetric) reduced approximate Newton equations for d:

$$\begin{pmatrix} H_F^B & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - z_F \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}. \tag{5.5}$$

Then, the expression $\Delta p = Nd$ implies that

$$\Delta p = \begin{pmatrix} \Delta x \\ \Delta x_1 \\ \Delta x_2 \\ \Delta s \\ \Delta s_1 \\ \Delta s_2 \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = Nd = \begin{pmatrix} E_F^T d_1 \\ d_1 \\ -d_1 \\ L_F^T d_2 \\ d_2 \\ -d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix}. \tag{5.6}$$

These identities allow us to write the equations (5.5) as

$$\begin{pmatrix} H_F^B & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - z_F \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}, \tag{5.7}$$

with $\Delta x = E_F^T \Delta x_F$, $\Delta s = L_F^T \Delta s_F$, $\Delta x_1 = \Delta x_F - (\ell^X - E_L x + x_1)$, $\Delta x_2 = -\Delta x_F + (u^X - E_U x - x_2)$, $\Delta s_1 = \Delta s_F - (\ell^S - L_L s + s_1)$ and $\Delta s_2 = -\Delta s_F + (u^S - L_U s - s_2)$, The shifted penalty-barrier equations (5.7) are the same as the path-following equations (3.4) except that $H(x,y)$ is replaced by H^B in the $(1,1)$ block.

6. The shifted primal-dual penalty-barrier direction

In this section we consider the solution of the shifted primal-dual penalty-barrier equations (5.7). Collecting terms and reordering the equations and unknowns, gives

$$\begin{pmatrix} D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & D_1^z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & D_2^z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T & 0 & 0 & L_F \\ -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 & 0 & H_F^B & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & -L_F^T & J_F & D_Y \end{pmatrix} \begin{pmatrix} \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_A(v - \pi^V) \\ D_1^z(z_1 - \pi_1^z) \\ D_2^z(z_2 - \pi_2^z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ L_F(y - w) \\ E_F(g - J^Ty - A^Tv - z) \\ D_Y(y - \pi^Y) \end{pmatrix}, \tag{6.1}$$

Consider the diagonal matrices

$$D_W = \big(L_L^T (D_1^W)^{-1} L_L + L_U^T (D_2^W)^{-1} L_U\big)^\dagger \quad \text{and} \quad D_Z = \big(E_L^T (D_1^z)^{-1} E_L + E_U^T (D_2^z)^{-1} E_U\big)^\dagger,$$

where $(\,\cdot\,)^\dagger$ denotes the Moore-Penrose pseudoinverse of a matrix. The identity $I_m = L_X^T L_X + L_F^T L_F$ implies that the $m \times m$ matrix D_W satisfies the identities

$$L_F^T L_F D_W = D_W = D_W L_F^T L_F, \quad \text{and} \quad L_X^T L_X D_W = 0.$$

In addition, the diagonal matrix $L_F D_W^\dagger L_F^T$ is nonsingular if every slack is either fixed or bounded above or below. If equations (6.1) are premultiplied by the matrix

$$\begin{pmatrix} I_A \\ 0 & I_{LF}^x \\ 0 & 0 & I_{UF}^x \\ 0 & 0 & 0 & I_{LF}^s \\ 0 & 0 & 0 & 0 & I_{UF}^s \\ 0 & 0 & 0 & L_{LF}^T(D_1^W)^{-1} & -L_{UF}^T(D_2^W)^{-1} & I_F^s \\ A_F^T D_A^{-1} & E_{LF}^T(D_1^z)^{-1} & -E_{UF}^T(D_2^z)^{-1} & 0 & 0 & 0 & I_F^x \\ 0 & 0 & 0 & D_W L_L^T (D_1^W)^{-1} & -D_W L_U^T (D_2^W)^{-1} & D_W L_F^T & 0 & I_m \end{pmatrix}$$

we obtain the block upper-triangular system

$$\begin{pmatrix} D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & D_2^Z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & L_F D_W^\dagger L_F^T & 0 & L_F \\ 0 & 0 & 0 & 0 & 0 & 0 & \widetilde{H}_F & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & 0 & J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_A(v - \pi^V) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ L_F(y - \pi^W) \\ E_F(g - J^T y - A^T \pi^V - \pi^Z) \\ D_W\big(y - \pi^W\big) + D_Y\big(y - \pi^Y\big) \end{pmatrix},$$

where $\widetilde{H}_F = H_F^B + A_F^T D_A^{-1} A_F + E_F D_Z^\dagger E_F^T$, $\pi^W = L_L^T \pi_1^W - L_U^T \pi_2^W$ and $\pi^Z = E_L^T \pi_1^Z - E_U^T \pi_2^Z$. Using block back-substitution, Δx_F and Δy can be computed by solving the equations

$$\begin{pmatrix} \widetilde{H}_F & -J_F^T \\ J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} E_F\big(\nabla f(x) - J(x)^T y - A^T \pi^V - \pi^Z\big) \\ D_W\big(y - \pi^W\big) + D_Y\big(y - \pi^Y\big) \end{pmatrix}.$$

Once Δx_F and Δy have been computed, the full vector Δx is given by $\Delta x = E_F^T \Delta x_F$. Similarly, substitution of the identity $\Delta s = L_F^T \Delta s_F$ in the sixth block of equations gives

$$\Delta s = -D_W(y + \Delta y - \pi^W).$$

There are several ways of computing Δw_1 and Δw_2. Instead of using the block upper-triangular system above, we use the last two blocks of equations of (3.4) to give

$$\Delta w_1 = -(S_1^\mu)^{-1}\big(w_1 \cdot (L_L(s + \Delta s) - \ell^S + \mu^B e) - \mu^B w_1^E\big) \;\text{ and }\; \Delta w_2 = -(S_2^\mu)^{-1}\big(w_2 \cdot (u^S - L_U(s + \Delta s) + \mu^B e) - \mu^B w_2^E\big).$$

Similarly, using (3.4) to solve for Δz_1 and Δz_2 yields

$$\Delta z_1 = -(X_1^\mu)^{-1}\big(z_1 \cdot (E_L(x + \Delta x) - \ell^X + \mu^B e) - \mu^B z_1^E\big) \;\text{ and }\; \Delta z_2 = -(X_2^\mu)^{-1}\big(z_2 \cdot (u^X - E_U(x + \Delta x) + \mu^B e) - \mu^B z_2^E\big).$$

Similarly, using the first block of equations (6.1) to solve for Δv gives $\Delta v = -(v - \widehat{\pi}^V)$, with $\widehat{\pi}^V = v^E - \frac{1}{\mu^A}\big(A(x + \Delta x) - b\big)$. Finally, the vectors Δw_X and Δz_X are recovered as $\Delta w_X = [\, y + \Delta y - w \,]_X$ and $\Delta z_X = [\, g + H \Delta x - J^T(y + \Delta y) - z \,]_X$, where $w = L_X^T w_X + L_L^T w_1 - L_U^T w_2$ and $z = E_X^T z_X + E_L^T z_1 - E_U^T z_2$.

