Equations for a Projected-Search Path-Following Method for Nonlinear Optimization

Philip E. Gill* Minxin Zhang*

UCSD Center for Computational Mathematics
Technical Report CCoM-22-2
June 2022

Abstract

In [2], Gill and Zhang propose a primal-dual path-following method for general nonlinearly constrained optimization that combines a shifted primal-dual path-following method with a projected-search method for bound-constrained optimization. The method involves the computation of an approximate Newton direction for a primal-dual penalty-barrier function that incorporates shifts on both the primal and dual variables. This note concerns the formulation of approximate Newton equations for a nonlinear optimization problem in general form. These equations may be used in conjunction with a projected-search 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. Nonlinearly constrained optimization, path-following methods, primal-dual methods, shifted penalty and barrier methods, projected-search methods, Armijo line search, augmented Lagrangian 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.

1. Introduction

This note concerns that derivation of the primal-dual 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 projected-search method for a problem with nonlinear inequality constraints is given by Gill and Zhang [2]. The equations are formulated for problems written in the general form:

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad \begin{cases} c(x)-s=0, & L_{\scriptscriptstyle X} s = h_{\scriptscriptstyle X}, \quad \ell^S \le L_{\scriptscriptstyle L} s, \quad L_{\scriptscriptstyle U} s \le u^S, \\ Ax-b=0, & E_{\scriptscriptstyle X} x = b_{\scriptscriptstyle X}, \quad \ell^X \le E_{\scriptscriptstyle L} x, \quad E_{\scriptscriptstyle U} x \le u^X, \end{cases} \tag{NLP}$$

where A denotes a constant $m_{\scriptscriptstyle A} \times n$ matrix, and b, $h_{\scriptscriptstyle X}$, $b_{\scriptscriptstyle X}$, ℓ^S, u^S, ℓ^X and u^X are fixed vectors of dimension $m_{\scriptscriptstyle A}$, $m_{\scriptscriptstyle X}$, $n_{\scriptscriptstyle X}$, $m_{\scriptscriptstyle L}$, $m_{\scriptscriptstyle U}$, $n_{\scriptscriptstyle L}$ and $n_{\scriptscriptstyle U}$, respectively. Similarly, $L_{\scriptscriptstyle X}$, $L_{\scriptscriptstyle L}$ and $L_{\scriptscriptstyle U}$ denote fixed matrices of dimension $m_{\scriptscriptstyle X} \times m$, $m_{\scriptscriptstyle L} \times m$ and $m_{\scriptscriptstyle U} \times m$, respectively, and $E_{\scriptscriptstyle X}$, $E_{\scriptscriptstyle L}$ and $E_{\scriptscriptstyle U}$ are fixed matrices of dimension $n_{\scriptscriptstyle X} \times n$, $n_{\scriptscriptstyle L} \times n$ and $n_{\scriptscriptstyle U} \times n$, 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 $E_{\scriptscriptstyle X}$ denotes an $n_{\scriptscriptstyle X} \times n$ matrix formed from $n_{\scriptscriptstyle X}$ independent rows of I_n, the identity matrix of order n. This implies that the equality constraints $E_{\scriptscriptstyle X} x = b_{\scriptscriptstyle X}$ fix $n_{\scriptscriptstyle X}$ components of x at the corresponding values of $b_{\scriptscriptstyle X}$. Similarly, $E_{\scriptscriptstyle L}$ and $E_{\scriptscriptstyle U}$ denote $n_{\scriptscriptstyle L} \times n$ and $n_{\scriptscriptstyle U} \times n$ matrices formed from subsets of rows of I_n such that $E_{\scriptscriptstyle X}^T E_{\scriptscriptstyle L} = 0$, $E_{\scriptscriptstyle X}^T E_{\scriptscriptstyle U} = 0$, 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 is not a row of $E_{\scriptscriptstyle X}$, $E_{\scriptscriptstyle L}$ or $E_{\scriptscriptstyle U}$. Analogous definitions hold for $L_{\scriptscriptstyle X}$, $L_{\scriptscriptstyle L}$ and $L_{\scriptscriptstyle U}$ 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 $E_{\scriptscriptstyle F}$ denote the matrix of rows of I_n that are not rows of $E_{\scriptscriptstyle X}$, and let $L_{\scriptscriptstyle F}$ denote the matrix of rows of I_m that are not rows of $L_{\scriptscriptstyle X}$. If $n_{\scriptscriptstyle F} = n - n_{\scriptscriptstyle X}$ and $m_{\scriptscriptstyle F} = m - m_{\scriptscriptstyle X}$, then $E_{\scriptscriptstyle F}$ and $L_{\scriptscriptstyle F}$ are $n_{\scriptscriptstyle F} \times n$ and $m_{\scriptscriptstyle F} \times m$ respectively. Note that $n_{\scriptscriptstyle L} + n_{\scriptscriptstyle U}$ may be less than $n_{\scriptscriptstyle F}$, but $m_{\scriptscriptstyle F}$ must equal $m_{\scriptscriptstyle L} + m_{\scriptscriptstyle U}$. The matrices $\begin{pmatrix} E_{\scriptscriptstyle X}^T & E_{\scriptscriptstyle F}^T \end{pmatrix}$ and $\begin{pmatrix} L_{\scriptscriptstyle X}^T & L_{\scriptscriptstyle F}^T \end{pmatrix}$ 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

$$P_x = \begin{pmatrix} E_{\scriptscriptstyle F} \\ E_{\scriptscriptstyle X} \end{pmatrix} \quad \text{and} \quad P_s = \begin{pmatrix} L_{\scriptscriptstyle F} \\ L_{\scriptscriptstyle X} \end{pmatrix},$$

with $E_{\scriptscriptstyle F} E_{\scriptscriptstyle F}^T = I_{\scriptscriptstyle F}^x$, $E_{\scriptscriptstyle X} E_{\scriptscriptstyle X}^T = I_{\scriptscriptstyle X}^x$, and $E_{\scriptscriptstyle F} E_{\scriptscriptstyle X}^T = 0$, and $L_{\scriptscriptstyle F} L_{\scriptscriptstyle F}^T = I_{\scriptscriptstyle F}^s$, $L_{\scriptscriptstyle X} L_{\scriptscriptstyle X}^T = I_{\scriptscriptstyle X}^s$, and $L_{\scriptscriptstyle F} L_{\scriptscriptstyle X}^T = 0$.

