NOTE ON THE FORMULATION OF A SHIFTED PRIMAL-DUAL PENALTY-BARRIER METHOD FOR NONLINEAR OPTIMIZATION

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

UCSD Center for Computational Mathematics
Technical Report CCoM-19-4
March 1, 2019

Abstract

The modified Newton equations for the shifted primal-dual penalty-barrier method are derived for a nonlinearly constrained problems with various constraint types.

Key words. Nonlinear programming, nonlinear constraints, augmented Lagrangian methods, interior methods, path-following methods, regularized methods, primal-dual 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 Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682 (daniel.p.robinson@jhu.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 derives the shifted primal-dual penalty-barrier merit functions and associated path-following equations for an optimization problem with constraints written in eight different ways:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; Ax=b, \quad \ell \le x \le u, \tag{1.1}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad s \ge 0, \tag{1.2}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad L_{\scriptscriptstyle X} s = h_{\scriptscriptstyle X}, \quad L_{\scriptscriptstyle F} s \ge 0, \tag{1.3}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad Ax=b, \quad x \ge 0, \quad s \ge 0, \tag{1.4}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad L_{\scriptscriptstyle X} s = h_{\scriptscriptstyle X}, \quad \ell \le L_{\scriptscriptstyle F} s \le u, \tag{1.5}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad s \ge 0, \quad E_{\scriptscriptstyle X} x = b_{\scriptscriptstyle X}, \quad \ell \le E_{\scriptscriptstyle F} x \le u. \tag{1.6}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad \ell^{\scriptscriptstyle X} \le x \le u^{\scriptscriptstyle X}, \quad \ell^{\scriptscriptstyle S} \le s \le u^{\scriptscriptstyle S}. \tag{1.7}$$

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad Ax=b, \quad \ell^{\scriptscriptstyle X} \le x \le u^{\scriptscriptstyle X}, \quad \ell^{\scriptscriptstyle S} \le s \le u^{\scriptscriptstyle S}. \tag{1.8}$$

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 linear constraints $Ax = b$ are imposed using a shifted primal-dual penalty method. In practice, the constraints involving A are used to temporarily fix a subset of the variables at their bounds, in which case the rows of A are rows of the identity matrix. The constraints $E_X x = b_X$, and $L_X s = h_X$ are also used to fix a subset of the variables and slacks. However, in this case, the constraints are imposed directly. All inequality constraints are imposed indirectly using a shifted primal-dual barrier function.

The equations for the eight problem formats are summarized in Sections 2.6, 3.6, 4.6, 5.6, 6.6, 7.6, 8.6 and 9.6 respectively. The structure of these equations allows us to write down the equations for the general problem

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

The equations and merit function for this general problem are given in Section 10.6.

Convergence results for Problem (1.2) are given by Gill, Kungurtsev and Robinson [1].

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 $\|\cdot\|$. The vector $g(x)$ is used to denote $\nabla f(x)$, the gradient of $f(x)$. 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 $H(x, y) = \nabla^2 f(x) - \sum_{i=1}^{m} y_i \nabla^2 c_i(x)$.

2. Linear Equality Constraints and Upper and Lower Bounds on the Variables

Next we consider methods for an optimization problem with linear equality constraints and upper and lower bounds on the variables.

2.1. Problem statement and optimality conditions

The problem has the form

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; f(x) \quad \text{subject to} \quad Ax = b, \quad \ell \le x \le u, \tag{2.1}$$

where $f:\mathbb{R}^n \mapsto \mathbb{R}$ is twice-continuously differentiable. The first-order KKT conditions for this problem are

$$Ax^* = b, \qquad x^* - \ell \ge 0, \qquad u - x^* \ge 0, \tag{2.2a}$$
$$g(x^*) - z_1^* + z_2^* - A^Tv^* = 0, \qquad z_1^* \ge 0, \qquad z_2^* \ge 0, \tag{2.2b}$$
$$z_1^* \cdot (x^* - \ell) = 0, \qquad z_2^* \cdot (u - x^*) = 0. \tag{2.2c}$$

The n-vectors z_1^* and z_2^* may be interpreted as Lagrange multipliers for the inequality constraints $x - \ell \ge 0$ and $u - x \ge 0$, respectively. The vector v^* is the multiplier vector for the linear equality constraints.

2.2. The path-following equations

Let z_1^E and z_2^E denote n-vectors of nonnegative estimates of the Lagrange multipliers for the inequality constraints $x - \ell \ge 0$ and $u - x \ge 0$, respectively. Let v^E denote an estimate of v^*. Given a small positive scalars μ^B and μ^A, consider the perturbed optimality conditions

$$Ax - b = \mu^A(v^E - v), \qquad x - \ell \ge 0, \qquad u - x \ge 0, \tag{2.3a}$$
$$g(x) - z_1 + z_2 - A^Tv = 0, \qquad z_1 \ge 0, \qquad z_2 \ge 0, \tag{2.3b}$$
$$z_1 \cdot (x - \ell) = \mu^B(z_1^E - z_1), \qquad z_2 \cdot (u - x) = \mu^B(z_2^E - z_2). \tag{2.3c}$$

Consider the primal-dual path parameterized by μ^B consisting of points (x, z_1, z_2) such that $F(x, v, z_1, z_2\,; \mu^B, \mu^A, v^E, z_1^E, z_2^E) = 0$, where

$$F(x, v, z_1, z_2\,; \mu^B, \mu^A, v^E, z_1^E, z_2^E) = \begin{pmatrix} g(x) - z_1 + z_2 - A^Tv \\ Ax - b + \mu^A(v - v^E) \\ z_1 \cdot (x - \ell) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u - x) + \mu^B(z_2 - z_2^E) \end{pmatrix}. \tag{2.4}$$

Any zero (x, v, z_1, z_2) of F that satisfies $z_1 > 0$ and $z_2 > 0$ approximates a point satisfying the optimality conditions (2.2), with the approximation becoming increasingly accurate as $\mu^A(v - v^E) \to 0$, $\mu^B(z_1 - z_1^E) \to 0$ and $\mu^B(z_2 - z_2^E) \to 0$. For any sequence of v^E, z_1^E and z_2^E such that $v^E \to v^*$, $z_1^E \to z_1^*$ and $z_2^E \to z_2^*$, and it must hold that solutions $(x,\, v,\, z_1,\, z_2)$ of (2.3) must satisfy $Ax - b \to 0$, $z_1 \cdot (x - \ell) \to 0$ and $z_2 \cdot (u - x) \to 0$. This implies that a solution $(x,\, v,\, z_1,\, z_2)$ of (2.3) will approximate a solution of (2.2) independently of the values of μ^A and μ^B (i.e., it is not necessary that μ^B, $\mu^A \to 0$).

If (x, v, z_1, z_2) is a given approximate zero of F such that $x - \ell + \mu^B e > 0$, $u - x + \mu^B e > 0$, $z_1 > 0$ and $z_2 > 0$, the Newton equations for the change in variables $(\Delta x,\, \Delta v,\, \Delta z_1,\, \Delta z_2)$ are given by

$$\begin{pmatrix} H & -A^T & -I & I \\ A & \mu^A I & 0 & 0 \\ Z_1 & 0 & X_1^\mu & 0 \\ -Z_2 & 0 & 0 & X_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \end{pmatrix} = - \begin{pmatrix} g - A^T v - z_1 + z_2 \\ Ax - b + \mu^A (v - v^E) \\ z_1 \cdot (x - \ell) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u - x) + \mu^B (z_2 - z_2^E) \end{pmatrix}, \tag{2.5}$$

where $X_1^\mu = \operatorname{diag}\bigl(x_j - \ell_j + \mu^B\bigr)$, $X_2^\mu = \operatorname{diag}\bigl(u_j - x_j + \mu^B\bigr)$, $Z_1 = \operatorname{diag}([z_1]_j)$, and $Z_2 = \operatorname{diag}([z_2]_j)$.

2.3. A shifted primal-dual penalty-barrier function

Problem (2.1) is equivalent to

$$\begin{aligned} \underset{x,x_1,x_2}{\text{minimize}} \;\; & f(x) \\ \text{subject to} \;\; & Ax = b, \qquad x - x_1 = \ell, \qquad x_1 \ge 0, \\ & x + x_2 = u, \qquad x_2 \ge 0. \end{aligned}$$

Consider the shifted primal-dual penalty-barrier problem

$$\begin{aligned} \underset{x,x_1,x_2,v,z_1,z_2}{\text{minimize}} \;\; & M(x, x_1, x_2, v, z_1, z_2\,; \mu^B, \mu^A, v^E, z_1^E, z_2^E) \\ \text{subject to} \;\; & x - x_1 = \ell, \qquad x_1 + \mu^B e > 0, \qquad z_1 > 0, \\ & x + x_2 = u, \qquad x_2 + \mu^B e > 0, \qquad z_2 > 0, \end{aligned} \tag{2.6}$$

where $M(x, x_1, x_2, v, z_1, z_2\,; \mu^B, \mu^A, v^E, z_1^E, z_2^E)$ is the penalty-barrier function

$$\begin{aligned} f(x) - (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A (v - v^E)\|^2 & \\ - \sum_{j=1}^n \Bigl\{ \mu^B [z_1^E]_j \ln\bigl([x_1]_j + \mu^B\bigr) + \mu^B [z_1^E]_j \ln\bigl([z_1]_j([x_1]_j + \mu^B)\bigr) - [z_1]_j([x_1]_j + \mu^B) \Bigr\} & \\ - \sum_{j=1}^n \Bigl\{ \mu^B [z_2^E]_j \ln\bigl([x_2]_j + \mu^B\bigr) + \mu^B [z_2^E]_j \ln\bigl([z_2]_j([x_2]_j + \mu^B)\bigr) - [z_2]_j\bigl([x_2]_j + \mu^B\bigr) \Bigr\}. & \end{aligned} \tag{2.7}$$

Differentiating $M(x, x_1, x_2, v, z_1, z_2)$ with respect to x, x_1, x_2, v, z_1, and z_2 gives

$$\nabla M(x, x_1, x_2, v, z_1, z_2) = \begin{pmatrix} g - A^T v^{\scriptscriptstyle E} + \frac{1}{\mu^A} A^T(Ax - b) + \frac{1}{\mu^A} A^T\big(Ax - b + \mu^{\scriptscriptstyle A}(v - v^{\scriptscriptstyle E})\big) \\ z_1 - 2\mu^{\scriptscriptstyle B}(X_1^\mu)^{-1} z_1^{\scriptscriptstyle E} \\ z_2 - 2\mu^{\scriptscriptstyle B}(X_2^\mu)^{-1} z_2^{\scriptscriptstyle E} \\ Ax - b + \mu^{\scriptscriptstyle A}(v - v^{\scriptscriptstyle E}) \\ x_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_1^{-1} z_1^{\scriptscriptstyle E} \\ x_2 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_2^{-1} z_2^{\scriptscriptstyle E} \end{pmatrix},$$

where $X_1^\mu = \operatorname{diag}\big(x_1 + \mu^{\scriptscriptstyle B} e\big) = \operatorname{diag}\big(x_j - \ell_j + \mu^{\scriptscriptstyle B}\big)$ and $X_2^\mu = \operatorname{diag}\big(x_2 + \mu^{\scriptscriptstyle B} e\big) = \operatorname{diag}\big(u_j - x_j + \mu^{\scriptscriptstyle B}\big)$.

Vectors of the form $x_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_1^{-1} z_1^{\scriptscriptstyle E}$ may be written as

$$\begin{aligned} x_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_1^{-1} z_1^{\scriptscriptstyle E} = Z_1^{-1}\big(Z_1(x_1 + \mu^{\scriptscriptstyle B} e) - \mu^{\scriptscriptstyle B} z_1^{\scriptscriptstyle E}\big) &= Z_1^{-1}\big(Z_1 x_1 + \mu^{\scriptscriptstyle B} z_1 - \mu^{\scriptscriptstyle B} z_1^{\scriptscriptstyle E}\big) \\ &= Z_1^{-1}\big(z_1 \cdot (x - \ell) + \mu^{\scriptscriptstyle B}(z_1 - z_1^{\scriptscriptstyle E})\big). \end{aligned} \tag{2.8}$$

Similarly,

$$\begin{aligned} x_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_1^{-1} z_1^{\scriptscriptstyle E} = Z_1^{-1}\big(Z_1(x_1 + \mu^{\scriptscriptstyle B} e) - \mu^{\scriptscriptstyle B} z_1^{\scriptscriptstyle E}\big) = Z_1^{-1}\big(X_1^\mu z_1 - \mu^{\scriptscriptstyle B} z_1^{\scriptscriptstyle E}\big) &= Z_1^{-1} X_1^\mu\big(z_1 - \mu^{\scriptscriptstyle B}(X_1^\mu)^{-1} z_1^{\scriptscriptstyle E}\big) \\ &= D_1^{\scriptscriptstyle Z}\big(z_1 - \pi_1^{\scriptscriptstyle Z}\big), \end{aligned} \tag{2.9}$$

where $D_1^{\scriptscriptstyle Z} = X_1^\mu Z_1^{-1} = Z_1^{-1} X_1^\mu$ and $\pi_1^{\scriptscriptstyle Z} = \mu^{\scriptscriptstyle B}(X_1^\mu)^{-1} z_1^{\scriptscriptstyle E}$. Analogous identities hold for $x_2 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_2^{-1} z_2^{\scriptscriptstyle E}$.

The identities above imply that the gradient may be written in several equivalent forms

$$\begin{aligned} \nabla M(x, x_1, x_2, z_1, z_2) &= \begin{pmatrix} g - A^T v^{\scriptscriptstyle E} + \frac{1}{\mu^A} A^T(Ax - b) + \frac{1}{\mu^A} A^T\big(Ax - b + \mu^{\scriptscriptstyle A}(v - v^{\scriptscriptstyle E})\big) \\ z_1 - 2\mu^{\scriptscriptstyle B}(X_1^\mu)^{-1} z_1^{\scriptscriptstyle E} \\ z_2 - 2\mu^{\scriptscriptstyle B}(X_2^\mu)^{-1} z_2^{\scriptscriptstyle E} \\ Ax - b + \mu^{\scriptscriptstyle A}(v - v^{\scriptscriptstyle E}) \\ x_1 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_1^{-1} z_1^{\scriptscriptstyle E} \\ x_2 + \mu^{\scriptscriptstyle B} e - \mu^{\scriptscriptstyle B} Z_2^{-1} z_2^{\scriptscriptstyle E} \end{pmatrix} = \begin{pmatrix} g - A^T\big(2(v^{\scriptscriptstyle E} + \frac{1}{\mu^A}(Ax - b)) - v\big) \\ (X_1^\mu)^{-1}\big(z_1 \cdot x_1 - \mu^{\scriptscriptstyle B} z_1^{\scriptscriptstyle E} + \mu^{\scriptscriptstyle B}(z_1 - z_1^{\scriptscriptstyle E})\big) \\ (X_2^\mu)^{-1}\big(z_2 \cdot x_2 - \mu^{\scriptscriptstyle B} z_2^{\scriptscriptstyle E} + \mu^{\scriptscriptstyle B}(z_2 - z_2^{\scriptscriptstyle E})\big) \\ -\mu^{\scriptscriptstyle A}\big(v^{\scriptscriptstyle E} - \frac{1}{\mu^A}(Ax - b) - v\big) \\ Z_1^{-1}\big(z_1 \cdot x_1 + \mu^{\scriptscriptstyle B}(z_1 - z_1^{\scriptscriptstyle E})\big) \\ Z_2^{-1}\big(z_2 \cdot x_2 + \mu^{\scriptscriptstyle B}(z_2 - z_2^{\scriptscriptstyle E})\big) \end{pmatrix} \\ &= \begin{pmatrix} g - A^T(2\pi^{\scriptscriptstyle A} - v) \\ -(2\pi_1^{\scriptscriptstyle Z} - z_1) \\ -(2\pi_2^{\scriptscriptstyle Z} - z_2) \\ -D_{\scriptscriptstyle A}(\pi^{\scriptscriptstyle A} - v) \\ -D_1^{\scriptscriptstyle Z}(\pi_1^{\scriptscriptstyle Z} - z_1) \\ -D_2^{\scriptscriptstyle Z}(\pi_2^{\scriptscriptstyle Z} - z_2) \end{pmatrix}, \end{aligned}$$

where

$$D_A = \mu^A I, \qquad \pi^A = v^E - \frac{1}{\mu^A}(Ax - b), \tag{2.10a}$$

$$D_1^z = X_1^\mu Z_1^{-1}, \qquad \pi_1^z = \mu^B (X_1^\mu)^{-1} z_1^E, \tag{2.10b}$$

$$D_2^z = X_2^\mu Z_2^{-1}, \qquad \pi_2^z = \mu^B (X_2^\mu)^{-1} z_2^E. \tag{2.10c}$$

Similarly, the Hessian of $M(x, x_1, x_2, v, z_1, z_2)$ is given by

$$\begin{pmatrix} H + \frac{2}{\mu^A}A^TA & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\mu^B (X_1^\mu)^{-2} Z_1^E & 0 & 0 & I & 0 \\ 0 & 0 & 2\mu^B (X_2^\mu)^{-2} Z_2^E & 0 & 0 & I \\ 0 & 0 & 0 & D_A & 0 & 0 \\ 0 & I & 0 & 0 & \mu^B Z_1^{-2} Z_1^E & 0 \\ 0 & 0 & I & 0 & 0 & \mu^B Z_2^{-2} Z_2^E \end{pmatrix},$$

where $H = \nabla^2 f$. Substituting $\mu^B Z_1^E = X_1^\mu \Pi_1^z$ and $\mu^B Z_2^E = X_2^\mu \Pi_2^z$ from (2.10) gives the Hessian

$$\begin{pmatrix} H + \frac{2}{\mu^A}A^TA & 0 & 0 & A^T & 0 & 0 \\ 0 & 2(X_1^\mu)^{-1} \Pi_1^z & 0 & 0 & I & 0 \\ 0 & 0 & 2(X_2^\mu)^{-1} \Pi_2^z & 0 & 0 & I \\ A & 0 & 0 & D_A & 0 & 0 \\ 0 & I & 0 & 0 & X_1^\mu Z_1^{-2} \Pi_1^z & 0 \\ 0 & 0 & I & 0 & 0 & X_2^\mu Z_2^{-2} \Pi_2^z \end{pmatrix}.$$

2.4. Derivation of the shifted primal-dual penalty-barrier direction

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

$$\underset{p \in \mathcal{I}}{\text{minimize}} \;\; M(p) \quad \text{subject to} \quad Cp = b_C, \tag{2.11}$$

where

$$\mathcal{I} = \{p : p = (x, x_1, x_2, v, z_1, z_2), \;\; \text{with } x_1 + \mu^B e > 0,\; x_2 + \mu^B e > 0,\; z_1 > 0,\; z_2 > 0\},$$

and

$$C = \begin{pmatrix} I & -I & 0 & 0 & 0 & 0 \\ I & 0 & I & 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad b_C = \begin{pmatrix} \ell \\ u \end{pmatrix}.$$

Let $p \in \mathcal{I}$ be given. As in the bounded slack case, assume that p is not necessarily feasible for the linear constraints, i.e., it may not hold that $x - x_1 = \ell$ and $x + x_2 = u$, in which case $b_C - Cp$ may not be zero. The Newton direction Δp is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T \nabla^2 M(p)\Delta p \quad \text{subject to} \;\; C\Delta p = b_C - Cp. \tag{2.12}$$

However, instead of solving (2.12), we define a linearly constrained modified Newton method by approximating the Hessian $\nabla^2 M(x, x_1, x_2, v, z_1, z_2)$ by a matrix $B(x, x_1, x_2, v, z_1, z_2)$. Consider the matrix defined by replacing π_1^z by z_1 and π_2^z by z_2 everywhere in the matrix $\nabla^2 M(x, x_1, x_2, v, z_1, z_2)$. This gives an approximate Hessian

$$B(x, x_1, x_2, v, z_1, z_2) = \begin{pmatrix} H + \frac{2}{\mu^A} A^T A & 0 & 0 & A^T & 0 & 0 \\ 0 & 2(X_1^\mu)^{-1} Z_1 & 0 & 0 & I & 0 \\ 0 & 0 & 2(X_2^\mu)^{-1} Z_2 & 0 & 0 & I \\ A & 0 & 0 & D_A & 0 & 0 \\ 0 & I & 0 & 0 & X_1^\mu Z_1^{-1} & 0 \\ 0 & 0 & I & 0 & 0 & X_2^\mu Z_2^{-1} \end{pmatrix}.$$

The definitions of D_A, D_1^z and D_2^z (2.10) may be used to write $B(x, x_1, x_2, v, z_1, z_2)$ in the form

$$\begin{pmatrix} H + 2A^T D_A^{-1} A & 0 & 0 & A^T & 0 & 0 \\ 0 & 2(D_1^z)^{-1} & 0 & 0 & I & 0 \\ 0 & 0 & 2(D_2^z)^{-1} & 0 & 0 & I \\ A & 0 & 0 & D_A & 0 & 0 \\ 0 & I & 0 & 0 & D_1^z & 0 \\ 0 & 0 & I & 0 & 0 & D_2^z \end{pmatrix}.$$

Given $B(p) = B(x, x_1, x_2, v, z_1, z_2)$, a modified Newton direction is given by the solution of the QP subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T B(p)\Delta p \quad \text{subject to} \;\; C\Delta p = b_C - Cp. \tag{2.13}$$

Let N denote a matrix whose columns form a basis for null(C), i.e., the columns of N are linearly independent and $CN = 0$. The vector

$$\Delta p_0 = \begin{pmatrix} 0 \\ -(\ell - x + x_1) \\ (u - x - x_2) \\ 0 \\ 0 \\ 0 \end{pmatrix} \triangleq \begin{pmatrix} 0 \\ -r_L \\ r_U \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.14}$$

satisfies $C\Delta p_0 = b_C - Cp$, and every feasible Δp may be written in the form

$$\Delta p = \Delta p_0 + Nd.$$

This implies that d satisfies the reduced equations

$$N^T B(p) N d = -N^T\big(\nabla M(p) + B(p)\Delta p_0\big).$$

Consider the null-space basis defined from the columns of

$$N = \begin{pmatrix} I & 0 & 0 & 0 \\ I & 0 & 0 & 0 \\ -I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}. \tag{2.15}$$

The definition of N of (2.15) gives the reduced Hessian

$$N^T B(p) N = \begin{pmatrix} H + 2A^T D_A^{-1} A + 2\big((D_1^z)^{-1} + (D_2^z)^{-1}\big) & A^T & I & -I \\ A & D_A & 0 & 0 \\ I & 0 & D_1^z & 0 \\ -I & 0 & 0 & D_2^z \end{pmatrix}.$$

Similarly,

$$N^T \nabla M(p) = N^T \begin{pmatrix} g - A^T(2\pi^A - v) \\ -(2\pi_1^z - z_1) \\ -(2\pi_2^z - z_2) \\ -D_A(\pi^A - v) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \end{pmatrix} = \begin{pmatrix} g - A^T(2\pi^A - v) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) \\ -D_A(\pi^A - v) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \end{pmatrix},$$

and

$$N^T B(p)\Delta p_0 = N^T \begin{pmatrix} 0 \\ -2(D_1^z)^{-1} r_L \\ 2(D_2^z)^{-1} r_U \\ 0 \\ -r_L \\ r_U \end{pmatrix} = \begin{pmatrix} -2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\ 0 \\ -r_L \\ r_U \end{pmatrix},$$

where $r_L = \ell - x + x_1$ and $r_U = u - x - x_2$. This gives the reduced gradient

$$N^T\big(\nabla M(p) + B(p)\Delta p_0\big) = \begin{pmatrix} g - A^T(2\pi^A - v) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) - 2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\ -D_A(\pi^A - v) \\ -D_1^z(\pi_1^z - z_1) - r_L \\ -D_2^z(\pi_2^z - z_2) + r_U \end{pmatrix}.$$

The reduced modified Newton equations $N^T B(p) N d = -N^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ are then

$$\begin{pmatrix} H + 2A^T D_A^{-1} A + 2\big((D_1^z)^{-1} + (D_2^z)^{-1}\big) & A^T & I & -I \\ A & D_A & 0 & 0 \\ I & 0 & D_1^z & 0 \\ -I & 0 & 0 & D_2^z \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}$$
$$= \begin{pmatrix} g - A^T(2\pi^A - v) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) - 2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\ -D_A(\pi^A - v) \\ -D_1^z(\pi_1^z - z_1) - r_L \\ -D_2^z(\pi_2^z - z_2) + r_U \end{pmatrix}.$$

Given any nonsingular matrix R, the direction d satisfies

$$RN^T B(p) N d = -RN^T\big(\nabla M(p) + B(p)\Delta p_0\big).$$

In particular, as Z_1 and Z_2 are positive definite, the block upper-triangular matrix

$$R = \begin{pmatrix} I & -2A^T D_A^{-1} & -2(D_1^z)^{-1} & 2(D_2^z)^{-1} \\ & I & 0 & 0 \\ & & Z_1 & 0 \\ & & & Z_2 \end{pmatrix},$$

is nonsingular, with

$$RN^T B(p) N = \begin{pmatrix} H & -A^T & -I & I \\ A & D_A & 0 & 0 \\ Z_1 & 0 & Z_1 D_1^z & 0 \\ -Z_2 & 0 & 0 & Z_2 D_2^z \end{pmatrix} = \begin{pmatrix} H & -A^T & -I & I \\ A & D_A & 0 & 0 \\ Z_1 & 0 & X_1^\mu & 0 \\ -Z_2 & 0 & 0 & X_2^\mu \end{pmatrix}.$$

Also, $RN^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ is given by

$$\begin{pmatrix} I & -2A^TD_A^{-1} & -2(D_1^z)^{-1} & 2(D_2^z)^{-1} \\ & I & 0 & 0 \\ & & Z_1 & 0 \\ & & & Z_2 \end{pmatrix} \begin{pmatrix} g - A^T(2\pi^A - v) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) - 2\big((D_1^z)^{-1}r_L + (D_2^z)^{-1}r_U\big) \\ -D_A(\pi^A - v) \\ -D_1^z(\pi_1^z - z_1) - r_L \\ -D_2^z(\pi_2^z - z_2) + r_U \end{pmatrix}$$
$$= \begin{pmatrix} g - A^Tv - z_1 + z_2 \\ -D_A(\pi^A - v) \\ -Z_1D_1^z(\pi_1^z - z_1) - Z_1r_L \\ -Z_2D_2^z(\pi_2^z - z_2) + Z_2r_U \end{pmatrix}.$$

This gives the following unsymmetric equations for d

$$\begin{pmatrix} H & -A^T & -I & I \\ A & D_A & 0 & 0 \\ Z_1 & 0 & X_1^\mu & 0 \\ -Z_2 & 0 & 0 & X_2^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = -\begin{pmatrix} g - A^Tv - z_1 + z_2 \\ Ax - b + \mu^A(v - v^E) \\ z_1 \cdot (x - \ell) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u - x) + \mu^B(z_2 - z_2^E) \end{pmatrix}. \tag{2.16}$$

Then, (2.14) implies that

$$\begin{pmatrix} \Delta x \\ \Delta x_1 \\ \Delta x_2 \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \end{pmatrix} = \Delta p = \Delta p_0 + Nd = \begin{pmatrix} d_1 \\ (d_1 - r_L) \\ -(d_1 - r_U) \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}.$$

These identities allow us to write equations (2.16) in the form

$$\begin{pmatrix} H & -A^T & -I & I \\ A & D_A & 0 & 0 \\ Z_1 & 0 & X_1^\mu & 0 \\ -Z_2 & 0 & 0 & X_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \end{pmatrix} = -\begin{pmatrix} g - A^Tv - z_1 + z_2 \\ Ax - b + \mu^A(v - v^E) \\ z_1 \cdot (x - \ell) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u - x) + \mu^B(z_2 - z_2^E) \end{pmatrix}, \tag{2.17}$$

from which we can compute $\Delta x_1 = \Delta x - (\ell - x + x_1)$ and $\Delta x_2 = -\Delta x + (u - x - x_2)$. If x_1 and x_2 satisfy $x - x_1 = \ell$ and $x + x_2 = u$ (i.e., they are feasible for (2.6)), then $\Delta x_1 = \Delta x$ and $\Delta x_2 = -\Delta x$. This assumption is made for the remainder of this section. Under this feasibility assumption, if X_1 and X_2 are written in terms of x, i.e., $X_1 = \mathrm{diag}\big(x_j - \ell_j\big)$ and $X_2 = \mathrm{diag}\big(u_j - x_j\big)$, respectively, then equations (2.5) are the Newton path-following equations (2.18) for a solution of the perturbed optimality conditions (2.3). The variables x_1 and x_2 may be computed implicitly for the line search, in which case the appropriate merit function is

$$
\begin{aligned}
f(x) - (Ax-b)^T v^{\scriptscriptstyle E} + \frac{1}{2\mu^{\scriptscriptstyle A}}\|Ax-b\|^2 + \frac{1}{2\mu^{\scriptscriptstyle A}}\|Ax-b+\mu^{\scriptscriptstyle A}(v-v^{\scriptscriptstyle E})\|^2 \\
- \sum_{j=1}^{n}\Big\{\mu^{\scriptscriptstyle B}[z_1^{\scriptscriptstyle E}]_j \ln\big(x_j-\ell_j+\mu^{\scriptscriptstyle B}\big) + \mu^{\scriptscriptstyle B}[z_1^{\scriptscriptstyle E}]_j \ln\big([z_1]_j(x_j-\ell_j+\mu^{\scriptscriptstyle B})\big) - [z_1]_j(x_j-\ell_j+\mu^{\scriptscriptstyle B})\Big\} \\
- \sum_{j=1}^{n}\Big\{\mu^{\scriptscriptstyle B}[z_2^{\scriptscriptstyle E}]_j \ln\big(u_j-x_j+\mu^{\scriptscriptstyle B}\big) + \mu^{\scriptscriptstyle B}[z_2^{\scriptscriptstyle E}]_j \ln\big([z_2]_j(u_j-x_j+\mu^{\scriptscriptstyle B})\big) - [z_2]_j(u_j-x_j+\mu^{\scriptscriptstyle B})\Big\}.
\end{aligned}
$$

2.5. Computation of the shifted primal-dual penalty-barrier direction

Next we consider the solution of the path-following modified Newton equations (2.5), which will be written in the form

$$
\begin{pmatrix} H & A^T & -I & I \\ A & -D_{\scriptscriptstyle A} & 0 & 0 \\ Z_1 & 0 & Z_1 D_1^z & 0 \\ -Z_2 & 0 & 0 & Z_2 D_2^z \end{pmatrix}
\begin{pmatrix} \Delta x \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \end{pmatrix}
= - \begin{pmatrix} g - A^T v - z_1 + z_2 \\ D_{\scriptscriptstyle A}(v - \pi^{\scriptscriptstyle A}) \\ Z_1 D_1^z (z_1 - \pi_1^z) \\ Z_2 D_2^z (z_2 - \pi_2^z) \end{pmatrix},
$$

which may be row-scaled to give

$$
\begin{pmatrix} H & A^T & -I & I \\ A & -D_{\scriptscriptstyle A} & 0 & 0 \\ I & 0 & D_1^z & 0 \\ -I & 0 & 0 & D_2^z \end{pmatrix}
\begin{pmatrix} \Delta x \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \end{pmatrix}
= - \begin{pmatrix} g - A^T v - z_1 + z_2 \\ D_{\scriptscriptstyle A}(v - \pi^{\scriptscriptstyle A}) \\ D_1^z (z_1 - \pi_1^z) \\ D_2^z (z_2 - \pi_2^z) \end{pmatrix}. \tag{2.18}
$$

The equations and variables can be rescaled and reordered to give

$$
\begin{pmatrix} I & 0 & 0 & (D_1^z)^{-1} \\ 0 & I & 0 & -(D_2^z)^{-1} \\ 0 & 0 & -I & D_{\scriptscriptstyle A}^{-1} A \\ -I & I & A^T & H \end{pmatrix}
\begin{pmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta v \\ \Delta x \end{pmatrix}
= - \begin{pmatrix} z_1 - \pi_1^z \\ z_2 - \pi_2^z \\ v - \pi^{\scriptscriptstyle A} \\ g - A^T v - z_1 + z_2 \end{pmatrix}. \tag{2.19}
$$

Applying the nonsingular matrix

$$
\begin{pmatrix} I & & & \\ 0 & I & & \\ 0 & 0 & I & \\ I & -I & A^T & I \end{pmatrix}
$$

to both sides of (2.19) gives the block upper-triangular system

$$\begin{pmatrix} I & 0 & 0 & (D_1^z)^{-1} \\ & I & 0 & -(D_2^z)^{-1} \\ & & I & D_A^{-1}A \\ & & & H + A^TD_A^{-1}A + D_z^{-1} \end{pmatrix} \begin{pmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta v \\ \Delta x \end{pmatrix} = - \begin{pmatrix} z_1 - \pi_1^z \\ z_2 - \pi_2^z \\ v - \pi^A \\ g - A^T\pi^A - \pi^z \end{pmatrix},$$

where $\pi^z = \pi_1^z - \pi_2^z$, and $D_z^{-1} = (D_1^z)^{-1} + (D_2^z)^{-1}$, for which $D_z = \big((D_1^z)^{-1} + (D_2^z)^{-1}\big)^{-1}$. It follows that the solution of the path-following equations is given by

$$\begin{aligned} \Delta v \;&=\; v^E - \frac{1}{\mu^A}(A(x+\Delta x) - b) - v, \\ \Delta z_1 &= -(X_1^\mu)^{-1}\big(z_1 \cdot (x + \Delta x - \ell + \mu^B e) - \mu^B z_1^E\big), \\ \Delta z_2 &= -(X_2^\mu)^{-1}\big(z_2 \cdot (u - x - \Delta x + \mu^B e) - \mu^B z_2^E\big), \end{aligned}$$

where Δx satisfies $\big(H + A^TD_A^{-1}A + D_z^{-1}\big)\Delta x = -(g - A^T\pi^A - \pi^z)$.