7. Summary: equations for the line-search direction

The results of the preceding section imply that the solution of the path-following equations $F'(v_P)\Delta v_P = -F(v_P)$ with F and F' given by (3.2) and (3.3) may be computed as follows. Let x and s be given primal variables and slack variables such that $E_X x = b_X$, $L_X s = h_X$ with $\ell^X - \mu^B < E_L x$, $E_U x < u^X + \mu^B$, $\ell^S - \mu^B < L_L s$, $L_U s < u^S + \mu^B$. Similarly, let z_1, z_2, w_1, w_2 and y denote dual variables such that $w_1 > 0$, $w_2 > 0$, $z_1 > 0$, and $z_2 > 0$. Consider the diagonal matrices $X_1^\mu = \operatorname{diag}(E_L x - \ell^X + \mu^B e)$, $X_2^\mu = \operatorname{diag}(u^X - E_U x + \mu^B e)$, $Z_1 = \operatorname{diag}(z_1)$, $Z_2 = \operatorname{diag}(z_2)$, $W_1 = \operatorname{diag}(w_1)$, $W_2 = \operatorname{diag}(w_2)$, $S_1^\mu = \operatorname{diag}(L_L s - \ell^S + \mu^B e)$ and $S_2^\mu = \operatorname{diag}(u^S - L_U s + \mu^B e)$. Consider the quantities

$$\begin{aligned}
D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c-s),\\
D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax-b),\\
(D_1^Z)^{-1} &= (X_1^\mu)^{-1} Z_1, & (D_1^W)^{-1} &= (S_1^\mu)^{-1} W_1,\\
(D_2^Z)^{-1} &= (X_2^\mu)^{-1} Z_2, & (D_2^W)^{-1} &= (S_2^\mu)^{-1} W_2,\\
D_Z &= \big(E_L^T (D_1^Z)^{-1} E_L + E_U^T (D_2^Z)^{-1} E_U\big)^\dagger, & D_W &= \big(L_L^T (D_1^W)^{-1} L_L + L_U^T (D_2^W)^{-1} L_U\big)^\dagger,\\
\pi_1^Z &= \mu^B \big(X_1^\mu\big)^{-1} z_1^E, & \pi_1^W &= \mu^B \big(S_1^\mu\big)^{-1} w_1^E,\\
\pi_2^Z &= \mu^B \big(X_2^\mu\big)^{-1} z_2^E, & \pi_2^W &= \mu^B \big(S_2^\mu\big)^{-1} w_2^E,\\
\pi^Z &= E_L^T \pi_1^Z - E_U^T \pi_2^Z, & \pi^W &= L_L^T \pi_1^W - L_U^T \pi_2^W.
\end{aligned}$$

Choose H_F^B so that H_F^B approximates $E_F H(x,y) E_F^T$ and the KKT matrix

$$\begin{pmatrix} H_F^B + A_F^T D_A^{-1} A_F + E_F D_Z^\dagger E_F^T & J_F^T \\ J_F & -(D_Y + D_W) \end{pmatrix}$$

is nonsingular with m negative eigenvalues. (A common choice of H_F^B is the matrix $E_F\big(H(x,y) + \sigma I_n\big)E_F^T$ for some nonnegative scalar σ.) Solve the KKT system

$$\begin{pmatrix} H_F^B + A_F^T D_A^{-1} A_F + E_F D_Z^\dagger E_F^T & -J_F^T \\ J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = -\begin{pmatrix} E_F\big(\nabla f(x) - J(x)^T y - A^T \pi^V - \pi^Z\big) \\ D_Y(y - \pi^Y) + D_W\big(y - \pi^W\big) \end{pmatrix}, \tag{7.1}$$

and set

$$
\begin{aligned}
\Delta x = E_F^T \Delta x_F \qquad \widehat{x} &= x + \Delta x, & \Delta z_1 &= -(X_1^{\mu})^{-1}\big(z_1 \cdot (E_L \widehat{x} - \ell^X + \mu^B e) - \mu^B z_1^E\big),\\
& & \Delta z_2 &= -(X_2^{\mu})^{-1}\big(z_2 \cdot (u^X - E_U \widehat{x} + \mu^B e) - \mu^B z_2^E\big),\\
\widehat{y} &= y + \Delta y, & \Delta s &= -D_W(\widehat{y} - \pi^W),\\
\widehat{s} &= s + \Delta s, & \Delta w_1 &= -(S_1^{\mu})^{-1}\big(w_1 \cdot (L_L \widehat{s} - \ell^S + \mu^B e) - \mu^B w_1^E\big),\\
& & \Delta w_2 &= -(S_2^{\mu})^{-1}\big(w_2 \cdot (u^S - L_U \widehat{s} + \mu^B e) - \mu^B w_2^E\big),\\
\widehat{\pi}^V &= v^E - \frac{1}{\mu^A}(A\widehat{x} - b), & \Delta v &= \widehat{\pi}^V - v,\\
w &= L_X^T w_X + L_L^T w_1 - L_U^T w_2, & z &= E_X^T z_X + E_L^T z_1 - E_U^T z_2,\\
\widehat{v} &= v + \Delta v, & \Delta w_X &= [\,\widehat{y} - w\,]_X,\\
& & \Delta z_X &= [\,\nabla f(x) + H^B(x,y)\Delta x - J(x)^T\widehat{y} - A^T\widehat{v} - z\,]_X.
\end{aligned}
$$