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 equalities are satisfied in the limit. The exception to this is when the constraints $E_{\scriptscriptstyle X} x = b_{\scriptscriptstyle X}$, and $L_{\scriptscriptstyle X} s = h_{\scriptscriptstyle X}$ are used to fix a subset of the variables and slacks. These bounds are enforced at every iterate.

[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.

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, the associated rows of A are rows of the identity matrix.

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 the projected-search method. The equations 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, whereas the subscript k is appended to a vector to denote its value during the kth iteration of an algorithm, e.g., x_k represents the value for x during the kth iteration, whereas $[\,x_k\,]_i$ denotes the ith component of the vector x_k. 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 $\nabla c_i(x)^{\mathrm{T}}$. Given a Lagrangian function $L(x, y) = f(x) - c(x)^{\mathrm{T}} y$ with y a m-vector of dual variables, the Hessian of the Lagrangian with respect to x is denoted by $H(x, y) = \nabla^2 f(x) - \sum_{i=1}^{m} y_i \nabla^2 c_i(x)$. The equations utilize the Moore-Penrose pseudoinverse of a diagonal matrix. In particular, if $D = \mathrm{diag}(d_1,\, d_2,\, \ldots,\, d_n)$, 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

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

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 $E_L x - \ell^X \ge 0$, $u^X - E_U x \ge 0$, $L_L s - \ell^S \ge 0$ and $u^S - L_U s \ge 0$, 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 expressions appearing in certain equations, in others they are regarded as independent variables associated with the problem

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

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 $x_1 \ge 0$, $x_2 \ge 0$, $s_1 \ge 0$, and $s_2 \ge 0$, respectively.

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

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

3. The path-following equations

Penalty and barrier methods are closely related to path-following methods. These methods approximate a continuous path that passes through a solution of (NLP). In the simplest case, the path is parameterized by a positive scalar parameter that may be interpreted as a perturbation for the optimality conditions for the problem (NLP).

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^*. Similarly, let v^E, x^E and s^E denote estimates of v^*, x^* and s^*. Given small positive scalars μ^P, μ^A and μ^B, consider the perturbed optimality conditions

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

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

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

where the first $n+m$ equations are written in terms of z and w such that $z = E_X^T z_X + E_L^T z_1 - E_U^T z_2$ and $w = L_X^T w_X + L_L^T w_1 - L_U^T w_2$. (To simplify the notation, the dependence of F on the parameters μ^A, μ^P, μ^B, x^E, s^E, 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 $\ell^X < E_L,\ E_U x < u^X,\ \ell^S < L_L s,\ L_U < u^S,\ z_1 > 0,\ z_2 > 0,\ w_1 > 0$, and $w_2 > 0$ approximates a point satisfying the optimality conditions (2.1), with the approximation becoming increasingly accurate as the terms $\mu^P(y - y^E)$, $\mu^A(v - v^E)$, $\mu^B(E_L x^E - E_L x)$, $\mu^B(E_U x^E - E_U x)$, $\mu^B(L_L s^E - L_L s)$, $\mu^B(L_U s - L^2_L)$, $\mu^B(z_1 - z_1^E)$, $\mu^B(z_2 - z_2^E)$, $\mu^B(w_1 - w_1^E)$ and $\mu^B(w_2 - w_2^E)$ approach zero. For any sequence of x^E, s^E, z_1^E, z_2^E, w_1^E, w_2^E, v^E and y^E such that $x^E \to x^*$, $s^E \to s^*$, $z_1^E \to z_1^*$, $z_2^E \to z_2^*$, $w_1^E \to w_1^*$, $w_2^E \to w_2^*$, $v^E \to v^*$ and $y^E \to y^*$, 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 $F(v_P)$ such that $\ell^X - \mu^B < E_L x$, $E_U x < u^X + \mu^B$, $\ell^S - \mu^B < L_L s$, $L_U s < u^S + \mu^B$, $z_1 > 0$, $z_2 > 0$, $w_1 > 0$, 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

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

where

$$\left.\begin{array}{llll} X_1^\mu = \mathrm{diag}(x_1 + \mu^B e), & X_2^\mu = \mathrm{diag}(x_2 + \mu^B e), & S_1^\mu = \mathrm{diag}(s_1 + \mu^B e), & S_2^\mu = \mathrm{diag}(s_2 + \mu^B e), \\ Z_1^\mu = \mathrm{diag}(z_1 + \mu^B e), & Z_2^\mu = \mathrm{diag}(z_2 + \mu^B e), & W_1^\mu = \mathrm{diag}(w_1 + \mu^B e), & W_2^\mu = \mathrm{diag}(w_2 + \mu^B e), \end{array}\right\} \tag{3.4}$$

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 L_F are the rows of I_m orthogonal to the rows of L_X, i.e., $L_F^T L_X = 0$. The vectors s_F and s_X are the components of s corresponding to the "free" and "fixed" components of s, respectively. The variables $L_L s$ and $L_U s$ form a subset of s_F. Throughout, we assume that s satisfies $L_X s - h_X = 0$, in which case $\Delta s_X = 0$ and Δs satisfies

$$\Delta s = L_F^T \Delta s_F + L_X^T \Delta s_X = L_F^T \Delta s_F.$$

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 components of x corresponding to the "free" and "fixed variables", respectively. The variables $E_L x$ and $E_U x$ form a subset of x_F. Throughout, we assume that x_X satisfies