2.6. Summary: linear equalities with upper and lower bounds

Define the quantities

$$\begin{aligned} D_A &= \mu^A I, & \pi^A &= v^E - \frac{1}{\mu^A}(Ax - b), \\ D_1^z &= X_1^\mu Z_1^{-1}, & \pi_1^z &= \mu^B (X_1^\mu)^{-1} z_1^E, \\ D_2^z &= X_2^\mu Z_2^{-1}, & \pi_2^z &= \mu^B (X_2^\mu)^{-1} z_2^E, \\ D_z &= \big((D_1^z)^{-1} + (D_2^z)^{-1}\big)^{-1}, & \pi^z &= \pi_1^z - \pi_2^z, \end{aligned}$$

then Δv, Δz_1 and Δz_1 are given by

$$\begin{aligned} \widehat{x} &= x + \Delta x, & \Delta z_1 &= -(X_1^\mu)^{-1}\big(z_1 \cdot (\widehat{x} - \ell + \mu^B e) - \mu^B z_1^E\big), \\ & & \Delta z_2 &= -(X_2^\mu)^{-1}\big(z_2 \cdot (u - \widehat{x} + \mu^B e) - \mu^B z_2^E\big), \\ \widehat{\pi}^A &= v^E - \frac{1}{\mu^A}(A\widehat{s} - b), & \Delta v \; &= \widehat{\pi}^A - v, \end{aligned}$$

where Δx is the solution of the equations

$$\big(H + A^TD_A^{-1}A + D_z^{-1}\big)\Delta x = -(g - A^T\pi^A - \pi^z).$$

The line-search merit function is

$$
\begin{aligned}
f(x) - (Ax-b)^T v^{E} + \frac{1}{2\mu^{A}}\|Ax-b\|^2 + \frac{1}{2\mu^{A}}\|Ax-b+\mu^{A}(v-v^{E})\|^2 \\
- \sum_{j=1}^{n} \Big\{ \mu^{B}[z_1^{E}]_j \ln\big([z_1]_j(x_j-\ell_j+\mu^{B})^2\big) - [z_1]_j(x_j-\ell_j+\mu^{B}) \Big\} \\
- \sum_{j=1}^{n} \Big\{ \mu^{B}[z_2^{E}]_j \ln\big([z_2]_j(u_j-x_j+\mu^{B})^2\big) - [z_2]_j(u_j-x_j+\mu^{B}) \Big\}.
\end{aligned}
\tag{2.20}
$$

3. Nonnegativity Constraints on the Slacks

We start by considering methods for an optimization problem with nonlinear equality constraints and non-negativity constraints on the slack variables only.

3.1. Problem statement and optimality conditions

The problem has the form

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \quad c(x)-s=0, \quad s\geq 0, \tag{3.1}$$

where $c:\mathbb{R}^n\mapsto\mathbb{R}^m$ and $f:\mathbb{R}^n\mapsto\mathbb{R}$ are twice-continuously differentiable. The first-order KKT conditions for this problem are

$$g(x^*)-J(x^*)^Ty^*=0, \tag{3.2a}$$
$$y^*-w^*=0, \qquad w^*\geq 0, \tag{3.2b}$$
$$c(x^*)-s^*=0, \qquad s^*\geq 0, \tag{3.2c}$$
$$w^*\cdot s^*=0. \tag{3.2d}$$

3.2. The path-following equations

Let y^E denote an estimate of the Lagrange multipliers y^* associated with the equality constraints $c(x)-s=0$. Similarly, let w^E denote a nonnegative estimate of the multipliers for the inequality constraints $s\geq 0$. Given small positive scalars μ^P and μ^B, consider the perturbed optimality conditions

$$g(x)-J(x)^Ty=0, \tag{3.3a}$$
$$y-w=0, \qquad w\geq 0, \tag{3.3b}$$
$$c(x)-s=\mu^P(y^E-y), \qquad s\geq 0, \tag{3.3c}$$
$$w\cdot s=\mu^B(w^E-w). \tag{3.3d}$$

Consider the following primal-dual path following equations given by $F(x,s,y,w\,;\mu^P,\mu^B,y^E,w^E)=0$, with

$$F(x,s,y,w\,;\mu^P,\mu^B,y^E,w^E)=\begin{pmatrix} g(x)-J(x)^Ty \\ y-w \\ c(x)-s+\mu^P(y-y^E) \\ w\cdot s+\mu^B(w-w^E)\end{pmatrix}. \tag{3.4}$$

Any zero (x,s,y,w) of F that satisfies $s>0$ and $w>0$ approximates a solution to problem (3.1), with the approximation becoming increasingly accurate as $\mu^P(y-y^E)\to 0$ and $\mu^B(w-w^E)\to 0$. For any sequence of y^E and w^E such that $y^E\to y^*$ and

$w^E \to w^*$, and it must hold that solutions (s, w) of (3.4) must satisfy $s \cdot w \to 0$. This implies that a solution (x, s, y, w) of (3.2) will approximate a solution of (3.4) independently of the values of μ^P and μ^B (i.e., it is not necessary that $\mu \to 0$).

If (x, s, y, w) is a given approximate zero of F such that $s + \mu^B e > 0$ and $w > 0$, the Newton equations for the change in variables $(\Delta x, \Delta s, \Delta y, \Delta w)$ are given by

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & I & -I \\ J & -I & \mu^P I & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - w \\ c - s + \mu^P (y - y^E) \\ w \cdot s + \mu^B (w - y^E) \end{pmatrix}, \tag{3.5}$$

where $S^\mu = \operatorname{diag}\big(s_i + \mu^B\big)$, $W = \operatorname{diag}(w_i)$, $c = c(x)$, $g = g(x)$, $J = J(x)$, and $H = H(x, y)$.

3.3. A shifted primal-dual penalty-barrier function

The shifted primal-dual problem associated with problem (3.1) is obtained by including the constraints $c(x) - s = 0$ with the objective using a shifted primal-dual augmented Lagrangian term, and using a shifted primal-dual penalty-barrier term for the simple bounds. This gives the problem

$$\underset{x,s,y,w}{\text{minimize}} \;\; M(x, s, y, w\,; \mu^P, \mu^B, y^E, w^E) \quad \text{subject to} \;\; s + \mu^B e > 0, \qquad w > 0, \tag{3.6}$$

where $M(x, s, y, w\,; \mu^P, \mu^B, y^E, w^E)$ is the shifted primal-dual penalty-barrier function

$$f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P (y - y^E)\|^2 \\ - \sum_{i=1}^m \Big\{ \mu^B w_i^E \ln\big(s_i + \mu^B\big) + \mu^B w_i^E \ln\big(w_i (s_i + \mu^B)\big) + \mu^B (w_i^E - w_i) - w_i s_i \Big\},$$

which is well defined for all $w > 0$ and s such that $s + \mu^B e > 0$. This function has the same gradient as

$$M(x, s, y, w\,; \mu^P, \mu^B, y^E, w^E) = f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P (y - y^E)\|^2 \\ - \sum_{i=1}^m \Big\{ \mu^B w_i^E \ln\big(s_i + \mu^B\big) + \mu^B w_i^E \ln\big(w_i (s_i + \mu^B)\big) - w_i (s_i + \mu^B) \Big\}. \tag{3.7}$$

Let c, g and J denote the quantities $c(x)$, $g(x)$ and $J(x)$. For clarity, the dependence of M on the parameters μ^P, μ^B, y^E and w^E, will be suppressed when appropriate, with $M(x, s, y, w)$ being used to denote $M(x, s, y, w\,; \mu^P, \mu^B, y^E, w^E)$. This function may be written in the form:

$$f-(c-s)^T y^E+\frac{1}{2\mu^P}\|c-s\|^2+\frac{1}{2\mu^P}\|c-s+\mu^P(y-y^E)\|^2-\sum_{i=1}^{m}\Big\{\mu^B w_i^E \ln\big(w_i(s_i+\mu^B)^2\big)-w_i(s_i+\mu^B)\Big\}. \tag{3.8}$$

Differentiating $M(x,s,y,w)$ with respect to x, s, y and w gives

$$\nabla M(x,s,y,w)=\begin{pmatrix} g-J^T\big(2(y^E-\frac{1}{\mu^P}(c-s))-y\big)\\ 2\big(y^E-\frac{1}{\mu^P}(c-s)\big)-y-\big(2\mu^B(S^\mu)^{-1}w^E-w\big)\\ c-s+\mu^P(y-y^E)\\ s+\mu^B e-\mu^B W^{-1}w^E\end{pmatrix},$$

with $S=\mathrm{diag}(s_1,\,s_2,\,\ldots,\,s_m)$ and $W=\mathrm{diag}(w_1,\,w_2,\,\ldots,\,w_m)$. The gradient may be written in several equivalent forms

$$\begin{aligned}\nabla M(x,s,y,w)=\begin{pmatrix} g-J^T\big(2(y^E-\frac{1}{\mu^P}(c-s))-y\big)\\ 2\big(y^E-\frac{1}{\mu^P}(c-s)\big)-y-\big(2\mu^B(S^\mu)^{-1}w^E-w\big)\\ c-s+\mu^P(y-y^E)\\ s+\mu^B e-\mu^B W^{-1}w^E\end{pmatrix} &=\begin{pmatrix} g-J^T\big(2(y^E-\frac{1}{\mu^P}(c-s))-y\big)\\ 2\big(y^E-\frac{1}{\mu^P}(c-s)\big)-y-\big(2\mu^B(S^\mu)^{-1}w^E-w\big)\\ \mu^P\Big(\frac{1}{\mu^P}(c-s)+y-y^E\Big)\\ W^{-1}\big(w\cdot s+\mu^B(w-w^E)\big)\end{pmatrix}\\ &=\begin{pmatrix} g-J^T\big(\pi^Y+(\pi^Y-y)\big)\\ \big(\pi^Y+(\pi^Y-y)\big)-\big(\pi^W+(\pi^W-w)\big)\\ -D_Y(\pi^Y-y)\\ -D_W(\pi^W-w)\end{pmatrix},\end{aligned}$$

where

$$D_Y=\mu^P I,\qquad \pi^Y=y^E-\frac{1}{\mu^P}(c-s), \tag{3.9a}$$

$$D_W=S^\mu W^{-1},\qquad \pi^W=\mu^B(S^\mu)^{-1}w^E. \tag{3.9b}$$

Similarly, the Hessian of $M(x,s,y,w)$ is given by

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P}J^T & J^T & 0\\ -\frac{2}{\mu^P}J & 2\big(D_Y^{-1}+\mu^B(S^\mu)^{-2}W^E\big) & -I & I\\ J & -I & \mu^P I & 0\\ 0 & I & 0 & \mu^B W^{-2}W^E\end{pmatrix},$$

where $H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^P} J^T J$. Substituting $\mu^B W^E = S^\mu \Pi^W$ from (3.9) gives the Hessian

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P} J^T & J^T & 0 \\ -\frac{2}{\mu^P} J & 2\big(D_Y^{-1} + (S^\mu)^{-1}\Pi^W\big) & -I & I \\ J & -I & \mu^P I & 0 \\ 0 & I & 0 & W^{-2}\Pi^W S^\mu \end{pmatrix},$$

where $H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^P} J^T J$.

3.4. Derivation of the shifted primal-dual penalty-barrier direction

The primal-dual penalty-barrier problem (3.6) may be written in the form

$$\underset{p \in \mathcal{I}}{\text{minimize}} \;\; M(p), \;\; \text{where } \mathcal{I} = \{p : p = (x, s, y, w), \;\; \text{with } s + \mu^B e > 0, \; w > 0\}.$$

Let $p \in \mathcal{I}$ be given. The Newton direction Δp is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2} \Delta p^T \nabla^2 M(p) \Delta p. \tag{3.10}$$

However, instead of solving (3.10), we define a modified subproblem by approximating the Hessian $\nabla^2 M(x, s, y, w)$ by a matrix $B(x, s, y, w)$. Consider the matrix defined by replacing π^Y by y and π^W by w everywhere in the matrix $\nabla^2 M(x, s, y, w)$. This gives an approximate Hessian $B(x, s, y, w)$ of the form

$$\begin{pmatrix} \widehat{H}_1 & -\frac{2}{\mu^P} J^T & J^T & 0 \\ -\frac{2}{\mu^P} J & 2\big(D_Y^{-1} + (S^\mu)^{-1} W\big) & -I & I \\ J & -I & \mu^P I & 0 \\ 0 & I & 0 & S^\mu W^{-1} \end{pmatrix},$$

where $\widehat{H}_1 = H(x, y) + \frac{2}{\mu^P} J^T J$. The definitions of D_Y and D_W may be used to write $B(x, s, y, w)$ in the form

$$\begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} & J^T & 0 \\ -2D_Y^{-1} J & 2\big(D_Y^{-1} + D_W^{-1}\big) & -I & I \\ J & -I & D_Y & 0 \\ 0 & I & 0 & D_W \end{pmatrix},$$

where $H = H(x,y)$. Given $B(p) = B(x,s,y,w)$, a modified Newton direction is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T B(p)\Delta p. \tag{3.11}$$

Given p, the modified Newton equations for this problem are given by

$$\begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} & J^T & 0 \\ -2D_Y^{-1} J & 2\big(D_Y^{-1} + D_W^{-1}\big) & -I & I \\ J & -I & D_Y & 0 \\ 0 & I & 0 & D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ (2\pi^Y - y) - (2\pi^W - w) \\ -D_Y(\pi^Y - y) \\ -D_W(\pi^W - w) \end{pmatrix}. \tag{3.12}$$

Consider the nonsingular block upper-triangular matrix

$$T = \begin{pmatrix} I & 0 & -2J^T D_Y^{-1} & 0 \\ & I & 2D_Y^{-1} & -2D_W^{-1} \\ & & I & 0 \\ & & & W \end{pmatrix}.$$

Applying T to both sides of (3.12) gives

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & I & -I \\ J & -I & D_Y & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - w \\ c - s + \mu^P(y - y^E) \\ s \cdot w + \mu^B(w - y^E) \end{pmatrix}. \tag{3.13}$$

Comparing these equations with the path-following Newton equations (3.5) implies that a solution of the path-following equations is also a solution of (3.13).

3.5. Computation of the shifted primal-dual penalty-barrier direction

The path-following Newton equations (3.5) may be written in symmetric form

$$\begin{pmatrix} H & 0 & J^T & 0 \\ 0 & 0 & -I & I \\ J & -I & -D_Y & 0 \\ 0 & I & 0 & -D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ -\Delta y \\ -\Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - w \\ c - s + \mu^P(y - y^E) \\ W^{-1}\big(w \cdot s + \mu^B(w - w^E)\big) \end{pmatrix},$$

where $D_Y = \mu^P I$ and $D_W = S^\mu W^{-1}$.

Consider the following reordered subset of equations and variables involving Δw, Δs, Δx and Δy:

$$\begin{pmatrix} D_W & I & 0 & 0 \\ -I & 0 & 0 & I \\ 0 & -I & J & D_Y \end{pmatrix} \begin{pmatrix} \Delta w \\ \Delta s \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} W^{-1}\big(w \cdot s + \mu^B (w - w^E)\big) \\ y - w \\ c - s + \mu^P (y - y^E) \end{pmatrix} = - \begin{pmatrix} -D_W(\pi^W - w) \\ y - w \\ -D_Y(\pi^Y - y) \end{pmatrix}.$$

This gives the equations

$$\begin{pmatrix} I & D_W^{-1} & 0 & 0 \\ -I & 0 & 0 & I \\ 0 & -I & J & D_Y \end{pmatrix} \begin{pmatrix} \Delta w \\ \Delta s \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} w - \pi^W \\ y - w \\ -D_Y(\pi^Y - y) \end{pmatrix}. \tag{3.14}$$

Applying the nonsingular matrix

$$\begin{pmatrix} I & & \\ I & I & \\ D_W & D_W & I \end{pmatrix}$$

on the left-hand side of (3.14) gives the block upper-trapezoidal system

$$\begin{pmatrix} I & D_W^{-1} & 0 & 0 \\ & D_W^{-1} & 0 & I \\ & & J & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta w \\ \Delta s \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} w - \pi^W \\ y - \pi^W \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix}.$$

The solution of this system of equations is given by

$$\begin{aligned} \Delta s &= -D_W\big(y + \Delta y - \pi^W\big) \\ \Delta w &= y + \Delta y - w, \end{aligned}$$

where Δx and Δy satisfy the KKT system

$$\begin{pmatrix} H & -J^T \\ J & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix}.$$

3.6. Summary

The results of Sections 3.1–3.5 imply that the solution of the path-following equations (3.5) may be computed as

$$\begin{aligned} \widehat{y} &= y + \Delta y, \qquad & \Delta s &= -D_W\big(\widehat{y} - \pi^W\big), \\ \widehat{s} &= s + \Delta s, \qquad & \Delta w &= y + \Delta y - w, \end{aligned}$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H(x,y) & -J(x)^T \\ J(x) & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g(x) - J(x)^T y \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix},$$

and D_Y, D_W, π^Y and π^W are given by

$$\begin{aligned} D_Y &= \mu^P I, \qquad & \pi^Y &= y^E - \frac{1}{\mu^P}\big(c(x) - s\big), \\ D_W &= S^\mu W^{-1}, \qquad & \pi^W &= \mu^B (S^\mu)^{-1} w^E. \end{aligned}$$

The associated line-search merit function $M(x, s, y, w\,; \mu^P, \mu^B, y^E, w^E)$ is given by

$$f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 - \sum_{i=1}^{m}\Big\{\mu^B w_i^E \ln\big(w_i(s_i + \mu^B)^2\big) - w_i(s_i + \mu^B)\Big\}.$$

4. Fixed and Nonnegative Slacks

Next we consider nonlinear equality constraints with slacks that are either fixed or nonnegative. The variables are not subject to bounds.

4.1. Problem statement and optimality conditions

The problem has the form

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x) - s = 0, \quad L_X s, \quad L_F s \ge 0, \tag{4.1}$$

where $c : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable and L_X and L_F are fixed matrices of dimension $m_F \times m$ and $m_X \times m$, respectively, with $m = m_F + m_X$. The matrices L_X and L_F are formed from rows of the identity matrix I_m in such a way that $L_X s$ and $L_F s$ give the fixed and "free" components of s. It follows that there is an $m \times m$ permutation matrix P such that

$$P = \begin{pmatrix} L_F \\ L_X \end{pmatrix},$$

with the matrices L_F and L_X satisfying the identities $L_F L_F^T = I_F$, $L_X L_X^T = I_X$, and $L_F L_X^T = 0$. The first-order KKT conditions for this problem are

$$g(x^*) - J(x^*)^T y^* = 0, \tag{4.2a}$$
$$c(x^*) - s^* = 0, \qquad L_X s^* = 0, \tag{4.2b}$$
$$y^* - L_X^T w_X^* - L_F^T w_F^* = 0, \tag{4.2c}$$
$$L_F s^* \ge 0, \qquad w_F^* \ge 0, \tag{4.2d}$$
$$w_F^* \cdot L_F s^* = 0, \tag{4.2e}$$

where y^* and w_X^* are the Lagrange multipliers for the equality constraints $c(x) - s = 0$ and $L_X s = 0$, and w_F^* may be interpreted as the Lagrange multipliers for the nonnegativity constraints $L_F s \ge 0$.

4.2. The path-following equations

Let y^E be an estimate of the Lagrange multipliers for the nonlinear equality constraints $c(x) - s = 0$. Similarly, let w^E denote a nonnegative estimate of the multipliers for the inequality constraints $L_F s \ge 0$. Given small positive scalars μ^P and μ^B, consider the perturbed optimality conditions

$$
\begin{aligned}
g(x) - J(x)^T y &= 0, \\
c(x) - s &= \mu^P (y^E - y), \qquad & L_X s &= 0, \\
y - L_X^T w_X - L_F^T w_F &= 0, \\
L_F s &\geq 0, & w_F &\geq 0, \\
w \cdot L_F s &= \mu^B (w^E - w).
\end{aligned}
$$

Consider the following primal-dual path following equations given by $F(x, s, y, w_X, w_F\,; \mu^P, \mu^B, y^E, w^E) = 0$, with

$$
F(x, s, y, w_X, w_F\,; \mu^P, \mu^B, y^E, w^E) = \begin{pmatrix} g(x) - J(x)^T y \\ y - L_X^T w_X - L_F^T w_F \\ c(x) - s + \mu^P (y - y^E) \\ L_X s \\ w_F \cdot L_F s + \mu^B (w_F - w^E) \end{pmatrix}. \tag{4.4}
$$

Any zero (x, s, y, w_X, w_F) of F satisfying $L_F s > 0$ and $w_F > 0$ approximates a point satisfying the optimality conditions (4.2), with the approximation becoming increasingly accurate as the terms $\mu^P (y - y^E)$ and $\mu^B (w_F - w^E)$ approach zero. For any sequence of y^E and w^E such that $y^E \to y^*$ and $w^E \to w_F^*$, it must hold that solutions (x, s, y, w_X, w_F) of (4.3) must satisfy $y \cdot (c(x) - s) \to 0$, $w_F \cdot (L_F s) \to 0$, and $w_F \cdot L_F s \to 0$. This implies that any solution (x, s, y, w_X, w_F) of (4.3) will approximate a solution of (4.2) independently of the values of μ^P and μ^B (i.e., it is not necessary that μ^P, $\mu^B \to 0$).

Given an approximate zero (x, s, y, w_X, w_F, w_2) of F such that $L_F s > 0$ and $w_F > 0$, the Newton equations for the change in variables $(\Delta x, \Delta s, \Delta y, \Delta w_X, \Delta w_F)$ are given by

$$
\begin{pmatrix} H & 0 & -J^T & 0 & 0 \\ 0 & 0 & I_m & -L_X^T & -L_F^T \\ J & -I_m & D_Y & 0 & 0 \\ 0 & L_X & 0 & 0 & 0 \\ 0 & W L_F & 0 & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w_X \\ \Delta w_F \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - L_X^T w_X - L_F^T w_F \\ c - s + \mu^P (y - y^E) \\ L_X s \\ w_F \cdot L_F s + \mu^B (w_F - w^E) \end{pmatrix}, \tag{4.5}
$$

where $D_Y = \mu^P I$, $W = \operatorname{diag}(w_F)$ and $S^\mu = \operatorname{diag}(s_i + \mu^B)$.

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. Throughout, we assume that s_X satisfies $L_X s = 0$, in which case the expansion of Δs satisfies

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

This identity allows us to write the equations (4.5) in the form

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix} \cdot \begin{pmatrix} \Delta x \\ \Delta s_F \\ \Delta y \\ \Delta w_F \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_F \\ c - s + \mu^P (y - y^E) \\ w_F \cdot L_F s + \mu^B (w_F - w^E) \end{pmatrix} \tag{4.6}$$

The vectors Δs and Δw_X are recovered as $\Delta s = L_F^T \Delta s_F$ and $\Delta w_X = [\, y + \Delta y - w \,]_X$.

4.3. A shifted primal-dual penalty-barrier function

Problem (4.1) is equivalent to

$$\begin{aligned} &\underset{x,s,s_F}{\text{minimize}} \;\; f(x) \\ &\text{subject to} \;\; c(x) - s = 0, \quad L_X s = 0, \quad L_F s - s_F = 0, \quad s_F \ge 0. \end{aligned}$$

Consider the shifted primal-dual penalty-barrier problem

$$\begin{aligned} &\underset{x,s,s_F,y,w_F}{\text{minimize}} \;\; M(x, s, s_F, y, w_F \,;\mu^P, \mu^B, y^E, w^E) \\ &\text{subject to} \;\; L_X s = 0, \quad L_F s - s_F = 0, \quad s_F + \mu^B e > 0, \quad w_F > 0, \end{aligned} \tag{4.7}$$

where $M(x, s, s_F, y, w_F \,;\mu^P, \mu^B, y^E, w^E)$ is the shifted primal-dual penalty-barrier function

$$\begin{aligned} f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P (y - y^E)\|^2 \\ - \sum_{i=1}^{n_F} \Big\{ \mu^B w_i^E \ln\big([\, s_F + \mu^B e \,]_i\big) + \mu^B w_i^E \ln\big([\, w_F \cdot (s_F + \mu^B e) \,]_i\big) - [\, w_F \cdot (s_F + \mu^B e) \,]_i \Big\}. \end{aligned} \tag{4.8}$$

Let c, g and J denote the quantities $c(x)$, $g(x)$ and $J(x)$. Differentiating $M(x, s, s_F, y, w_F, w_2)$ with respect to x, s, s_F, y and w_F gives

$$\nabla M(x, s, s_F, y, w_F) = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c - s)) - y\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c - s)\big) - y \\ w_F - 2\mu^B (S^\mu)^{-1} w^E \\ c - s + \mu^P (y - y^E) \\ s_F + \mu^B e - \mu^B W^{-1} w^E \end{pmatrix}.$$

The gradient may be written in several equivalent forms

$$\nabla M(x,s,s_F,y,w_F) = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y \\ w_F - 2\mu^B(S^\mu)^{-1}w^E \\ c - s + \mu^P(y - y^E) \\ s_F + \mu^B e - \mu^B W^{-1} w^E \end{pmatrix} = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y \\ (S^\mu)^{-1}\big(w_F \cdot s_F + \mu^B w^E + \mu^B(w_F - w^E)\big) \\ c - s + \mu^P(y - y^E) \\ W^{-1}\big(w_F \cdot s_F + \mu^B(w_F - w^E)\big) \end{pmatrix}$$

$$= \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi^Y + (\pi^Y - y) \\ -\big(\pi^W + (\pi^W - w_F)\big) \\ -D_Y(\pi^Y - y) \\ -D_W(\pi^W - w_F) \end{pmatrix},$$

where

$$D_Y = \mu^P I_m, \qquad \pi^Y = y^E - \frac{1}{\mu^P}(c-s), \tag{4.9a}$$

$$D_W = S^\mu W^{-1}, \qquad \pi^W = \mu^B (S^\mu)^{-1} w^E. \tag{4.9b}$$

Similarly, the Hessian of $M(x,s,s_1,s_2,y,w_1,w_2)$ is given by

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P}J^T & 0 & J^T & 0 \\ -\frac{2}{\mu^P}J & \frac{2}{\mu^P}I_m & 0 & -I_m & 0 \\ 0 & 0 & 2\mu^B(S^\mu)^{-2}W^E & 0 & I_F \\ J & -I_m & 0 & \mu^P I_m & 0 \\ 0 & 0 & I_F & 0 & \mu^B W^{-2} W^E \end{pmatrix},$$

where $H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^P}J^TJ$. Substituting $\mu^B W^E = S^\mu \Pi^W$ from (4.9) gives the Hessian

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P}J^T & 0 & J^T & 0 \\ -\frac{2}{\mu^P}J & \frac{2}{\mu^P}I_m & 0 & -I_m & 0 \\ 0 & 0 & 2(S^\mu)^{-1}\Pi^W & 0 & I_F \\ J & -I_m & 0 & \mu^P I_m & 0 \\ 0 & 0 & I_F & 0 & S^\mu W^{-2}\Pi^W \end{pmatrix}$$

4.4. Derivation of the shifted primal-dual penalty-barrier direction

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

$$\underset{p\in\mathcal{I}}{\text{minimize}} \;\; M(p) \quad \text{subject to} \quad Cp = 0, \tag{4.10}$$

where

$$\mathcal{I} = \{p : p = (x, s, s_F, y, w_F), \;\; \text{with } s_F + \mu^B e > 0, \; w_F > 0\},$$

with

$$C = \begin{pmatrix} 0 & L_X & 0 & 0 & 0 \\ 0 & L_F & -I_F & 0 & 0 \end{pmatrix}.$$

Let $p \in \mathcal{I}$ be given. The Newton direction Δp is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T \nabla^2 M(p) \Delta p \quad \text{subject to} \quad C\Delta p = -Cp. \tag{4.11}$$

However, instead of solving (4.11), we define a linearly constrained modified Newton method by approximating the Hessian $\nabla^2 M(x, s, s_F, y, w_F)$ by a matrix $B(x, s, s_F, y, w_F)$. Consider the matrix defined by replacing π^Y by y and π^W by w_F, everywhere in the matrix $\nabla^2 M(x, s, s_F, y, w_F)$. This gives an approximate Hessian $B(x, s, s_F, y, w_F)$ of the form

$$\begin{pmatrix} \widehat{H}_1 & -\frac{2}{\mu^P} J^T & 0 & J^T & 0 \\ -\frac{2}{\mu^P} J & \frac{2}{\mu^P} I_m & 0 & -I_m & 0 \\ 0 & 0 & 2(S^\mu)^{-1} W & 0 & I_F \\ J & -I_m & 0 & \mu^P I_m & 0 \\ 0 & 0 & I_F & 0 & S^\mu W^{-1} \end{pmatrix}$$

where $\widehat{H}_1 = H(x,y) + 2J^T D_Y^{-1} J$. The definitions of D_Y and D_W may be used to write $B(x, s, s_F, y, w_F)$ as

$$\begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} & 0 & J^T & 0 \\ -2D_Y^{-1} J & 2D_Y^{-1} & 0 & -I_m & 0 \\ 0 & 0 & 2D_W^{-1} & 0 & I_F \\ J & -I_m & 0 & D_Y & 0 \\ 0 & 0 & I_F & 0 & D_W \end{pmatrix}$$

where $H = H(x,y)$. Given $B(p) = B(x, s, s_F, y, w_F)$, a modified Newton direction is given by the solution of the QP subproblem

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

If $p = (x, s, s_F, y, w_F)$ is feasible for the constraints then $Cp = 0$ and $L_X s = 0$ and $L_F s - s_F = 0$. In this case every feasible Δp may be written in the form $\Delta p = Nd$, where N denote a matrix whose columns form a basis for null(C), i.e., $CN = 0$ and $\begin{pmatrix} C^T & N \end{pmatrix}$ is nonsingular. This implies that d must satisfy the reduced equations

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

Consider the particular null-space basis given by

$$N = \begin{pmatrix} I_n & 0 & 0 & 0 \\ 0 & L_F^T & 0 & 0 \\ 0 & I_F & 0 & 0 \\ 0 & 0 & I_m & 0 \\ 0 & 0 & 0 & I_F \end{pmatrix}. \tag{4.12}$$

This definition of N gives the reduced Hessian

$$N^T B(p) N = \begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} L_F^T & J^T & 0 \\ -2L_F D_Y^{-1} J & 2\big(L_F D_Y^{-1} L_F^T + D_W^{-1}\big) & -L_F & I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & I_F & 0 & D_W \end{pmatrix}.$$

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

$$N^T \nabla M(p) = \begin{pmatrix} I_n & 0 & 0 & 0 & 0 \\ 0 & L_F & I_F & 0 & 0 \\ 0 & 0 & 0 & I_m & 0 \\ 0 & 0 & 0 & 0 & I_F \end{pmatrix} \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi^Y + (\pi^Y - y) \\ -\big(\pi^W + (\pi^W - w_F)\big) \\ -D_Y(\pi^Y - y) \\ -D_W(\pi^W - w_F) \end{pmatrix} = \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi_F^Y + (\pi_F^Y - y_F) - \big(\pi^W + (\pi^W - w_F)\big) \\ -D_Y(\pi^Y - y) \\ -D_W(\pi^W - w_F) \end{pmatrix}.$$

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

$$\begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} L_F^T & J^T & 0 \\ -2L_F D_Y^{-1} J & 2\big(L_F D_Y^{-1} L_F^T + D_W^{-1}\big) & -L_F & I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & I_F & 0 & D_W \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi_F^Y + (\pi_F^Y - y_F) - \big(\pi^W + (\pi^W - w_F)\big) \\ -D_Y(\pi^Y - y) \\ -D_W(\pi^W - w_F) \end{pmatrix}.$$