As $(x,s) \to (x^*,s^*)$ it holds that $\|D_W^\dagger\|$ and $\|D_Z^\dagger\|$ are bounded, but $\|D_W\| \to \infty$ and $\|A_F^T D_A^{-1} A_F\| \to \infty$. This implies that the matrix and right-hand side of (7.1) goes to infinity. In the situation where $A_F^T D_A^{-1} A_F$ is diagonal, then the KKT system can be rescaled so that the equations to be solved are bounded. If $\widetilde{D}_Z$ and $\widetilde{D}_W$ denote diagonal matrices such that $\widetilde{D}_Z^2 = (A_F^T D_A^{-1} A_F)^{-1}$ and $\widetilde{D}_W^2 = (L_X^T L_X + D_W)^{-1}$, then $\|\widetilde{D}_Z\|$ and $\|\widetilde{D}_W\|$ are bounded as $(x,s) \to (x^*,s^*)$. The equations (7.1) may be written in the form

$$
\begin{pmatrix} \widetilde{D}_Z H_F^B(x,y)\widetilde{D}_Z + \widetilde{D}_Z^2 E_F D_Z^\dagger E_F^T + I_F^x & -(\widetilde{D}_W J_F(x)\widetilde{D}_Z)^T \\ \widetilde{D}_W J_F(x)\widetilde{D}_Z & \widetilde{D}_W^2 D_Y + L_F^T L_F \end{pmatrix} \begin{pmatrix} \Delta\widetilde{x}_F \\ \Delta\widetilde{y} \end{pmatrix} = -\begin{pmatrix} \widetilde{D}_Z E_F\big(\nabla f(x) - J(x)^T y - A^T\pi^V - \pi^Z\big) \\ \widetilde{D}_W\big(D_Y(y - \pi^Y) + D_W(y - \pi^W)\big) \end{pmatrix},
$$

with $\Delta x_F = \widetilde{D}_Z \Delta\widetilde{x}_F$ and $\Delta y = \widetilde{D}_W \Delta\widetilde{y}$. In this case, the scaled KKT matrix remains bounded if $H(x,y)$ is bounded. Similarly, the right-hand side remains bounded if $\|\widetilde{D}_W D_W\big(y - \pi^W\big)\|$ is bounded.

The associated line-search merit function (4.2) can be written as

$$
\begin{aligned}
f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\
- (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A(v - v^E)\|^2 \\
- \sum_{j=1}^{n_L}\Big\{\mu^B[\,z_1^E\,]_j \ln\big([\,z_1\,]_j[\,E_L x - \ell^X + \mu^B e\,]_j^2\big) - [\,z_1\cdot(E_L x - \ell^X + \mu^B e)\,]_j\Big\} \\
- \sum_{j=1}^{n_U}\Big\{\mu^B[\,z_2^E\,]_j \ln\big([\,z_2\,]_j[\,u^X - E_U x + \mu^B e\,]_j^2\big) - [\,z_2\cdot(u^X - E_U x + \mu^B e)\,]_j\Big\} \\
- \sum_{i=1}^{m_L}\Big\{\mu^B[\,w_1^E\,]_i \ln\big([\,w_1\,]_i[\,L_L s - \ell^S + \mu^B e\,]_i^2\big) - [\,w_1\cdot(L_L s - \ell^S + \mu^B e)\,]_i\Big\} \\
- \sum_{i=1}^{m_U}\Big\{\mu^B[\,w_2^E\,]_i \ln\big([\,w_2\,]_i[\,u^S - L_U s + \mu^B e\,]_i^2\big) - [\,w_2\cdot(u^S - L_U s + \mu^B e)\,]_i\Big\}.
\end{aligned}
$$

8. The primal-dual trust-region direction

Given a vector of primal-dual variables $p = (x, x_1, x_2, s, s_1, s_2, y, v, z_1, z_2, w_1, w_2)$, each iteration of a trust-region method for solving (NLP) involves finding a vector Δp of the form $\Delta p = Nd$, where N is a basis for the null-space of the matrix C of (5.1), and d is an approximate solution of the subproblem

$$\underset{d}{\text{minimize}}\;\; g_N^T d + \tfrac{1}{2} d^T B_N(p) d \quad \text{subject to} \;\; \|d\|_T \le \delta, \tag{8.1}$$

where g_N and B_N are the reduced gradient and reduced Hessian $g_N = N^T\nabla M$ and $B_N(p) = N^T B(p) N$, $\|d\|_T = (d^T T d)^{1/2}$, δ is the trust-region radius, and T is positive-definite. The subproblem (8.1) may be written as

$$\underset{\Delta v_M}{\text{minimize}}\;\; g_N^T T^{-1/2}\Delta v_M + \tfrac{1}{2}\Delta v_M^T T^{-1/2} B_N(p) T^{-1/2}\Delta v_M \quad \text{subject to} \;\; \|\Delta v_M\|_2 \le \delta, \tag{8.2}$$

where $\Delta v_M = T^{1/2} d$. The application of the method of Moré and Sorensen [8] to solve the subproblem (8.2) requires the solution of the so-called *secular equations*, which have the form

$$(\bar{B}_N + \sigma I)\Delta v_M = -\bar{g}_N, \tag{8.3}$$

with σ a nonnegative scalar, $\bar{B}_N = T^{-1/2} B_N(p) T^{-1/2}$, and $\bar{g}_N = T^{-1/2} g_N$. In this note we consider the solution of the related equations

$$(B_N + \sigma T)d = -g_N, \tag{8.4}$$

from which the solution of the secular equations (8.3) may be computed as $\Delta v_M = T^{1/2} d$.

The identity (5.6) allows the solution of the approximate Newton equations $B_N(p)d = -g_N$ (5.4) to be written in terms of the change in the variables (x, s, s_1, s_2, y, v, z_1, z_2, w_1, w_2). In particular, we have

$$\begin{pmatrix} \widehat{H}_F & -2J_F^T D_Y^{-1} L_F^T & J_F^T & A_F^T & E_{LF}^T & -E_{UF}^T & 0 & 0 \\ -2L_F D_Y^{-1} J_F & 2L_F(D_Y^{-1} + D_W^{\dagger})L_F^T & -L_F & 0 & 0 & 0 & L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix}$$
$$= -\begin{pmatrix} g_F - A_F^T\big(2\pi^V - v\big) - J_F^T\big(2\pi^Y - y\big) - E_{LF}^T(2\pi_1^Z - z_1) + E_{UF}^T(2\pi_2^Z - z_2) \\ 2\pi_F^Y - y_F - L_{LF}^T(2\pi_1^W - w_1) + L_{UF}^T(2\pi_2^W - w_2) \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}, \tag{8.5}$$