$E_X x - b_X = 0$, in which case $\Delta x_X = 0$ and Δx satisfies

$$\Delta x = E_F^T \Delta x_F + E_X^T \Delta x_X = E_F^T \Delta x_F.$$

After premultiplying the first and fifth blocks of equations of (3.3) by E_F and L_F respectively, and substituting $\Delta x = E_F^T \Delta x_F$ and $\Delta s = L_F^T \Delta s_F$, the equations (3.3) can be written in the reduced form $\widehat{F}'(v_F)\Delta v_F = -\widehat{F}(v_F)$, where $\Delta v_F = (\Delta x_F, \Delta s_F, \Delta y, \Delta v, \Delta z_1, \Delta z_2, \Delta w_1, \Delta w_2)$,

$$\begin{pmatrix} H_F & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ Z_1^\mu E_{LF} & 0 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2^\mu E_{UF} & 0 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1^\mu L_{LF} & 0 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2^\mu L_{UF} & 0 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - E_{LF}^T z_1 + E_{UF}^T z_2 \\ y_F - L_{LF}^T w_1 + L_{UF}^T w_2 \\ c(x) - s + \mu^P (y - y^E) \\ Ax - b + \mu^A (v - v^E) \\ z_1 \cdot (E_L x - \ell^X) + \mu^B (z_1 - z_1^E) + \mu^B (E_L x - E_L x^E) \\ z_2 \cdot (u^X - E_U x) + \mu^B (z_2 - z_2^E) + \mu^B (E_U x^E - E_U x) \\ w_1 \cdot (L_L s - \ell^S) + \mu^B (w_1 - w_1^E) + \mu^B (L_L s - L_L s^E) \\ w_2 \cdot (u^S - L_U s) + \mu^B (w_2 - w_2^E) + \mu^B (L_U s^E - L_U s) \end{pmatrix},$$

where $H_F = E_F H E_F^T$, $J_F = J(x) E_F^T$, $A_F = A E_F^T$, $g_F = E_F \nabla f(x)$, $E_{LF} = E_L E_F^T$, $E_{UF} = E_U E_F^T$, $y_F = L_F y$, $L_{LF} = L_L L_F^T$ and $L_{UF} = L_U L_F^T$. The matrices J_F, A_F, E_{LF} and E_{UF} are the columns of $J(x)$, A, E_L and E_U associated with the "free" components of x. The matrices L_{LF} and L_{UF} are the columns of L_L and L_U associated with the "free" components of s. Then scaling the last four blocks of equations by (respectively) $(Z_1^\mu)^{-1}$, $(Z_2^\mu)^{-1}$, $(W_1^\mu)^{-1}$ and $(W_2^\mu)^{-1}$ gives

$$\begin{pmatrix} H_F & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - E_{LF}^T z_1 + E_{UF}^T z_2 \\ y_F - L_{LF}^T w_1 + L_{UF}^T w_2 \\ c(x) - s + \mu^P (y - y^E) \\ Ax - b + \mu^A (v - v^E) \\ D_1^Z (z_1 - \pi_1^Z) \\ D_2^Z (z_2 - \pi_2^Z) \\ D_1^W (w_1 - \pi_1^W) \\ D_2^W (w_2 - \pi_2^W) \end{pmatrix}, \tag{3.5}$$

where $A_F = A E_F^T$ are the columns of A associated with the "free" components of x, and

$$\begin{aligned} D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c - s), & D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax - b), \\ D_1^W &= S_1^\mu (W_1^\mu)^{-1}, & \pi_1^W &= \mu^B (S_1^\mu)^{-1}(w_1^E - s_1 + s_1^E), & D_1^Z &= X_1^\mu (Z_1^\mu)^{-1}, & \pi_1^Z &= \mu^B (X_1^\mu)^{-1}(z_1^E - x_1 + x_1^E), \\ D_2^W &= S_2^\mu (W_2^\mu)^{-1}, & \pi_2^W &= \mu^B (S_2^\mu)^{-1}(w_2^E - s_2 + s_2^E), & D_2^Z &= X_2^\mu (Z_2^\mu)^{-1}, & \pi_2^Z &= \mu^B (X_2^\mu)^{-1}(z_2^E - x_2 + x_2^E), \end{aligned}$$

with auxiliary quantities

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

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 $\Delta x = L_F^T \Delta x_F$ and $\Delta z_X = [\, g + H\Delta x - J^T(y + \Delta y) - z\,]_X$.

4. A shifted primal-dual penalty-barrier function

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

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

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

$$\begin{aligned}
f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\
- (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A(v - v^E)\|^2 \\
- \sum_{j=1}^{n_L} \Big\{ \mu^B([\, z_1^E\,]_j + [\, x_1^E\,]_j + \mu^B) \ln\big([\, z_1 + \mu^B e\,]_j [\, x_1 + \mu^B e\,]_j^2\big) - [\, z_1 \cdot (x_1 + \mu^B e)\,]_j - 2\mu^B [\, x_1\,]_j \Big\} \\
- \sum_{j=1}^{n_U} \Big\{ \mu^B([\, z_2^E\,]_j + [\, x_2^E\,]_j + \mu^B) \ln\big([\, z_2 + \mu^B e\,]_j [\, x_2 + \mu^B e\,]_j^2\big) - [\, z_2 \cdot (x_2 + \mu^B e)\,]_j - 2\mu^B [\, x_2\,]_j \Big\} \\
- \sum_{i=1}^{m_L} \Big\{ \mu^B([\, w_1^E\,]_i + [\, s_1^E\,]_i + \mu^B) \ln\big([\, w_1 + \mu^B\,]_i [\, s_1 + \mu^B e\,]_i^2\big) - [\, w_1 \cdot (s_1 + \mu^B e)\,]_i - 2\mu^B [\, s_1\,]_i \Big\} \\
- \sum_{i=1}^{m_U} \Big\{ \mu^B([\, w_2^E\,]_i + [\, s_2^E\,] + \mu^B) \ln\big([\, w_2 + \mu^B\,]_i [\, s_2 + \mu^B e\,]_i^2\big) - [\, w_2 \cdot (s_2 + \mu^B e)\,]_i - 2\mu^B [\, s_2\,]_i \Big\}. \quad (4.1)
\end{aligned}$$