Given any nonsingular matrix R, the direction d satisfies

$$RN^T B(p)Nd = -RN^T \nabla M(p).$$

In particular, consider

$$R = \begin{pmatrix} I_n & 0 & -2J^T D_Y^{-1} & 0 \\ & I_F & 2L_F D_Y^{-1} & -2D_W^{-1} \\ & & I_m & 0 \\ & & & W \end{pmatrix},$$

which is nonsingular if W is positive definite, with

$$R^{-1} = \begin{pmatrix} I_n & 0 & 2J^T D_Y^{-1} & 0 \\ & I_F & -2L_F D_Y^{-1} & 2W^{-1} D_W^{-1} \\ & & I_m & 0 \\ & & & W^{-1} \end{pmatrix} = \begin{pmatrix} I_n & 0 & 2J^T D_Y^{-1} & 0 \\ & I_F & -2L_F D_Y^{-1} & 2(S^\mu)^{-1} \\ & & I_m & 0 \\ & & & W^{-1} \end{pmatrix}.$$

For this R, the product $RN^T B(p)N$ is given by

$$\begin{pmatrix} I_n & 0 & -2J^T D_Y^{-1} & 0 \\ & I_F & 2L_F D_Y^{-1} & -2D_W^{-1} \\ & & I_m & 0 \\ & & & W \end{pmatrix} \begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} L_F^T & J^T & 0 \\ -2L_F D_Y^{-1} J & 2\big(L_F D_Y^{-1} L_F^T + D_W^{-1}\big) & -L_F & I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & I_F & 0 & D_W \end{pmatrix}$$

$$= \begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & W & 0 & WD_W \end{pmatrix} = \begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix}.$$

Similarly, the transformed right-hand side $RN^T \nabla M(p)$ is given by

$$\begin{pmatrix} g - J^T y \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -WD_W(\pi^W - w_F) \end{pmatrix}.$$

Putting all this together gives the following transformed unsymmetric reduced modified Newton equations for the vector d:

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -WD_W(\pi^W - w_F) \end{pmatrix},$$

or, equivalently,

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_F \\ c - s + \mu^P (y - y^E) \\ w_F \cdot L_F s + \mu^B (w_F - w^E) \end{pmatrix}. \tag{4.13}$$

Then, the definition of N from (4.12) implies that

$$\begin{pmatrix} \Delta x \\ \Delta s \\ \Delta s_F \\ \Delta y \\ \Delta w_F \end{pmatrix} = \Delta p = N d = \begin{pmatrix} d_1 \\ L_F^T d_2 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix}.$$

These identities allow us to write equations (4.13) as

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & W & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s_F \\ \Delta y \\ \Delta w_F \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_F \\ c - s + \mu^P (y - y^E) \\ w_F \cdot L_F s + \mu^B (w_F - w^E) \end{pmatrix}. \tag{4.14}$$

If S is written in terms of s, i.e., $S = \operatorname{diag}\big(L_F s\big)$, then the equations (4.14) are the Newton equations for the solution of the perturbed optimality conditions (4.3). The variables s_F may be computed implicitly for the line search, in which case the appropriate merit function is

$$f - (c-s)^T y^E + \frac{1}{2\mu^P}\|c - s\|^2 + \frac{1}{2\mu^P}\|c - s + \mu^P (y - y^E)\|^2$$
$$- \sum_{i=1}^{n_F} \Big\{ \mu^B w_i^E \ln\big([\, s_F + \mu^B e\,]_i\big) + \mu^B w_i^E \ln\big([\, w \cdot (s_F + \mu^B e)\,]_i\big) - [\, w \cdot (s_F + \mu^B e)\,]_i \Big\}.$$

4.5. Computation of the shifted primal-dual penalty-barrier direction

Next we consider the solution of the modified Newton equations (4.14), which are written in the form

$$\begin{pmatrix} H & 0 & -J^T & 0 \\ 0 & 0 & L_F & -I_F \\ J & -L_F^T & D_Y & 0 \\ 0 & D_W^{-1} & 0 & I_F \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s_F \\ \Delta y \\ \Delta w_F \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_F \\ -D_Y(\pi^Y - y) \\ -(\pi^W - w_F) \end{pmatrix} \tag{4.15}$$

using the identities $D_W = S^\mu W^{-1}$ and $w_F \cdot s_F + \mu^B(w_F - w^E) = -S^\mu(\pi^W - w_F)$. Consider the following reordered set of equations and variables involving (in order) Δw_F, Δs_F, Δx and Δy:

$$\begin{pmatrix} I_F & D_W^{-1} & 0 & 0 \\ -I_F & 0 & 0 & L_F \\ 0 & -L_F^T & J & D_Y \\ 0 & 0 & H & -J^T \end{pmatrix} \begin{pmatrix} \Delta w_F \\ \Delta s_F \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} w_F - \pi^W \\ y_F - w_F \\ D_Y(y - \pi^Y) \\ g - J^T y \end{pmatrix}. \tag{4.16}$$

Applying the nonsingular matrix

$$\begin{pmatrix} I_F & & & \\ I_F & I_F & & \\ L_F^T D_W & L_F^T D_W & I_m & \\ & & & I_n \end{pmatrix}$$

on the left- and right-hand side of (4.16) yields the block upper-triangular system of equations

$$\begin{pmatrix} I_F & D_W^{-1} & 0 & 0 \\ & D_W^{-1} & 0 & L_F \\ & & J & D_Y + L_F^T D_W L_F \\ & & H & -J^T \end{pmatrix} \begin{pmatrix} \Delta w_F \\ \Delta s_F \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} w_F - \pi^W \\ y_F - \pi^W \\ D_Y(y - \pi^Y) + L_F^T D_W \big(y_F - \pi^W\big) \\ g - J^T y \end{pmatrix}. \tag{4.17}$$

Solving (4.17) while using the last block equation of (4.14) as an alternative definition of Δw_F gives the solution of the path-following equations as

$$\begin{aligned} \Delta s_F &= -\bar{D}_W\big(y_F + \Delta y_F - \pi^W\big), \\ \Delta s &= \ L_F^T \Delta s_F, \\ \Delta w_F &= -(S^\mu)^{-1}\big(w_F \cdot (L_F(s + \Delta s) - s_F + \mu^B e) - \mu^B w^E\big), \end{aligned}$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H & -J^T \\ J & D_Y + L_F^T D_W L_F \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ D_Y(y - \pi^Y) + L_F^T D_W \big(y_F - \pi^W\big) \end{pmatrix}.$$

4.6. Summary: bounded slacks

Consider the quantities

$$
\begin{aligned}
D_Y &= \mu^P I, & \pi^Y &= y^E - \frac{1}{\mu^P}\big(c(x) - s\big),\\
D_W &= S^\mu W^{-1}, & \pi^W &= \mu^B (S^\mu)^{-1} w^E,
\end{aligned}
$$

then Δs, Δs_F and Δw_F are given by

$$
\begin{aligned}
\widehat{y} = y + \Delta y, \qquad \Delta s_F &= -D_W\big(\widehat{y}_F - \pi^W\big),\\
\Delta s &= L_F^T \Delta s_F,\\
\Delta w_X &= [\,\widehat{y} - w\,]_X,\\
\widehat{s} = s + \Delta s, \qquad \Delta w_F &= -(S^\mu)^{-1}\big(w_F \cdot (L_F \widehat{s} + \mu^B e) - \mu^B w^E\big),
\end{aligned}
$$

where Δx and Δy satisfy the equations

$$
\begin{pmatrix} H & -J^T \\ J & D_Y + L_F^T D_W L_F \end{pmatrix}
\begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}
= -\begin{pmatrix} g - J^T y \\ D_Y(y - \pi^Y) + L_F^T D_W\big(y_F - \pi^W\big) \end{pmatrix}.
$$

The associated line-search merit function is given by

$$
\begin{aligned}
f - (c - s)^T y^E &+ \frac{1}{2\mu^P}\|c - s\|^2 + \frac{1}{2\mu^P}\|c - s + \mu^P(y - y^E)\|^2\\
&- \sum_{i=1}^{n_F}\Big\{\mu^B w_i^E \ln\big([\,s_F + \mu^B e\,]_i\big) + \mu^B w_i^E \ln\big([\,w \cdot (s_F + \mu^B e)\,]_i\big) - [\,w \cdot (s_F + \mu^B e)\,]_i\Big\}. \qquad (4.18)
\end{aligned}
$$

5. Nonnegativity Constraints on the Variables and Slacks

Next we consider methods for an optimization problem with nonlinear equality constraints and non-negativity constraints on the variables and slacks.

5.1. Problem statement and optimality conditions

The problem has the form

$$\underset{x\in\mathbb{R}^n,s\in\mathbb{R}^m}{\text{minimize}}\;\; f(x) \quad \text{subject to} \quad c(x)-s=0, \quad Ax=b, \quad x\ge 0, \quad s\ge 0, \tag{5.1}$$

where $c:\mathbb{R}^n\mapsto\mathbb{R}^m$ and $f:\mathbb{R}^n\mapsto\mathbb{R}$ are twice-continuously differentiable. The first-order KKT conditions for this problem are

$$g(x^*)-A^Tv^*-J(x^*)^Ty^*-z^*=0, \qquad z^*\ge 0, \tag{5.2a}$$
$$y^*-w^*=0, \qquad w^*\ge 0, \tag{5.2b}$$
$$c(x^*)-s^*=0, \qquad Ax^*=b, \qquad x^*\ge 0, \qquad s^*\ge 0, \tag{5.2c}$$
$$z^*\cdot x^*=0, \qquad w^*\cdot s^*=0. \tag{5.2d}$$

The vectors y^*, z^* and w^* constitute the Lagrange multipliers for the equality constraints $c(x)-s=0$ and the nonnegativity constraints $x\ge 0$ and $s\ge 0$, respectively. A vector (x,y,z,w) is said to constitute a *primal-dual estimate* of the quantities (x^*,y^*,z^*,w^*) satisfying the optimality conditions for (5.1).

5.2. The path-following equations

Let v^E and y^E denote estimates of the Lagrange multipliers for the equality constraints $Ax=b$ and $c(x)-s=0$. Similarly, let z^E and w^E denote nonnegative estimates of the multipliers for the inequality constraints $x\ge 0$ and $s\ge 0$. Given small positive scalars μ^A, μ^P and μ^B, consider the perturbed optimality conditions

$$g(x)-A^Tv-J(x)^Ty-z=0, \qquad z\ge 0, \tag{5.3a}$$
$$y-w=0, \qquad w\ge 0, \tag{5.3b}$$
$$c(x)-s=\mu^P(y^E-y), \qquad Ax-b=\mu^A(v^E-v), \qquad x\ge 0, \qquad s\ge 0, \tag{5.3c}$$
$$z\cdot x=\mu^B(z^E-z), \qquad w\cdot s=\mu^B(w^E-w). \tag{5.3d}$$

Consider the primal-dual path-following equations $F(x, s, v, y, z, w\,; \mu^A, \mu^P, \mu^B, y^E, z^E, w^E) = 0$, with

$$F(x, s, v, y, z, w\,; \mu^A, \mu^P, \mu^B, v^E, y^E, w^E) = \begin{pmatrix} g(x) - J(x)^T y - A^T v - z \\ y - w \\ Ax - b - \mu^A(v - v^E) \\ c(x) - s + \mu^P(y - y^E) \\ z \cdot x + \mu^B(z - z^E) \\ w \cdot s + \mu^B(w - w^E) \end{pmatrix}. \tag{5.4}$$

Any zero (x, s, v, y, z, w) of F that satisfies $x > 0$, $s > 0$, $z > 0$, and $w > 0$ approximates a point satisfying the optimality conditions (5.2), with the approximation becoming increasingly accurate as $\mu^P(y - y^E) \to 0$, $\mu^A(v - v^E) \to 0$, $\mu^B(z - z^E) \to 0$, and $\mu^B(w - w^E) \to 0$. For any sequence of v^E, y^E, z^E and w^E such that $v^E \to v^*$, $y^E \to y^*$, $z^E \to z^*$ and $w^E \to w^*$, it must hold that solutions (s, w) of (5.4) must satisfy $z \cdot x \to 0$ and $w \cdot s \to 0$. This implies that a solution (x, s, v, y, z, w) of (5.4) will approximate a solution of (5.2) independently of the values of μ^A, μ^P and μ^B (i.e., it is not necessary that the parameters μ^A, μ^P and μ^B go to zero).

If $v = (x, s, y, w)$ is a given approximate zero of F such that $x + \mu^B e > 0$, $s + \mu^B e > 0$, $z > 0$, and $w > 0$, the Newton equations for the change in variables $(\Delta x, \Delta s, \Delta v, \Delta y, \Delta z, \Delta w)$ are given by

$$\begin{pmatrix} H & 0 & -A^T & -J^T & -I & 0 \\ 0 & 0 & 0 & I & 0 & -I \\ A & 0 & D_A & 0 & 0 & 0 \\ J & -I & 0 & D_Y & 0 & 0 \\ Z & 0 & 0 & 0 & X^\mu & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta v \\ \Delta y \\ \Delta z \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - J^T y - A^T v - z \\ y - w \\ Ax - b + \mu^A(v - v^E) \\ c - s + \mu^P(y - y^E) \\ z \cdot x + \mu^B(z - z^E) \\ w \cdot s + \mu^B(w - w^E) \end{pmatrix}, \tag{5.5}$$

where $D_A = \mu^A I$, $D_Y = \mu^P I$, $X^\mu = \mathrm{diag}\big(x_j + \mu^B\big)$, $S^\mu = \mathrm{diag}\big(s_i + \mu^B\big)$, $Z = \mathrm{diag}(z_j)$, and $W = \mathrm{diag}(w_i)$.

5.3. A shifted primal-dual penalty-barrier function

Consider the shifted primal-dual penalty-barrier function

$$
\begin{aligned}
f(x) - (Ax-b)^T v^E + \frac{1}{2\mu^A}\|Ax-b\|^2 + \frac{1}{2\mu^A}\|Ax-b+\mu^A(v-v^E)\|^2 \\
- \big(c(x)-s\big)^T y^E + \frac{1}{2\mu^P}\|c(x)-s\|^2 + \frac{1}{2\mu^P}\|c(x)-s+\mu^P(y-y^E)\|^2 \\
- \sum_{j=1}^{n}\Big\{\mu^B z_j^E \ln\big(x_j+\mu^B\big) + \mu^B z_j^E \ln\big(z_j(x_j+\mu^B)\big) + \mu^B(z_j^E - z_j) - z_j x_j\Big\} \\
- \sum_{i=1}^{m}\Big\{\mu^B w_i^E \ln\big(s_i+\mu^B\big) + \mu^B w_i^E \ln\big(w_i(s_i+\mu^B)\big) + \mu^B(w_i^E - w_i) - w_i s_i\Big\},
\end{aligned}
$$

which is well defined for all x and s such that $x+\mu^B e > 0$ and $s+\mu^B e > 0$. This function has the same gradient as the function $M(x,s,v,y,z,w\,;\mu^A,\mu^P,\mu^B,v^E,y^E,z^E,w^E)$ given by

$$
\begin{aligned}
f(x) - (Ax-b)^T v^E + \frac{1}{2\mu^A}\|Ax-b\|^2 + \frac{1}{2\mu^A}\|Ax-b+\mu^A(v-v^E)\|^2 \\
- \big(c(x)-s\big)^T y^E + \frac{1}{2\mu^P}\|c(x)-s\|^2 + \frac{1}{2\mu^P}\|c(x)-s+\mu^P(y-y^E)\|^2 \\
- \sum_{j=1}^{n}\Big\{\mu^B z_j^E \ln\big(x_j+\mu^B\big) + \mu^B z_j^E \ln\big(z_j(x_j+\mu^B)\big) - z_j(x_j+\mu^B)\Big\} \\
- \sum_{i=1}^{m}\Big\{\mu^B w_i^E \ln\big(s_i+\mu^B\big) + \mu^B w_i^E \ln\big(w_i(s_i+\mu^B)\big) - w_i(s_i+\mu^B)\Big\}.
\end{aligned}
\tag{5.6}
$$

Let c, g and J denote the quantities $c(x)$, $g(x)$ and $J(x)$. For clarity, the dependence of M on the parameters μ^P, μ^B, y^E, z^E, and w^E, will be suppressed, with $M(x,s,v,y,z,w)$ being used to denote $M(x,s,v,y,z,w\,;\mu^A,\mu^P,\mu^B,v^E,y^E,z^E,w^E)$. This function may be written in the form:

$$\begin{aligned} M(x,s,v,y,z,w) = f &- (Ax-b)^T v^{\scriptscriptstyle E} + \frac{1}{2\mu^{\scriptscriptstyle A}}\|Ax-b\|^2 + \frac{1}{2\mu^{\scriptscriptstyle A}}\|Ax-b+\mu^{\scriptscriptstyle A}(v-v^{\scriptscriptstyle E})\|^2 \\ &- (c-s)^T y^{\scriptscriptstyle E} + \frac{1}{2\mu^{\scriptscriptstyle P}}\|c-s\|^2 + \frac{1}{2\mu^{\scriptscriptstyle P}}\|c-s+\mu^{\scriptscriptstyle P}(y-y^{\scriptscriptstyle E})\|^2 \\ &- \sum_{j=1}^{n}\Big\{\mu^{\scriptscriptstyle B} z_j^{\scriptscriptstyle E}\ln\big(z_j(x_j+\mu^{\scriptscriptstyle B})^2\big) - z_j(x_j+\mu^{\scriptscriptstyle B})\Big\} - \sum_{i=1}^{m}\Big\{\mu^{\scriptscriptstyle B} w_i^{\scriptscriptstyle E}\ln\big(w_i(s_i+\mu^{\scriptscriptstyle B})^2\big) - w_i(s_i+\mu^{\scriptscriptstyle B})\Big\}. \end{aligned} \tag{5.7}$$

Differentiating $M(x,s,v,y,z,w)$ with respect to x, s, v, y, z and w gives

$$\nabla M(x,s,v,y,z,w) = \begin{pmatrix} g - A^T\big(2(v^{\scriptscriptstyle E} + \frac{1}{\mu^{\scriptscriptstyle A}}(Ax-b)) - v\big) - J^T\big(2(y^{\scriptscriptstyle E} - \frac{1}{\mu^{\scriptscriptstyle P}}(c-s)) - y\big) - 2\mu^{\scriptscriptstyle B}(X^{\mu})^{-1}z^{\scriptscriptstyle E} + z \\ 2\big(y^{\scriptscriptstyle E} - \frac{1}{\mu^{\scriptscriptstyle P}}(c-s)\big) - y - 2\mu^{\scriptscriptstyle B}(S^{\mu})^{-1}w^{\scriptscriptstyle E} + w \\ Ax - b + \mu^{\scriptscriptstyle A}(v-v^{\scriptscriptstyle E}) \\ c - s + \mu^{\scriptscriptstyle P}(y-y^{\scriptscriptstyle E}) \\ x + \mu^{\scriptscriptstyle B}e - \mu^{\scriptscriptstyle B}Z^{-1}z^{\scriptscriptstyle E} \\ s + \mu^{\scriptscriptstyle B}e - \mu^{\scriptscriptstyle B}W^{-1}w^{\scriptscriptstyle E} \end{pmatrix},$$

with $X = \mathrm{diag}(x_1,\ x_2,\ \ldots,\ x_n)$, $S = \mathrm{diag}(s_1,\ s_2,\ \ldots,\ s_m)$, $Z = \mathrm{diag}(z_1,\ z_2,\ \ldots,\ z_n)$ and $W = \mathrm{diag}(w_1,\ w_2,\ \ldots,\ w_m)$. The gradient may be written in several equivalent forms

$$\nabla M(x,s,y,z,w) = \begin{pmatrix} g - A^T\big(2(v^E + \frac{1}{\mu^A}(Ax-b)) - v\big) - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) - 2\mu^B(X^\mu)^{-1}z^E + z \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y - 2\mu^B(S^\mu)^{-1}w^E + w \\ Ax - b + \mu^A(v - v^E) \\ c - s + \mu^P(y - y^E) \\ x + \mu^B e - \mu^B Z^{-1} z^E \\ s + \mu^B e - \mu^B W^{-1} w^E \end{pmatrix}$$

$$= \begin{pmatrix} g - A^T\big(2(v^E + \frac{1}{\mu^A}(Ax-b)) - v\big) - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) - 2\mu^B(X^\mu)^{-1}z^E + z \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y - 2\mu^B(S^\mu)^{-1}w^E + w \\ Ax - b + \mu^A(v - v^E) \\ c - s + \mu^P(y - y^E) \\ Z^{-1}\big(z \cdot x + \mu^B(z - z^E)\big) \\ W^{-1}\big(w \cdot s + \mu^B(w - w^E)\big) \end{pmatrix}$$

$$= \begin{pmatrix} g - A^T\big(\pi^A + (\pi^A - v)\big) - J^T\big(\pi^Y + (\pi^Y - y)\big) - \big(\pi^Z + (\pi^Z - z)\big) \\ \big(\pi^Y + (\pi^Y - y)\big) - \big(\pi^W + (\pi^W - w)\big) \\ -D_A(\pi^A - v) \\ -D_Y(\pi^Y - y) \\ -D_Z(\pi^Z - z) \\ -D_W(\pi^W - w) \end{pmatrix},$$

where

$$D_A = \mu^A I, \qquad \pi^A = v^E - \frac{1}{\mu^A}(Ax - b), \tag{5.8a}$$

$$D_Y = \mu^P I, \qquad \pi^Y = y^E - \frac{1}{\mu^P}(c - s), \tag{5.8b}$$

$$D_Z = X^\mu Z^{-1}, \qquad \pi^Z = \mu^B (X^\mu)^{-1} z^E, \tag{5.8c}$$

$$D_W = S^\mu W^{-1}, \qquad \pi^W = \mu^B (S^\mu)^{-1} w^E. \tag{5.8d}$$

Similarly, the Hessian of $M(x,s,v,y,z,w)$ is given by

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P}J^T & A^T & J^T & I & 0 \\ -\frac{2}{\mu^P}J & 2\big(\frac{1}{\mu^P}I + \mu^B(S^\mu)^{-2}W^E\big) & 0 & -I & 0 & I \\ A & 0 & \mu^A I & 0 & 0 & 0 \\ J & -I & 0 & \mu^P I & 0 & 0 \\ I & 0 & 0 & 0 & \mu^B Z^{-2}Z^E & 0 \\ 0 & I & 0 & 0 & 0 & \mu^B W^{-2}W^E \end{pmatrix},$$

where $H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^A}A^TA + \frac{2}{\mu^P}J^TJ + 2\mu^B(X^\mu)^{-2}Z^E$. Substituting $\mu^B Z^E = X^\mu \Pi^Z$ and $\mu^B W^E = S^\mu \Pi^W$ from (5.8) gives the Hessian

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P}J^T & A^T & J^T & I & 0 \\ -\frac{2}{\mu^P}J & 2\big(\frac{1}{\mu^P}I + (S^\mu)^{-1}\Pi^W\big) & 0 & -I & 0 & I \\ A & 0 & \mu^A I & 0 & 0 & 0 \\ J & -I & 0 & \mu^P I & 0 & 0 \\ I & 0 & 0 & 0 & Z^{-2}\Pi^Z X^\mu & 0 \\ 0 & I & 0 & 0 & 0 & W^{-2}\Pi^W S^\mu \end{pmatrix},$$

where $H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^A}A^TA + \frac{2}{\mu^P}J^TJ + 2(X^\mu)^{-1}\Pi^Z$.

5.4. Derivation of the shifted primal-dual penalty-barrier direction

Now consider the matrix defined by replacing π^Y by y, π^Z by z, and π^W by w, everywhere in $\nabla^2 M(x,s,v,y,z,w)$. This gives an approximate Hessian $B(x,s,v,y,z,w)$ of the form

$$\begin{pmatrix} \widehat{H}_1 & -\frac{2}{\mu^P}J^T & A^T & J^T & I & 0 \\ -\frac{2}{\mu^P}J & 2\big(\frac{1}{\mu^P}I + (S^\mu)^{-1}W\big) & 0 & -I & 0 & I \\ A & 0 & \mu^A I & 0 & 0 & 0 \\ J & -I & 0 & \mu^P I & 0 & 0 \\ I & 0 & 0 & 0 & Z^{-1}X^\mu & 0 \\ 0 & I & 0 & 0 & 0 & W^{-1}S^\mu \end{pmatrix},$$

where $\widehat{H}_1 = H(x,y) + \frac{2}{\mu^A} A^T A + \frac{2}{\mu^P} J^T J + 2(X^\mu)^{-1} Z$. The definitions of D_Y, D_Z, and D_W may be used to write $B(x,s,v,y,z,w)$ in the form

$$\begin{pmatrix} H + 2A^T D_A^{-1} A + 2J^T D_Y^{-1} J + 2D_Z^{-1} & -2J^T D_Y^{-1} & A^T & J^T & I & 0 \\ -2D_Y^{-1} J & 2(D_Y^{-1} + D_W^{-1}) & 0 & -I & 0 & I \\ A & 0 & D_A & 0 & 0 & 0 \\ J & -I & 0 & D_Y & 0 & 0 \\ I & 0 & 0 & 0 & D_Z & 0 \\ 0 & I & 0 & 0 & 0 & D_W \end{pmatrix},$$

where $H = H(x,y)$. A modified Newton direction satisfies

$$B(x,s,v,y,z,w)d = -\nabla M(x,s,v,y,z,w).$$

Given any nonsingular matrix R, the modified Newton direction also satisfies

$$RB(x,s,v,y,z,w)d = -R\nabla M(x,s,v,y,z,w).$$

In particular, consider the block upper-triangular matrix

$$R = \begin{pmatrix} I & 0 & -2A^T D_A^{-1} & -2J^T D_Y^{-1} & -2D_Z^{-1} & 0 \\ & I & 0 & 2D_Y^{-1} & 0 & -2D_W^{-1} \\ & & I & 0 & 0 & 0 \\ & & & I & 0 & 0 \\ & & & & Z & 0 \\ & & & & & W \end{pmatrix},$$

which is nonsingular if Z and W are positive definite. For this R, the product $RB(x,s,v,y,z,w)$ is given by

$$\begin{pmatrix} I & 0 & -2A^T D_A^{-1} & -2J^T D_Y^{-1} & -2D_Z^{-1} & 0 \\ & I & 0 & 2D_Y^{-1} & 0 & -2D_W^{-1} \\ & & I & 0 & 0 & 0 \\ & & & I & 0 & 0 \\ & & & & Z & 0 \\ & & & & & W \end{pmatrix} \begin{pmatrix} H + 2A^T D_A^{-1} A + 2J^T D_Y^{-1} J + 2D_Z^{-1} & -2J^T D_Y^{-1} & A^T & J^T & I & 0 \\ -2D_Y^{-1} J & 2(D_Y^{-1} + D_W^{-1}) & 0 & -I & 0 & I \\ A & 0 & D_A & 0 & 0 & 0 \\ J & -I & 0 & D_Y & 0 & 0 \\ I & 0 & 0 & 0 & D_Z & 0 \\ 0 & I & 0 & 0 & 0 & D_W \end{pmatrix}$$

$$= \begin{pmatrix} H & 0 & -A^T & -J^T & -I & 0 \\ 0 & 0 & 0 & I & 0 & -I \\ A & 0 & D_A & 0 & 0 & 0 \\ J & -I & 0 & D_Y & 0 & 0 \\ Z & 0 & 0 & 0 & X^\mu & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu \end{pmatrix}.$$

Similarly, for the right-hand side vector $R\nabla M(x,s,v,y,z,w)$ we obtain

$$\begin{pmatrix} I & 0 & -2A^TD_A^{-1} & -2J^TD_Y^{-1} & -2D_Z^{-1} & 0 \\ & I & 0 & 2D_Y^{-1} & 0 & -2D_W^{-1} \\ & & I & 0 & 0 & 0 \\ & & & I & 0 & 0 \\ & & & & Z & 0 \\ & & & & & W \end{pmatrix} \begin{pmatrix} g - A^T\big(\pi^A + (\pi^A - v)\big) - J^T\big(\pi^Y + (\pi^Y - y)\big) - \big(\pi^Z + (\pi^Z - z)\big) \\ \big(\pi^Y + (\pi^Y - y)\big) - \big(\pi^W + (\pi^W - w)\big) \\ -D_A(\pi^A - v) \\ -D_Y(\pi^Y - y) \\ -D_Z(\pi^Z - z) \\ -D_W(\pi^W - w) \end{pmatrix}$$

$$= \begin{pmatrix} g - A^Tv - J^Ty - z \\ y - w \\ -D_A(\pi^A - v) \\ -D_Y(\pi^Y - y) \\ -ZD_Z(\pi^Z - z) \\ -WD_W(\pi^W - w) \end{pmatrix} = \begin{pmatrix} g - A^Tv - J^Ty - z \\ y - w \\ Ax - b + \mu^A(v - v^E) \\ c - s + \mu^P(y - y^E) \\ z \cdot x + \mu^B(z - z^E) \\ w \cdot s + \mu^B(w - w^E) \end{pmatrix}.$$

This gives the (unsymmetric) transformed modified Newton equations

$$\begin{pmatrix} H & 0 & -A^T & -J^T & -I & 0 \\ 0 & 0 & 0 & I & 0 & -I \\ A & 0 & D_A & 0 & 0 & 0 \\ J & -I & 0 & D_Y & 0 & 0 \\ Z & 0 & 0 & 0 & X^\mu & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta v \\ \Delta y \\ \Delta z \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g - A^Tv - J^Ty - z \\ y - w \\ Ax - b + \mu^A(v - v^E) \\ c - s + \mu^P(y - y^E) \\ z \cdot x + \mu^B(z - z^E) \\ w \cdot s + \mu^B(w - w^E) \end{pmatrix},$$

which are equivalent to the path-following equations (5.5) associated with the perturbed optimality conditions (5.3).

5.5. Computation of the shifted primal-dual penalty-barrier direction

The path-following equations (5.5) may be written in symmetric form

$$\begin{pmatrix} H & 0 & A^T & J^T & I & 0 \\ 0 & 0 & 0 & -I & 0 & I \\ A & 0 & -D_A & 0 & 0 & 0 \\ J & -I & 0 & -D_Y & 0 & 0 \\ I & 0 & 0 & 0 & -D_Z & 0 \\ 0 & I & 0 & 0 & 0 & -D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ -\Delta v \\ -\Delta y \\ -\Delta z \\ -\Delta w \end{pmatrix} = - \begin{pmatrix} g - A^Tv - J^Ty - z \\ y - w \\ Ax - b + \mu^A(v - v^E) \\ c - s + \mu^P(y - y^E) \\ Z^{-1}\big(z \cdot x + \mu^B(z - z^E)\big) \\ W^{-1}\big(w \cdot s + \mu^B(w - w^E)\big) \end{pmatrix},$$

where $D_A = \mu^A I$, $D_Y = \mu^P I$, $D_Z = X^\mu Z^{-1}$ and $D_W = S^\mu W^{-1}$ from (5.8).