where

$$\widehat{H}_F = E_F\big(H(x,y) + J^T D_Y^{-1} J + A^T D_A^{-1} A + D_Z^{\dagger}\big)E_F^T,$$

with

$$\begin{array}{llll} D_Y = \mu^P I_m, & \pi^Y = y^E - \dfrac{1}{\mu^P}(c - s), & D_A = \mu^A I_A, & \pi^V = v^E - \dfrac{1}{\mu^A}(Ax - b), \\ D_1^W = S_1^{\mu} W_1^{-1}, & \pi_1^W = \mu^B (S_1^{\mu})^{-1} w_1^E, & D_1^Z = X_1^{\mu} Z_1^{-1}, & \pi_1^Z = \mu^B (X_1^{\mu})^{-1} z_1^E, \\ D_2^W = S_2^{\mu} W_2^{-1}, & \pi_2^W = \mu^B (S_2^{\mu})^{-1} w_2^E, & D_2^Z = X_2^{\mu} Z_2^{-1}, & \pi_2^Z = \mu^B (X_2^{\mu})^{-1} z_2^E, \\ & \pi^W = L_L^T \pi_1^W - L_U^T \pi_2^W & & \pi^Z = E_L^T \pi_1^Z - E_U^T \pi_2^Z. \end{array}$$

Note that in the trust-region case we make no assumption that B_N is positive definite.

The first step in the formulation of the trust-region equations (8.4) and their solution is to write the (reduced) gradient and approximate Hessian of equations (8.5) in terms of vectors $\vec{x}$ and $\vec{y}$ that combine the primal variables (x, s) and dual variables $(y,\, v,\, z_1,\, z_2,\, w_1,\, w_2)$. Let $\vec{g}$, $\vec{H}$, $\vec{J}$ and $\vec{D}$ denote the quantities

$$\vec{g} = \begin{pmatrix} g_F \\ 0 \end{pmatrix}, \quad \vec{H} = \begin{pmatrix} H_F & 0 \\ 0 & 0 \end{pmatrix}, \quad \vec{J} = \begin{pmatrix} J_F & -L_F^{\mathrm{T}} \\ A_F & 0 \\ E_{LF} & 0 \\ -E_{UF} & 0 \\ 0 & L_{LF} \\ 0 & -L_{UF} \end{pmatrix} \quad \text{and} \quad \vec{D} = \begin{pmatrix} D_Y & 0 & 0 & 0 & 0 & 0 \\ 0 & D_A & 0 & 0 & 0 & 0 \\ 0 & 0 & D_1^Z & 0 & 0 & 0 \\ 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix},$$

where $g_F = E_F \nabla f(x)$, $J_F = J(x)E_F^{\mathrm{T}}$, $H_F = E_F H(x,y) E_F^{\mathrm{T}}$, and $A_F = AE_F^{\mathrm{T}}$. Similarly, let $\vec{T}_x = \operatorname{diag}(T^x, T^s)$ and $\vec{T}_y = \operatorname{diag}(T^y,$ $T^v,\, T_1^z,\, T_2^z,\, T_1^w,\, T_2^w)$. The trust-region equations associated with the modified Newton equations (8.5) are $(B_N + \sigma T)\Delta p = -g_N$, which may be written in the form

$$\begin{pmatrix} \vec{H} + 2\vec{J}^{\mathrm{T}}\vec{D}^{-1}\vec{J} + \sigma\vec{T}_x & \vec{J}^{\mathrm{T}} \\ \vec{J} & \vec{D} + \sigma\vec{T}_y \end{pmatrix} \begin{pmatrix} \Delta\vec{x} \\ \Delta\vec{y} \end{pmatrix} = -\begin{pmatrix} \vec{g} - \vec{J}^{\mathrm{T}}\vec{\pi} - \vec{J}^{\mathrm{T}}(\vec{\pi} - \vec{y}) \\ -\vec{D}(\vec{\pi} - \vec{y}) \end{pmatrix}, \tag{8.6}$$

where

$$\vec{y} = \begin{pmatrix} y \\ v \\ z_1 \\ z_2 \\ w_1 \\ w_2 \end{pmatrix}, \quad \vec{\pi} = \begin{pmatrix} \pi^Y \\ \pi^V \\ \pi_1^Z \\ \pi_2^Z \\ \pi_1^W \\ \pi_2^W \end{pmatrix}, \quad \Delta\vec{x} = \begin{pmatrix} \Delta x_F \\ \Delta s_F \end{pmatrix}, \quad \text{and} \quad \Delta\vec{y} = \begin{pmatrix} \Delta y \\ \Delta z \\ \Delta w \end{pmatrix}.$$

Applying the nonsingular matrix $\begin{pmatrix} I & -2\vec{J}^{\mathrm{T}}\vec{D}^{-1} \\ & I \end{pmatrix}$ to both sides of (8.6) gives the equivalent system

$$\begin{pmatrix} \vec{H} + \sigma\vec{T}_x & -\vec{J}^{\mathrm{T}}(I + 2\sigma\vec{D}^{-1}\vec{T}_y) \\ \vec{J} & \vec{D} + \sigma\vec{T}_y \end{pmatrix} \begin{pmatrix} \Delta\vec{x} \\ \Delta\vec{y} \end{pmatrix} = -\begin{pmatrix} \vec{g} - \vec{J}^{\mathrm{T}}\vec{y} \\ \vec{D}(\vec{y} - \vec{\pi}) \end{pmatrix}.$$