The gradient may be written as

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

where X_1^μ, X_2^μ, S_1^μ, S_2^μ, Z_1^μ, Z_2^μ, W_1^μ and W_2^μ are defined in (3.4). Equivalently,

$$\nabla M = \begin{pmatrix} \nabla f(x) - A^T\big(\pi^V + (\pi^V - v)\big) - J(x)^T\big(\pi^Y + (\pi^Y - y)\big) \\ z_1 - 2\pi_1^Z \\ z_2 - 2\pi_2^Z \\ \pi^Y + (\pi^Y - y) \\ w_1 - 2\pi_1^W \\ w_2 - 2\pi_2^W \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}.$$

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

$$\begin{pmatrix} H_1 & 0 & 0 & -2J^{\mathrm{T}}D_{\scriptscriptstyle Y}^{-1} & 0 & 0 & J^{\mathrm{T}} & A^{\mathrm{T}} & 0 & 0 & 0 & 0 \\ 0 & 2G_1^{\scriptscriptstyle X} & 0 & 0 & 0 & 0 & -I_m & 0 & I_{\scriptscriptstyle L}^x & 0 & 0 & 0 \\ 0 & 0 & 2G_2^{\scriptscriptstyle X} & 0 & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle U}^x & 0 & 0 \\ -2D_{\scriptscriptstyle Y}^{-1}J & 0 & 0 & 2D_{\scriptscriptstyle Y}^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2G_1^{\scriptscriptstyle S} & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle L}^s & 0 \\ 0 & 0 & 0 & 0 & 0 & 2G_2^{\scriptscriptstyle S} & 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle U}^s \\ J & 0 & 0 & -I_m & 0 & 0 & D_{\scriptscriptstyle Y} & 0 & 0 & 0 & 0 & 0 \\ A & 0 & 0 & 0 & 0 & 0 & 0 & D_{\scriptscriptstyle A} & 0 & 0 & 0 & 0 \\ 0 & I_{\scriptscriptstyle L}^x & 0 & 0 & 0 & 0 & 0 & 0 & G_1^{\scriptscriptstyle Z} & 0 & 0 & 0 \\ 0 & 0 & I_{\scriptscriptstyle U}^x & 0 & 0 & 0 & 0 & 0 & 0 & G_2^{\scriptscriptstyle Z} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{\scriptscriptstyle L}^s & 0 & 0 & 0 & 0 & 0 & G_1^{\scriptscriptstyle W} & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{\scriptscriptstyle U}^s & 0 & 0 & 0 & 0 & 0 & G_2^{\scriptscriptstyle W} \end{pmatrix},$$

where $H_1 = H(x, 2\pi^{\scriptscriptstyle Y} - y) + \frac{2}{\mu^{\scriptscriptstyle A}} A^{\mathrm{T}}A + \frac{2}{\mu^{\scriptscriptstyle P}} J(x)^{\mathrm{T}} J(x)$, and $I_{\scriptscriptstyle L}^x$, $I_{\scriptscriptstyle L}^x$, $I_{\scriptscriptstyle L}^s$, $I_{\scriptscriptstyle U}^s$ are identity matrices of size $n_{\scriptscriptstyle L}$, $n_{\scriptscriptstyle U}$, $m_{\scriptscriptstyle L}$, $m_{\scriptscriptstyle U}$ respectively. In addition

$$\begin{aligned} G_1^{\scriptscriptstyle X} &= (X_1^{\mu})^{-1}\big(\Pi_1^{\scriptscriptstyle Z} + \mu^{\scriptscriptstyle B} I\big), & G_2^{\scriptscriptstyle X} &= (X_2^{\mu})^{-1}\big(\Pi_2^{\scriptscriptstyle Z} + \mu^{\scriptscriptstyle B} I\big), \\ G_1^{\scriptscriptstyle S} &= (S_1^{\mu})^{-1}\big(\Pi_1^{\scriptscriptstyle W} + \mu^{\scriptscriptstyle B} I\big), & G_2^{\scriptscriptstyle S} &= (S_2^{\mu})^{-1}\big(\Pi_1^{\scriptscriptstyle W} + \mu^{\scriptscriptstyle B} I\big), \\ G_1^{\scriptscriptstyle Z} &= (Z_1^{\mu})^{-1}\big(\Pi_1^{\scriptscriptstyle X} + \mu^{\scriptscriptstyle B} I\big), & G_2^{\scriptscriptstyle Z} &= (Z_2^{\mu})^{-1}\big(\Pi_2^{\scriptscriptstyle X} + \mu^{\scriptscriptstyle B} I\big), \\ G_1^{\scriptscriptstyle W} &= (W_1^{\mu})^{-1}\big(\Pi_1^{\scriptscriptstyle S} + \mu^{\scriptscriptstyle B} I\big), & G_2^{\scriptscriptstyle W} &= (W_2^{\mu})^{-1}\big(\Pi_2^{\scriptscriptstyle S} + \mu^{\scriptscriptstyle B} I\big), \end{aligned}$$

with $\Pi_1^{\scriptscriptstyle Z} = \operatorname{diag}(\pi_1^{\scriptscriptstyle Z})$, $\Pi_2^{\scriptscriptstyle Z} = \operatorname{diag}(\pi_2^{\scriptscriptstyle Z})$, $\Pi_1^{\scriptscriptstyle W} = \operatorname{diag}(\pi_1^{\scriptscriptstyle W})$, $\Pi_2^{\scriptscriptstyle W} = \operatorname{diag}(\pi_2^{\scriptscriptstyle W})$, $X_1^{\scriptscriptstyle E} = \operatorname{diag}(x_1^{\scriptscriptstyle E})$, $X_2^{\scriptscriptstyle E} = \operatorname{diag}(x_2^{\scriptscriptstyle E})$, $S_1^{\scriptscriptstyle E} = \operatorname{diag}(s_1^{\scriptscriptstyle E})$, $W_1^{\scriptscriptstyle E} = \operatorname{diag}(w_1^{\scriptscriptstyle E})$, $W_2^{\scriptscriptstyle E} = \operatorname{diag}(w_2^{\scriptscriptstyle E})$, $Z_1^{\scriptscriptstyle E} = \operatorname{diag}(z_1^{\scriptscriptstyle E})$ and $Z_2^{\scriptscriptstyle E} = \operatorname{diag}(z_2^{\scriptscriptstyle E})$.