The solution of this system of equations is given by

$$\begin{aligned}
\Delta w &= y - w + \Delta y \\
\Delta s &= -W^{-1}\big(s \cdot (y + \Delta y) + \mu^B(y + \Delta y - w^E)\big) \\
\Delta v &= -D_A^{-1}\big(A(x + \Delta x) + \mu^A(v - v^E)\big) \\
\Delta z &= -(X^\mu)^{-1}\big(z \cdot (x + \Delta x) + \mu^B(z - z^E)\big),
\end{aligned}$$

where Δx and Δy satisfy the KKT system

$$\begin{pmatrix} H + A^T D_A^{-1} A + D_Z^{-1} & -J^T \\ J & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g - J^T y - z - A^T \pi^A + (X^\mu)^{-1}\big(z \cdot x + \mu^B(z - z^E)\big) \\ c - s + \mu^P(y - y^E) + W^{-1}\big(s \cdot y + \mu^B(y - w^E)\big) \end{pmatrix}.$$

The right-hand side may be simplified using the identity

$$\begin{aligned}
(X^\mu)^{-1}\big(z \cdot x + \mu^B(z - z^E)\big) &= (X + \mu^B I)^{-1}\big((X + \mu^B I)z - \mu^B z^E \\
&= z - \mu^B(X + \mu^B I)^{-1} z^E \\
&= z - \pi^Z.
\end{aligned}$$

Similarly,

$$\begin{aligned}
W^{-1}\big(s \cdot y + \mu^B(y - w^E)\big) &= W^{-1}\big((S + \mu^B I)y - \mu^B w^E\big) \\
&= (S + \mu^B I)W^{-1}\big(y - \mu^B(S + \mu^B I)^{-1} w^E\big) \\
&= D_W\big(y - \mu^B(S + \mu^B I)^{-1} w^E\big) \\
&= D_W(w - \pi^W).
\end{aligned}$$

It follows that the right-hand side is given by

$$\begin{aligned}
\begin{pmatrix} g - J^T y - z - A^T \pi^A + (X^\mu)^{-1}\big(z \cdot x + \mu^B(z - z^E)\big) \\ c - s + \mu^P(y - y^E) + W^{-1}\big(s \cdot y + \mu^B(y - w^E)\big) \end{pmatrix} &= \begin{pmatrix} g - J^T y - \pi^Z - A^T \pi^A \\ c - s + \mu^P(y - y^E) + W^{-1}\big(s \cdot y + \mu^B(y - w^E)\big) \end{pmatrix} \\
&= \begin{pmatrix} g - J^T y - \pi^Z - A^T \pi^A \\ D_Y(y - y^E) + D_W(w - \pi^W) \end{pmatrix}.
\end{aligned}$$

5.6. Summary

The results of Sections 5.1–5.5 imply that the solution of the path-following equations (5.5) may be computed as

$$
\begin{aligned}
\widehat{y} &= y + \Delta y, & \Delta s &= -D_W\big(\widehat{y} - \pi^W\big),\\
\widehat{s} &= s + \Delta s, & \Delta w &= -(S^\mu)^{-1}\big(w \cdot \widehat{s} + \mu^B(w - w^E)\big),\\
\widehat{x} &= x + \Delta x, & \Delta z &= -(X^\mu)^{-1}\big(z \cdot \widehat{x} + \mu^B(z - z^E)\big),\\
& & \Delta v &= \widehat{\pi}^A - v,
\end{aligned}
$$

where Δx and Δy satisfy the equations

$$
\begin{pmatrix} H(x,y) + A^T D_A^{-1} A + D_Z^{-1} & -J(x)^T \\ J(x) & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T y - A^T \pi^A - \pi^Z \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix},
$$

and D_A, D_Y, D_Z, D_W, π^Y, π^Z, π^W, π^A and $\widehat{\pi}^A$ are given by

$$
\begin{aligned}
D_A &= \mu^A I, & \pi^A &= v^E - \frac{1}{\mu^A}\big(Ax - b\big),\\
D_Y &= \mu^P I, & \pi^Y &= y^E - \frac{1}{\mu^P}\big(c(x) - s\big),\\
D_Z &= X^\mu Z^{-1}, & \pi^Z &= \mu^B (X^\mu)^{-1} z^E,\\
D_W &= S^\mu W^{-1}, & \pi^W &= \mu^B (S^\mu)^{-1} w^E,\\
& & \widehat{\pi}^A &= v^E - \frac{1}{\mu^A}\big(A\widehat{x} - b\big).
\end{aligned}
$$

The associated line-search merit function $M(x, s, v, y, z, w\,; \mu^A, \mu^P, \mu^B, v^E, y^E, z^E, w^E)$ is given by

$$
\begin{aligned}
f(x) - (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A(v - v^E)\|^2 \\
- \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\
- \sum_{i=1}^{m} \Big\{\mu^B w_i^E \ln\big(w_i(s_i + \mu^B)^2\big) - w_i(s_i + \mu^B)\Big\} - \sum_{j=1}^{n} \Big\{\mu^B z_j^E \ln\big(z_j(x_j + \mu^B)^2\big) - z_j(x_j + \mu^B)\Big\}.
\end{aligned}
$$

6. Fixed and Bounded Slacks with Linear Constraints

Next we consider nonlinear equality constraints and upper and lower bounds on the slacks. The variables are not subject to bounds.

6.1. Problem statement and optimality conditions

The problem has the form

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad L_X s = h_X, \quad \ell \le L_F s \le u, \tag{6.1}$$

where $c:\mathbb{R}^n\mapsto\mathbb{R}^m$ and $f:\mathbb{R}^n\mapsto\mathbb{R}$ are twice-continuously differentiable and L_X and L_F are fixed matrices of dimension $m_F\times m$ and $m_X\times m$, respectively, with $m=m_F+m_X$. The matrices L_X and L_F are formed from rows of the identity matrix I_m in such a way that $L_X s$ and $L_F s$ give the fixed and "free" components of s. It follows that there is an $m\times m$ permutation matrix P such that

$$P=\begin{pmatrix} L_F \\ L_X \end{pmatrix},$$

with the matrices L_F and L_X satisfying the identities $L_F L_F^T = I_F$, $L_X L_X^T = I_X$, and $L_F L_X^T = 0$. The first-order KKT conditions for this problem are

$$\begin{aligned} g(x^*) - J(x^*)^T y^* &= 0, & & \qquad (6.2\text{a})\\ c(x^*) - s^* &= 0, & L_X s^* - h_X &= 0, \qquad (6.2\text{b})\\ y^* - L_X^T w_X^* - L_F^T w_1^* + L_F^T w_2^* &= 0, & & \qquad (6.2\text{c})\\ L_F s^* - \ell &\ge 0, & u - L_F s^* &\ge 0, \qquad (6.2\text{d})\\ w_1^* &\ge 0, & w_2^* &\ge 0, \qquad (6.2\text{e})\\ w_1^* \cdot (L_F s^* - \ell) &= 0, & w_2^* \cdot (u - L_F s^*) &= 0, \qquad (6.2\text{f}) \end{aligned}$$

where y^* and w_X^* are the Lagrange multipliers for the equality constraints $c(x)-s=0$ and $L_X s = h_X$, and w_1^* and w_2^* may be interpreted as the Lagrange multipliers for the inequality constraints $L_F s - \ell \ge 0$ and $u - L_F s \ge 0$, respectively. Given any $s\ge 0$, we define the index set $\mathcal{F}$ of indices from 1, 2, …, m that define the rows of L_F.

6.2. The path-following equations

Let y^E be an estimate of the Lagrange multipliers for the nonlinear equality constraints $c(x)-s=0$. Similarly, let w_1^E and w_2^E denote nonnegative estimates of the multipliers for the inequality constraints $L_F s - \ell \ge 0$ and $u - L_F s \ge 0$, respectively. Given small positive scalars μ^P and μ^B, consider the perturbed optimality conditions

$$\begin{aligned} g(x) - J(x)^T y &= 0, & & \\ c(x) - s &= \mu^P(y^E - y), & L_X s - h_X &= 0, \\ y - L_X^T w_X - L_F^T w_1 + L_F^T w_2 &= 0, & & \\ L_F s - \ell &\geq 0, & u - L_F s &\geq 0, \\ w_1 &\geq 0, & w_2 &\geq 0, \\ w_1 \cdot (L_F s - \ell) &= \mu^B(w_1^E - w_1), & w_2 \cdot (u - L_F s) &= \mu^B(w_2^E - w_2). \end{aligned}$$

Consider the following primal-dual path following equations given by $F(x, s, y, w_X, w_1, w_2\,; \mu^P, \mu^B, y^E, w_1^E, w_2^E) = 0$, with

$$F(x, s, y, w_X, w_1, w_2\,; \mu^P, \mu^B, y^E, w_1^E, w_2^E) = \begin{pmatrix} g(x) - J(x)^T y \\ y - L_X^T w_X - L_F^T w_1 + L_F^T w_2 \\ c(x) - s + \mu^P(y - y^E) \\ L_X s - h_X \\ w_1 \cdot (L_F s - \ell) + \mu^B(w_1 - w_1^E) \\ w_2 \cdot (u - L_F s) + \mu^B(w_2 - w_2^E) \end{pmatrix}. \tag{6.4}$$

Any zero (x, s, y, w_X, w_1, w_2) of F satisfying $\ell < L_F s < u$, $w_1 > 0$, and $w_2 > 0$ approximates a point satisfying the optimality conditions (6.2), with the approximation becoming increasingly accurate as the terms $\mu^P(y - y^E)$, $\mu^B(w_1 - w_1^E)$ and $\mu^B(w_2 - w_2^E)$ approach zero. For any sequence of y^E and w_2^E such that $y^E \to y^*$, $w_1^E \to w_1^*$ and $w_2^E \to w_2^*$, it must hold that solutions (x, s, y, w_X, w_1, w_2) of (6.3) must satisfy $y \cdot (c(x) - s) \to 0$, $w_1 \cdot (L_F s - \ell) \to 0$, and $w_2 \cdot (u - L_F s) \to 0$. This implies that any solution (x, s, y, w_X, w_1, w_2) of (6.3) will approximate a solution of (6.2) independently of the values of μ^P and μ^B (i.e., it is not necessary that μ^P, $\mu^B \to 0$).

Given an approximate zero (x, s, y, w_X, w_1, w_2) of F such that $\ell < L_F s < u$, $w_1 > 0$, and $w_2 > 0$, the Newton equations for the change in variables $(\Delta x, \Delta s, \Delta y, \Delta w_X, \Delta w_1, \Delta w_2)$ are given by

$$\begin{pmatrix} H & 0 & -J^T & 0 & 0 & 0 \\ 0 & 0 & I_m & -L_X^T & -L_F^T & L_F^T \\ J & -I_m & D_Y & 0 & 0 & 0 \\ 0 & L_X & 0 & 0 & 0 & 0 \\ 0 & W_1 L_F & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 L_F & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta w_X \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y - L_X^T w_X - L_F^T w_1 + L_F^T w_2 \\ c - s + \mu^P(y - y^E) \\ L_X s - h_X \\ w_1 \cdot (L_F s - \ell) + \mu^B(w_1 - w_1^E) \\ w_2 \cdot (u - L_F s) + \mu^B(w_2 - w_2^E) \end{pmatrix}, \tag{6.5}$$

where $D_Y = \mu^P I$, $W_1 = \operatorname{diag}([w_1]_i)$, $W_2 = \operatorname{diag}([w_2]_i)$, $S_1^\mu = \operatorname{diag}(e_i^T s - \ell_i + \mu^B)$, and $S_2^\mu = \operatorname{diag}(u_i - e_i^T s + \mu^B)$.

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. Throughout, we assume that s_X satisfies $L_X s = h_X$, in which case the expansion of Δs satisfies

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

This identity allows us to write the equations (6.5) in the form

$$\begin{pmatrix} H & 0 & -J^T & 0 & 0 \\ 0 & 0 & L_F & -I_F & I_F \\ J & -L_F^T & D_Y & 0 & 0 \\ 0 & W_1 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s_F \\ \Delta y \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_1 + w_2 \\ c - s + \mu^P (y - y^E) \\ w_1 \cdot (L_F s - \ell) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u - L_F s) + \mu^B (w_2 - w_2^E) \end{pmatrix}. \tag{6.6}$$

The vectors Δs and Δw_X are recovered as $\Delta s = L_F^T \Delta s_F$ and $\Delta w_X = [\, y + \Delta y - w \,]_X$.

6.3. A shifted primal-dual penalty-barrier function

Problem (6.1) may be written in the equivalent form

$$\begin{aligned} \underset{x,s,s_1,s_2}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) - s = 0, \qquad L_F s - s_1 = \ell, \qquad s_1 \ge 0, \\ & L_X s - h_X = 0, \qquad L_F s + s_2 = u, \qquad s_2 \ge 0. \end{aligned}$$

The nonlinear equality constraints and bounds may be treated using shifted primal-dual penalty-barrier and augmented Lagrangian terms, which gives the approximate problem

$$\begin{aligned} \underset{x,s,s_1,s_2,y,w_1,w_2}{\text{minimize}} \quad & M(x, s, s_1, s_2, y, w_1, w_2\,; \mu^P, \mu^B, y^E, w_1^E, w_2^E) \\ \text{subject to} \quad & L_X s - h_X = 0, \\ & L_F s - s_1 = \ell, \qquad s_1 + \mu^B e > 0, \qquad w_1 > 0, \\ & L_F s + s_2 = u, \qquad s_2 + \mu^B e > 0, \qquad w_2 > 0, \end{aligned} \tag{6.7}$$

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

$$\begin{aligned} f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\ - \sum_{i=1}^{n_L} \Big\{ \mu^B [w_1^E]_i \ln\big([\,s_1 + \mu^B e\,]_i\big) + \mu^B [w_1^E]_i \ln\big([\,w_1 \cdot (s_1 + \mu^B e)\,]_i\big) - [\,w_1 \cdot (s_1 + \mu^B e)\,]_i \Big\} \\ - \sum_{i=1}^{n_U} \Big\{ \mu^B [w_2^E]_i \ln\big([\,s_2 + \mu^B e\,]_i\big) + \mu^B [w_2^E]_i \ln\big([\,w_2 \cdot (s_2 + \mu^B e)\,]_i\big) - [\,w_2 \cdot (s_2 + \mu^B e)\,]_i \Big\}. \end{aligned} \tag{6.8}$$

Let c, g and J denote the quantities $c(x)$, $g(x)$ and $J(x)$. Differentiating $M(x, s, s_1, s_2, y, w_1, w_2)$ with respect to x, s, s_1, s_2, y, w_1 and w_2 gives

$$\nabla M(x, s, s_1, s_2, y, w_1, w_2) = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c - s)) - y\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c - s)\big) - y \\ w_1 - 2\mu^B (S_1^\mu)^{-1} w_1^E \\ w_2 - 2\mu^B (S_2^\mu)^{-1} w_2^E \\ c - s + \mu^P(y - y^E) \\ s_1 + \mu^B e - \mu^B W_1^{-1} w_1^E \\ s_2 + \mu^B e - \mu^B W_2^{-1} w_2^E \end{pmatrix}.$$

The gradient may be written in several equivalent forms

$$\nabla M(x,s,s_1,s_2,y,w_1,w_2) = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y \\ w_1 - 2\mu^B(S_1^\mu)^{-1}w_1^E \\ w_2 - 2\mu^B(S_2^\mu)^{-1}w_2^E \\ c - s + \mu^P(y - y^E) \\ s_1 + \mu^B e - \mu^B W_1^{-1}w_1^E \\ s_2 + \mu^B e - \mu^B W_2^{-1}w_2^E \end{pmatrix} = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y \\ (S_1^\mu)^{-1}\big(w_1 \cdot s_1 + \mu^B w_1^E + \mu^B(w_1 - w_1^E)\big) \\ (S_2^\mu)^{-1}\big(w_2 \cdot s_2 + \mu^B w_2^E + \mu^B(w_2 - w_2^E)\big) \\ c - s + \mu^P(y - y^E) \\ W_1^{-1}\big(w_1 \cdot s_1 + \mu^B(w_1 - w_1^E)\big) \\ W_2^{-1}\big(w_2 \cdot s_2 + \mu^B(w_2 - w_2^E)\big) \end{pmatrix}$$

$$= \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi^Y + (\pi^Y - y) \\ -\big(\pi_1^W + (\pi_1^W - w_1)\big) \\ -\big(\pi_2^W + (\pi_2^W - w_2)\big) \\ -D_Y(\pi^Y - y) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix},$$

where

$$D_Y = \mu^P I_m, \qquad \pi^Y = y^E - \frac{1}{\mu^P}(c-s), \tag{6.9a}$$

$$D_1^W = S_1^\mu W_1^{-1}, \qquad \pi_1^W = \mu^B(S_1^\mu)^{-1}w_1^E, \tag{6.9b}$$

$$D_2^W = S_2^\mu W_2^{-1}, \qquad \pi_2^W = \mu^B(S_2^\mu)^{-1}w_2^E. \tag{6.9c}$$

Similarly, the Hessian of $M(x,s,s_1,s_2,y,w_1,w_2)$ is given by

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^P}J^T & 0 & 0 & J^T & 0 & 0 \\ -\frac{2}{\mu^P}J & \frac{2}{\mu^P}I_m & 0 & 0 & -I_m & 0 & 0 \\ 0 & 0 & 2\mu^B(S_1^\mu)^{-2}W_1^E & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 2\mu^B(S_2^\mu)^{-2}W_2^E & 0 & 0 & I_F \\ J & -I_m & 0 & 0 & \mu^P I_m & 0 & 0 \\ 0 & 0 & I_F & 0 & 0 & \mu^B W_1^{-2}W_1^E & 0 \\ 0 & 0 & 0 & I_F & 0 & 0 & \mu^B W_2^{-2}W_2^E \end{pmatrix},$$

where $H_1 = H(x, 2\pi^{\scriptscriptstyle Y} - y) + \frac{2}{\mu^{\scriptscriptstyle P}} J^T J$. Substituting $\mu^{\scriptscriptstyle B} W_1^{\scriptscriptstyle E} = S_1^\mu \Pi_1^{\scriptscriptstyle W}$ and $\mu^{\scriptscriptstyle B} W_2^{\scriptscriptstyle E} = S_2^\mu \Pi_2^{\scriptscriptstyle W}$ from (6.9) gives the Hessian

$$\begin{pmatrix} H_1 & -\frac{2}{\mu^{\scriptscriptstyle P}} J^T & 0 & 0 & J^T & 0 & 0 \\ -\frac{2}{\mu^{\scriptscriptstyle P}} J & \frac{2}{\mu^{\scriptscriptstyle P}} I_m & 0 & 0 & -I_m & 0 & 0 \\ 0 & 0 & 2(S_1^\mu)^{-1}\Pi_1^{\scriptscriptstyle W} & 0 & 0 & I_{\scriptscriptstyle F} & 0 \\ 0 & 0 & 0 & 2(S_2^\mu)^{-1}\Pi_2^{\scriptscriptstyle W} & 0 & 0 & I_{\scriptscriptstyle F} \\ J & -I_m & 0 & 0 & \mu^{\scriptscriptstyle P} I_m & 0 & 0 \\ 0 & 0 & I_{\scriptscriptstyle F} & 0 & 0 & S_1^\mu W_1^{-2}\Pi_1^{\scriptscriptstyle W} & 0 \\ 0 & 0 & 0 & I_{\scriptscriptstyle F} & 0 & 0 & S_2^\mu W_2^{-2}\Pi_2^{\scriptscriptstyle W} \end{pmatrix}.$$

6.4. Derivation of the shifted primal-dual penalty-barrier direction

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

$$\underset{p\in\mathcal{I}}{\text{minimize}} \;\; M(p) \quad \text{subject to} \;\; Cp = b_{\scriptscriptstyle C}, \tag{6.10}$$

where

$$\mathcal{I} = \{p : p = (x, s, s_1, s_2, y, w_1, w_2), \;\; \text{with } s_1 + \mu^{\scriptscriptstyle B} e > 0,\; s_2 + \mu^{\scriptscriptstyle B} e > 0,\; w_1 > 0,\; w_2 > 0\},$$

with

$$C = \begin{pmatrix} 0 & L_X & 0 & 0 & 0 & 0 & 0 \\ 0 & L_{\scriptscriptstyle F} & -I_{\scriptscriptstyle F} & 0 & 0 & 0 & 0 \\ 0 & L_{\scriptscriptstyle F} & 0 & I_{\scriptscriptstyle F} & 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad b_{\scriptscriptstyle C} = \begin{pmatrix} \ell \\ u \end{pmatrix}.$$

Let $p \in \mathcal{I}$ be given. For the moment, assume that p is not necessarily feasible for the linear constraints, i.e., it may not hold that $L_{\scriptscriptstyle F} s - s_1 = \ell$ and $L_{\scriptscriptstyle F} s + s_2 = u$, in which case $b_{\scriptscriptstyle C} - Cp$ may not be zero. The Newton direction Δp is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2} \Delta p^T \nabla^2 M(p) \Delta p \quad \text{subject to} \;\; C\Delta p = b_{\scriptscriptstyle C} - Cp. \tag{6.11}$$

However, instead of solving (6.11), we define a linearly constrained modified Newton method by approximating the Hessian $\nabla^2 M(x, s, s_1, s_2, y, w_1, w_2)$ by a matrix $B(x, s, s_1, s_2, y, w_1, w_2)$. Consider the matrix defined by replacing $\pi^{\scriptscriptstyle Y}$ by y, $\pi_1^{\scriptscriptstyle W}$ by w_1, and $\pi_2^{\scriptscriptstyle W}$ by w_2, everywhere in the matrix $\nabla^2 M(x, s, s_1, s_2, y, w_1, w_2)$. This gives an approximate Hessian $B(x, s, s_1, s_2, y, w_1, w_2)$

of the form

$$\begin{pmatrix} \widehat{H}_1 & -\frac{2}{\mu^P}J^T & 0 & 0 & J^T & 0 & 0 \\ -\frac{2}{\mu^P}J & \frac{2}{\mu^P}I_m & 0 & 0 & -I_m & 0 & 0 \\ 0 & 0 & 2(S_1^\mu)^{-1}W_1 & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 2(S_2^\mu)^{-1}W_2 & 0 & 0 & I_F \\ J & -I_m & 0 & 0 & \mu^P I_m & 0 & 0 \\ 0 & 0 & I_F & 0 & 0 & S_1^\mu W_1^{-1} & 0 \\ 0 & 0 & 0 & I_F & 0 & 0 & S_2^\mu W_2^{-1} \end{pmatrix},$$

where $\widehat{H}_1 = H(x,y) + 2J^T D_Y^{-1} J$. The definitions of D_Y, D_1^W, and D_2^W may be used to write $B(x,s,s_1,s_2,y,w_1,w_2)$ as

$$\begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} & 0 & 0 & J^T & 0 & 0 \\ -2D_Y^{-1}J & 2D_Y^{-1} & 0 & 0 & -I_m & 0 & 0 \\ 0 & 0 & 2(D_1^W)^{-1} & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 2(D_2^W)^{-1} & 0 & 0 & I_F \\ J & -I_m & 0 & 0 & D_Y & 0 & 0 \\ 0 & 0 & I_F & 0 & 0 & D_1^W & 0 \\ 0 & 0 & 0 & I_F & 0 & 0 & D_2^W \end{pmatrix},$$

where $H = H(x,y)$. Given $B(p) = B(x,s,s_1,s_2,y,w_1,w_2)$, a modified Newton direction is given by the solution of the QP subproblem

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

Let N denote a matrix whose columns form a basis for null(C), i.e., $CN = 0$ and $\begin{pmatrix} C^T & N \end{pmatrix}$ is nonsingular. The vector

$$\Delta p_0 = \begin{pmatrix} 0 \\ 0 \\ -(\ell - s + s_1) \\ (u - s - s_2) \\ 0 \\ 0 \\ 0 \end{pmatrix} \triangleq \begin{pmatrix} 0 \\ 0 \\ -r_L \\ r_U \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{6.12}$$

satisfies $C\Delta p_0 = b_C - Cp$, and it follows that every feasible Δp may be written in the form

$$\Delta p = \Delta p_0 + Nd.$$

This implies that d must satisfy the reduced equations

$$N^T B(p) N d = -N^T\big(\nabla M(p) + B(p)\Delta p_0\big).$$

Consider the particular null-space basis given by

$$N = \begin{pmatrix} I_n & 0 & 0 & 0 & 0 \\ 0 & L_F^T & 0 & 0 & 0 \\ 0 & I_F & 0 & 0 & 0 \\ 0 & -I_F & 0 & 0 & 0 \\ 0 & 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 0 & I_F \end{pmatrix}. \tag{6.13}$$

The definition of N of (6.13) gives the reduced Hessian

$$N^T B(p) N = \begin{pmatrix} H + 2J^T D_Y^{-1} J & -2J^T D_Y^{-1} L_F^T & J^T & 0 & 0 \\ -2L_F D_Y^{-1} J & 2\big(L_F D_Y^{-1} L_F^T + \bar{D}_W^{-1}\big) & -L_F & I_F & -I_F \\ J & -L_F^T & D_Y & 0 & 0 \\ 0 & I_F & 0 & D_1^W & 0 \\ 0 & -I_F & 0 & 0 & D_2^W \end{pmatrix},$$

where $\bar{D}_W^{-1} = (D_1^W)^{-1} + (D_2^W)^{-1}$. Similarly, the reduced gradient is

$$N^T \nabla M(p) = \begin{pmatrix} I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_F & I_F & -I_F & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_F \end{pmatrix} \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi^Y + (\pi^Y - y) \\ -\big(\pi_1^W + (\pi_1^W - w_1)\big) \\ -\big(\pi_2^W + (\pi_2^W - w_2)\big) \\ -D_Y(\pi^Y - y) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}$$

$$= \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi_F^Y + (\pi_F^Y - y_F) - \big(\pi_1^W + (\pi_1^W - w_1)\big) + \big(\pi_2^W + (\pi_2^W - w_2)\big) \\ -D_Y(\pi^Y - y) \\ -D_1^W(\pi_1^W - w_1) \\ -D_2^W(\pi_2^W - w_2) \end{pmatrix}.$$

Moreover

$$B(p)\Delta p_0 = \begin{pmatrix} 0 \\ 0 \\ -2(D_1^W)^{-1}(\ell - L_F s + s_1) \\ 2(D_2^W)^{-1}(u - L_F s - s_2) \\ 0 \\ -(\ell - L_F s + s_1) \\ (u - L_F s - s_2) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -2(D_1^W)^{-1} r_L \\ 2(D_2^W)^{-1} r_U \\ 0 \\ -r_L \\ r_U \end{pmatrix},$$

where $r_L = \ell - L_F s + s_1$ and $r_U = u - L_F s - s_2$. This implies that $N^T B(p)\Delta p_0$ is given by

$$\begin{pmatrix} I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_F & I_F & -I_F & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_F \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ -2(D_1^W)^{-1} r_L \\ 2(D_2^W)^{-1} r_U \\ 0 \\ -r_L \\ r_U \end{pmatrix} = \begin{pmatrix} 0 \\ -2\big((D_1^W)^{-1} r_L + (D_2^W)^{-1} r_U\big) \\ 0 \\ -r_L \\ r_U \end{pmatrix}.$$

This gives $N^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ such that

$$\begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi_F^Y + (\pi_F^Y - y_F) - \big(\pi_1^W + (\pi_1^W - w_1)\big) + \big(\pi_2^W + (\pi_2^W - w_2)\big) - 2\big((D_1^W)^{-1} r_L + (D_2^W)^{-1} r_U\big) \\ -D_Y(\pi^Y - y) \\ -D_1^W(\pi_1^W - w_1) - r_L \\ -D_2^W(\pi_2^W - w_2) + r_U \end{pmatrix}.$$

The reduced modified equations $N^TB(p)Nd = -N^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ are then

$$\begin{pmatrix} H + 2J^TD_Y^{-1}J & -2J^TD_Y^{-1}L_F^T & J^T & 0 & 0 \\ -2L_FD_Y^{-1}J & 2\big(L_FD_Y^{-1}L_F^T + \bar{D}_W^{-1}\big) & -L_F & I_F & -I_F \\ J & -L_F^T & D_Y & 0 & 0 \\ 0 & I_F & 0 & D_1^W & 0 \\ 0 & -I_F & 0 & 0 & D_2^W \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix}$$
$$= \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ \pi_F^Y + (\pi_F^Y - y_F) - \big(\pi_1^W + (\pi_1^W - w_1)\big) + \big(\pi_2^W + (\pi_2^W - w_2)\big) - 2\big((D_1^W)^{-1}r_L + (D_2^W)^{-1}r_U\big) \\ -D_Y(\pi^Y - y) \\ -D_1^W(\pi_1^W - w_1) - r_L \\ -D_2^W(\pi_2^W - w_2) + r_U \end{pmatrix}.$$

Given any nonsingular matrix R, the direction d satisfies

$$RN^TB(p)Nd = -RN^T\big(\nabla M(p) + B(p)\Delta p_0\big).$$

In particular, consider

$$R = \begin{pmatrix} I_n & 0 & -2J^TD_Y^{-1} & 0 & 0 \\ & I_F & 2L_FD_Y^{-1} & -2(D_1^W)^{-1} & 2(D_2^W)^{-1} \\ & & I_m & 0 & 0 \\ & & & W_1 & 0 \\ & & & & W_2 \end{pmatrix},$$

which is nonsingular if W_1 and W_2 are positive definite, with

$$R^{-1} = \begin{pmatrix} I_n & 0 & 2J^TD_Y^{-1} & 0 & 0 \\ & I_F & -2L_FD_Y^{-1} & 2(S_1^\mu)^{-1} & -2(S_2^\mu)^{-1} \\ & & I_m & 0 & 0 \\ & & & W_1^{-1} & 0 \\ & & & & W_2^{-1} \end{pmatrix}.$$

For this R, the product $RN^TB(p)N$ is given by

$$
\begin{pmatrix}
I_n & 0 & -2J^TD_Y^{-1} & 0 & 0 \\
 & I_F & 2L_FD_Y^{-1} & -2(D_1^W)^{-1} & 2(D_2^W)^{-1} \\
 & & I_m & 0 & 0 \\
 & & & W_1 & 0 \\
 & & & & W_2
\end{pmatrix}
\begin{pmatrix}
H + 2J^TD_Y^{-1}J & -2J^TD_Y^{-1}L_F^T & J^T & 0 & 0 \\
-2L_FD_Y^{-1}J & 2\big(L_FD_Y^{-1}L_F^T + \bar{D}_W^{-1}\big) & -L_F & I_F & -I_F \\
J & -L_F^T & D_Y & 0 & 0 \\
0 & I_F & 0 & D_1^W & 0 \\
0 & -I_F & 0 & 0 & D_2^W
\end{pmatrix}
$$

$$
= \begin{pmatrix}
H & 0 & -J^T & 0 & 0 \\
0 & 0 & L_F & -I_F & I_F \\
J & -L_F^T & D_Y & 0 & 0 \\
0 & W_1 & 0 & W_1D_1^W & 0 \\
0 & -W_2 & 0 & 0 & W_2D_2^W
\end{pmatrix}
= \begin{pmatrix}
H & 0 & -J^T & 0 & 0 \\
0 & 0 & L_F & -I_F & I_F \\
J & -L_F^T & D_Y & 0 & 0 \\
0 & W_1 & 0 & S_1^\mu & 0 \\
0 & -W_2 & 0 & 0 & S_2^\mu
\end{pmatrix}.
$$

Similarly $RN^T\nabla M(p)$ is given by

$$
\begin{pmatrix}
g - J^Ty \\
y_F - w_1 + w_2 \\
-D_Y(\pi^Y - y) \\
-W_1D_1^W(\pi_1^W - w_1) \\
-W_2D_2^W(\pi_2^W - w_2)
\end{pmatrix},
$$

and $RN^TB(p)\varDelta p_0$ is

$$
\begin{pmatrix}
I_n & 0 & -2J^TD_Y^{-1} & 0 & 0 \\
 & I_F & 2L_FD_Y^{-1} & -2(D_1^W)^{-1} & 2(D_2^W)^{-1} \\
 & & I_m & 0 & 0 \\
 & & & W_1 & 0 \\
 & & & & W_2
\end{pmatrix}
\begin{pmatrix}
0 \\
-2\big((D_1^W)^{-1}r_L + (D_2^W)^{-1}r_U\big) \\
0 \\
0 \\
-r_L \\
r_U
\end{pmatrix}
= \begin{pmatrix}
0 \\ 0 \\ 0 \\ 0 \\ -W_1r_L \\ W_2r_U
\end{pmatrix}.
$$

Putting all this together gives the following transformed unsymmetric reduced modified Newton equations for d

$$
\begin{pmatrix}
H & 0 & -J^T & 0 & 0 \\
0 & 0 & L_F & -I_F & I_F \\
J & -L_F^T & D_Y & 0 & 0 \\
0 & W_1 & 0 & S_1^\mu & 0 \\
0 & -W_2 & 0 & 0 & S_2^\mu
\end{pmatrix}
\begin{pmatrix}
d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5
\end{pmatrix}
= - \begin{pmatrix}
g - J^Ty \\
y_F - w_1 + w_2 \\
-D_Y(\pi^Y - y) \\
-W_1\big(D_1^W(\pi_1^W - w_1) + r_L\big) \\
-W_2\big(D_2^W(\pi_2^W - w_2) - r_U\big)
\end{pmatrix},
$$

or, equivalently,

$$\begin{pmatrix} H & 0 & -J^T & 0 & 0 \\ 0 & 0 & L_F & -I_F & I_F \\ J & -L_F^T & D_Y & 0 & 0 \\ 0 & W_1 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_1 + w_2 \\ c - s + \mu^P (y - y^E) \\ w_1 \cdot (L_F s - \ell) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u - L_F s) + \mu^B (w_2 - w_2^E) \end{pmatrix}. \tag{6.14}$$

Then, (6.12) and (6.13) implies that

$$\begin{pmatrix} \Delta x \\ \Delta s \\ \Delta s_1 \\ \Delta s_2 \\ \Delta y \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = \Delta p = \Delta p_0 + N d = \begin{pmatrix} 0 \\ 0 \\ -r_L \\ r_U \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} d_1 \\ L_F^T d_2 \\ d_2 \\ -d_2 \\ d_3 \\ d_4 \\ d_5 \end{pmatrix} = \begin{pmatrix} d_1 \\ L_F^T d_2 \\ (d_2 - r_L) \\ -(d_2 - r_U) \\ d_3 \\ d_4 \\ d_5 \end{pmatrix}.$$

These identities allow us to write equations (6.14) as

$$\begin{pmatrix} H & 0 & -J^T & 0 & 0 \\ 0 & 0 & L_F & -I_F & I_F \\ J & -L_F^T & D_Y & 0 & 0 \\ 0 & W_1 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s_F \\ \Delta y \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ y_F - w_1 + w_2 \\ c - s + \mu^P (y - y^E) \\ w_1 \cdot (L_F s - \ell) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u - L_F s) + \mu^B (w_2 - w_2^E) \end{pmatrix}, \tag{6.15}$$

with $\Delta s = L_F^T \Delta s_F$, $\Delta s_1 = \Delta s - (\ell - L_F s + s_1)$ and $\Delta s_2 = -\Delta s + (u - L_F s - s_2)$. The Newton equations have been derived for arbitrary interior s_1 and s_2, i.e., it is not assumed that s_1 and s_2 satisfy the linear constraints $L_F s - s_1 = \ell$ and $L_F s + s_2 = u$. However, unless an extra term is added to the objective function of (6.10) that forces the linear constraints to become feasible, it is necessary to choose feasible s_1 and s_2. In this case, $L_F s - s_1 = \ell$ and $L_F s + s_2 = u$, and it follows that $\Delta s_1 = \Delta s_F$ and $\Delta s_2 = -\Delta s_F$. This assumption is made for the remainder of this section.