As in Gertz and Gill [3], we set $\vec{T}_x = I$ and $\vec{T}_y = \vec{D}$. With this choice, the associated vectors $\Delta\vec{x}$ and $\Delta\vec{y}$ satisfy the equations

$$\begin{pmatrix} \vec{H} + \sigma I & -\vec{J}^{\mathrm{T}} \\ \vec{J} & \sigma\vec{D} \end{pmatrix} \begin{pmatrix} \Delta\vec{x} \\ (1 + 2\sigma)\Delta\vec{y} \end{pmatrix} = -\begin{pmatrix} \vec{g} - \vec{J}^{\mathrm{T}}\vec{y} \\ \vec{D}(\vec{y} - \vec{\pi}) \end{pmatrix}, \tag{8.7}$$

where $\bar\sigma = (1+\sigma)/(1+2\sigma)$. In terms of the original variables, the unsymmetric equations (8.7) are

$$\begin{pmatrix} H_F+\sigma I_F^x & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & \sigma I_F^s & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & \bar\sigma D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & \bar\sigma D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & \bar\sigma D_1^z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & \bar\sigma D_2^z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & \bar\sigma D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & \bar\sigma D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ (1+2\sigma)\Delta y \\ (1+2\sigma)\Delta v \\ (1+2\sigma)\Delta z_1 \\ (1+2\sigma)\Delta z_2 \\ (1+2\sigma)\Delta w_1 \\ (1+2\sigma)\Delta w_2 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - z_F \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}, \quad (8.8)$$

where $\bar\sigma = (1+\sigma)/(1+2\sigma)$. These equations are equivalent to (5.5) when $\sigma = 0$ and $\bar\sigma = 1$. Collecting terms and reordering the equations and unknowns, we obtain

$$\begin{pmatrix} \bar\sigma D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & \bar\sigma D_1^z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & \bar\sigma D_2^z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & \bar\sigma D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar\sigma D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T & \sigma I_F^s & 0 & L_F \\ -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 & 0 & H_F+\sigma I_F^x & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & -L_F^T & J_F & \bar\sigma D_Y \end{pmatrix} \begin{pmatrix} \Delta\widetilde v \\ \Delta\widetilde z_1 \\ \Delta\widetilde z_2 \\ \Delta\widetilde w_1 \\ \Delta\widetilde w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta\widetilde y \end{pmatrix} = - \begin{pmatrix} D_A(v-\pi^V) \\ D_1^z(z_1-\pi_1^z) \\ D_2^z(z_2-\pi_2^z) \\ D_1^W(w_1-\pi_1^W) \\ D_2^W(w_2-\pi_2^W) \\ L_F\big(y-w\big) \\ E_F\big(g - J^T y - A^T v - z\big) \\ D_Y(y-\pi^Y) \end{pmatrix}, \quad (8.9)$$

where $\bar D_A = \bar\sigma D_A$, $\bar D_1^W = \bar\sigma D_1^W$, $\bar D_2^W = \bar\sigma D_2^W$, $\bar D_1^z = \bar\sigma D_1^z$, $\bar D_2^z = \bar\sigma D_2^z$, $\bar D_Y = \bar\sigma D_Y$, $\Delta\widetilde y = (1+2\sigma)\Delta y$, $\Delta\widetilde v = (1+2\sigma)\Delta v$, $\Delta\widetilde z_1 = (1+2\sigma)\Delta z_1$, $\Delta\widetilde z_2 = (1+2\sigma)\Delta z_2$, $\Delta\widetilde w_1 = (1+2\sigma)\Delta w_1$, and $\Delta\widetilde w_2 = (1+2\sigma)\Delta w_2$. We define

$$\bar D_W = \big(L_L^T(\bar D_1^W)^{-1}L_L + L_U^T(\bar D_2^W)^{-1}L_U\big)^\dagger = \bar\sigma\big(L_L^T(D_1^W)^{-1}L_L + L_U^T(D_2^W)^{-1}L_U\big)^\dagger = \bar\sigma D_W,$$

with $D_W = \big(L_{LF}^T(D_1^W)^{-1}L_{LF} + L_{UF}^T(D_2^W)^{-1}L_{UF}\big)^\dagger$. Similarly, define

$$\breve D_W = \big(D_W^\dagger + \sigma\bar\sigma L_F^T L_F\big)^\dagger.$$

Premultiplying the equations (8.9) by the matrix

$$\begin{pmatrix} I_A \\ 0 & I^x_{LF} \\ 0 & 0 & I^x_{UF} \\ 0 & 0 & 0 & I^s_{LF} \\ 0 & 0 & 0 & 0 & I^s_{UF} \\ 0 & 0 & 0 & \frac{1}{\bar\sigma}L^T_{LF}(D^W_1)^{-1} & -\frac{1}{\bar\sigma}L^T_{UF}(D^W_2)^{-1} & I^s_F \\ \frac{1}{\bar\sigma}A^T_F D_A^{-1} & \frac{1}{\bar\sigma}E^T_{LF}(D^Z_1)^{-1} & -\frac{1}{\bar\sigma}E^T_{UF}(D^Z_2)^{-1} & 0 & 0 & 0 & I^x_F \\ 0 & 0 & 0 & \breve D_W L^T_L(D^W_1)^{-1} & -\breve D_W L^T_U(D^W_2)^{-1} & \bar\sigma \breve D_W L^T_F & 0 & I_m \end{pmatrix}$$

gives the block upper-triangular system

$$\begin{pmatrix} \bar\sigma D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & \bar\sigma D^Z_1 & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & \bar\sigma D^Z_2 & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & \bar\sigma D^W_1 & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar\sigma D^W_2 & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\bar\sigma}L_F\breve D^\dagger_W L^T_F & 0 & L_F \\ 0 & 0 & 0 & 0 & 0 & 0 & \widetilde H_F + \sigma I^x_F & -J^T_F \\ 0 & 0 & 0 & 0 & 0 & 0 & J_F & \bar\sigma(D_Y + \breve D_W) \end{pmatrix} \begin{pmatrix} \Delta\widetilde v \\ \Delta\widetilde z_1 \\ \Delta\widetilde z_2 \\ \Delta\widetilde w_1 \\ \Delta\widetilde w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta\widetilde y \end{pmatrix}$$

$$= - \begin{pmatrix} D_A(v - \pi^V) \\ D^Z_1(z_1 - \pi^Z_1) \\ D^Z_2(z_2 - \pi^Z_2) \\ D^W_1(w_1 - \pi^W_1) \\ D^W_2(w_2 - \pi^W_2) \\ L_F\Big(y - w + \dfrac{1}{\bar\sigma}\big[w - \pi^W\big]\Big) \\ E_F\Big(g - J^T y - A^T v - z + \dfrac{1}{\bar\sigma}\big[A^T(v - \pi^V) + z - \pi^Z\big]\Big) \\ D_Y(y - \pi^Y) + \breve D_W\big(\bar\sigma(y - w) + w - \pi^W\big) \end{pmatrix},$$

where

$$\widetilde H_F = E_F\Big(H + \frac{1}{\bar\sigma}A^T D_A^{-1} A + \frac{1}{\bar\sigma}D^\dagger_Z\Big)E^T_F,$$