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

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

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

where

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

and

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

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

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

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 $N^T \nabla^2 M(p) N d = -N^T \nabla M(p)$. However, instead of solving (5.2), we formulate a linearly constrained *approximate* Newton method by approximating the Hessian $\nabla^2 M(p)$ by a matrix $B(p)$ such that $N^T B(p) N$ is positive definite with $N^T B(p) N \approx N^T \nabla^2 M(p) N$. Consider the matrix B obtained by replacing π^Y by y, π_1^Z by z_1, π_2^Z by z_2, π_1^W by w_1, π_2^W by w_2, x_1^E by x_1, x_2^E by x_2, s_1^E by s_1, s_2^E by s_2, z_1^E by z_1, z_2^E by z_2, w_1^E by w_1 and w_2^E by w_2 in $\nabla^2 M(x, x_1, x_2, s, s_1, s_2, y, v, z_1, z_2, w_1, w_2)$. 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

$$\begin{pmatrix} H^B + \frac{2}{\mu^A} A^T A + \frac{2}{\mu^P} J^T J & 0 & 0 & -2J^T D_Y^{-1} & 0 & 0 & J^T & A^T & 0 & 0 & 0 & 0 \\ 0 & 2(D_1^Z)^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & I_L^x & 0 & 0 & 0 \\ 0 & 0 & -2(D_2^Z)^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & I_U^x & 0 & 0 \\ -2D_Y^{-1} J & 0 & 0 & 2D_Y^{-1} & 0 & 0 & -I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2(D_1^W)^{-1} & 0 & 0 & 0 & 0 & 0 & I_L^s & 0 \\ 0 & 0 & 0 & 0 & 0 & 2(D_2^W)^{-1} & 0 & 0 & 0 & 0 & 0 & I_U^s \\ J & 0 & 0 & -I_m & 0 & 0 & D_Y & 0 & 0 & 0 & 0 & 0 \\ A & 0 & 0 & 0 & 0 & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ 0 & I_L^x & 0 & 0 & 0 & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ 0 & 0 & I_U^x & 0 & 0 & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & 0 & 0 & 0 & I_L^s & 0 & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & 0 & 0 & 0 & 0 & I_U^s & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix},$$

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

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

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 Δp may be written in the form $\Delta p = Nd$. This implies that d satisfies the reduced equations $N^{\mathrm{T}} B(p) N d = -N^{\mathrm{T}} \nabla M(p)$. Consider the null-space basis defined from the columns of

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

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

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

where $J_F = J(x)E_F^T$, $A_F = AE_F^T$, $\widehat{H}_F = E_F H^B E_F^T + \frac{2}{\mu^A} A_F^T A_F + \frac{2}{\mu^P} J_F^T J_F + 2\big(E_{LF}^T (D_1^z)^{-1} E_{LF} + E_{UF}^T (D_2^z)^{-1} E_{UF}\big)$ and $D_W = \big((L_L^T (D_1^W)^{-1} L_L + L_U^T (D_2^W)^{-1} L_U)\big)^\dagger$. Similarly, the reduced gradient $N^T \nabla M(p)$ is given by

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

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 $N^T B(p) N d = -N^T \nabla M(p)$ are then

$$\begin{pmatrix} \widehat{H}_F & -2J_F^T D_Y^{-1} L_F^T & J_F^T & A_F^T & E_{LF}^T & -E_{UF}^T & 0 & 0 \\ -2L_F D_Y^{-1} J_F & 2L_F(D_Y^{-1} + D_W^\dagger)L_F^T & -L_F & 0 & 0 & 0 & L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \\ d_8 \end{pmatrix}$$
$$= - \begin{pmatrix} g_F - A_F^T\big(2\pi^V - v\big) - J_F^T\big(2\pi^Y - y\big) - E_{LF}(2\pi_1^z - z_1) + E_{UF}(2\pi_2^z - z_2) \\ 2\pi_F^Y - y_F - L_{LF}(2\pi_1^W - w_1) + L_{UF}(2\pi_2^W - w_2) \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}. \quad (5.4)$$

Given any nonsingular matrix R, the direction d satisfies $RN^{\mathrm{T}}B(p)Nd = -RN^{\mathrm{T}}\nabla M(p)$. In particular, consider 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 again, 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

$$RN^{\mathrm{T}}B(p)N = \begin{pmatrix} E_F H^B E_F^{\mathrm{T}} & 0 & -J_F^{\mathrm{T}} & -A_F^{\mathrm{T}} & -E_{LF}^{\mathrm{T}} & E_{UF}^{\mathrm{T}} & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^{\mathrm{T}} & L_{UF}^{\mathrm{T}} \\ J_F & -L_F^{\mathrm{T}} & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix}$$

Also,

$$RN^{\mathrm{T}}\nabla M(p) = \begin{pmatrix} g_F - J_F^{\mathrm{T}}y - A_F^{\mathrm{T}}v - z_1 + z_2 \\ y_F - w_1 + w_2 \\ -D_Y(\pi^Y - y) \\ -D_A(\pi^V - v) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}.$$

This gives the following (unsymmetric) reduced approximate Newton equations for d:

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

Then, the identity $\Delta p = Nd$ 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 equations (5.5) in the form

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

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 (3.5) except for the (1, 1) block, where H_F is replaced by $E_F H^B E_F^T$.