Under the feasibility assumption, if S_1 and S_2 are written in terms of s, i.e., $S_1 = \operatorname{diag}\big(e_i^T s - \ell_i\big)$ and $S_2 = \operatorname{diag}\big(u_i - e_i^T s\big)$, then the equations (6.15) are the Newton equations for the solution of the perturbed optimality conditions (6.3). The variables

s_1 and s_2 may be computed implicitly for the line search, in which case the appropriate merit function is

$$
\begin{aligned}
f - (c-s)^T y^{E} + \frac{1}{2\mu^{P}}\|c-s\|^2 + \frac{1}{2\mu^{P}}\|c-s+\mu^{P}(y-y^{E})\|^2 \\
- \sum_{i\in\mathcal{F}} \Big\{ \mu^{B}[w_1^{E}]_i \ln\big(s_i - \ell_i + \mu^{B}\big) + \mu^{B}[w_1^{E}]_i \ln\big([w_1]_i(s_i-\ell_i+\mu^{B})\big) - [w_1]_i(s_i-\ell_i+\mu^{B}) \Big\} \\
- \sum_{i\in\mathcal{F}} \Big\{ \mu^{B}[w_2^{E}]_i \ln\big(u_i - s_i + \mu^{B}\big) + \mu^{B}[w_2^{E}]_i \ln\big([w_2]_i(u_i-s_i+\mu^{B})\big) - [w_2]_i(u_i-s_i+\mu^{B}) \Big\},
\end{aligned}
$$

where $\mathcal{F}$ denotes the index set of slacks with upper and lower bounds.

6.5. Computation of the shifted primal-dual penalty-barrier direction

Next we consider the solution of the modified Newton equations (6.15), which are written in the form

$$
\begin{pmatrix}
H & 0 & -J^T & 0 & 0 \\
0 & 0 & L_{F} & -I_{F} & I_{F} \\
J & -L_{F}^T & D_{Y} & 0 & 0 \\
0 & W_1 & 0 & W_1 D_1^{W} & 0 \\
0 & -W_2 & 0 & 0 & W_2 D_2^{W}
\end{pmatrix}
\begin{pmatrix} \Delta x \\ \Delta s_F \\ \Delta y \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix}
= -
\begin{pmatrix}
g - J^T y \\
y_{F} - w_1 + w_2 \\
-D_{Y}(\pi^{Y} - y) \\
-W_1\big(D_1^{W}(\pi_1^{W} - w_1)\big) \\
-W_2\big(D_2^{W}(\pi_2^{W} - w_2)\big)
\end{pmatrix}. \tag{6.16}
$$

Consider the following reordered set of equations and variables involving (in order) Δw_1, Δw_2, Δs_F, Δx and Δy:

$$
\begin{pmatrix}
I_{F} & 0 & (D_1^{W})^{-1} & 0 & 0 \\
0 & I_{F} & -(D_2^{W})^{-1} & 0 & 0 \\
-I_{F} & I_{F} & 0 & 0 & L_{F} \\
0 & 0 & -L_{F}^T & J & D_{Y} \\
 & & & H & -J^T
\end{pmatrix}
\begin{pmatrix} \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x \\ \Delta y \end{pmatrix}
= -
\begin{pmatrix}
w_1 - \pi_1^{W} \\
w_2 - \pi_2^{W} \\
y_{F} - w_1 + w_2 \\
D_{Y}(y - \pi^{Y}) \\
g - J^T y
\end{pmatrix}. \tag{6.17}
$$

If, as above, $\bar{D}_W$ denotes the matrix $\bar{D}_W = \big((D_1^{W})^{-1} + (D_2^{W})^{-1}\big)^{-1}$, then applying the nonsingular matrix

$$
\begin{pmatrix}
I_{F} & & & & \\
0 & I_{F} & & & \\
I_{F} & -I_{F} & I_{F} & & \\
L_{F}^T \bar{D}_W & -L_{F}^T \bar{D}_W & L_{F}^T \bar{D}_W & I_m & \\
 & & & & I_n
\end{pmatrix}
$$

on the left and right-hand side of (6.17) gives the block upper-trapezoidal system

$$\begin{pmatrix} I_F & 0 & (D_1^W)^{-1} & 0 & 0 \\ & I_F & -(D_2^W)^{-1} & 0 & 0 \\ & & \bar{D}_W^{-1} & 0 & L_F \\ & & & J & D_Y + L_F^T \bar{D}_W L_F \\ & & & H & -J^T \end{pmatrix} \begin{pmatrix} \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} w_1 - \pi_1^W \\ w_2 - \pi_2^W \\ y_F - \pi^W \\ D_Y(y - \pi^Y) + L_F^T \bar{D}_W \big(y_F - \pi^W\big) \\ g - J^T y \end{pmatrix}, \tag{6.18}$$

where $\pi^W = \pi_1^W - \pi_2^W$. Solving (6.18) for Δy and Δs_F, and using the last two block equations of (6.15) for Δw_1 and Δw_2 gives the solution of the path-following equations as

$$\begin{aligned} \Delta s_F &= -\bar{D}_W \big(y_F + \Delta y_F - \pi^W\big), \\ \Delta s &= L_F^T \Delta s_F, \\ \Delta w_1 &= -(S_1^\mu)^{-1} \big(w_1 \cdot (L_F(s + \Delta s) - \ell + \mu^B e) - \mu^B w_1^E\big), \\ \Delta w_2 &= -(S_2^\mu)^{-1} \big(w_2 \cdot (u - L_F(s + \Delta s) + \mu^B e) - \mu^B w_2^E\big), \end{aligned}$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H & -J^T \\ J & D_Y + L_F^T \bar{D}_W L_F \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g - J^T y \\ D_Y(y - \pi^Y) + L_F^T \bar{D}_W \big(y_F - \pi^W\big) \end{pmatrix}.$$

6.6. Summary: bounded slacks

Consider the quantities

$$\begin{aligned} D_Y &= \mu^P I, & \pi^Y &= y^E - \frac{1}{\mu^P}\big(c(x) - s\big), \\ D_1^W &= S_1^\mu W_1^{-1}, & \pi_1^W &= \mu^B (S_1^\mu)^{-1} w_1^E, \\ D_2^W &= S_2^\mu W_2^{-1}, & \pi_2^W &= \mu^B (S_2^\mu)^{-1} w_2^E, \\ \bar{D}_W &= \big((D_1^W)^{-1} + (D_2^W)^{-1}\big)^{-1}, & \pi^W &= \pi_1^W - \pi_2^W, \end{aligned}$$

then Δs, Δs_1, Δs_2, Δw_1 and Δw_1 are given by

$$\begin{aligned} \widehat{y} = y + \Delta y, \qquad \Delta s_{\scriptscriptstyle F} &= -\bar{D}_W\big(\widehat{y}_{\scriptscriptstyle F} - \pi^{\scriptscriptstyle W}\big), \\ \Delta s &= L_{\scriptscriptstyle F}^T \Delta s_{\scriptscriptstyle F}, \\ \Delta w_X &= [\,\widehat{y} - w\,]_X, \\ \widehat{s} = s + \Delta s, \qquad \Delta w_1 &= -(S_1^\mu)^{-1}\big(w_1 \cdot (L_{\scriptscriptstyle F}\widehat{s} - \ell + \mu^{\scriptscriptstyle B} e) - \mu^{\scriptscriptstyle B} w_1^{\scriptscriptstyle E}\big), \\ \Delta w_2 &= -(S_2^\mu)^{-1}\big(w_2 \cdot (u - L_{\scriptscriptstyle F}\widehat{s} + \mu^{\scriptscriptstyle B} e) - \mu^{\scriptscriptstyle B} w_2^{\scriptscriptstyle E}\big), \end{aligned}$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H & -J^T \\ J & D_{\scriptscriptstyle Y} + L_{\scriptscriptstyle F}^T \bar{D}_{\scriptscriptstyle W} L_{\scriptscriptstyle F} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g - J^T y \\ D_{\scriptscriptstyle Y}(y - \pi^{\scriptscriptstyle Y}) + L_{\scriptscriptstyle F}^T \bar{D}_{\scriptscriptstyle W}\big(y_{\scriptscriptstyle F} - \pi^{\scriptscriptstyle W}\big) \end{pmatrix}.$$

The associated line-search merit function is given by

$$\begin{aligned} f(x) - \big(c(x) - s\big)^T y^{\scriptscriptstyle E} + \frac{1}{2\mu^{\scriptscriptstyle P}}\|c(x) - s\|^2 + \frac{1}{2\mu^{\scriptscriptstyle P}}\|c(x) - s + \mu^{\scriptscriptstyle P}(y - y^{\scriptscriptstyle E})\|^2 \\ - \sum_{i \in \mathcal{F}} \Big\{ \mu^{\scriptscriptstyle B}[w_1^{\scriptscriptstyle E}]_i \ln\big([w_1]_i(e_i^T s - \ell_i + \mu^{\scriptscriptstyle B})^2\big) - [w_1]_i(e_i^T s - \ell_i + \mu^{\scriptscriptstyle B})\Big\} \\ - \sum_{i \in \mathcal{F}} \Big\{ \mu^{\scriptscriptstyle B}[w_2^{\scriptscriptstyle E}]_i \ln\big([w_2]_i(u_i - e_i^T s + \mu^{\scriptscriptstyle B})^2\big) - [w_2]_i(u_i - e_i^T s + \mu^{\scriptscriptstyle B})\Big\}. \qquad (6.19) \end{aligned}$$

7. Fixed and Bounded Variables

Next we consider nonlinear equality constraints and upper and lower bounds on the variables but only nonnegativity constraints for the slacks.

7.1. Problem statement and optimality conditions

The problem has the form

$$\underset{x\in\mathbb{R}^n,s\in\mathbb{R}^m}{\text{minimize}}\; f(x) \quad \text{subject to} \quad c(x)-s=0, \quad s\geq 0, \quad E_X x = b_X, \quad \ell \leq E_F x \leq u, \tag{7.1}$$

where $c:\mathbb{R}^n \mapsto \mathbb{R}^m$ and $f:\mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable and E_X and E_F are fixed matrices of dimension $n_F \times n$ and $n_X \times n$, respectively, with $n = n_F + n_X$. The matrices E_X and E_F are formed from rows of the identity matrix I_n in such a way that $E_X x$ and $E_F x$ give the fixed and "free" components of x. It follows that there is an n by n permutation matrix P such that

$$P = \begin{pmatrix} E_F \\ E_X \end{pmatrix},$$

with the matrices E_F and E_X satisfying the identities $E_F E_F^T = I_F$, $E_X E_X^T = I_X$, and $E_F E_X^T = 0$. The first-order KKT conditions for this problem are

$$\begin{aligned}
g(x^*) - J(x^*)^T y^* - E_X^T z_X^* - E_F^T z_1^* + E_F^T z_2^* &= 0, & z_1^* &\geq 0, \quad z_2^* \geq 0, && \text{(7.2a)}\\
y^* - w^* &= 0, & w^* &\geq 0, && \text{(7.2b)}\\
c(x^*) - s^* &= 0, & s^* &\geq 0, && \text{(7.2c)}\\
E_F x^* - \ell &\geq 0, & u - E_F x^* &\geq 0, && \text{(7.2d)}\\
z_1^* \cdot (E_F x^* - \ell) &= 0, & z_2^* \cdot (u - E_F x^*) &= 0, && \text{(7.2e)}\\
w^* \cdot s^* &= 0, & & && \text{(7.2f)}\\
E_X x^* - b_X &= 0, & & && \text{(7.2g)}
\end{aligned}$$

where y^* and z_X^* are the multipliers for the equality constraints $c(x)-s=0$ and $E_X x = b_X$, and z_1^*, z_2^* and w^* may be interpreted as the Lagrange multipliers for the constraints $E_F x - \ell \geq 0$, $u - E_F x \geq 0$, and $s \geq 0$ respectively.

7.2. The path-following equations

Let y^E, z_1^E, z_2^E, and w^E denote nonnegative estimates of the Lagrange multipliers for the inequality constraints $E_F x - \ell \geq 0$, $u - E_F x \geq 0$, and $s \geq 0$, respectively. Given small positive scalars μ^P and μ^B, consider the perturbed optimality conditions

$$\begin{aligned}
g(x) - J(x)^T y - E_X^T z_X - E_F^T z_1 + E_F^T z_2 &= 0, & z_1 &\geq 0, \quad z_2 \geq 0, && \text{(7.3a)}\\
y - w &= 0, & w &\geq 0, && \text{(7.3b)}\\
c(x) - s &= \mu^P (y^E - y), & s &\geq 0, && \text{(7.3c)}\\
E_F x - \ell &\geq 0, & u - E_F x &\geq 0, && \text{(7.3d)}\\
z_1 \cdot (E_F x - \ell) &= \mu^B (z_1^E - z_1), & z_2 \cdot (u - E_F x) &= \mu^B (z_2^E - z_2), && \text{(7.3e)}\\
w \cdot s &= \mu^B (w^E - w), & & && \text{(7.3f)}\\
E_X x - b_X &= 0. & & && \text{(7.3g)}
\end{aligned}$$

Consider the following primal-dual path following equations given by $F(x, s, y, z_X, z_1, z_2, w\,; \mu^P, \mu^B, y^E, z_1^E, z_2^E, w^E) = 0$, with

$$F(x, s, y, z_X, z_1, z_2, w\,; \mu^P, \mu^B, y^E, z_1^E, z_2^E, w^E) = \begin{pmatrix} g(x) - J(x)^T y - E_X^T z_X - E_F^T z_1 + E_F^T z_2 \\ y - w \\ c(x) - s + \mu^P (y - y^E) \\ z_1 \cdot (E_F x - \ell) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u - E_F x) + \mu^B (z_2 - z_2^E) \\ w \cdot s + \mu^B (w - w^E) \\ E_X x - b_X \end{pmatrix}. \tag{7.4}$$

Any zero $(x, s, y, z_X, z_1, z_2, w)$ of F that satisfies $\ell < E_F x < u$, $z_1 > 0$, $z_2 > 0$, and $w > 0$ approximates a point satisfying the optimality conditions (7.2), with the approximation becoming increasingly accurate as the terms $\mu^P (y - y^E)$, $\mu^B (z_1 - z_1^E)$, $\mu^B (z_2 - z_2^E)$, and $\mu^B (w - w^E)$ approach zero. For any sequence of z_1^E, z_2^E, w^E and y^E such that $z_1^E \to z_1^*$, $z_2^E \to z_2^*$, $w^E \to w^*$, and $y^E \to y^*$, it must hold that solutions $(x, s, y, z_X, z_1, z_2, w)$ of (7.3) must satisfy $z_1 \cdot (E_F x - \ell) \to 0$, $z_2 \cdot (u - E_F x) \to 0$, and $w \cdot s \to 0$, This implies that any solution $(x, s, y, z_X, z_1, z_2, w)$ of (7.3) will approximate a solution of (7.2) independently of the values of μ^P and μ^B (i.e., it is not necessary that $\mu^P \to 0$ and $\mu^B \to 0$).

If $(x, s, y, z_X, z_1, z_2, w)$ is a given approximate zero of F such that $\ell - \mu^B e < E_F x < u + \mu^B e$, $s + \mu^B e > 0$, $z_1 > 0$, $z_2 > 0$, and

$w > 0$, the Newton equations for the change in variables $(\Delta x, \Delta s, \Delta y, \Delta z_X, \Delta z_1, \Delta z_2, \Delta w)$ are given by

$$\begin{pmatrix} H & 0 & -J^T & -E_F^T & E_F^T & 0 & -E_X^T \\ 0 & 0 & I_m & 0 & 0 & -I_m & 0 \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ Z_1 E_F & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 E_F & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu & 0 \\ E_X & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w \\ \Delta z_X \end{pmatrix} = - \begin{pmatrix} g - J^T y - E_X^T z_X - E_F^T z_1 + E_F^T z_2 \\ y - w \\ c - s + \mu^P (y - y^E) \\ z_1 \cdot (E_F x - \ell) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u - E_F x) + \mu^B (z_2 - z_2^E) \\ w \cdot s + \mu^B (w - w^E) \\ E_X x - b_X \end{pmatrix}, \tag{7.5}$$

where $D_Y = \mu^P I$, $W = \operatorname{diag}(w_i)$, $X_1^\mu = \operatorname{diag}(e_j^T x - \ell_j + \mu^B)$, $X_2^\mu = \operatorname{diag}(u_j - e_j^T x + \mu^B)$, and $S^\mu = \operatorname{diag}\big(s_i + \mu^B\big)$.

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. Throughout, we assume that x_X satisfies $E_X x = b_X$, in which case the expansion of Δx satisfies

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

This identity allows us to write the equations (7.5) in the form

$$\begin{pmatrix} H_F & 0 & -J_F^T & -I_F & I_F & 0 \\ 0 & 0 & I_m & 0 & 0 & -I_m \\ J_F & -I_m & D_Y & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - z_1 + z_2 \\ y - w \\ c - s + \mu^P (y - y^E) \\ z_1 \cdot (E_F x - \ell) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u - E_F x) + \mu^B (z_2 - z_2^E) \\ w \cdot s + \mu^B (w - w^E) \end{pmatrix}, \tag{7.6}$$

where H_F and J_F denote the "free" rows and columns of H and the "free" columns of J, i.e., $H_F = E_F H E_F^T$ and $J_F = J E_F^T$. Once these equations are solved, Δx and Δz_X are recovered as $\Delta x = E_F^T \Delta x_F$ and $\Delta z_X = [\, g + H \Delta x - J^T (y + \Delta y) \,]_X - z_X$.

7.3. A shifted primal-dual penalty-barrier function

Problem (7.1) is equivalent to

$$\begin{array}{lll} \underset{x, x_1, x_2, s}{\text{minimize}} & f(x) & \\ \text{subject to} & c(x) - s = 0, & s \ge 0, \\ & E_X x - b_X = 0, & \\ & E_F x - x_1 = \ell, & x_1 \ge 0, \\ & E_F x + x_2 = u, & x_2 \ge 0. \end{array}$$

Consider the shifted primal-dual penalty-barrier problem

$$\begin{array}{llllll}\underset{x,x_1,x_2,s,y,z_1,z_2,w}{\text{minimize}} & M(x,x_1,x_2,s,y,z_1,z_2,w\,;\mu^P,\mu^B,y^E,z_1^E,z_2^E,w^E) & & & \\ \text{subject to} & E_X x = b_X, & E_F x - x_1 = \ell, & x_1 + \mu^B e > 0, & z_1 > 0, \\ & & E_F x + x_2 = u, & x_2 + \mu^B e > 0, & z_2 > 0,\end{array} \tag{7.7}$$

where $M(x,x_1,x_2,s,y,z_1,z_2,w\,;\mu^P,\mu^B,y^E,z_1^E,z_2^E,w^E)$ is the barrier function

$$\begin{aligned} f(x) - \big(c(x)-s\big)^T y^E + \frac{1}{2\mu^P}\|c(x)-s\|^2 + \frac{1}{2\mu^P}\|c(x)-s+\mu^P(y-y^E)\|^2 \\ - \sum_{j\in\mathcal{F}}\Big\{\mu^B[z_1^E]_j\ln\big([x_1]_j+\mu^B\big) + \mu^B[z_1^E]_j\ln\big([z_1]_j([x_1]_j+\mu^B)\big) - [z_1]_j([x_1]_j+\mu^B)\Big\} \\ - \sum_{j\in\mathcal{F}}\Big\{\mu^B[z_2^E]_j\ln\big([x_2]_j+\mu^B\big) + \mu^B[z_2^E]_j\ln\big([z_2]_j([x_2]_j+\mu^B)\big) - [z_2]_j([x_2]_j+\mu^B)\Big\} \\ - \sum_{i=1}^{m}\Big\{\mu^B w_i^E\ln\big(s_i+\mu^B\big) + \mu^B w_i^E\ln\big(w_i(s_i+\mu^B)\big) - w_i(s_i+\mu^B)\Big\}. \end{aligned} \tag{7.8}$$

Let c, g and J denote the quantities $c(x)$, $g(x)$ and $J(x)$. Differentiating $M(x,x_1,x_2,s,y,z_1,z_2,w)$ with respect to x, x_1, x_2, s, y, z_1, z_2, and w gives

$$\nabla M(x,x_1,x_2,s,y,z_1,z_2,w) = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ z_1 - 2\mu^B(X_1^\mu)^{-1}z_1^E \\ z_2 - 2\mu^B(X_2^\mu)^{-1}z_2^E \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y - 2\mu^B(S^\mu)^{-1}w^E + w \\ c - s + \mu^P(y-y^E) \\ x_1 + \mu^B e - \mu^B Z_1^{-1}z_1^E \\ x_2 + \mu^B e - \mu^B Z_2^{-1}z_2^E \\ s + \mu^B e - \mu^B W^{-1}w^E \end{pmatrix}.$$

The gradient may be written in several equivalent forms

$$\nabla M(x,x_1,x_2,s,y,z_1,z_2,w) = \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ z_1 - 2\mu^B (X_1^\mu)^{-1} z_1^E \\ z_2 - 2\mu^B (X_2^\mu)^{-1} z_2^E \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y - 2\mu^B (S^\mu)^{-1} w^E + w \\ c - s + \mu^P (y - y^E) \\ x_1 + \mu^B e - \mu^B Z_1^{-1} z_1^E \\ x_2 + \mu^B e - \mu^B Z_2^{-1} z_2^E \\ s + \mu^B e - \mu^B W^{-1} w^E \end{pmatrix}$$

$$= \begin{pmatrix} g - J^T\big(2(y^E - \frac{1}{\mu^P}(c-s)) - y\big) \\ (X_1^\mu)^{-1}\big(z_1 \cdot x_1 + \mu^B z_1^E + \mu^B (z_1 - z_1^E)\big) \\ (X_2^\mu)^{-1}\big(z_2 \cdot x_2 + \mu^B z_2^E + \mu^B (z_2 - z_2^E)\big) \\ 2\big(y^E - \frac{1}{\mu^P}(c-s)\big) - y - 2\mu^B (S + \mu^B I)^{-1} w^E + w \\ c - s + \mu^P (y - y^E) \\ Z_1^{-1}\big(z_1 \cdot x_1 + \mu^B (z_1 - z_1^E)\big) \\ Z_2^{-1}\big(z_2 \cdot x_2 + \mu^B (z_2 - z_2^E)\big) \\ W^{-1}\big(w \cdot s + \mu^B (w - w^E)\big) \end{pmatrix} = \begin{pmatrix} g - J^T\big(\pi^Y + (\pi^Y - y)\big) \\ -(2\pi_1^Z - z_1) \\ -(2\pi_2^Z - z_2) \\ (2\pi^Y - y) - (2\pi^W - w) \\ -D_Y(\pi^Y - y) \\ -D_1^Z(\pi_1^Z - z_1) \\ -D_2^Z(\pi_2^Z - z_2) \\ -D_W(\pi^W - w) \end{pmatrix},$$

where now, $X_1^\mu = \mathrm{diag}([x_1]_j)$ and $X_2^\mu = \mathrm{diag}([x_2]_j)$, with

$$D_Y = \mu^P I, \qquad \pi^Y = y^E - \frac{1}{\mu^P}(c-s), \tag{7.9a}$$

$$D_1^Z = X_1^\mu Z_1^{-1}, \qquad \pi_1^Z = \mu^B (X_1^\mu)^{-1} z_1^E, \tag{7.9b}$$

$$D_2^Z = X_2^\mu Z_2^{-1}, \qquad \pi_2^Z = \mu^B (X_2^\mu)^{-1} z_2^E, \tag{7.9c}$$

$$D_W = S^\mu W^{-1}, \qquad \pi^W = \mu^B (S^\mu)^{-1} w^E. \tag{7.9d}$$

Similarly, the Hessian of $M(x, x_1, x_2, s, y, z_1, z_2, w)$ is given by

$$\begin{pmatrix} H_1 & 0 & 0 & -\frac{2}{\mu^P}J^T & J^T & 0 & 0 & 0 \\ 0 & 2\mu^B(X_1^\mu)^{-2}Z_1^E & 0 & 0 & 0 & I_F & 0 & 0 \\ 0 & 0 & 2\mu^B(X_2^\mu)^{-2}Z_2^E & 0 & 0 & 0 & I_F & 0 \\ -\frac{2}{\mu^P}J & 0 & 0 & 2\big(\frac{1}{\mu^P}I + \mu^B(S^\mu)^{-2}W^E\big) & -I_m & 0 & 0 & I_m \\ J & 0 & 0 & -I_m & \mu^P I_m & 0 & 0 & 0 \\ 0 & I_F & 0 & 0 & 0 & \mu^B Z_1^{-2}Z_1^E & 0 & 0 \\ 0 & 0 & I_F & 0 & 0 & 0 & \mu^B Z_2^{-2}Z_2^E & 0 \\ 0 & 0 & 0 & I_m & 0 & 0 & 0 & \mu^B W^{-2}W^E \end{pmatrix},$$

where $H_1 = H(x, 2\pi^Y - y) + \frac{2}{\mu^P}J^TJ$. Substituting $\mu^B Z_1^E = X_1^\mu \Pi_1^Z$, $\mu^B Z_2^E = (X_2 + \mu^B I)\Pi_2^Z$, and $\mu^B W^E = S^\mu \Pi^W$ from (7.9) gives the Hessian

$$\begin{pmatrix} H_1 & 0 & 0 & -\frac{2}{\mu^P}J^T & J^T & 0 & 0 & 0 \\ 0 & 2(X_1^\mu)^{-1}\Pi_1^Z & 0 & 0 & 0 & I_F & 0 & 0 \\ 0 & 0 & 2(X_2^\mu)^{-1}\Pi_2^Z & 0 & 0 & 0 & I_F & 0 \\ -\frac{2}{\mu^P}J & 0 & 0 & 2\big(\frac{1}{\mu^P}I + (S^\mu)^{-1}\Pi^W\big) & -I_m & 0 & 0 & I_m \\ J & 0 & 0 & -I_m & \mu^P I_m & 0 & 0 & 0 \\ 0 & I_F & 0 & 0 & 0 & X_1^\mu Z_1^{-2}\Pi_1^Z & 0 & 0 \\ 0 & 0 & I_F & 0 & 0 & 0 & X_2^\mu Z_2^{-2}\Pi_2^Z & 0 \\ 0 & 0 & 0 & I_m & 0 & 0 & 0 & S^\mu W^{-2}\Pi^W \end{pmatrix}.$$

7.4. Derivation of the shifted primal-dual penalty-barrier direction

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

$$\underset{p\in\mathcal{I}}{\text{minimize}} \;\; M(p) \quad \text{subject to} \quad Cp = b_C, \tag{7.10}$$

where

$$\mathcal{I} = \{p : p = (x, x_1, x_2, s, y, z_1, z_2, w),\;\; \text{with } x_1 + \mu^B e > 0,\; x_2 + \mu^B e > 0,\; s + \mu^B e > 0,\; z_1 > 0,\; z_2 > 0,\; w > 0\},$$

and

$$C = \begin{pmatrix} E_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_F & -I_F & 0 & 0 & 0 & 0 & 0 & 0 \\ E_F & 0 & I_F & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad b_C = \begin{pmatrix} b_X \\ \ell \\ u \end{pmatrix}.$$

Let $p \in \mathcal{I}$ be given. Assume that x is feasible for the equality constraints $E_X x = b_X$, but not necessarily for the linear inequality constraints, i.e., it may not hold that $E_F x - x_1 = \ell$ and $E_F x + x_2 = u$. The Newton direction Δp is given by the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T \nabla^2 M(p) \Delta p \quad \text{subject to} \;\; C\Delta p = b_C - Cp. \tag{7.11}$$

However, instead of solving (7.11), we define a linearly constrained modified Newton method by approximating the Hessian $\nabla^2 M(x, x_1, x_2, s, y, z_1, z_2, w)$ by a matrix $B(x, x_1, x_2, s, y, z_1, z_2, w)$. Consider the matrix defined by replacing π^Y by y, π_1^z by z_1, π_2^z by z_2, and π^W by w everywhere in the matrix $\nabla^2 M(x, x_1, x_2, s, y, z_1, z_2, w)$. This gives an approximate Hessian $B(x, x_1, x_2, s, y, z_1, z_2, w)$ of the form

$$\begin{pmatrix} \widehat{H}_1 & 0 & 0 & -\frac{2}{\mu^P} J^T & J^T & 0 & 0 & 0 \\ 0 & 2(X_1^\mu)^{-1} Z_1 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 2(X_2^\mu)^{-1} Z_2 & 0 & 0 & 0 & I & 0 \\ -\frac{2}{\mu^P} J & 0 & 0 & 2\big(\frac{1}{\mu^P} I + (S^\mu)^{-1} W\big) & -I & 0 & 0 & I \\ J & 0 & 0 & -I & \mu^P I & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & (X_1^\mu) Z_1^{-1} & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & X_2^\mu Z_2^{-1} & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & S^\mu W^{-1} \end{pmatrix},$$

where $\widehat{H}_1 = H(x, y) + \frac{2}{\mu^P} J^T J$. The definitions of D_Y, D_1^z, and D_2^z may be used to write $B(x, x_1, x_2, s, y, z_1, z_2, w)$ in the form

$$\begin{pmatrix} H + 2J^T D_Y^{-1} J & 0 & 0 & -2J^T D_Y^{-1} & J^T & 0 & 0 & 0 \\ 0 & 2(D_1^z)^{-1} & 0 & 0 & 0 & I_F & 0 & 0 \\ 0 & 0 & 2(D_2^z)^{-1} & 0 & 0 & 0 & I_F & 0 \\ -2D_Y^{-1} J & 0 & 0 & 2\big(D_Y^{-1} + D_W^{-1}\big) & -I_m & 0 & 0 & I_m \\ J & 0 & 0 & -I & D_Y & 0 & 0 & 0 \\ 0 & I_F & 0 & 0 & 0 & D_1^z & 0 & 0 \\ 0 & 0 & I_F & 0 & 0 & 0 & D_2^z & 0 \\ 0 & 0 & 0 & I_m & 0 & 0 & 0 & D_W \end{pmatrix},$$

where $H = H(x, y)$. Given $B(p) = B(x, x_1, x_2, s, y, z_1, z_2, w)$, a modified Newton direction is given by the solution of the QP subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T B(p) \Delta p \quad \text{subject to} \;\; C\Delta p = b_C - Cp. \tag{7.12}$$

Let N denote a matrix whose columns form a basis for null(C), i.e., the columns of N are linearly independent and $CN = 0$. The vector

$$\Delta p_0 = \begin{pmatrix} 0 \\ -(\ell - E_F x + x_1) \\ (u - E_F x - x_2) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \triangleq \begin{pmatrix} 0 \\ -r_L \\ r_U \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{7.13}$$

satisfies $C\Delta p_0 = b_C - Cp$, and every feasible Δp may be written in the form

$$\Delta p = \Delta p_0 + Nd.$$

This implies that d satisfies the reduced equations

$$N^T B(p) N d = -N^T\big(\nabla M(p) + B(p)\Delta p_0\big).$$

Consider the null-space basis defined from the columns of

$$N = \begin{pmatrix} E_F^T & 0 & 0 & 0 & 0 & 0 \\ I_F & 0 & 0 & 0 & 0 & 0 \\ -I_F & 0 & 0 & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 & 0 & 0 \\ 0 & 0 & I_m & 0 & 0 & 0 \\ 0 & 0 & 0 & I_F & 0 & 0 \\ 0 & 0 & 0 & 0 & I_F & 0 \\ 0 & 0 & 0 & 0 & 0 & I_m \end{pmatrix}. \tag{7.14}$$