$w = L_X^T w_X + L_L^T w_1 - L_U^T w_2$, $z = E_X^T z_X + E_L^T z_1 - E_U^T z_2$, $\pi^W = L_L^T \pi_1^W - L_U^T \pi_2^W$ and $\pi^Z = E_L^T \pi_1^Z - E_U^T \pi_2^Z$. Using block back-substitution, Δx_F and Δy may be computed by solving the equations

$$\begin{pmatrix} \widetilde{H}_F + \sigma I_F^x & -J_F^T \\ J_F & \bar{\sigma}(D_Y + \breve{D}_W) \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta \widetilde{y} \end{pmatrix} = -\begin{pmatrix} E_F\left(g - J^T y - A^T v - z + \dfrac{1}{\bar{\sigma}}\left[A^T(v - \pi^V) + z - \pi^Z\right]\right) \\ D_Y\,(y - \pi^Y) + \breve{D}_W\,(\bar{\sigma}(y - w) + w - \pi^W) \end{pmatrix}.$$

Once Δx_F and $\Delta \widetilde{y}$ are known, the full vector Δx is computed as $\Delta x = E_F^T \Delta x_F$. Using the identity $\Delta s = L_F^T \Delta s_F$ in the sixth block of equations gives

$$\Delta s = -\bar{\sigma}\breve{D}_W\left(y + (1 + 2\sigma)\Delta y - w + \frac{1}{\bar{\sigma}}\left[w - \pi^W\right]\right).$$

There are several ways of computing Δw_1 and Δw_2. Instead of using the block upper-triangular system above, we use the last two blocks of equations of (8.8) to give

$$\Delta w_1 = -\frac{1}{1+\sigma}(S_1^\mu)^{-1}\big(w_1 \cdot (L_L(s + \Delta s) - \ell^S + \mu^B e) - \mu^B w_1^E\big) \quad \text{and}$$
$$\Delta w_2 = -\frac{1}{1+\sigma}(S_2^\mu)^{-1}\big(w_2 \cdot (u^S - L_U(s + \Delta s) + \mu^B e) - \mu^B w_2^E\big).$$

Similarly, using (8.8) to solve for Δz_1 and Δz_2 yields

$$\Delta z_1 = -\frac{1}{1+\sigma}(X_1^\mu)^{-1}\big(z_1 \cdot (E_L(x + \Delta x) - \ell^X + \mu^B e) - \mu^B z_1^E\big) \quad \text{and}$$
$$\Delta z_2 = -\frac{1}{1+\sigma}(X_2^\mu)^{-1}\big(z_2 \cdot (u^X - E_U(x + \Delta x) + \mu^B e) - \mu^B z_2^E\big).$$

Similarly, using the first block of equations (8.9) to solve for Δv gives $\Delta v = -(v - \widehat{\pi}^V)/(1+\sigma)$, with $\widehat{\pi}^V = v^E - \frac{1}{\mu^A}\big(A(x + \Delta x) - b\big)$. Finally, the vectors Δw_X and Δz_X are recovered as $\Delta w_X = [\,y + \Delta y - w\,]_X$ and $\Delta z_X = [\,g + H\Delta x - J^T(y + \Delta y) - A^T(v + \Delta v) - z\,]_X$.

9. Summary: equations for the trust-region direction

The results of the preceding section imply that the solution of the trust-region equations $(B_N + \sigma T)\Delta v_M = -g_N$, with σ a nonnegative scalar, may be computed as follows. Let x and s be given primal variables and slack variables such that $E_X x = b_X$, $L_X s = h_X$ with $\ell^X - \mu^B < E_L x$, $E_U x < u^X + \mu^B$, $\ell^S - \mu^B < L_L s$, $L_U s < u^S + \mu^B$. Similarly, let z_1, z_2, w_1, w_2 and y denotes dual variables such that $w_1 > 0$, $w_2 > 0$, $z_1 > 0$, and $z_2 > 0$. Consider the diagonal matrices $X_1^\mu = \operatorname{diag}(E_L x - \ell^X + \mu^B e)$, $X_2^\mu = \operatorname{diag}(u^X - E_U x + \mu^B e)$, $Z_1 = \operatorname{diag}(z_1)$, $Z_2 = \operatorname{diag}(z_2)$, $W_1 = \operatorname{diag}(w_1)$, $W_2 = \operatorname{diag}(w_2)$, $S_1^\mu = \operatorname{diag}(L_L s - \ell^S + \mu^B e)$ and $S_2^\mu = \operatorname{diag}(u^S - L_U s + \mu^B e)$. Given the quantities

$$
\begin{aligned}
D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c - s),\\
D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax - b),\\
(D_1^Z)^{-1} &= (X_1^\mu)^{-1} Z_1, & (D_1^W)^{-1} &= (S_1^\mu)^{-1} W_1,\\
(D_2^Z)^{-1} &= (X_2^\mu)^{-1} Z_2, & (D_2^W)^{-1} &= (S_2^\mu)^{-1} W_2,\\
D_Z &= \big(E_L^T (D_1^Z)^{-1} E_L + E_U^T (D_2^Z)^{-1} E_U\big)^\dagger, & D_W &= \big(L_L^T (D_1^W)^{-1} L_L + L_U^T (D_2^W)^{-1} L_U\big)^\dagger,\\
& & \breve{D}_W &= \big(D_W^\dagger + \sigma\bar{\sigma} L_F L_F^T\big)^\dagger,\\
\pi_1^Z &= \mu^B (X_1^\mu)^{-1} z_1^E, & \pi_1^W &= \mu^B (S_1^\mu)^{-1} w_1^E,\\
\pi_2^Z &= \mu^B (X_2^\mu)^{-1} z_2^E, & \pi_2^W &= \mu^B (S_2^\mu)^{-1} w_2^E,\\
\pi^Z &= E_L^T \pi_1^Z - E_U^T \pi_2^Z, & \pi^W &= L_L^T \pi_1^W - L_U^T \pi_2^W,
\end{aligned}
$$

solve the KKT system

$$
\begin{pmatrix}
E_F\Big(H(x,y) + \sigma I_n + \dfrac{1}{\bar{\sigma}} A^T D_A^{-1} A + \dfrac{1}{\bar{\sigma}} D_Z^\dagger\Big) E_F^T & -J_F^T\\
J_F & \bar{\sigma}\big(D_Y + \breve{D}_W\big)
\end{pmatrix}
\begin{pmatrix}
\Delta x_F\\
\Delta \widetilde{y}
\end{pmatrix}
= -\begin{pmatrix}
E_F\Big(\nabla f(x) - J(x)^T y - A^T v - z + \dfrac{1}{\bar{\sigma}}\big[A^T(v - \pi^V) + z - \pi^Z\big]\Big)\\
D_Y\big(y - \pi^Y\big) + \breve{D}_W\big(\bar{\sigma}(y - w) + w - \pi^W\big)
\end{pmatrix}.
$$