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, we obtain

$$\begin{pmatrix} D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & D_2^Z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T & 0 & 0 & L_F \\ -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 & 0 & E_F H^B E_F^T & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & -L_F^T & J_F & D_Y \end{pmatrix} \begin{pmatrix} \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_A(v - \pi^V) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ y_F - L_{LF}^T w_1 + L_{UF}^T w_2 \\ g_F - J_F^T y - A_F^T v - E_{LF}^T z_1 + E_{UF}^T z_2 \\ D_Y(y - \pi^Y) \end{pmatrix}. \tag{6.1}$$

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

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

If equations (6.1) are premultiplied by the matrix

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

gives the block upper-triangular system

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

where $\widetilde{H}_F = E_F H^B E_F^T + A_F^T D_A^{-1} A_F + E_F D_Z^{\dagger} E_F^T$, $\pi_F^W = L_{LF}^T \pi_1^W - L_{UF}^T \pi_2^W$ and $\pi_F^Z = E_{LF}^T \pi_1^Z - E_{UF}^T \pi_2^Z$. Using block back-substitution, Δx_F and Δy can be computed by solving the equations

$$\begin{pmatrix} \widetilde{H}_F & -J_F^T \\ J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T \pi^V - \pi_F^Z \\ D_W\big(y - \pi^W\big) + D_Y\big(y - \pi^Y\big) \end{pmatrix}.$$

Once Δx_F and Δy 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

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

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 (3.5) to give

$$\Delta w_1 = -(S_1^\mu)^{-1}\big(w_1 \cdot (L_L(s+\Delta s) - \ell^S + \mu^B e) - \mu^B w_1^E + \mu^B L_L(s - s^E + \Delta s)\big),$$

and

$$\Delta w_2 = -(S_2^\mu)^{-1}\big(w_2 \cdot (u^S - L_U(s+\Delta s) + \mu^B e) - \mu^B w_2^E + \mu^B L_U(s^E - s - \Delta s)\big).$$

Similarly, using (3.5) to solve for Δz_1 and Δz_2 yields

$$\Delta z_1 = -(X_1^\mu)^{-1}\big(z_1 \cdot (E_L(x+\Delta x) - \ell^X + \mu^B e) - \mu^B z_1^E + \mu^B E_L(x - x^E + \Delta x)\big),$$

and

$$\Delta z_2 = -(X_2^\mu)^{-1}\big(z_2 \cdot (u^X - E_U(x+\Delta x) + \mu^B e) - \mu^B z_2^E + \mu^B E_U(x^E - x - \Delta x)\big).$$

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 primal-dual 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 $X_1^\mu = \text{diag}(E_L x - \ell^X + \mu^B e)$, $X_2^\mu = \text{diag}(u^X - E_U x + \mu^B e)$, $Z_1 = \text{diag}(z_1)$, $Z_2 = \text{diag}(z_2)$, $W_1 = \text{diag}(w_1)$, $W_2 = \text{diag}(w_2)$, $S_1^\mu = \text{diag}(L_L s - \ell^S + \mu^B e)$ and

$S_2^\mu = \operatorname{diag}(u^S - L_U s + \mu^B e)$. Consider the quantities

$$\begin{aligned} D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c-s), \\ D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax-b), \\ (D_1^Z)^{-1} &= (X_1^\mu)^{-1} Z_1^\mu, & (D_1^W)^{-1} &= (S_1^\mu)^{-1} W_1^\mu, \\ (D_2^Z)^{-1} &= (X_2^\mu)^{-1} Z_2^\mu, & (D_2^W)^{-1} &= (S_2^\mu)^{-1} W_2^\mu, \\ D_Z &= \big(E_L^T (D_1^Z)^{-1} E_L + E_U^T (D_2^Z)^{-1} E_U\big)^\dagger, & D_W &= \big(L_L^T (D_1^W)^{-1} L_L + L_U^T (D_2^W)^{-1} L_U\big)^\dagger, \\ \pi_1^Z &= \mu^B (X_1^\mu)^{-1}(z_1^E - x_1 + x_1^E), & \pi_1^W &= \mu^B (S_1^\mu)^{-1}(w_1^E - s_1 + s_1^E), \\ \pi_2^Z &= \mu^B (X_2^\mu)^{-1}(z_2^E - x_2 + x_2^E), & \pi_2^W &= \mu^B (S_2^\mu)^{-1}(w_2^E - s_2 + s_2^E), \\ \pi^Z &= E_L^T \pi_1^Z - E_U^T \pi_2^Z, & \pi^W &= L_L^T \pi_1^W - L_U^T \pi_2^W. \end{aligned}$$

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

$$\begin{pmatrix} H_F^B + A_F^T D_A^{-1} A_F + E_F D_Z^\dagger E_F^T & -J_F^T \\ J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g_F - J_F^T y - A_F^T \pi^V - \pi_F^Z \\ D_W\big(y_F - \pi_F^W\big) + D_Y\big(y - \pi^Y\big) \end{pmatrix},$$

and set

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

The associated merit function (4.1) can be written as

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

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 $g_N = \nabla M$ 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 [3] 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

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

and recover the solution of the secular equations (8.3) from the computed vector 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 $(x,\ x_1,\ x_2,\ s,\ s_1,\ s_2,\ y,\ v,\ z_1,\ z_2,\ w_1,\ w_2)$. In particular, we have

$$\begin{pmatrix} \widehat{H}_F & -2J_F^T D_Y^{-1} L_F^T & J_F^T & A_F^T & E_{LF}^T & -E_{UF}^T & 0 & 0 \\ -2L_F D_Y^{-1} J_F & 2L_F(D_Y^{-1} + D_W^\dagger) L_F^T & -L_F & 0 & 0 & 0 & L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix}$$

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

where

$$\widehat{H}_F = E_F H(x,y) E_F^T + \frac{2}{\mu^A} A_F^T A_F + \frac{2}{\mu^P} J_F^T J_F + 2\big(E_{LF}^T (D_1^Z)^{-1} E_{LF} + E_{UF}^T (D_2^Z)^{-1} E_{UF}\big),$$

with $H(x,y)$ the Hessian of the Lagrangian function, and

$$\begin{aligned} D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c - s), & D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax - b), \\ D_1^W &= S_1^\mu (W_1^\mu)^{-1}, & \pi_1^W &= \mu^B (S_1^\mu)^{-1}(w_1^E - s_1 + s_1^E), & D_1^Z &= X_1^\mu (Z_1^\mu)^{-1}, & \pi_1^Z &= \mu^B (X_1^\mu)^{-1}(z_1^E - x_1 + x_1^E), \\ D_2^W &= S_2^\mu (W_2^\mu)^{-1}, & \pi_2^W &= \mu^B (S_2^\mu)^{-1}(w_2^E - s_2 + s_2^E), & D_2^Z &= X_2^\mu (Z_2^\mu)^{-1}, & \pi_2^Z &= \mu^B (X_2^\mu)^{-1}(z_2^E - x_2 + x_2^E). \end{aligned}$$