The definition of N of (7.14) gives the reduced Hessian $N^T B(p) N$ such that

$$\begin{pmatrix} H_F + 2J_F^T D_Y^{-1} J_F + 2\big((D_1^z)^{-1} + (D_2^z)^{-1}\big) & -2J_F^T D_Y^{-1} & J_F^T & I_F & -I_F & 0 \\ -2D_Y^{-1} J_F & 2\big(D_Y^{-1} + D_W^{-1}\big) & -I_m & 0 & 0 & I_m \\ J_F & -I_m & D_Y & 0 & 0 & 0 \\ I_F & 0 & 0 & D_1^z & 0 & 0 \\ -I_F & 0 & 0 & 0 & D_2^z & 0 \\ 0 & I_m & 0 & 0 & 0 & D_W \end{pmatrix}.$$

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

$$N^T \begin{pmatrix} g_F - J_F^T\big(\pi^Y + (\pi^Y - y)\big) \\ -(2\pi_1^z - z_1) \\ -(2\pi_2^z - z_2) \\ (2\pi^Y - y) - (2\pi^W - w) \\ -D_Y(\pi^Y - y) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \\ -D_W(\pi^W - w) \end{pmatrix} = \begin{pmatrix} g_F - J_F^T\big(\pi^Y + (\pi^Y - y)\big) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) \\ (2\pi^Y - y) - (2\pi^W - w) \\ -D_Y(\pi^Y - y) \\ -D_1^z(\pi_1^z - z_1) \\ -D_2^z(\pi_2^z - z_2) \\ -D_W(\pi^W - w) \end{pmatrix}.$$

Moreover

$$B\Delta p_0 = \begin{pmatrix} 0 \\ -2(D_1^z)^{-1} r_L \\ 2(D_2^z)^{-1} r_U \\ 0 \\ 0 \\ -r_L \\ r_U \\ 0 \end{pmatrix},$$

where $r_L = \ell - E_F x + x_1$ and $r_U = u - E_F x - x_2$. This implies that $N^T B(p)\Delta p_0$ is given by

$$\begin{pmatrix} -2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\ 0 \\ 0 \\ -r_L \\ r_U \\ 0 \end{pmatrix}.$$

This gives the reduced gradient $N^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ such that

$$N^T\big(\nabla M(p) + B(p)\Delta p_0\big) = \begin{pmatrix} g_F - J_F^T\big(\pi^Y + (\pi^Y - y)\big) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) - 2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\ (2\pi^Y - y) - (2\pi^W - w) \\ -D_Y(\pi^Y - y) \\ -D_1^z(\pi_1^z - z_1) - r_L \\ -D_2^z(\pi_2^z - z_2) + r_U \\ -D_W(\pi^W - w) \end{pmatrix}.$$

The reduced modified Newton equations $N^TB(p)Nd = -N^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ are then

$$
\begin{pmatrix}
H_F + 2J_F^T D_Y^{-1} J_F + 2\big((D_1^z)^{-1} + (D_2^z)^{-1}\big) & -2J_F^T D_Y^{-1} & J_F^T & I_F & -I_F & 0 \\
-2D_Y^{-1} J & 2\big(D_Y^{-1} + D_W^{-1}\big) & -I_m & 0 & 0 & I_m \\
J_F & -I_m & D_Y & 0 & 0 & 0 \\
I_F & 0 & 0 & D_1^z & 0 & 0 \\
-I_F & 0 & 0 & 0 & D_2^z & 0 \\
0 & I_m & 0 & 0 & 0 & D_W
\end{pmatrix}
\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{pmatrix}
$$

$$
= -\begin{pmatrix}
g_F - J_F^T\big(\pi^Y + (\pi^Y - y)\big) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) - 2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\
(2\pi^Y - y) - (2\pi^W - w) \\
-D_Y(\pi^Y - y) \\
-D_1^z(\pi_1^z - z_1) - r_L \\
-D_2^z(\pi_2^z - z_2) + r_U \\
-D_W(\pi^W - w)
\end{pmatrix}.
$$

Given any nonsingular matrix R, the direction d satisfies

$$
RN^TB(p)Nd = -RN^T\big(\nabla M(p) + B(p)\Delta p_0\big).
$$

In particular, if R is the block upper-triangular matrix R such that

$$
R = \begin{pmatrix}
I_F & 0 & -2J_F^T D_Y^{-1} & -2(D_1^z)^{-1} & 2(D_2^z)^{-1} & 0 \\
 & I_m & 2D_Y^{-1} & 0 & 0 & -2D_W^{-1} \\
 & & I_m & 0 & 0 & 0 \\
 & & & Z_1 & 0 & 0 \\
 & & & & Z_2 & 0 \\
 & & & & & W
\end{pmatrix},
$$

then R is nonsingular because Z_1, Z_2 and W are positive definite, and

$$
RN^TB(p)N = \begin{pmatrix}
H_F & 0 & -J_F^T & -I_F & I_F & 0 \\
0 & 0 & I_m & 0 & 0 & -I_m \\
J_F & -I_m & D_Y & 0 & 0 & 0 \\
Z_1 & 0 & 0 & Z_1 D_1^z & 0 & 0 \\
-Z_2 & 0 & 0 & 0 & Z_2 D_2^z & 0 \\
0 & W & 0 & 0 & 0 & W D_W
\end{pmatrix}
= \begin{pmatrix}
H_F & 0 & -J_F^T & -I_F & I_F & 0 \\
0 & 0 & I_m & 0 & 0 & -I_m \\
J_F & -I_m & D_Y & 0 & 0 & 0 \\
Z_1 & 0 & 0 & X_1^\mu & 0 & 0 \\
-Z_2 & 0 & 0 & 0 & X_2^\mu & 0 \\
0 & W & 0 & 0 & 0 & S^\mu
\end{pmatrix}.
$$

Also, $RN^T\big(\nabla M(p) + B(p)\Delta p_0\big)$ is given by

$$\begin{pmatrix} I_F & 0 & -2J_F^T D_Y^{-1} & -2(D_1^z)^{-1} & 2(D_2^z)^{-1} & 0 \\ & I_m & 2D_Y^{-1} & 0 & 0 & -2D_W^{-1} \\ & & I_m & 0 & 0 & 0 \\ & & & Z_1 & 0 & 0 \\ & & & & Z_2 & 0 \\ & & & & & W \end{pmatrix} \begin{pmatrix} g_F - J_F^T\big(\pi^Y + (\pi^Y - y)\big) - (2\pi_1^z - z_1) + (2\pi_2^z - z_2) - 2\big((D_1^z)^{-1} r_L + (D_2^z)^{-1} r_U\big) \\ (2\pi^Y - y) - (2\pi^W - w) \\ -D_Y(\pi^Y - y) \\ -D_1^z(\pi_1^z - z_1) - r_L \\ -D_2^z(\pi_2^z - z_2) + r_U \\ -D_W(\pi^W - w) \end{pmatrix}$$

$$= \begin{pmatrix} g_F - J_F^T y - z_1 + z_2 \\ y - w \\ c - s + \mu^P(y - y^E) \\ z_1 \cdot (E_F x - \ell) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u - E_F x) + \mu^B(z_2 - z_2^E) \\ w \cdot s + \mu^B(w - w^E) \end{pmatrix}.$$

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

$$\begin{pmatrix} H_F & 0 & -J_F^T & -I_F & I_F & 0 \\ 0 & 0 & I_m & 0 & 0 & -I_m \\ J_F & -I_m & D_Y & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - z_1 + z_2 \\ y - w \\ c - s + \mu^P(y - y^E) \\ z_1 \cdot (E_F x - \ell) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u - E_F x) + \mu^B(z_2 - z_2^E) \\ w \cdot s + \mu^B(w - w^E) \end{pmatrix}. \tag{7.15}$$

Then, (7.13) implies that

$$\begin{pmatrix} \Delta x \\ \Delta x_1 \\ \Delta x_2 \\ \Delta s \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w \end{pmatrix} = \Delta p = \Delta p_0 + Nd = \begin{pmatrix} E_F^T d_1 \\ (d_1 - r_L) \\ -(d_1 - r_U) \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{pmatrix}.$$

These identities allow us to write equations (7.15) in the form

$$\begin{pmatrix} H_F & 0 & -J_F^T & -I_F & I_F & 0 \\ 0 & 0 & I_m & 0 & 0 & -I_m \\ J_F & -I_m & D_Y & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 \\ 0 & W & 0 & 0 & 0 & S^\mu \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - z_1 + z_2 \\ y - w \\ c - s + \mu^P (y - y^E) \\ z_1 \cdot (E_F x - \ell) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u - E_F x) + \mu^B (z_2 - z_2^E) \\ w \cdot s + \mu^B (w - w^E) \end{pmatrix}, \tag{7.16}$$

with $\Delta x = E_F^T \Delta x_F$, $\Delta x_1 = \Delta x - (\ell - E_F x + x_1)$ and $\Delta x_2 = -\Delta x + (u - E_F x - x_2)$.

As in the bounded slack case, it is necessary to choose feasible x_1 and x_2, which gives $E_F x - x_1 = \ell$ and $E_F x + x_2 = u$, and it follows that $\Delta x_1 = \Delta x$ and $\Delta x_2 = -\Delta x$. (This assumption is made for the remainder of this section.) Under this feasibility assumption, if X_1 and X_2 are written in terms of x, i.e., $X_1 = \operatorname{diag}\big(e_j^T x - \ell_j\big)$ and $X_2 = \operatorname{diag}\big(u_j - e_j^T x\big)$, respectively, then equations (7.16) are the Newton equations for a solution of the perturbed optimality conditions (7.3). The variables x_1 and x_2 may be computed implicitly for the line search, in which case the appropriate merit function is

$$\begin{aligned} f - (c-s)^T y^E + \frac{1}{2\mu^P}\|c - s\|^2 + \frac{1}{2\mu^P}\|c - s + \mu^P (y - y^E)\|^2 & \\ - \sum_{j\in\mathcal{F}} \Big\{ \mu^B [z_1^E]_j \ln\big(x_j - \ell_j + \mu^B\big) + \mu^B [z_1^E]_j \ln\big([z_1]_j (x_j - \ell_j + \mu^B)\big) - [z_1]_j (x_j - \ell_j + \mu^B) \Big\} & \\ - \sum_{j\in\mathcal{F}} \Big\{ \mu^B [z_2^E]_j \ln\big(u_j - x_j + \mu^B\big) + \mu^B [z_2^E]_j \ln\big([z_2]_j (u_j - x_j + \mu^B)\big) - [z_2]_j (u_j - x_j + \mu^B) \Big\} & \\ - \sum_{i=1}^{m} \Big\{ \mu^B w_i^E \ln\big(s_i + \mu^B\big) + \mu^B w_i^E \ln\big(w_i (s_i + \mu^B)\big) - w_i (s_i + \mu^B) \Big\}. & \end{aligned}$$

7.5. Computation of the shifted primal-dual penalty-barrier direction

Next we consider the solution of the path-following modified Newton equations (7.16), which may be written in the form

$$\begin{pmatrix} H_F & 0 & -J_F^T & -I_F & I_F & 0 \\ 0 & 0 & I_m & 0 & 0 & -I_m \\ J_F & -I_m & D_Y & 0 & 0 & 0 \\ I_F & 0 & 0 & D_1^Z & 0 & 0 \\ -I_F & 0 & 0 & 0 & D_2^Z & 0 \\ 0 & I_m & 0 & 0 & 0 & D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - z_1 + z_2 \\ y - w \\ D_Y (y - \pi^Y) \\ D_1^Z (z_1 - \pi_1^Z) \\ D_2^Z (z_2 - \pi_2^Z) \\ D_W (w - \pi^W) \end{pmatrix}.$$

Consider the following reordered equations and variables:

$$\begin{pmatrix} D_1^z & 0 & 0 & 0 & I_F & 0 \\ 0 & D_2^z & 0 & 0 & -I_F & 0 \\ 0 & 0 & D_W & I_m & 0 & 0 \\ 0 & 0 & -I_m & 0 & 0 & I_m \\ 0 & 0 & 0 & -I_m & J & D_Y \\ -I_F & I_F & 0 & 0 & H_F & -J_F^T \end{pmatrix} \begin{pmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta w \\ \Delta s \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_1^z(z_1 - \pi_1^z) \\ D_2^z(z_2 - \pi_2^z) \\ D_W(w - \pi^W) \\ y - w \\ D_Y(y - \pi^Y) \\ g_F - J_F^T y - z_1 + z_2 \end{pmatrix}. \tag{7.17}$$

Applying the nonsingular matrix

$$\begin{pmatrix} I_F & & & & & \\ 0 & I_F & & & & \\ 0 & & I_m & & & \\ & & D_W^{-1} & I_m & & \\ & & I_m & D_W & I_m & \\ (D_1^z)^{-1} & -(D_2^z)^{-1} & & & & I_m \end{pmatrix}$$

to both sides of (7.17) gives the block upper-trapezoidal system

$$\begin{pmatrix} D_1^z & 0 & 0 & 0 & I_F & 0 \\ 0 & D_2^z & 0 & 0 & -I_F & 0 \\ 0 & 0 & D_W & I_m & 0 & 0 \\ 0 & 0 & 0 & D_W^{-1} & 0 & I_m \\ 0 & 0 & 0 & 0 & J & D_Y + D_W \\ 0 & 0 & 0 & 0 & H_F + D_Z^{-1} & -J_F^T \end{pmatrix} \begin{pmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta w \\ \Delta s \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_1^z(z_1 - \pi_1^z) \\ D_2^z(z_2 - \pi_2^z) \\ D_W(w - \pi^W) \\ y - \pi^W \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \\ g_F - J_F^T y - \pi^Z \end{pmatrix}.$$

with $\pi^Z = \pi_1^z - \pi_2^z$, and $D_Z^{-1} = (D_1^z)^{-1} + (D_2^z)^{-1}$. It follows that the solution of the Newton path-following equations (7.5) is given by

$$\begin{aligned} \Delta x &= E_F^T \Delta x_F, \\ \Delta z_X &= [\, g + H\Delta x - J^T(y + \Delta y)\,]_X - z_X \\ \Delta w &= y + \Delta y - w, \\ \Delta s &= -W^{-1}\big((w + \Delta w) \cdot s + \mu^B(w + \Delta w - w^E)\big), \\ \Delta z_1 &= -(X_1^\mu)^{-1}\big(z_1 \cdot (E_F(x + \Delta x) - \ell + \mu^B e) - \mu^B z_1^E\big), \\ \Delta z_2 &= -(X_2^\mu)^{-1}\big(z_2 \cdot (u - E_F(x + \Delta x) + \mu^B e) - \mu^B z_2^E\big), \end{aligned}$$

where Δx_F and Δy satisfy the equations

$$\begin{pmatrix} H_F + D_Z^{-1} & -J_F^T \\ J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g_F - J_F^T y - \pi^Z \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix}.$$

7.6. Summary: bounded variables

Define the quantities

$$\begin{aligned} D_Y &= \mu^P I, & \pi^Y &= y^E - \frac{1}{\mu^P}\big(c(x) - s\big), \\ D_1^Z &= X_1^\mu Z_1^{-1}, & \pi_1^Z &= \mu^B (X_1^\mu)^{-1} z_1^E, \\ D_2^Z &= X_2^\mu Z_2^{-1}, & \pi_2^Z &= \mu^B (X_2^\mu)^{-1} z_2^E, \\ D_Z &= \big((D_1^Z)^{-1} + (D_2^Z)^{-1}\big)^{-1}, & \pi^Z &= \pi_1^Z - \pi_2^Z, \\ D_W &= S^\mu W^{-1}, & \pi^W &= \mu^B (S^\mu)^{-1} w^E, \end{aligned}$$

then Δs, Δw, Δx_1, Δx_2, Δz_1 and Δz_1 are given by

$$\begin{aligned} \Delta x &= E_F^T \Delta x_F, & & \\ \widehat{y} &= y + \Delta y, & \Delta w &= \widehat{y} - w, \\ \widehat{w} &= w + \Delta w, & \Delta s &= -W^{-1}\big(\widehat{w} \cdot s + \mu^B(\widehat{w} - w^E)\big), \\ \widehat{x} &= x + \Delta x, & \Delta z_1 &= -(X_1^\mu)^{-1}\big(z_1 \cdot (\widehat{x} - \ell + \mu^B e) - \mu^B z_1^E\big), \\ & & \Delta z_2 &= -(X_2^\mu)^{-1}\big(z_2 \cdot (u - \widehat{x} + \mu^B e) - \mu^B z_2^E\big), \end{aligned}$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H_F + D_Z^{-1} & -J^T \\ J_F & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g_F - J_F^T y - \pi^Z \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix}.$$

The line-search merit function is

$$
\begin{aligned}
f - (c-s)^T y^{E} + \frac{1}{2\mu^{P}}\|c-s\|^2 + \frac{1}{2\mu^{P}}\|c - s + \mu^{P}(y - y^{E})\|^2 \\
- \sum_{j\in\mathcal{F}} \Big\{ \mu^{B}[z_1^{E}]_j \ln\big(x_j - \ell_j + \mu^{B}\big) + \mu^{B}[z_1^{E}]_j \ln\big([z_1]_j(x_j - \ell_j + \mu^{B})\big) - [z_1]_j(x_j - \ell_j + \mu^{B}) \Big\} \\
- \sum_{j\in\mathcal{F}} \Big\{ \mu^{B}[z_2^{E}]_j \ln\big(u_j - x_j + \mu^{B}\big) + \mu^{B}[z_2^{E}]_j \ln\big([z_2]_j(u_j - x_j + \mu^{B})\big) - [z_2]_j(u_j - x_j + \mu^{B}) \Big\} \\
- \sum_{i=1}^{m} \Big\{ \mu^{B} w_i^{E} \ln\big(s_i + \mu^{B}\big) + \mu^{B} w_i^{E} \ln\big(w_i(s_i + \mu^{B})\big) - w_i(s_i + \mu^{B}) \Big\}. \quad (7.18)
\end{aligned}
$$

8. Upper and Lower Bounds on all Variables and Slacks

Next we consider the case with upper and lower bounds on all the variables and slacks.

8.1. Problem statement and optimality conditions

The definition of the problem is

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \;\; f(x) \quad \text{subject to} \;\; c(x)-s=0, \quad \ell^X \le x \le u^X, \quad \ell^S \le s \le u^S, \tag{8.1}$$

where $c:\mathbb{R}^n \mapsto \mathbb{R}^m$ and $f:\mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable. The first-order KKT conditions for this problem are

$$\begin{aligned}
g(x^*) - J(x^*)^T y^* - z_1^* + z_2^* &= 0, & z_1^* &\ge 0, & z_2^* &\ge 0, & &(8.2a)\\
y^* - w_1^* + w_2^* &= 0, & w_1^* &\ge 0, & w_2^* &\ge 0, & &(8.2b)\\
c(x^*) - s^* &= 0, & & & & & &(8.2c)\\
x^* - \ell^X &\ge 0, & u^X - x^* &\ge 0, & & & &(8.2d)\\
s^* - \ell^S &\ge 0, & u^S - s^* &\ge 0, & & & &(8.2e)\\
z_1^* \cdot (x^* - \ell^X) &= 0, & z_2^* \cdot (u^X - x^*) &= 0, & & & &(8.2f)\\
w_1^* \cdot (s^* - \ell^S) &= 0, & w_2^* \cdot (u^S - s^*) &= 0, & & & &(8.2g)
\end{aligned}$$

where y^* are the multipliers for the equality constraints $c(x)-s=0$, and z_1^*, z_2^*, w_1^*, and w_2^* may be interpreted as the Lagrange multipliers for the inequality constraints $x-\ell^X \ge 0$, $u^X - x \ge 0$, $s - \ell^S \ge 0$ and $u^S - s \ge 0$, respectively.

8.2. The path-following equations

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

$$\begin{aligned}
g(x) - J(x)^T y - z_1 + z_2 &= 0, & z_1 &\ge 0, & z_2 &\ge 0, & &(8.3a)\\
y - w_1 + w_2 &= 0, & w_1 &\ge 0, & w_2 &\ge 0, & &(8.3b)\\
c(x) - s &= \mu^P(y^E - y), & & & & & &(8.3c)\\
x - \ell^X &\ge 0, & u^X - x &\ge 0, & & & &(8.3d)\\
s - \ell^S &\ge 0, & u^S - s &\ge 0, & & & &(8.3e)\\
z_1 \cdot (x - \ell^X) &= \mu^B(z_1^E - z_1), & z_2 \cdot (u^X - x) &= \mu^B(z_2^E - z_2), & & & &(8.3f)\\
w_1 \cdot (s - \ell^S) &= \mu^B(w_1^E - w_1), & w_2 \cdot (u^S - s) &= \mu^B(w_2^E - w_2), & & & &(8.3g)
\end{aligned}$$

Consider the primal-dual path-following equations $F(x,s,y,z_1,z_2,w_1,w_2\,;\mu^P,\mu^B,y^E,z_1^E,z_2^E,w_1^E,w_2^E)=0$, with

$$F(x,s,y,z_1,z_2,w_1,w_2\,;\mu^P,\mu^B,y^E,z_1^E,z_2^E,w_1^E,w_2^E)=\begin{pmatrix} g(x)-J(x)^Ty-z_1+z_2\\ y-w_1+w_2\\ c(x)-s+\mu^P(y-y^E)\\ z_1\cdot(x-\ell^X)+\mu^B(z_1-z_1^E)\\ z_2\cdot(x-\ell^X)+\mu^B(z_2-z_2^E)\\ w_1\cdot(s-\ell^S)+\mu^B(w_1-w_1^E)\\ w_2\cdot(s-\ell^S)+\mu^B(w_2-w_2^E)\end{pmatrix}. \tag{8.4}$$

Any zero (x,s,y,z_1,z_2,w_1,w_2) of F that satisfies $\ell^X<x<u^X$, $\ell^S<s<u^S$, $z_1>0$, $z_2>0$, $w_1>0$, and $w_2>0$ approximates a point satisfying the optimality conditions (8.2), with the approximation becoming increasingly accurate as the terms $\mu^P(y-y^E)$, $\mu^B(z_1-z_1^E)$, $\mu^B(z_2-z_2^E)$, $\mu^B(w_1-w_1^E)$ 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 and y^E such that $z_1^E\to z_1^*$, $z_2^E\to z_2^*$, $w_1^E\to w_1^*$, $w_2^E\to w_2^*$, and $y^E\to y^*$, and it must hold that solutions (x,s,y,z_1,z_2,w_1,w_2) of (8.3) 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,z_1,z_2,w_1,w_2) of (8.3) will approximate a solution of (8.2) independently of the values of μ^P and μ^B (i.e., it is not necessary that $\mu^P\to 0$ and $\mu^B\to 0$).

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

$$\begin{pmatrix} H & 0 & -J^T & -I & I & 0 & 0\\ 0 & 0 & I & 0 & 0 & -I & I\\ J & -I & D_Y & 0 & 0 & 0 & 0\\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 & 0\\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 & 0\\ 0 & W_1 & 0 & 0 & 0 & S_1^\mu & 0\\ 0 & -W_2 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix}\begin{pmatrix}\Delta x\\ \Delta s\\ \Delta y\\ \Delta z_1\\ \Delta z_2\\ \Delta w_1\\ \Delta w_2\end{pmatrix}=-\begin{pmatrix} g-J^Ty-z_1+z_2\\ y-w_1+w_2\\ c-s+\mu^P(y-y^E)\\ z_1\cdot(x-\ell^X)+\mu^B(z_1-z_1^E)\\ z_2\cdot(u^X-x)+\mu^B(z_2-z_2^E)\\ w_1\cdot(s-\ell^S)+\mu^B(w_1-w_1^E)\\ w_2\cdot(u^S-s)+\mu^B(w_2-w_2^E)\end{pmatrix}, \tag{8.5}$$

where $D_Y=\mu^P I$, $X_1^\mu=\operatorname{diag}(x_j-\ell_j^X+\mu^B)$, $X_2^\mu=\operatorname{diag}(u_j^X-x_j+\mu^B)$, $Z_1=\operatorname{diag}([z_1]_j)$, $Z_2=\operatorname{diag}([z_2]_j)$, $W_1=\operatorname{diag}([w_1]_i)$, $W_2=\operatorname{diag}([w_2]_i)$, $S_1^\mu=\operatorname{diag}(s_i-\ell_i^S+\mu^B)$, and $S_2^\mu=\operatorname{diag}(u_i^S-s_i+\mu^B)$.

8.3. A shifted primal-dual penalty-barrier function

Problem (8.1) is equivalent to

$$\begin{array}{ll} \underset{x,x_1,x_2,s,s_1,s_2}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) - s = 0, \\ & x - x_1 = \ell^{X}, \quad s - s_1 = \ell^{S}, \qquad x_1 \geq 0, \quad s_1 \geq 0, \\ & x + x_2 = u^{X}, \quad s + s_2 = u^{S}, \qquad x_2 \geq 0, \quad s_2 \geq 0. \end{array}$$

Consider the shifted primal-dual penalty-barrier problem

$$\begin{array}{lllllll} \underset{x,x_1,x_2,s,s_1,s_2,y,z_1,z_2,w_1,w_2}{\text{minimize}} & M(x,x_1,x_2,s,s_1,s_2,y,w_1,w_2\,; \mu^{P},\mu^{B},y^{E},w_1^{E},w_2^{E}) \\ \text{subject to} & x - x_1 = \ell^{X}, & s - s_1 = \ell^{S}, & x_1 + \mu^{B} e > 0, & z_1 > 0, & s_1 + \mu^{B} e > 0, & w_1 > 0, \\ & x + x_2 = u^{X}, & s + s_2 = u^{S}, & x_2 + \mu^{B} e > 0, & z_2 > 0, & s_2 + \mu^{B} e > 0, & w_2 > 0, \end{array} \tag{8.6}$$

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

$$\begin{aligned} f(x) &- \big(c(x)-s\big)^T y^{E} + \frac{1}{2\mu^{P}}\|c(x)-s\|^2 + \frac{1}{2\mu^{P}}\|c(x)-s+\mu^{P}(y-y^{E})\|^2 \\ &- \sum_{j=1}^{n}\Big\{\mu^{B}[z_1^{E}]_j \ln\big([\,x_1+\mu^{B}e\,]_j\big) + \mu^{B}[z_1^{E}]_j \ln\big([\,z_1\cdot(x_1+\mu^{B}e)\,]_j\big) - [\,z_1\cdot(x_1+\mu^{B}e)\,]_j\Big\} \\ &- \sum_{j=1}^{n}\Big\{\mu^{B}[z_2^{E}]_j \ln\big([\,x_2+\mu^{B}e\,]_j\big) + \mu^{B}[z_2^{E}]_j \ln\big([\,z_2\cdot(x_2+\mu^{B}e)\,]_j\big) - [\,z_2\cdot(x_2+\mu^{B}e)\,]_j\Big\} \\ &- \sum_{i=1}^{m}\Big\{\mu^{B}[w_1^{E}]_i \ln\big([\,s_1+\mu^{B}e\,]_i\big) + \mu^{B}[w_1^{E}]_i \ln\big([\,w_1\cdot(s_1+\mu^{B}e)\,]_i\big) - [\,w_1\cdot(s_1+\mu^{B}e)\,]_i\Big\} \\ &- \sum_{i=1}^{m}\Big\{\mu^{B}[w_2^{E}]_i \ln\big([\,s_2+\mu^{B}e\,]_i\big) + \mu^{B}[w_2^{E}]_i \ln\big([\,w_2\cdot(s_2+\mu^{B}e)\,]_i\big) - [\,w_2\cdot(s_2+\mu^{B}e)\,]_i\Big\}. \end{aligned} \tag{8.7}$$

Let c, g and J denote the quantities $c(x)$, $g(x)$ and $J(x)$, The gradient of the merit function as a function of x, x_1, x_2, s, s_1, s_2 y, z_1, z_2, w_1, and w_2, is

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

Similarly, the Hessian of $M(x, x_1, x_2, s, s_1, s_2, y, w_1, w_2)$ is given by

$$\begin{pmatrix} H_1 & 0 & 0 & -\frac{2}{\mu^P}J^T & 0 & 0 & J^T & 0 & 0 & 0 & 0 \\ 0 & 2\mu^B(X_1^\mu)^{-2}Z_1^E & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 \\ 0 & 0 & 2\mu^B(X_2^\mu)^{-2}Z_2^E & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 \\ -\frac{2}{\mu^P}J & 0 & 0 & \frac{2}{\mu^P}I_m & 0 & 0 & -I_m & 0 & I_n & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\mu^B(S_1^\mu)^{-2}W_1^E & 0 & 0 & 0 & 0 & I_m & 0 \\ 0 & 0 & 0 & 0 & 0 & 2\mu^B(S_2^\mu)^{-2}W_2^E & 0 & 0 & 0 & 0 & I_m \\ J & 0 & 0 & -I_m & 0 & 0 & \mu^P I_m & 0 & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 & 0 & 0 & 0 & \mu^B Z_1^{-2}Z_1^E & 0 & 0 & 0 \\ 0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 & \mu^B Z_2^{-2}Z_2^E & 0 & 0 \\ 0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & \mu^B W_1^{-2}W_1^E & 0 \\ 0 & 0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & \mu^B W_2^{-2}W_2^E \end{pmatrix},$$

where $H_1 = H(x, 2\pi^{\scriptscriptstyle Y} - y) + \frac{2}{\mu^{\scriptscriptstyle P}} J^T J$. Substituting $\mu^{\scriptscriptstyle B} Z_1^{\scriptscriptstyle E} = X_1^\mu \Pi_1^{\scriptscriptstyle Z}$, $\mu^{\scriptscriptstyle B} Z_2^{\scriptscriptstyle E} = X_2^\mu \Pi_2^{\scriptscriptstyle Z}$, $\mu^{\scriptscriptstyle B} W_1^{\scriptscriptstyle E} = S_1^\mu \Pi_1^{\scriptscriptstyle W}$ and $\mu^{\scriptscriptstyle B} W_2^{\scriptscriptstyle E} = S_2^\mu \Pi_2^{\scriptscriptstyle W}$ gives

$$\begin{pmatrix}
H_1 & 0 & 0 & -\frac{2}{\mu^{\scriptscriptstyle P}} J^T & 0 & 0 & J^T & 0 & 0 & 0 & 0 \\
0 & 2(X_1^\mu)^{-1}\Pi_1^{\scriptscriptstyle Z} & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 \\
0 & 0 & 2(X_2^\mu)^{-1}\Pi_2^{\scriptscriptstyle Z} & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 \\
-\frac{2}{\mu^{\scriptscriptstyle P}} J & 0 & 0 & \frac{2}{\mu^{\scriptscriptstyle P}} I_m & 0 & 0 & -I_m & 0 & I_n & 0 & 0 \\
0 & 0 & 0 & 0 & 2(S_1^\mu)^{-1}\Pi_1^{\scriptscriptstyle W} & 0 & 0 & 0 & 0 & I_m & 0 \\
0 & 0 & 0 & 0 & 0 & 2(S_2^\mu)^{-1}\Pi_2^{\scriptscriptstyle W} & 0 & 0 & 0 & 0 & I_m \\
J & 0 & 0 & -I_m & 0 & 0 & \mu^{\scriptscriptstyle P} I_m & 0 & 0 & 0 & 0 \\
0 & I_n & 0 & 0 & 0 & 0 & 0 & X_1^\mu Z_1^{-2}\Pi_1^{\scriptscriptstyle Z} & 0 & 0 & 0 \\
0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 & X_2^\mu Z_2^{-2}\Pi_2^{\scriptscriptstyle Z} & 0 & 0 \\
0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & S_1^\mu W_1^{-2}\Pi_1^{\scriptscriptstyle W} & 0 \\
0 & 0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & S_2^\mu W_2^{-2}\Pi_2^{\scriptscriptstyle W}
\end{pmatrix}.$$

8.4. Derivation of the shifted primal-dual penalty-barrier direction

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

$$\underset{p \in \mathcal{I}}{\text{minimize}} \;\; M(p) \quad \text{subject to} \;\; Cp = 0, \tag{8.8}$$

where $\mathcal{I}$ is the set of vectors $p = (x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$ such that $x_1 + \mu^{\scriptscriptstyle B} e > 0$, $x_2 + \mu^{\scriptscriptstyle B} e > 0$, $s_1 + \mu^{\scriptscriptstyle B} e > 0$, $s_2 + \mu^{\scriptscriptstyle B} e > 0$, $z_1 > 0$, $z_2 > 0 > 0$, $w_1 > 0$, and $w_2 > 0$, and

$$C = \begin{pmatrix}
I_n & -I_n & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
I_n & 0 & -I_n & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & I_m & -I_m & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & I_m & 0 & -I_m & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.$$