Then

$$
\begin{array}{lll}
\Delta x = E_F^{\mathrm{T}} \Delta x_F, & \widehat{x} = x + \Delta x, & \Delta z_1 = -\dfrac{1}{1+\sigma}(X_1^{\mu})^{-1}\big(z_1 \cdot (E_L \widehat{x} - \ell^X + \mu^B e) - \mu^B z_1^E\big), \\
 & & \Delta z_2 = -\dfrac{1}{1+\sigma}(X_2^{\mu})^{-1}\big(z_2 \cdot (u^X - E_U \widehat{x} + \mu^B e) - \mu^B z_2^E\big), \\
\Delta y = \Delta \widetilde{y}/(1+2\sigma), & \widehat{y} = y + \Delta y, & \Delta s = -\bar{\sigma} \check{D}_W \Big(y + (1+2\sigma)\Delta y - w + \dfrac{1}{\bar{\sigma}}\big[w - \pi^W\big]\Big), \\
 & \widehat{s} = s + \Delta s, & \Delta w_1 = -\dfrac{1}{1+\sigma}(S_1^{\mu})^{-1}\big(w_1 \cdot (L_L \widehat{s} - \ell^S + \mu^B e) - \mu^B w_1^E\big), \\
 & & \Delta w_2 = -\dfrac{1}{1+\sigma}(S_2^{\mu})^{-1}\big(w_2 \cdot (u^S - L_U \widehat{s} + \mu^B e) - \mu^B w_2^E\big), \\
 & \widehat{\pi}^V = v^E - \dfrac{1}{\mu^A}(A\widehat{x} - b), & \Delta v = -\dfrac{1}{1+\sigma}\big(v - \widehat{\pi}^V\big), \\
 & w = L_X^{\mathrm{T}} w_X + L_L^{\mathrm{T}} w_1 - L_U^{\mathrm{T}} w_2, & z = E_X^{\mathrm{T}} z_X + E_L^{\mathrm{T}} z_1 - E_U^{\mathrm{T}} z_2, \\
 & \widehat{v} = v + \Delta v, & \Delta w_X = [\,\widehat{y} - w\,]_X, \\
 & & \Delta z_X = [\,\nabla f(x) + H(x,y)\Delta x - J(x)^{\mathrm{T}}\widehat{y} - A^{\mathrm{T}}\widehat{v} - z\,]_X.
\end{array}
$$

10. Solution of the trust-region equations with an arbitrary right-hand-side

Moré and Sorensen define a routine $z_{\mathrm{null}}(\cdot)$ that uses the Cholesky factors of $\bar{B}_N + \sigma I$ and the condition estimator proposed by Cline, Moler, Stewart and Wilkinson [2]. As the method of Gill, Kungurtsev and Robinson does not compute an explicit factorization of $\bar{B}_N + \sigma I$, we define $z_{\mathrm{null}}(\cdot)$ using the condition estimator `DLACON` supplied with LAPACK [1]. This routine, which generates an approximate null vector using Higham's [7] modification of Hager's algorithm [6], uses matrix-vector products with $((\bar{B}_N + \sigma I)^{-1})^{-1}$, rather than a matrix factorization, to estimate $\|(\bar{B}_N + \sigma I)^{-1}\|_1$. By-products of the computation of $\|(\bar{B}_N + \sigma I)^{-1}\|_1$ are vectors v and w such that $w = (\bar{B}_N + \sigma I)^{-1} v$, $\|v\|_1 = 1$ and

$$\|(\bar{B}_N + \sigma I)^{-1} v\|_1 = \|w\|_1 \approx \|(\bar{B}_N + \sigma I)^{-1}\|_1 = \max_{\|u\|_1 = 1} \|(\bar{B}_N + \sigma I)^{-1} u\|_1.$$

Thus, unless $\|w\| = 0$, the vector $y = w/\|w\|$ is a unit approximate null vector from which we determine an appropriate z such that $\|\Delta v_M + z\|_T = \delta$.

The reduced trust-region equations with a general right-hand side analogous to (8.9) are given by

$$\begin{pmatrix} \bar\sigma D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & \bar\sigma D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & \bar\sigma D_2^Z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & \bar\sigma D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar\sigma D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & -L_{LF}^{\mathrm{T}} & L_{UF}^{\mathrm{T}} & \sigma I_F^s & 0 & L_F \\ -A_F^{\mathrm{T}} & -E_{LF}^{\mathrm{T}} & E_{UF}^{\mathrm{T}} & 0 & 0 & 0 & H_F + \sigma I_F^x & -J_F^{\mathrm{T}} \\ 0 & 0 & 0 & 0 & 0 & -L_F^{\mathrm{T}} & J_F & \bar\sigma D_Y \end{pmatrix} \begin{pmatrix} \widetilde{q}_A \\ \widetilde{q}_Z^{(1)} \\ \widetilde{q}_Z^{(2)} \\ \widetilde{q}_W^{(1)} \\ \widetilde{q}_W^{(2)} \\ \widetilde{q}_S \\ \widetilde{q}_X \\ \widetilde{q}_Y \end{pmatrix} = \begin{pmatrix} r_A \\ r_1^Z \\ r_2^Z \\ r_1^W \\ r_2^W \\ L_F r_S \\ E_F r_X \\ r_Y \end{pmatrix}.$$