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 Hessian of the merit function in terms of the 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

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

Similarly, let $\vec{T}_x = \mathrm{diag}(T^x, T^s)$ and $\vec{T}_y = \mathrm{diag}(T^y,\, T^v,\, T_1^z,\, T_2^z,\, T_1^w,\, T_2^w)$. The equations $(B_N + \sigma T)\Delta p = -g_N$ may be written in the form

$$\begin{pmatrix} \vec{H} + 2\vec{J}^{\mathrm{T}}\vec{D}^{-1}\vec{J} + \sigma\vec{T}_x & \vec{J}^{\mathrm{T}} \\ \vec{J} & \vec{D} + \sigma\vec{T}_y \end{pmatrix} \begin{pmatrix} \Delta\vec{x} \\ \Delta\vec{y} \end{pmatrix} = -\begin{pmatrix} \vec{g} - \vec{J}^{\mathrm{T}}\vec{\pi} - \vec{J}^{\mathrm{T}}(\vec{\pi} - \vec{y}) \\ -\vec{D}(\vec{\pi} - \vec{y}) \end{pmatrix}, \tag{8.5}$$

where

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

Applying the nonsingular matrix $\begin{pmatrix} I & -2\vec{J}^{\mathrm{T}}\vec{D}^{-1} \\ & I \end{pmatrix}$ to both sides of (8.5) 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 [1], 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

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

where $\bar{\sigma} = (1+\sigma)/(1+2\sigma)$. In terms of the original variables, the unsymmetric equations (8.6) are

$$\begin{pmatrix} H_F + \sigma I_F^x & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & \sigma I_F^s & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & \bar{\sigma} D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & \bar{\sigma} D_A & 0 & 0 & 0 & 0 \\ E_{LF} & 0 & 0 & 0 & \bar{\sigma} D_1^Z & 0 & 0 & 0 \\ -E_{UF} & 0 & 0 & 0 & 0 & \bar{\sigma} D_2^Z & 0 & 0 \\ 0 & L_{LF} & 0 & 0 & 0 & 0 & \bar{\sigma} D_1^W & 0 \\ 0 & -L_{UF} & 0 & 0 & 0 & 0 & 0 & \bar{\sigma} D_2^W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ (1+2\sigma)\Delta y \\ (1+2\sigma)\Delta v \\ (1+2\sigma)\Delta z_1 \\ (1+2\sigma)\Delta z_2 \\ (1+2\sigma)\Delta w_1 \\ (1+2\sigma)\Delta w_2 \end{pmatrix}$$
$$= - \begin{pmatrix} E_F\big(g - J^T y - A^T v - z\big) \\ L_F\big(y - w\big) \\ c(x) - s + \mu^P (y - y^E) \\ Ax - b + \mu^A (v - v^E) \\ (Z_1^\mu)^{-1}\big(z_1 \cdot x_1 + \mu^B (z_1 - z_1^E + x_1 - x_1^E)\big) \\ (Z_2^\mu)^{-1}\big(z_2 \cdot x_2 + \mu^B (z_2 - z_2^E + x_2 - x_2^E)\big) \\ (W_1^\mu)^{-1}\big(w_1 \cdot s_1 + \mu^B (w_1 - w_1^E + s_1 - s_1^E)\big) \\ (W_2^\mu)^{-1}\big(w_2 \cdot s_2 + \mu^B (w_2 - w_2^E + s_2 - s_2^E)\big) \end{pmatrix}, \tag{8.7}$$

where $\bar{\sigma} = (1+\sigma)/(1+2\sigma)$. Collecting terms and reordering the equations and unknowns, we obtain

$$\begin{pmatrix} \bar{\sigma} D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & \bar{\sigma} D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & \bar{\sigma} D_2^Z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & \bar{\sigma} D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{\sigma} D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T & \sigma I_F^s & 0 & L_F \\ -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 & 0 & H_F + \sigma I_F^x & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & -L_F^T & J_F & \bar{\sigma} D_Y \end{pmatrix} \begin{pmatrix} \Delta\widetilde{v} \\ \Delta\widetilde{z}_1 \\ \Delta\widetilde{z}_2 \\ \Delta\widetilde{w}_1 \\ \Delta\widetilde{w}_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta\widetilde{y} \end{pmatrix} = - \begin{pmatrix} D_A(v - \pi^V) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ L_F\big(y - w\big) \\ E_F\big(g - J^T y - A^T v - z\big) \\ D_Y(y - \pi^Y) \end{pmatrix}, \tag{8.8}$$

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 (8.8) by the block lower-triangular matrix

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

gives the block upper-triangular system

$$\begin{pmatrix} \bar\sigma D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & \bar\sigma D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & \bar\sigma D_2^Z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & \bar\sigma D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar\sigma D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\bar\sigma} L_F \breve{D}_W^\dagger L_F^T & 0 & L_F \\ 0 & 0 & 0 & 0 & 0 & 0 & \widetilde{H}_F + \sigma I_F^x & -J_F^T \\ 0 & 0 & 0 & 0 & 0 & 0 & J_F & \bar\sigma(D_Y + \breve{D}_W) \end{pmatrix} \begin{pmatrix} \Delta\widetilde{v} \\ \Delta\widetilde{z}_1 \\ \Delta\widetilde{z}_2 \\ \Delta\widetilde{w}_1 \\ \Delta\widetilde{w}_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta\widetilde{y} \end{pmatrix}$$

$$= -\begin{pmatrix} D_A(v - \pi^V) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ L_F\Big(y - w + \dfrac{1}{\bar\sigma}\big[w - \pi^W\big]\Big) \\ E_F\Big(g - J^T y - A^T v - z + \dfrac{1}{\bar\sigma}\big[A^T(v - \pi^V) + z - \pi^Z\big]\Big) \\ D_Y(y - \pi^Y) + \breve{D}_W\big(\bar\sigma(y - w) + w - \pi^W\big) \end{pmatrix},$$

where

$$\widetilde{H}_F = E_F\Big(H(x,y) + \frac{1}{\bar{\sigma}}A^T D_A^{-1} A + \frac{1}{\bar{\sigma}} D_Z^{\dagger}\Big)E_F^T,$$