Given $p \in \mathcal{I}$, the Newton direction Δp is the solution of the subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2} \Delta p^T \nabla^2 M(p) \Delta p \quad \text{subject to} \;\; C\Delta p = -Cp. \tag{8.9}$$

However, instead of solving (8.9), we define a linearly constrained modified Newton method by approximating the Hessian $\nabla^2 M(x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$ by a matrix $B(x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$. 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 and $\pi_2^{\scriptscriptstyle W}$ by w_2 everywhere in the matrix $\nabla^2 M(x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$. This gives an *approximate* Hessian $B(x, s, s_F, y, w_F)$ of the form

$$
\begin{pmatrix}
H_1 & 0 & 0 & -\frac{2}{\mu^P} J^T & 0 & 0 & J^T & 0 & 0 & 0 & 0 \\
0 & 2(X_1^\mu)^{-1} Z_1 & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 \\
0 & 0 & 2(X_2^\mu)^{-1} Z_2 & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 \\
-\frac{2}{\mu^P} J & 0 & 0 & \frac{2}{\mu^P} I_m & 0 & 0 & -I_m & 0 & I_n & 0 & 0 \\
0 & 0 & 0 & 0 & 2(S_1^\mu)^{-1} W_1 & 0 & 0 & 0 & 0 & I_m & 0 \\
0 & 0 & 0 & 0 & 0 & 2(S_2^\mu)^{-1} W_2 & 0 & 0 & 0 & 0 & I_m \\
J & 0 & 0 & -I_m & 0 & 0 & \mu^P I_m & 0 & 0 & 0 & 0 \\
0 & I_n & 0 & 0 & 0 & 0 & 0 & X_1^\mu Z_1^{-1} & 0 & 0 & 0 \\
0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 & X_2^\mu Z_2^{-1} & 0 & 0 \\
0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & S_1^\mu W_1^{-1} & 0 \\
0 & 0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & S_2^\mu W_2^{-1}
\end{pmatrix},
$$

where

$$
\begin{aligned}
D_1^Z &= X_1^\mu Z_1^{-1}, & \pi_1^Z &= \mu^B (X_1^\mu)^{-1} z_1^E, & \text{(8.10a)} \\
D_2^Z &= X_2^\mu Z_2^{-1}, & \pi_2^Z &= \mu^B (X_2^\mu)^{-1} z_2^E, & \text{(8.10b)} \\
D_1^W &= X_1^\mu W_1^{-1}, & \pi_1^W &= \mu^B (S_1^\mu)^{-1} w_1^E, & \text{(8.10c)} \\
D_2^W &= X_2^\mu W_2^{-1}, & \pi_2^W &= \mu^B (S_2^\mu)^{-1} w_2^E. & \text{(8.10d)}
\end{aligned}
$$

These definitions of D_1^Z, D_2^Z, D_1^W and D_2^W can be used to write $B(x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$ in the form

$$
\begin{pmatrix}
H + 2J^T D_Y^{-1} J & 0 & 0 & -2J^T D_Y^{-1} & 0 & 0 & J^T & 0 & 0 & 0 & 0 \\
0 & 2(D_1^Z)^{-1} & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 & 0 \\
0 & 0 & 2(D_2^Z)^{-1} & 0 & 0 & 0 & 0 & 0 & I_n & 0 & 0 \\
-2D_Y^{-1} J & 0 & 0 & 2D_Y^{-1} & 0 & 0 & -I_m & 0 & I_n & 0 & 0 \\
0 & 0 & 0 & 0 & 2(D_1^W)^{-1} & 0 & 0 & 0 & 0 & I_m & 0 \\
0 & 0 & 0 & 0 & 0 & 2(D_2^W)^{-1} & 0 & 0 & 0 & 0 & I_m \\
J & 0 & 0 & -I_m & 0 & 0 & D_Y & 0 & 0 & 0 & 0 \\
0 & I_n & 0 & 0 & 0 & 0 & 0 & D_1^Z & 0 & 0 & 0 \\
0 & 0 & I_n & 0 & 0 & 0 & 0 & 0 & D_2^Z & 0 & 0 \\
0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & D_1^W & 0 \\
0 & 0 & 0 & 0 & 0 & I_m & 0 & 0 & 0 & 0 & D_2^W
\end{pmatrix},
$$

where $H = H(x,y)$. Given $B(p) = B(x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$, a modified Newton direction is given by the solution of the QP subproblem

$$\underset{\Delta p}{\text{minimize}} \;\; \nabla M(p)^T \Delta p + \tfrac{1}{2}\Delta p^T B(p)\Delta p \quad \text{subject to} \;\; C\Delta p = -Cp. \tag{8.11}$$

If $p = (x, x_1, x_2, s, s_1, s_2, y, z_1, z_2, w_1, w_2)$ is feasible for the constraints then $Cp = 0$. In this case every feasible Δp may be written in the form $\Delta p = Nd$, where N is a matrix whose columns form a basis for null(C), i.e., $CN = 0$ and $\begin{pmatrix} C^T & N \end{pmatrix}$ is nonsingular. This implies that d must satisfy the reduced equations

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

Consider the particular null-space basis given by

$$N = \begin{pmatrix} I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ -I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & -I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_m & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_n & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_n & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_m & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_m \end{pmatrix}. \tag{8.12}$$

With this definition, the reduced modified Newton equations $N^TB(p)Nd = -N^T\nabla M(p)$ for the linearly constrained problem (8.6) are

$$\begin{pmatrix} H + 2J^TD_Y^{-1}J + 2D_Z^{-1} & -2J^TD_Y^{-1} & J^T & I_n & -I_n & 0 & 0 \\ -2D_Y^{-1}J & 2\big(D_Y^{-1} + D_W^{-1}\big) & -I_m & 0 & 0 & I_m & -I_m \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ I_n & 0 & 0 & D_1^Z & 0 & 0 & 0 \\ -I_n & 0 & 0 & 0 & D_2^Z & 0 & 0 \\ 0 & I_m & 0 & 0 & 0 & D_1^W & 0 \\ 0 & -I_m & 0 & 0 & 0 & 0 & D_2^W \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \end{pmatrix}$$
$$= - \begin{pmatrix} g - J^T(2\pi^Y - y) - (2\pi_1^Z - z_1) + (2\pi_2^Z - z_2) \\ (2\pi^Y - y) - (2\pi_1^W - w_1) + (2\pi_2^W - w_2) \\ D_Y(y - \pi^Y) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \end{pmatrix},$$

where $D_W = \big((D_1^W)^{-1} + (D_2^W)^{-1}\big)^{-1}$, and $D_Z = \big((D_1^Z)^{-1} + (D_2^Z)^{-1}\big)^{-1}$. Given any nonsingular matrix R, the direction d satisfies

$$RN^TB(p)Nd = -RN^T\nabla M(p).$$

In particular, if R is the block upper-triangular matrix

$$R = \begin{pmatrix} I & 0 & -2J^TD_Y^{-1} & -2(D_1^Z)^{-1} & 2(D_2^Z)^{-1} & 0 & 0 \\ & I & 2D_Y^{-1} & 0 & 0 & -2(D_1^W)^{-1} & 2(D_2^W)^{-1} \\ & & I & 0 & 0 & 0 & 0 \\ & & & Z_1 & 0 & 0 & 0 \\ & & & & Z_2 & 0 & 0 \\ & & & & & W_1 & 0 \\ & & & & & & W_2 \end{pmatrix},$$

then

$$RN^TB(p)N = \begin{pmatrix} H & 0 & J^T & -I_n & I_n & 0 & 0 \\ 0 & 0 & -I_m & 0 & 0 & -I_m & I_m \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ Z_1 & 0 & 0 & Z_1D_1^Z & 0 & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & Z_2D_2^Z & 0 & 0 \\ 0 & W_1 & 0 & 0 & 0 & W_1D_1^W & 0 \\ 0 & -W_2 & 0 & 0 & 0 & 0 & W_2D_2^W \end{pmatrix} = \begin{pmatrix} H & 0 & J^T & -I_n & I_n & 0 & 0 \\ 0 & 0 & -I_m & 0 & 0 & -I_m & I_m \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix}.$$

and

$$RN^T\nabla M(p) = -\begin{pmatrix} g - J^Ty - z_1 + z_2 \\ y - w_1 + w_2 \\ D_Y(y - \pi^Y) \\ X_1^\mu(z_1 - \pi_1^Z) \\ X_2^\mu(z_2 - \pi_2^Z) \\ S_1^\mu(w_1 - \pi_1^W) \\ S_2^\mu(w_2 - \pi_2^W) \end{pmatrix},$$

Giving the transformed modified Newton system $RN^TB(p)Nd = -RN^T\nabla M(p)$ as

$$\begin{pmatrix} H & 0 & J^T & -I_n & I_n & 0 & 0 \\ 0 & 0 & -I_m & 0 & 0 & -I_m & I_m \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \end{pmatrix} = -\begin{pmatrix} g - J^Ty - z_1 + z_2 \\ y - w_1 + w_2 \\ D_Y(y - \pi^Y) \\ X_1^\mu(z_1 - \pi_1^Z) \\ X_2^\mu(z_2 - \pi_2^Z) \\ S_1^\mu(w_1 - \pi_1^W) \\ S_2^\mu(w_2 - \pi_2^W) \end{pmatrix}.$$

Identities of the form

$$X_1^\mu\big(z_1 - \pi_1^Z\big) = X_1^\mu\big(z_1 - \mu^B(X_1^\mu)^{-1}z_1^E\big) = Z_1(x_1 + \mu^B e) - \mu^B z_1^E = z_1 \cdot (x - \ell) + \mu^B(z_1 - z_1^E)$$

for each of the terms $X_1^\mu(z_1 - \pi_1^Z)$, $X_2^\mu(z_2 - \pi_2^Z)$, $S_1^\mu(w_1 - \pi_1^W)$, $S_2^\mu(w_2 - \pi_2^W)$ give

$$\begin{pmatrix} H & 0 & J^T & -I_n & I_n & 0 & 0 \\ 0 & 0 & -I_m & 0 & 0 & -I_m & I_m \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \end{pmatrix} = - \begin{pmatrix} g - J^T y - z_1 + z_2 \\ y - w_1 + w_2 \\ D_Y(y - \pi^Y) \\ z_1 \cdot (x - \ell^X) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u^X - x) + \mu^B(z_1 - z_1^E) \\ w_1 \cdot (s - \ell^S) + \mu^B(w_1 - w_1^E) \\ w_2 \cdot (u^S - s) + \mu^B(w_2 - w_2^E) \end{pmatrix}. \tag{8.13}$$

Then, the definition of N from (8.12) implies that

$$\begin{pmatrix} \Delta x \\ \Delta x_1 \\ \Delta x_2 \\ \Delta s \\ \Delta s_1 \\ \Delta s_2 \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = \Delta p = Nd = \begin{pmatrix} I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ -I_n & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & -I_m & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I_m & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_n & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_n & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_m & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I_m \end{pmatrix} = \begin{pmatrix} d_1 \\ d_1 \\ -d_1 \\ d_2 \\ d_2 \\ -d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \\ d_7 \end{pmatrix}.$$

Using these identities to substitute for the components of d in (8.13) yields

$$\begin{pmatrix} H & 0 & J^T & -I_n & I_n & 0 & 0 \\ 0 & 0 & -I_m & 0 & 0 & -I_m & I_m \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 \\ Z_1 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ \Delta y \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g - J^T y - z_1 + z_2 \\ y - w_1 + w_2 \\ D_Y(y - \pi^Y) \\ z_1 \cdot (x - \ell^X) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u^X - x) + \mu^B(z_1 - z_1^E) \\ w_1 \cdot (s - \ell^S) + \mu^B(w_1 - w_1^E) \\ w_2 \cdot (u^S - s) + \mu^B(w_2 - w_2^E) \end{pmatrix}. \tag{8.14}$$

This system is identical to the Newton equations (8.5) for a solution of the path-following equations (8.3).

8.5. Computation of the shifted primal-dual penalty-barrier direction

The symmetric path-following equations are

$$\begin{pmatrix} H & 0 & J^T & I & -I & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & I & -I \\ J & -I & -D_Y & 0 & 0 & 0 & 0 \\ I & 0 & 0 & -Z_1^{-1}X_1^\mu & 0 & 0 & 0 \\ -I & 0 & 0 & 0 & -Z_2^{-1}X_2^\mu & 0 & 0 \\ 0 & I & 0 & 0 & 0 & -W_1^{-1}S_1^\mu & 0 \\ 0 & -I & 0 & 0 & 0 & 0 & -W_2^{-1}S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta s \\ -\Delta y \\ -\Delta z_1 \\ -\Delta z_2 \\ -\Delta w_1 \\ -\Delta w_2 \end{pmatrix} = - \begin{pmatrix} g - J^T y - z_1 + z_2 \\ y - w_1 + w_2 \\ c - s + \mu^P (y - y^E) \\ Z_1^{-1}\big(z_1 \cdot (x - \ell^X) + \mu^B (z_1 - z_1^E)\big) \\ Z_2^{-1}\big(z_2 \cdot (u^X - x) + \mu^B (z_2 - z_2^E)\big) \\ W_1^{-1}\big(w_1 \cdot (s - \ell^S) + \mu^B (w_1 - w_1^E)\big) \\ W_2^{-1}\big(w_2 \cdot (u^S - s) + \mu^B (w_2 - w_2^E)\big) \end{pmatrix}. \tag{8.15}$$

After collecting terms and reordering the equations and unknowns, we obtain

$$\begin{pmatrix} D_1^Z & 0 & 0 & 0 & 0 & I_n & 0 \\ 0 & D_2^Z & 0 & 0 & 0 & -I_n & 0 \\ 0 & 0 & D_1^W & 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & D_2^W & -I_m & 0 & 0 \\ 0 & 0 & -I_m & I_m & 0 & 0 & I_m \\ 0 & 0 & 0 & 0 & -I_m & J & D_Y \\ -I_n & I_n & 0 & 0 & 0 & H & -J^T \end{pmatrix} \begin{pmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_1^Z (z_1 - \pi_1^Z) \\ D_2^Z (z_2 - \pi_2^Z) \\ D_1^W (w_1 - \pi_1^W) \\ D_2^W (w_2 - \pi_2^W) \\ y - w_1 + w_2 \\ D_Y (y - \pi^Y) \\ g - J^T y - z_1 + z_2 \end{pmatrix}, \tag{8.16}$$

where

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

If we define $\bar{H} = H + (D_1^Z)^{-1} + (D_2^Z)^{-1}$, $\bar{D}_Y = D_Y + D_W$ and $D_W = \big((D_1^W)^{-1} + (D_2^W)^{-1}\big)^{-1}$, then premultiplying the equations (8.16) by the matrix

$$\begin{pmatrix} I_n & & & & & & \\ 0 & I_n & & & & & \\ 0 & 0 & I_m & & & & \\ 0 & 0 & 0 & I_m & & & \\ 0 & 0 & (D_1^W)^{-1} & -(D_2^W)^{-1} & I_m & & \\ 0 & 0 & D_W (D_1^W)^{-1} & -D_W (D_2^W)^{-1} & D_W & I_n & \\ (D_1^Z)^{-1} & -(D_2^Z)^{-1} & 0 & 0 & 0 & 0 & I_m \end{pmatrix}$$

gives the block upper-triangular system

$$\begin{pmatrix} D_1^Z & 0 & 0 & 0 & 0 & I_n & 0 \\ 0 & D_2^Z & 0 & 0 & 0 & -I_n & 0 \\ 0 & 0 & D_1^W & 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & D_2^W & -I_m & 0 & 0 \\ 0 & 0 & 0 & 0 & I_m & 0 & D_W \\ 0 & 0 & 0 & 0 & 0 & J & \bar{D}_Y \\ 0 & 0 & 0 & 0 & 0 & \bar{H}_F & -J^T \end{pmatrix} \begin{pmatrix} \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s \\ \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ D_W(y - \pi^W) \\ D_W(y - \pi^W) + D_Y(y - \pi^Y) \\ g - J^T y - \pi^Z \end{pmatrix},$$

where $\pi^W = \pi_1^W - \pi_2^W$ and $\pi^Z = \pi_1^Z - \pi_2^Z$. Using block back substitution, we may compute Δx and Δy by solving the equations

$$\begin{pmatrix} \bar{H}(x,y) + D_Z^{-1} & -J(x)^T \\ J(x) & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = - \begin{pmatrix} g(x) - J(x)^T y - \pi^Z \\ D_W(y - \pi^W) + D_Y(y - \pi^Y) \end{pmatrix}, \tag{8.17}$$

with $D_Z = \big((D_1^Z)^{-1} + (D_2^Z)^{-1}\big)^{-1}$ and $D_W = \big((D_1^W)^{-1} + (D_2^W)^{-1}\big)^{-1}$. The fifth block of equations gives

$$\Delta s = -D_W(y + \Delta y - w_1 + w_2).$$

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 (8.5) to give

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

Similarly, using (8.5) to solve for Δz_1 and Δz_2 yields

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

The variables x_1, x_2, s_1 and s_2 may be computed implicitly for the line search, and the appropriate merit function is

$$
\begin{aligned}
f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\
- \sum_{j=1}^{n} \Big\{ \mu^B [z_1^E]_j \ln\big(x_j - \ell_j^X + \mu^B\big) + \mu^B [z_1^E]_j \ln\big([z_1]_j (x_j - \ell_j^X + \mu^B)\big) - [z_1]_j (x_j - \ell_j^X + \mu^B) \Big\} \\
- \sum_{j=1}^{n} \Big\{ \mu^B [z_2^E]_j \ln\big(u_j^X - x_j + \mu^B\big) + \mu^B [z_2^E]_j \ln\big([z_2]_j (u_j^X - x_j + \mu^B)\big) - [z_2]_j (u_j^X - x_j + \mu^B) \Big\} \\
- \sum_{i=1}^{m} \Big\{ \mu^B [w_1^E]_i \ln\big(s_i - \ell_i^S + \mu^B\big) + \mu^B [w_1^E]_i \ln\big([w_1]_i (s_i - \ell_i^S + \mu^B)\big) - [w_1]_i (s_i - \ell_i^S + \mu^B) \Big\} \\
- \sum_{i=1}^{m} \Big\{ \mu^B [w_2^E]_i \ln\big(u_i^S - s_i + \mu^B\big) + \mu^B [w_2^E]_i \ln\big([w_2]_i (u_i^S - s_i + \mu^B)\big) - [w_2]_i (u_i^S - s_i + \mu^B) \Big\}.
\end{aligned}
\tag{8.18}
$$

8.6. Summary: upper and lower bounds on all variables and slacks

The results of Sections 6.5 and 7.5 imply that the solution of the path-following equations (8.5) may be computed as follows. Let x and s be given primal variables such that

$$\ell^X - \mu^B < x < u^X + \mu^B, \quad \text{and} \quad \ell^S - \mu^B < s < u^S + \mu^B,$$

and dual variables y, w_1, w_2, z_1, and z_2 such that

$$w_1 > 0, \quad w_2 > 0, \quad z_1 > 0, \quad \text{and} \quad z_2 > 0.$$

Let X_1, X_2, S_1, and S_2 denote the matrices $\operatorname{diag}\big(x_j - \ell^X_j\big)$, $\operatorname{diag}\big(u^X_j - x_j\big)$, $\operatorname{diag}\big(s_i - \ell^S_i\big)$ and $\operatorname{diag}\big(u^S_i - s_i\big)$, respectively. If D^Z_1, D^Z_2, D^W_1, D^W_2, D_Y, D_Z, D_W, π^Y, π^Z_1, π^Z_2, π^W_1 and π^W_2 denote the quantities

$$\begin{aligned}
D_Y &= \mu^P I, & \pi^Y &= y^E - \frac{1}{\mu^P}\big(c(x) - s\big), & & \\
D^Z_1 &= Z_1^{-1} X^\mu_1, & D^Z_2 &= Z_2^{-1} X^\mu_2, & & \\
D^W_1 &= W_1^{-1} S^\mu_1, & D^W_2 &= W_2^{-1} S^\mu_2, & & \\
D_Z &= \big((D^Z_1)^{-1} + (D^Z_2)^{-1}\big)^{-1}, & \pi^Z_1 &= \mu^B (X^\mu_1)^{-1} z^E_1, & \pi^Z_2 &= \mu^B (X^\mu_2)^{-1} z^E_2, \\
D_W &= \big((D^W_1)^{-1} + (D^W_2)^{-1}\big)^{-1}, & \pi^W_1 &= \mu^B (S^\mu_1)^{-1} w^E_1, & \pi^W_2 &= \mu^B (S^\mu_2)^{-1} w^E_2,
\end{aligned}$$

then Δx, Δy, Δs, Δw_1, Δw_2, Δz_1 and Δz_2, can be computed using the equations

$$\begin{aligned}
\pi^W &= \pi^W_1 - \pi^W_2, & \widehat{y} &= y + \Delta y, & \Delta s &= -D_W\big(\widehat{y} - \pi^W\big), \\
\pi^Z &= \pi^Z_1 - \pi^Z_2, & \widehat{x} &= x + \Delta x, & \Delta z_1 &= -(X^\mu_1)^{-1}\big(z_1 \cdot (\widehat{x} - \ell^X) + \mu^B(z_1 - z^E_1)\big), \\
& & & & \Delta z_2 &= -(X^\mu_2)^{-1}\big(z_2 \cdot (u^X - \widehat{x}) + \mu^B(z_2 - z^E_2)\big), \\
& & \widehat{s} &= s + \Delta s, & \Delta w_1 &= -(S^\mu_1)^{-1}\big(w_1 \cdot (\widehat{s} - \ell^S) + \mu^B(w_1 - w^E_1)\big), \\
& & & & \Delta w_2 &= -(S^\mu_2)^{-1}\big(w_2 \cdot (u^S - \widehat{s}) + \mu^B(w_2 - w^E_2)\big),
\end{aligned}$$

where Δx and Δy satisfy the equations

$$\begin{pmatrix} H(x,y) + D_Z^{-1} & -J(x)^T \\ J(x) & D_Y + D_W \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = -\begin{pmatrix} g(x) - J(x)^T y - \pi^Z \\ D_Y(y - \pi^Y) + D_W(y - \pi^W) \end{pmatrix}. \tag{8.19}$$

As $(x, s) \to (x^*, s^*)$ it holds that $\|\bar{D}_W\| \to \infty$, which implies that the matrix and right-hand side of this system goes to infinity. If $\widehat{D}_W$ is the diagonal matrix such that $\widehat{D}_W^2 = D_W^{-1}$, equations (8.19) may be written in the form

$$\begin{pmatrix} H(x,y) + D_Z^{-1} & -(\widehat{D}_W J(x))^T \\ \widehat{D}_W J(x) & \widehat{D}_W^2 D_Y + I \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta \widehat{y} \end{pmatrix} = -\begin{pmatrix} g(x) - J(x)^T y - \pi^Z \\ \widehat{D}_W\big(D_Y(y - \pi^Y) + D_W(y - \pi^W)\big) \end{pmatrix}, \qquad \Delta y = \widehat{D}_W \Delta\widehat{y}. \tag{8.20}$$

In this case, the scaled KKT matrix remains bounded if $H(x,y)$ is bounded. Similarly, the right-hand side remains bounded if $\|\widehat{D}_W D_W(y-\pi^W)\|$ is bounded.

The associated line-search merit function $M(x,s,y,z_1,z_2,w_1,w_2)$ of (8.18) can be written as

$$
\begin{aligned}
f(x) - \big(c(x)-s\big)^T y^E + \frac{1}{2\mu^P}\|c(x)-s\|^2 + \frac{1}{2\mu^P}\|c(x)-s+\mu^P(y-y^E)\|^2 \\
- \sum_{j=1}^{n}\Big\{\mu^B[z_1^E]_j \ln\big([z_1]_j(x_j-\ell_j^X+\mu^B)^2\big) - [z_1]_j(x_j-\ell_j^X+\mu^B)\Big\} \\
- \sum_{j=1}^{n}\Big\{\mu^B[z_2^E]_j \ln\big([z_2]_j(u_j^X-x_j+\mu^B)^2\big) - [z_2]_j(u_j^X-x_j+\mu^B)\Big\} \\
- \sum_{i=1}^{m}\Big\{\mu^B[w_1^E]_i \ln\big([w_1]_i(s_i-\ell_i^S+\mu^B)^2\big) - [w_1]_i(s_i-\ell_i^S+\mu^B)\Big\} \\
- \sum_{i=1}^{m}\Big\{\mu^B[w_2^E]_i \ln\big([w_2]_i(u_i^S-s_i+\mu^B)^2\big) - [w_2]_i(u_i^S-s_i+\mu^B)\Big\},
\end{aligned}
$$

for which the gradient $\nabla M(x,s,y,z_1,z_2,w_1,w_2)$ can be written as

$$
\begin{pmatrix}
g - J^T\big(\pi^Y + (\pi^Y - y)\big) - \big(\pi^Z + (\pi^Z - z)\big) \\
\pi^Y + (\pi^Y - y) - \big(\pi^W + (\pi^W - w)\big) \\
-D_Y(\pi^Y - y) \\
-D_1^Z(\pi_1^Z - z_1) \\
-D_2^Z(\pi_2^Z - z_2) \\
-D_1^W(\pi_1^W - w_1) \\
-D_2^W(\pi_2^W - w_2)
\end{pmatrix},
$$

where $z = z_1 - z_2$, $w = w_1 - w_2$.

The residuals of the unsymmetric path-following equations (8.5) may be written as

$$r = \begin{pmatrix} g - J^T y - z \\ y - w \\ c - s + \mu^P (y - y^E) \\ z_1 \cdot (x - \ell^X) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u^X - x) + \mu^B (z_2 - z_2^E) \\ w_1 \cdot (s - \ell^S) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u^S - s) + \mu^B (w_2 - w_2^E) \end{pmatrix} = \begin{pmatrix} g - J^T y - z \\ y - w \\ \mu^P (y - \pi^Y) \\ X_1^\mu (z_1 - \pi_1^Z) \\ X_2^\mu (z_2 - \pi_2^Z) \\ S_1^\mu (w_1 - \pi_1^W) \\ S_2^\mu (w_2 - \pi_2^W) \end{pmatrix},$$

with $z = z_1 - z_2$ and $w = w_1 - w_2$.

9. General case: upper and lower Bounds on all variables

Next we consider the case with upper and lower bounds on both the variables and slacks, together with both implicit and explicit bounds on the variables.

9.1. Problem statement and optimality conditions

The definition of the problem is

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad \begin{cases} c(x)-s=0, & L_X s = h_X, \quad \ell^S \le L_F s \le u^S, \\ Ax-b=0, & E_X x = b_X, \quad \ell^X \le E_F x \le u^X, \end{cases} \tag{9.1}$$

where $c:\mathbb{R}^n \mapsto \mathbb{R}^m$ and $f:\mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable. 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 matrices associated with the linear constraints $E_X x = b_X$ and $Ax = b$ have linearly independent rows. The matrices L_X and L_F are formed from rows of the identity matrix I_m in such a way that $s_X = L_X s$ and $s_F = L_F s$ are the vectors of "fixed" and "free" components of s. It follows that there is an $m\times m$ permutation matrix P such that

$$P_s = \begin{pmatrix} L_F \\ L_X \end{pmatrix},$$

with the matrices L_F and L_X satisfying the identities $L_F L_F^T = I_F$, $L_X L_X^T = I_X$, and $L_F L_X^T = 0$. The matrices E_X and E_F define an analogous partition of x into fixed and free components x_F and x_X of x. The bound constraints involving E_X and L_X are enforced exactly. The linear constraints $Ax-b=0$ are imposed using a shifted primal-dual penalty method.

The first-order KKT conditions for problem (9.1) are

$$\begin{aligned}
g(x^*) - J(x^*)^T y^* - A^T v^* - E_X^T z_X^* - E_F^T z_1^* + E_F^T z_2^* &= 0, & z_1^* &\ge 0, & z_2^* &\ge 0, && \text{(9.2a)}\\
y^* - L_X^T w_X^* - L_F^T w_1^* + L_F^T w_2^* &= 0, & w_1^* &\ge 0, & w_2^* &\ge 0, && \text{(9.2b)}\\
c(x^*) - s^* &= 0, & & & L_X s^* - h_X &= 0, && \text{(9.2c)}\\
& & Ax^* - b &= 0, & E_X x^* - b_X &= 0, && \text{(9.2d)}\\
E_F x^* - \ell^X &\ge 0, & u^X - E_F x^* &\ge 0, & & && \text{(9.2e)}\\
L_F s^* - \ell^S &\ge 0, & u^S - L_F s^* &\ge 0, & & && \text{(9.2f)}\\
z_1^* \cdot (E_F x^* - \ell^X) &= 0, & z_2^* \cdot (u^X - E_F x^*) &= 0, & & && \text{(9.2g)}\\
w_1^* \cdot (L_F s^* - \ell^S) &= 0, & w_2^* \cdot (u^S - L_F s^*) &= 0, & & && \text{(9.2h)}
\end{aligned}$$

where y^* are the multipliers for the equality constraints $c(x)-s=0$, and z_1^*, z_2^*, w_1^*, and w_2^* may be interpreted as the Lagrange multipliers for the inequality constraints $E_F x - \ell^X \ge 0$, $u^X - E_F x \ge 0$, $L_F s - \ell^S \ge 0$ and $u^S - L_F s \ge 0$, respectively. The components of v^* are the multipliers for the linear equality constraints $Ax = b$. If $x_1 = E_F x - \ell^X$, $x_2 = u^X - E_F x$, $s_1 = L_F s - \ell^S$, and $s_2 = u^S - L_F s$, then z_1^*, z_2^*, w_1^*, and w_2^* 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.

9.2. The path-following equations

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

$$
\begin{aligned}
g(x) - J(x)^T y - A^T v - E_X^T z_X - E_F^T z_1 + E_F^T z_2 &= 0, & z_1 &\geq 0, & z_2 &\geq 0, && \text{(9.3a)}\\
y - L_X^T w_X - L_F^T w_1 + L_F^T w_2 &= 0, & w_1 &\geq 0, & w_2 &\geq 0, && \text{(9.3b)}\\
c(x) - s &= \mu^P (y^E - y), & E_X x^* - b_X &= 0, & L_X s^* - h_X &= 0, && \text{(9.3c)}\\
Ax - b &= \mu^A (v^E - v), & & & & && \text{(9.3d)}\\
E_F x - \ell^X &\geq 0, & u^X - E_F x &\geq 0, & & && \text{(9.3e)}\\
L_F s - \ell^S &\geq 0, & u^S - L_F s &\geq 0, & & && \text{(9.3f)}\\
z_1 \cdot (E_F x - \ell^X) &= \mu^B (z_1^E - z_1), & z_2 \cdot (u^X - E_F x) &= \mu^B (z_2^E - z_2), & & && \text{(9.3g)}\\
w_1 \cdot (L_F s - \ell^S) &= \mu^B (w_1^E - w_1), & w_2 \cdot (u^S - L_F s) &= \mu^B (w_2^E - w_2), & & && \text{(9.3h)}
\end{aligned}
$$

Consider the primal-dual path-following equations $F(x,\, s,\, y,\, v,\, w_X,\, z_X,\, z_1,\, z_2,\, w_1,\, w_2\,;\mu^A,\, \mu^P,\, \mu^B,\, y^E,\, v^E,\, z_1^E,\, z_2^E,\, w_1^E,\, w_2^E) = 0$, with

$$
F(x, s, y, v, z_1, z_2, w_1, w_2\,; \mu^P, \mu^B, y^E, v^E, z_1^E, z_2^E, w_1^E, w_2^E) =
\begin{pmatrix}
g(x) - J(x)^T y - A^T v - E_X^T z_X - E_F^T z_1 + E_F^T z_2 \\
y - L_X^T w_X - L_F^T w_1 + L_F^T w_2 \\
c(x) - s + \mu^P (y - y^E) \\
Ax - b + \mu^A (v - v^E) \\
L_X s - h_X \\
z_1 \cdot (E_F x - \ell^X) + \mu^B (z_1 - z_1^E) \\
z_2 \cdot (u^X - E_F x) + \mu^B (z_2 - z_2^E) \\
w_1 \cdot (L_F s - \ell^S) + \mu^B (w_1 - w_1^E) \\
w_2 \cdot (u^S - L_F s) + \mu^B (w_2 - w_2^E) \\
E_X x - b_X
\end{pmatrix}. \tag{9.4}
$$

Any zero $(x,\, s,\, y,\, v,\, w_X,\, z_X,\, z_1,\, z_2,\, w_1,\, w_2)$ of F such that $\ell^X < E_F x < u^X$, $\ell^S < L_F s < u^S$, $z_1 > 0$, $z_2 > 0$, $w_1 > 0$, and $w_2 > 0$ approximates a point satisfying the optimality conditions (9.2), 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 (9.3) 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 (9.3) will approximate a solution of (9.2) 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 $\mu^B \to 0$).