Premultiplying these equations by

$$\begin{pmatrix} I_A & & & & & & & \\ 0 & I_{LF}^x & & & & & & \\ 0 & 0 & I_{UF}^x & & & & & \\ 0 & 0 & 0 & I_{LF}^s & & & & \\ 0 & 0 & 0 & 0 & I_{UF}^s & & & \\ 0 & 0 & 0 & \frac{1}{\bar\sigma} L_{LF}^{\mathrm{T}} (D_1^W)^{-1} & -\frac{1}{\bar\sigma} L_{UF}^{\mathrm{T}} (D_2^W)^{-1} & I_F^s & & \\ \frac{1}{\bar\sigma} A_F^{\mathrm{T}} D_A^{-1} & \frac{1}{\bar\sigma} E_{LF}^{\mathrm{T}} (D_1^Z)^{-1} & -\frac{1}{\bar\sigma} E_{UF}^{\mathrm{T}} (D_2^Z)^{-1} & 0 & 0 & 0 & I_F^x & \\ 0 & 0 & 0 & \breve{D}_W L_L^{\mathrm{T}} (D_1^W)^{-1} & -\breve{D}_W L_U^{\mathrm{T}} (D_2^W)^{-1} & \bar\sigma \breve{D}_W L_F^{\mathrm{T}} & 0 & I_m \end{pmatrix}$$

gives the block upper-triangular equations

$$\begin{pmatrix} \bar\sigma D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & \bar\sigma D_1^z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & \bar\sigma D_2^z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & \bar\sigma D_1^w & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar\sigma D_2^w & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\bar\sigma} L_F \breve D_W^\dagger L_F^T & 0 & L_F \\ 0 & 0 & 0 & 0 & 0 & 0 & \widetilde H_F + \sigma I_F^x & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & 0 & J_F & \bar\sigma(D_Y + \breve D_W) \end{pmatrix} \begin{pmatrix} \widetilde q_A \\ \widetilde q_Z^{(1)} \\ \widetilde q_Z^{(2)} \\ \widetilde q_W^{(1)} \\ \widetilde q_W^{(2)} \\ \widetilde q_S \\ \widetilde q_X \\ \widetilde q_Y \end{pmatrix}$$

$$= \begin{pmatrix} r_A \\ r_Z^{(1)} \\ r_Z^{(2)} \\ r_W^{(1)} \\ r_W^{(2)} \\ \frac{1}{\bar\sigma} L_F \left(L_L^T (D_1^w)^{-1} r_W^{(1)} - L_U^T (D_2^w)^{-1} r_W^{(2)} + \bar\sigma r_S \right) \\ \frac{1}{\bar\sigma} E_F \left(A^T D_A^{-1} r_A + E_L^T (D_1^z)^{-1} r_Z^{(1)} - E_U^T (D_2^z)^{-1} r_Z^{(2)} + \bar\sigma r_X \right) \\ \breve D_W \left(L_L^T (D_1^w)^{-1} r_W^{(1)} - L_U^T (D_2^w)^{-1} r_W^{(2)} + \bar\sigma r_S \right) + r_Y \end{pmatrix},$$

with $\widetilde H_F = E_F \Big(H(x,y) + \dfrac{1}{\bar\sigma} A^T D_A^{-1} A + \dfrac{1}{\bar\sigma} D_Z^\dagger \Big) E_F^T$. Using block back-substitution, $\widetilde q_X$ and $\widetilde q_Y$ can be computed by solving the equations

$$\begin{pmatrix} \widetilde H_F & -J_F^T \\ J_F & \bar\sigma\big(D_Y + \breve D_W\big) \end{pmatrix} \begin{pmatrix} \widetilde q_X \\ \widetilde q_Y \end{pmatrix} = \begin{pmatrix} \frac{1}{\bar\sigma} E_F \left(A^T D_A^{-1} r_A + E_L^T (D_1^z)^{-1} r_Z^{(1)} - E_U^T (D_2^z)^{-1} r_Z^{(2)} + \bar\sigma r_X \right) \\ \breve D_W \left(L_L^T (D_1^w)^{-1} r_W^{(1)} - L_U^T (D_2^w)^{-1} r_W^{(2)} + \bar\sigma r_S \right) + r_Y \end{pmatrix},$$

with the remaining vectors computed as

$$
\begin{aligned}
\widetilde{q}_S &= \breve{D}_W \left(L_L^{\mathrm{T}} (D_1^W)^{-1} r_W^{(1)} - L_U^{\mathrm{T}} (D_2^W)^{-1} r_W^{(2)} + \bar{\sigma}(r_S - \widetilde{q}_Y) \right) \\
\widetilde{q}_W^{(2)} &= \frac{1}{\bar{\sigma}} (D_2^W)^{-1} \big(r_W^{(2)} - L_U L_F^{\mathrm{T}} \widetilde{q}_S \big) \\
\widetilde{q}_W^{(1)} &= \frac{1}{\bar{\sigma}} (D_1^W)^{-1} \big(r_W^{(1)} - L_L L_F^{\mathrm{T}} \widetilde{q}_S \big) \\
\widetilde{q}_Z^{(2)} &= \frac{1}{\bar{\sigma}} (D_2^Z)^{-1} \big(r_Z^{(2)} - E_U E_F \widetilde{q}_X \big) \\
\widetilde{q}_Z^{(1)} &= \frac{1}{\bar{\sigma}} (D_1^Z)^{-1} \big(r_Z^{(1)} - E_L E_F \widetilde{q}_X \big) \\
\widetilde{q}_A &= \frac{1}{\bar{\sigma}} (D_A)^{-1} \big(r_A - A_F \widetilde{q}_X \big).
\end{aligned}
$$

References

[1] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, third edition, 1999. 28

[2] A. K. Cline, C. B. Moler, G. W. Stewart, and J. H. Wilkinson. An estimate for the condition number of a matrix. *SIAM J. Numer. Anal.*, 16(2):368–375, 1979. 28

[3] E. M. Gertz and P. E. Gill. A primal-dual trust-region algorithm for nonlinear programming. *Math. Program., Ser. B*, 100:49–94, 2004. 23

[4] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A shifted primal-dual penalty-barrier method for nonlinear optimization. *SIAM J. Optim.*, 30(2):1067–1093, 2020. 2, 3, 5

[5] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A trust-region shifted primal-dual penalty-barrier method for nonlinear optimization. Center for Computational Mathematics Report CCoM 21-01, University of California, San Diego, 2021. 2

[6] W. W. Hager. Condition estimates. *SIAM J. Sci. Statist. Comput.*, 5(2):311–316, 1984. 28

[7] N. J. Higham. FORTRAN codes for estimating the one-norm of a real or complex matrix, with applications to condition estimation. *ACM Trans. Math. Software*, 14:381–396, 1988. 28

[8] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM J. Sci. and Statist. Comput.*, 4:553–572, 1983. 21

Index