$w = L_X^T w_X + L_L^T w_1 - L_U^T w_2$, $z = E_X^T z_X + E_L^T z_1 - E_U^T z_2$, $\pi^W = L_L^T \pi_1^W - L_U^T \pi_2^W$ 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

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

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 (8.7) to give

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

Similarly, using (8.7) to solve for Δz_1 and Δz_2 yields

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

Similarly, using the first block of equations (8.8) 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 $\Delta z_X = [\,g + H\Delta x - J^T(y + \Delta y) - z\,]_X$.

9. Summary: equations for the trust-region direction

The results of the preceding section implies that the solution of the secular equations $(\bar{B}_N + \sigma I)\Delta v_M = -\bar{g}_N$, with σ a nonnegative scalar, $\bar{B}_N = T^{-1/2} B_N(p) T^{-1/2}$, and $\bar{g}_N = T^{-1/2} g_N$ 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

$$
\begin{aligned}
D_Y &= \mu^P I_m, & \pi^Y &= y^E - \frac{1}{\mu^P}(c - s),\\
D_A &= \mu^A I_A, & \pi^V &= v^E - \frac{1}{\mu^A}(Ax - b),\\
(D_1^Z)^{-1} &= (X_1^\mu)^{-1} Z_1^\mu, & (D_1^W)^{-1} &= (S_1^\mu)^{-1} W_1^\mu,\\
(D_2^Z)^{-1} &= (X_2^\mu)^{-1} Z_2^\mu, & (D_2^W)^{-1} &= (S_2^\mu)^{-1} W_2^\mu,\\
D_Z &= \big(E_L^T (D_1^Z)^{-1} E_L + E_U^T (D_2^Z)^{-1} E_U\big)^\dagger, & D_W &= \big(L_L^T (D_1^W)^{-1} L_L + L_U^T (D_2^W)^{-1} L_U\big)^\dagger,\\
& & \breve{D}_W &= \big(D_W^\dagger + \sigma L_F^T L_F\big)^\dagger,\\
\pi_1^Z &= \mu^B (X_1^\mu)^{-1}(z_1^E - x_1 + x_1^E), & \pi_1^W &= \mu^B (S_1^\mu)^{-1}(w_1^E - s_1 + s_1^E),\\
\pi_2^Z &= \mu^B (X_2^\mu)^{-1}(z_2^E - x_2 + x_2^E), & \pi_2^W &= \mu^B (S_2^\mu)^{-1}(w_2^E - s_2 + s_2^E),\\
\pi^Z &= E_L^T \pi_1^Z - E_U^T \pi_2^Z, & \pi^W &= L_L^T \pi_1^W - L_U^T \pi_2^W,
\end{aligned}
$$

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

$$\Delta x = E_F^{\mathrm{T}} \Delta x_F, \qquad \widehat{x} = x + \Delta x, \qquad \Delta z_1 = -\frac{1}{1+\sigma}(X_1^{\mu})^{-1}\big(z_1 \cdot (E_L \widehat{x} - \ell^X + \mu^B e) - \mu^B z_1^E + \mu^B L_L (s - s^E + \Delta s)\big),$$

$$\Delta z_2 = -\frac{1}{1+\sigma}(X_2^{\mu})^{-1}\big(z_2 \cdot (u^X - E_U \widehat{x} + \mu^B e) - \mu^B z_2^E + \mu^B L_U (s^E - s - \Delta s)\big),$$

$$\Delta y = \Delta \widetilde{y}/(1+2\sigma), \qquad \widehat{y} = y + \Delta y, \qquad \Delta s = -\bar{\sigma} \breve{D}_W \Big(y + (1+2\sigma)\Delta y - w + \frac{1}{\bar{\sigma}}\big[w - \pi^W\big]\Big),$$

$$\widehat{s} = s + \Delta s, \qquad \Delta w_1 = -\frac{1}{1+\sigma}(S_1^{\mu})^{-1}\big(w_1 \cdot (L_L \widehat{s} - \ell^S + \mu^B e) - \mu^B w_1^E + \mu^B L_L (s - s^E + \Delta s)\big),$$

$$\Delta w_2 = -\frac{1}{1+\sigma}(S_2^{\mu})^{-1}\big(w_2 \cdot (u^S - L_U \widehat{s} + \mu^B e) - \mu^B w_2^E + \mu^B L_U (s^E - s - \Delta s)\big),$$

$$\widehat{\pi}^V = v^E - \frac{1}{\mu^A}(A\widehat{x} - b), \qquad \Delta v = -\frac{1}{1+\sigma}\big(v - \widehat{\pi}^V\big),$$

$$w = L_X^{\mathrm{T}} w_X + L_L^{\mathrm{T}} w_1 - L_U^{\mathrm{T}} w_2, \qquad z = E_X^{\mathrm{T}} z_X + E_L^{\mathrm{T}} z_1 - E_U^{\mathrm{T}} z_2,$$

$$\widehat{v} = v + \Delta v, \qquad \Delta w_X = [\,\widehat{y} - w\,]_X,$$

$$\Delta z_X = [\,g + H\Delta x - J^{\mathrm{T}}\widehat{y} - z\,]_X.$$

References

[1] 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

[2] P. E. Gill and M. Zhang. A projected-search path-following method for nonlinear optimization. Center for Computational Mathematics Report CCoM 22-01, Center for Computational Mathematics, University of California, San Diego, La Jolla, CA, 2022. 1, 2

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