9.3. A shifted primal-dual penalty-barrier function

Problem (9.1) is equivalent to

$$\begin{array}{lllll}
\underset{x,x_1,x_2,s,s_1,s_2}{\text{minimize}} & f(x) & & & \\
\text{subject to} & c(x) - s = 0, & Ax - b = 0, & & \\
& E_F x - x_1 = \ell^X, & L_F s - s_1 = \ell^S, & x_1 \ge 0, & s_1 \ge 0, \\
& E_F x + x_2 = u^X, & L_F s + s_2 = u^S, & x_2 \ge 0, & s_2 \ge 0, \\
& E_X x - b_X = 0, & L_X s - h_X = 0. & &
\end{array}$$

Consider the primal-dual shifted penalty-barrier problem

$$\begin{array}{lllllll}
\underset{x,x_1,x_2,s,s_1,s_2,y,v,z_1,z_2,w_1,w_2}{\text{minimize}} & \multicolumn{6}{l}{M(x, x_1, x_2, s, s_1, s_2, y, v, w_1, w_2\,; \mu^P, \mu^B, y^E, v^E, w_1^E, w_2^E)} \\
\text{subject to} & E_F x - x_1 = \ell^X, & L_F s - s_1 = \ell^S, & x_1 + \mu^B e > 0, & z_1 > 0, & s_1 + \mu^B e > 0, & w_1 > 0, \\
& E_F x + x_2 = u^X, & L_F s + s_2 = u^S, & x_2 + \mu^B e > 0, & z_2 > 0, & s_2 + \mu^B e > 0, & w_2 > 0, \\
& E_X x - b_X = 0, & L_X s - h_X = 0, & & & &
\end{array} \tag{9.5}$$

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 penalty-barrier function

$$
\begin{aligned}
f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\
- (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A(v - v^E)\|^2 \\
- \sum_{j=1}^{n_F}\Big\{\mu^B[\,z_1^E\,]_j \ln\big([\,z_1\,]_j[\,x_1 + \mu^B e\,]_j^2\big) - [\,z_1 \cdot (x_1 + \mu^B e)\,]_j\Big\} \\
- \sum_{j=1}^{n_F}\Big\{\mu^B[\,z_2^E\,]_j \ln\big([\,z_2\,]_j[\,x_2 + \mu^B e\,]_j^2\big) - [\,z_2 \cdot (x_2 + \mu^B e)\,]_j\Big\} \\
- \sum_{i=1}^{m_F}\Big\{\mu^B[\,w_1^E\,]_i \ln\big([\,w_1\,]_i[\,s_1 + \mu^B e\,]_i^2\big) - [\,w_1 \cdot (s_1 + \mu^B e)\,]_i\Big\} \\
- \sum_{i=1}^{m_F}\Big\{\mu^B[\,w_2^E\,]_i \ln\big([\,w_2\,]_i[\,s_2 + \mu^B e\,]_i^2\big) - [\,w_2 \cdot (s_2 + \mu^B e)\,]_i\Big\}.
\end{aligned} \tag{9.6}
$$

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

The gradient may be written in several equivalent forms

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

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

9.4. Derivation of the shifted primal-dual penalty-barrier direction

9.5. Computation of the shifted primal-dual penalty-barrier direction

Next we consider the solution of the path-following Newton equations (9.4). If $v = (x,\, s,\, y,\, v,\, w_X,\, z_X,\, z_1,\, z_2,\, w_1,\, w_2)$ is a given approximate zero of $F(v)$ such that $\ell^X - \mu^B < E_F x < u^X + \mu^B$, $\ell^S - \mu^B < L_F s < u^S + \mu^B$, $z_1 > 0$, $z_2 > 0$, $w_1 > 0$, and

$w_2 > 0$, the Newton equations for the change in variables $\Delta v = (\Delta x,\ \Delta s,\ \Delta y,\ \Delta v,\ \Delta w_X,\ \Delta z_X,\ \Delta z_1,\ \Delta z_2,\ \Delta w_1,\ \Delta w_2)$ are given by $F'(v)\Delta v = -F(v)$, where

$$F(v) = \begin{pmatrix} g(x) - J(x)^T y - A^T v - E_X^T z_X - E_F^T z_1 + E_F^T z_2 \\ y - L_X^T w_X - L_F^T w_1 + L_F^T w_2 \\ c(x) - s + \mu^P (y - y^E) \\ Ax - b + \mu^A (v - v^E) \\ L_X s - h_X \\ E_X x - b_X \\ z_1 \cdot (E_F x - \ell^X) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u^X - E_F x) + \mu^B (z_2 - z_2^E) \\ w_1 \cdot (L_F s - \ell^S) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u^S - L_F s) + \mu^B (w_2 - w_2^E) \end{pmatrix} \tag{9.7}$$

and

$$F'(v) = \begin{pmatrix} H & 0 & -J^T & -A^T & 0 & -E_X^T & -E_F^T & E_F^T & 0 & 0 \\ 0 & 0 & I_m & 0 & -L_X^T & 0 & 0 & 0 & -L_F^T & L_F^T \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A & 0 & 0 & D_A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_F & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Z_1 E_F & 0 & 0 & 0 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 E_F & 0 & 0 & 0 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 L_F & 0 & 0 & 0 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 L_F & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix}, \tag{9.8}$$

where $D_Y = \mu^P I_m$, $D_A = \mu^A I_A$, $X_1^\mu = \operatorname{diag}(E_F x - \ell^X + \mu^B e)$, $X_2^\mu = \operatorname{diag}(u^X - E_F x + \mu^B e)$, $Z_1 = \operatorname{diag}(z_1)$, $Z_2 = \operatorname{diag}(z_2)$, $W_1 = \operatorname{diag}([w_1]_i)$, $W_2 = \operatorname{diag}([w_2]_i)$, $S_1^\mu = \operatorname{diag}(L_F s - \ell^S + \mu^B e)$, and $S_2^\mu = \operatorname{diag}(u^S - L_F s + \mu^B e)$.

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. 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. 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 fourth block of equations by L_F and A_F respectively, these identities allow us to write the equations (7.5) in the reduced form $\widehat{F}'\Delta v_F = -\widehat{F}$, where $\Delta v_F = (\Delta x_F,\, \Delta s_F,\, \Delta y,\, \Delta v,\, \Delta w_X,\, \Delta z_X,\, \Delta z_1,\, \Delta z_2,\, \Delta w_1,\, \Delta w_2)$,

$$\begin{pmatrix} H_F & 0 & -J_F^T & -A_F^T & -I_F^x & I_F^x & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -I_F^s & I_F^s \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ Z_1 & 0 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 & 0 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 & 0 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 & 0 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} g_F - J_F^T y - A_F^T v - z_1 + z_2 \\ y_F - w_1 + w_2 \\ c(x) - s + \mu^P(y - y^E) \\ Ax - b + \mu^A(v - v^E) \\ z_1 \cdot (E_F x - \ell^X) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u^X - E_F x) + \mu^B(z_2 - z_2^E) \\ w_1 \cdot (L_F s - \ell^S) + \mu^B(w_1 - w_1^E) \\ w_2 \cdot (u^S - L_F s) + \mu^B(w_2 - w_2^E) \end{pmatrix}, \tag{9.9}$$

where $A_F = AE_F^T$ are the columns of A associated with the "free" components of x. 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 $\Delta z_X = [\,g + H\Delta x - J^T(y + \Delta y) - z\,]_X$. 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

$$\begin{pmatrix} D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & D_1^Z & 0 & 0 & 0 & 0 & I_F^x & 0 \\ 0 & 0 & D_2^Z & 0 & 0 & 0 & -I_F^x & 0 \\ 0 & 0 & 0 & D_1^W & 0 & I_F^s & 0 & 0 \\ 0 & 0 & 0 & 0 & D_2^W & -I_F^s & 0 & 0 \\ 0 & 0 & 0 & -I_F^s & I_F^s & 0 & 0 & L_F \\ 0 & 0 & 0 & 0 & 0 & -L_F^T & J_F & D_Y \\ -A_F^T & -I_F^x & I_F^x & 0 & 0 & 0 & H_F & -J_F^T \end{pmatrix} \begin{pmatrix} \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_A(v - \pi^A) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ y_F - w_1 + w_2 \\ D_Y(y - \pi^Y) \\ g_F - J_F^T y - A_F^T v - z_1 + z_2 \end{pmatrix}, \tag{9.10}$$

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

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

If we define $\bar{H}_F = H_F + A_F^T D_A^{-1} A_F + (D_1^Z)^{-1} + (D_2^Z)^{-1}$, $\bar{D}_Y = D_Y + L_F^T D_W L_F$ and $D_W = \big((D_1^W)^{-1} + (D_2^W)^{-1}\big)^{-1}$, then premultiplying the equations (9.10) by the matrix

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

gives the block upper-triangular system

$$\begin{pmatrix}
D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\
0 & D_1^z & 0 & 0 & 0 & 0 & I_F^x & 0 \\
0 & 0 & D_2^z & 0 & 0 & 0 & -I_F^x & 0 \\
0 & 0 & 0 & D_1^W & 0 & I_F^s & 0 & 0 \\
0 & 0 & 0 & 0 & D_2^W & -I_F^s & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & I_F^s & 0 & D_W L_F \\
0 & 0 & 0 & 0 & 0 & 0 & J_F & \bar{D}_Y \\
0 & 0 & 0 & 0 & 0 & 0 & \bar{H}_F & -J_F^T
\end{pmatrix}
\begin{pmatrix}
\Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta y
\end{pmatrix}
= -
\begin{pmatrix}
D_A(v - \pi^A) \\
D_1^z(z_1 - \pi_1^z) \\
D_2^z(z_2 - \pi_2^z) \\
D_1^W(w_1 - \pi_1^W) \\
D_2^W(w_2 - \pi_2^W) \\
D_W(y_F - \pi_F^W) \\
L_F^T D_W(y_F - \pi_F^W) + D_Y(y - \pi^Y) \\
g_F - J_F^T y - A_F^T \pi^A - \pi_F^z
\end{pmatrix},$$

where $\pi^W = L_F^T \pi_1^W - L_F^T \pi_2^W$ and $\pi^z = E_F^T \pi_1^z - E_F^T \pi_2^z$. Using block back substitution, we may compute Δx and Δy by solving the equations

$$\begin{pmatrix}
\bar{H}_F(x,y) + A_F^T D_A^{-1} A_F + D_Z^{-1} & -J_F(x)^T \\
J_F(x) & D_Y + L_F^T D_W L_F
\end{pmatrix}
\begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix}
= -
\begin{pmatrix}
g_F(x) - J_F(x)^T y - A_F^T \pi^A - \pi_F^z \\
L_F^T D_W(y_F - \pi_F^W) + D_Y(y - \pi^Y)
\end{pmatrix}, \tag{9.11}$$

with $D_Z = \big((D_1^z)^{-1} + (D_2^z)^{-1}\big)^{-1}$ and $D_W = \big((D_1^W)^{-1} + (D_2^W)^{-1}\big)^{-1}$. The full vector Δx is then computed as $\Delta x = E_F^T \Delta x_F$. Using the identity $\Delta s = L_F^T \Delta s_F$ in the fifth block of equations gives

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

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 (9.9) to give

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

Similarly, using (9.9) to solve for Δz_1 and Δz_2 yields

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

Similarly, using the fourth and fifth block of equations of the Newton equations for a zero of (9.7) to solve for Δv gives $\Delta v = -(v - \widehat{\pi}^A)$, with $\widehat{\pi}^A = v^E - \frac{1}{\mu^A}\big(A(x+\Delta x) - b\big)$. Finally, the vectors Δw_X and Δz_X are recovered as $\Delta w_X = [\, y + \Delta y - w\,]_X$ and $\Delta z_X = [\, g + H\Delta x - J^T(y+\Delta y) - z\,]_X$.

9.6. Summary: computations associated with the general problem

The results of the preceding section implies that the solution of the path-following equations $F'(v)\Delta v = -F(v)$ with F and F' given by (9.7) and (9.8) may be computed as follows. Let x and s be given primal variables such that $E_X x = b_X$, $L_X s = h_X$, with

$$\ell^X - \mu^B e < E_F x < u^X + \mu^B e, \quad \text{and} \quad \ell^S - \mu^B e < L_F s < u^S + \mu^B e,$$

and dual variables y, w_1, w_2, z_1, and z_2 such that $w_1 > 0$, $w_2 > 0$, $z_1 > 0$, and $z_2 > 0$. Let X_1, X_2, S_1, and S_2 denote the matrices $\operatorname{diag}\big([\, x_F - \ell^X\,]_i\big)$, $\operatorname{diag}\big([\, u^X - x_F\,]_i\big)$, $\operatorname{diag}\big([\, s_F - \ell^S\,]_i\big)$ and $\operatorname{diag}\big([\, u^S - s_F\,]_i\big)$, respectively, and define the quantities

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

Solve the KKT system

$$\begin{pmatrix} H_F(x,y) + A_F^T D_A^{-1} A_F + D_Z^{-1} & -J_F(x)^T \\ J_F(x) & D_Y + L_F^T D_W L_F \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} E_F\big(g(x) - J(x)^T y - A^T \pi^A - \pi^Z\big) \\ L_F^T D_W L_F\big(y - \pi^W\big) + D_Y(y - \pi^Y) \end{pmatrix}. \tag{9.12}$$

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

As $(x,s) \to (x^*, s^*)$ it holds that $\|D_Z^{-1}\|$ is 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 this system 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 $\widehat{D}_Z$ and $\widehat{D}_W$ denote diagonal matrices such that $\widehat{D}_Z^2 = (A_F^T D_A^{-1} A_F)^{-1}$ and $\widehat{D}_W^2 = (L_F^T D_W L_F)^{-1}$, then $\|\widehat{D}_Z\|$ and $\|\widehat{D}_W\|$ are bounded as $(x,s) \to (x^*, s^*)$. The equations (9.12) may be written in the form

$$\begin{pmatrix} \widehat{D}_Z H_F(x,y)\widehat{D}_Z + \widehat{D}_Z^2 D_Z^{-1} + I & -(\widehat{D}_W J_F(x)\widehat{D}_Z)^T \\ \widehat{D}_W J_F(x)\widehat{D}_Z & D_Y + L_F^T D_W L_F \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = -\begin{pmatrix} \widehat{D}_Z E_F\big(g(x) - J(x)^T y - A^T\pi^A - \pi^Z\big) \\ \widehat{D}_W\big(L_F^T D_W L_F\big(y - \pi^W\big) + D_Y(y - \pi^Y)\big) \end{pmatrix}, \tag{9.13}$$

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

The associated line-search merit function (9.6) can be written as

$$
\begin{aligned}
f(x) - \big(c(x)-s\big)^T y^{E} + \frac{1}{2\mu^{P}}\|c(x)-s\|^2 + \frac{1}{2\mu^{P}}\|c(x)-s+\mu^{P}(y-y^{E})\|^2 \\
- (Ax-b)^T v^{E} + \frac{1}{2\mu^{A}}\|Ax-b\|^2 + \frac{1}{2\mu^{A}}\|Ax-b+\mu^{A}(v-v^{E})\|^2 \\
- \sum_{j=1}^{n_F}\Big\{\mu^{B}[\,z_1^{E}\,]_j \ln\big([\,z_1\,]_j[\,x_F-\ell^{X}+\mu^{B}e\,]_j^2\big) - [\,z_1\cdot(x_F-\ell^{X}+\mu^{B}e)\,]_j\Big\} \\
- \sum_{j=1}^{n_F}\Big\{\mu^{B}[\,z_2^{E}\,]_j \ln\big([\,z_2\,]_j[\,u^{X}-x_F+\mu^{B}e\,]_j^2\big) - [\,z_2\cdot(u^{X}-x_F+\mu^{B}e)\,]_j\Big\} \\
- \sum_{i=1}^{m_F}\Big\{\mu^{B}[\,w_1^{E}\,]_i \ln\big([\,w_1\,]_i[\,s_F-\ell^{S}+\mu^{B}e\,]_i^2\big) - [\,w_1\cdot(s_F-\ell^{S}+\mu^{B}e)\,]_i\Big\} \\
- \sum_{i=1}^{m_F}\Big\{\mu^{B}[\,w_2^{E}\,]_i \ln\big([\,w_2\,]_i[\,u^{S}-s_F+\mu^{B}e\,]_i^2\big) - [\,w_2\cdot(u^{S}-s_F+\mu^{B}e)\,]_i\Big\}.
\end{aligned} \tag{9.14}
$$

The residuals of the unsymmetric path-following equations may be written as

$$
r = \begin{pmatrix} g - J^T y - z \\ y - w \\ c - s + \mu^{P}(y-y^{E}) \\ z_1\cdot(x-\ell^{X}) + \mu^{B}(z_1 - z_1^{E}) \\ z_2\cdot(u^{X}-x) + \mu^{B}(z_2 - z_2^{E}) \\ w_1\cdot(s-\ell^{S}) + \mu^{B}(w_1 - w_1^{E}) \\ w_2\cdot(u^{S}-s) + \mu^{B}(w_2 - w_2^{E}) \end{pmatrix} = \begin{pmatrix} g - J^T y - z \\ y - w \\ \mu^{P}(y-\pi^{Y}) \\ X_1^{\mu}(z_1-\pi_1^{Z}) \\ X_2^{\mu}(z_2-\pi_2^{Z}) \\ S_1^{\mu}(w_1-\pi_1^{W}) \\ S_2^{\mu}(w_2-\pi_2^{W}) \end{pmatrix},
$$

with $z = z_1 - z_2$ and $w = w_1 - w_2$.

10. General case: upper and lower bounds on some of the variables

Finally, we assume that the problem has nonlinear equality constraints $c(x) - s = 0$, where s is the vector of slack variables. In addition, it is assumed that a subset of the components of x and s are fixed and that a subset of the other components are subject to upper and lower bounds.

10.1. Problem statement and optimality conditions

The problem of interest has the form

$$\underset{x\in\mathbb{R}^n, s\in\mathbb{R}^m}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad \begin{cases} c(x) - s = 0, & L_X s = h_X, \quad \ell^S \le L_L s, \quad L_U s \le u^S, \\ Ax - b = 0, & E_X x = b_X, \quad \ell^X \le E_L x, \quad E_U x \le u^X, \end{cases} \tag{10.1}$$

where $c : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ are twice-continuously differentiable. 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 matrices formed from subsets 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 a variable x_j need not be subject to a lower or upper bound, or may be bounded below and above, in which case e_j 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. The shifted primal-dual penalty-barrier equations can be derived without this restriction, but the derivation is beyond the scope of this note. In addition, E_F and L_F denote rows of I_n and I_m such that $\begin{pmatrix} E_X^T & E_F^T \end{pmatrix}$ and $\begin{pmatrix} E_X^T & E_F^T \end{pmatrix}$ are column permutations of I_n and I_m. It follows that the rows of E_L and E_U are a subset of the rows of E_F, and that L_F is formed from the rows of L_L and L_U. The bound constraints involving E_X and L_X are enforced explicitly. The linear constraints $Ax - b = 0$ are imposed using the shifted primal-dual augmented Lagrangian.

The first-order KKT conditions for problem (10.1) are

$$\begin{aligned}
g(x^*) - J(x^*)^T y^* - A^T v^* - E_X^T z_X^* - E_L^T z_1^* + E_U^T z_2^* &= 0, & z_1^* &\ge 0, & z_2^* &\ge 0, && \text{(10.2a)}\\
y^* - L_X^T w_X^* - L_L^T w_1^* + L_U^T w_2^* &= 0, & w_1^* &\ge 0, & w_2^* &\ge 0, && \text{(10.2b)}\\
c(x^*) - s^* &= 0, & & & L_X s^* - h_X &= 0, && \text{(10.2c)}\\
& & Ax^* - b &= 0, & E_X x^* - b_X &= 0, && \text{(10.2d)}\\
E_L x^* - \ell^X &\ge 0, & u^X - E_U x^* &\ge 0, & & && \text{(10.2e)}\\
L_L s^* - \ell^S &\ge 0, & u^S - L_U s^* &\ge 0, & & && \text{(10.2f)}\\
z_1^* \cdot (E_L x^* - \ell^X) &= 0, & z_2^* \cdot (u^X - E_U x^*) &= 0, & & && \text{(10.2g)}\\
w_1^* \cdot (L_L s^* - \ell^S) &= 0, & w_2^* \cdot (u^S - L_U s^*) &= 0, & & && \text{(10.2h)}
\end{aligned}$$

where y^* are the multipliers for the equality constraints $c(x)-s=0$, 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 $Ax=b$. If $x_1 = E_L x - \ell^X$, $x_2 = u^X - E_U x$, $s_1 = L_L s - \ell^S$, and $s_2 = u^S - L_U s$, then z_1^*, z_2^*, w_1^*, and w_2^* 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

$$z = E_X^T z_X + E_L^T z_1 - E_U^T z_2, \quad \text{and} \quad w = L_X^T w_X + L_L^T w_1 - L_U^T w_2. \tag{10.3}$$

10.2. The path-following equations

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

$$\begin{aligned}
g(x) - J(x)^T y - A^T v - E_X^T z_X - E_L^T z_1 + E_U^T z_2 &= 0, & z_1 &\ge 0, & z_2 &\ge 0, & &(10.4a)\\
y - L_X^T w_X - L_L^T w_1 + L_U^T w_2 &= 0, & w_1 &\ge 0, & w_2 &\ge 0, & &(10.4b)\\
c(x) - s &= \mu^P(y^E - y), & E_X x^* - b_X &= 0, & L_X s^* - h_X &= 0, & &(10.4c)\\
Ax - b &= \mu^A(v^E - v), & & & & & &(10.4d)\\
E_L x - \ell^X &\ge 0, & u^X - E_U x &\ge 0, & & & &(10.4e)\\
L_L s - \ell^S &\ge 0, & u^S - L_U s &\ge 0, & & & &(10.4f)\\
z_1 \cdot (E_L x - \ell^X) &= \mu^B(z_1^E - z_1), & z_2 \cdot (u^X - E_U x) &= \mu^B(z_2^E - z_2), & & & &(10.4g)\\
w_1 \cdot (L_L s - \ell^S) &= \mu^B(w_1^E - w_1), & w_2 \cdot (u^S - L_U s) &= \mu^B(w_2^E - w_2), & & & &(10.4h)
\end{aligned}$$

Consider the primal-dual path-following equations $F(x,\, s,\, y,\, v,\, w_X,\, z_X,\, z_1,\, z_2,\, w_1,\, w_2\,;\mu^A,\, \mu^P,\, \mu^B,\, y^E,\, v^E,\, z_1^E,\, z_2^E,\, w_1^E,\, w_2^E) = 0$, with

$$F(x,s,y,v,z_1,z_2,w_1,w_2\,;\mu^P,\mu^B,y^E,v^E,z_1^E,z_2^E,w_1^E,w_2^E) = \begin{pmatrix} g(x) - J(x)^T y - A^T v - E_X^T z_X - E_L^T z_1 + E_U^T z_2 \\ y - L_X^T w_X - L_L^T w_1 + L_U^T w_2 \\ c(x) - s + \mu^P(y - y^E) \\ Ax - b + \mu^A(v - v^E) \\ E_X x - b_X \\ L_X s - h_X \\ z_1 \cdot (E_L x - \ell^X) + \mu^B(z_1 - z_1^E) \\ z_2 \cdot (u^X - E_U x) + \mu^B(z_2 - z_2^E) \\ w_1 \cdot (L_L s - \ell^S) + \mu^B(w_1 - w_1^E) \\ w_2 \cdot (u^S - L_U s) + \mu^B(w_2 - w_2^E) \end{pmatrix}. \tag{10.5}$$

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 (10.2), 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 (10.4) 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 (10.4) will approximate a solution of (10.2) 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 $\mu^B \to 0$).

10.3. A shifted primal-dual penalty-barrier function

Problem (10.1) is equivalent to

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

Consider the shifted primal-dual penalty-barrier problem

$$\begin{array}{lllllll}
\underset{x,x_1,x_2,s,s_1,s_2,y,v,z_1,z_2,w_1,w_2}{\text{minimize}} & \multicolumn{6}{l}{M(x, x_1, x_2, s, s_1, s_2, y, v, w_1, w_2\,; \mu^P, \mu^B, y^E, v^E, w_1^E, w_2^E)} \\
\text{subject to} & E_L x - x_1 = \ell^X, & L_L s - s_1 = \ell^S, & x_1 + \mu^B e > 0, & z_1 > 0, & s_1 + \mu^B e > 0, & w_1 > 0, \\
& E_U x + x_2 = u^X, & L_U s + s_2 = u^S, & x_2 + \mu^B e > 0, & z_2 > 0, & s_2 + \mu^B e > 0, & w_2 > 0, \\
& E_X x - b_X = 0, & L_X s - h_X = 0, & & & &
\end{array} \tag{10.6}$$

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

$$\begin{aligned} f(x) - \big(c(x) - s\big)^T y^E + \frac{1}{2\mu^P}\|c(x) - s\|^2 + \frac{1}{2\mu^P}\|c(x) - s + \mu^P(y - y^E)\|^2 \\ - (Ax - b)^T v^E + \frac{1}{2\mu^A}\|Ax - b\|^2 + \frac{1}{2\mu^A}\|Ax - b + \mu^A(v - v^E)\|^2 \\ - \sum_{j=1}^{n_F}\Big\{\mu^B[\,z_1^E\,]_j \ln\big([\,z_1\,]_j[\,x_1 + \mu^B e\,]_j^2\big) - [\,z_1 \cdot (x_1 + \mu^B e)\,]_j\Big\} \\ - \sum_{j=1}^{n_F}\Big\{\mu^B[\,z_2^E\,]_j \ln\big([\,z_2\,]_j[\,x_2 + \mu^B e\,]_j^2\big) - [\,z_2 \cdot (x_2 + \mu^B e)\,]_j\Big\} \\ - \sum_{i=1}^{m_F}\Big\{\mu^B[\,w_1^E\,]_i \ln\big([\,w_1\,]_i[\,s_1 + \mu^B e\,]_i^2\big) - [\,w_1 \cdot (s_1 + \mu^B e)\,]_i\Big\} \\ - \sum_{i=1}^{m_F}\Big\{\mu^B[\,w_2^E\,]_i \ln\big([\,w_2\,]_i[\,s_2 + \mu^B e\,]_i^2\big) - [\,w_2 \cdot (s_2 + \mu^B e)\,]_i\Big\}. \end{aligned} \tag{10.7}$$

The gradient of M may be defined in terms of the quantities $X_1^\mu = \mathrm{diag}(E_L x - \ell^X + \mu^B e)$, $X_2^\mu = \mathrm{diag}(u^X - E_U x + \mu^B e)$, $Z_1 = \mathrm{diag}(z_1)$, $Z_2 = \mathrm{diag}(z_2)$, $W_1 = \mathrm{diag}([w_1]_i)$, $W_2 = \mathrm{diag}([w_2]_i)$, $S_1^\mu = \mathrm{diag}(L_L s - \ell^S + \mu^B e)$ and $S_2^\mu = \mathrm{diag}(u^S - L_U s + \mu^B e)$. In particular,

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

The gradient may be written in several equivalent forms

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

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

where $D_Y = \mu^P I_m$, $D_A = \mu^A I_A$, $D_1^Z = X_1^\mu Z_1^{-1}$, $D_2^Z = X_2^\mu Z_2^{-1}$, $\pi_1^Z = \mu^B (X_1^\mu)^{-1} z_1^E$, and $\pi_2^Z = \mu^B (X_2^\mu)^{-1} z_2^E$.

10.4. Derivation of the shifted primal-dual penalty-barrier direction

10.5. Computation of the shifted primal-dual penalty-barrier direction

Next we consider the solution of the path-following Newton equations (10.5). If $v = (x,\, s,\, y,\, v,\, w_X,\, z_X,\, z_1,\, z_2,\, w_1,\, w_2)$ is a given approximate zero of $F(v)$ 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 $\Delta v = (\Delta x,\, \Delta s,\, \Delta y,\, \Delta v,\, \Delta w_X,\, \Delta z_X,\, \Delta z_1,\, \Delta z_2,\, \Delta w_1,\, \Delta w_2)$ are given by $F'(v)\Delta v = -F(v)$, where

$$F(v) = \begin{pmatrix} g(x) - J(x)^T y - A^T v - z \\ y - w \\ c(x) - s + \mu^P (y - y^E) \\ Ax - b + \mu^A (v - v^E) \\ L_X s - h_X \\ E_X x - b_X \\ z_1 \cdot (E_L x - \ell^X) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u^X - E_U x) + \mu^B (z_2 - z_2^E) \\ w_1 \cdot (L_L s - \ell^S) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u^S - L_U s) + \mu^B (w_2 - w_2^E) \end{pmatrix} \tag{10.8}$$

and

$$F'(v) = \begin{pmatrix} H & 0 & -J^T & -A^T & 0 & -E_X^T & -E_L^T & E_U^T & 0 & 0 \\ 0 & 0 & I_m & 0 & -L_X^T & 0 & 0 & 0 & -L_L^T & L_U^T \\ J & -I_m & D_Y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ A & 0 & 0 & D_A & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & L_F & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ E_X & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ Z_1 E_L & 0 & 0 & 0 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 E_U & 0 & 0 & 0 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 L_L & 0 & 0 & 0 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 L_U & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \tag{10.9}$$

(recall that $z = E_X^T z_X + E_L^T z_1 - E_U^T z_2$ and $w = L_X^T w_X + L_L^T w_1 - L_U^T w_2$. 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 (10.9) 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 (10.9) can be written in the reduced form $\widehat{F}' \Delta v_F = -\widehat{F}$, where $\Delta v_F = (\Delta x_F,\, \Delta s_F,\, \Delta y,\, \Delta v,\, \Delta z_1,\, \Delta z_2,\, \Delta w_1,\, \Delta w_2)$,

$$\begin{pmatrix} H_F & 0 & -J_F^T & -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 \\ 0 & 0 & L_F & 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T \\ J_F & -L_F^T & D_Y & 0 & 0 & 0 & 0 & 0 \\ A_F & 0 & 0 & D_A & 0 & 0 & 0 & 0 \\ Z_1 E_{LF} & 0 & 0 & 0 & X_1^\mu & 0 & 0 & 0 \\ -Z_2 E_{UF} & 0 & 0 & 0 & 0 & X_2^\mu & 0 & 0 \\ 0 & W_1 L_{LF} & 0 & 0 & 0 & 0 & S_1^\mu & 0 \\ 0 & -W_2 L_{UF} & 0 & 0 & 0 & 0 & 0 & S_2^\mu \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta s_F \\ \Delta y \\ \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \end{pmatrix} = - \begin{pmatrix} E_F\big(g - J^T y - A^T v - z\big) \\ L_F\big(y - w\big) \\ c(x) - s + \mu^P (y - y^E) \\ Ax - b + \mu^A (v - v^E) \\ z_1 \cdot (E_L x - \ell^X) + \mu^B (z_1 - z_1^E) \\ z_2 \cdot (u^X - E_U x) + \mu^B (z_2 - z_2^E) \\ w_1 \cdot (L_L s - \ell^S) + \mu^B (w_1 - w_1^E) \\ w_2 \cdot (u^S - L_U s) + \mu^B (w_2 - w_2^E) \end{pmatrix}, \quad (10.10)$$

where $H_F = E_F H E_F^T$, $J_F = J E_F^T$, $A_F = A E_F^T$, $g_F = E_F g$, $E_{LF} = E_L E_F^T$, $E_{UF} = E_U E_F^T$, $y_F = L_F y$, $L_{LF} = L_L L_F^T$ and $L_{UF} = L_U L_F^T$. The matrices J_F, A_F, E_{LF} and E_{UF} are the columns of J, 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. Given the definitions (10.3), 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 $\Delta z_X = [\, g + H\Delta x - J^T(y + \Delta y) - z \,]_X$. 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

$$\begin{pmatrix} D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0 \\ 0 & D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0 \\ 0 & 0 & D_2^Z & 0 & 0 & 0 & -E_{UF} & 0 \\ 0 & 0 & 0 & D_1^W & 0 & L_{LF} & 0 & 0 \\ 0 & 0 & 0 & 0 & D_2^W & -L_{UF} & 0 & 0 \\ 0 & 0 & 0 & -L_{LF}^T & L_{UF}^T & 0 & 0 & L_F \\ 0 & 0 & 0 & 0 & 0 & -L_F^T & J_F & D_Y \\ -A_F^T & -E_{LF}^T & E_{UF}^T & 0 & 0 & 0 & H_F & -J_F^T \end{pmatrix} \begin{pmatrix} \Delta v \\ \Delta z_1 \\ \Delta z_2 \\ \Delta w_1 \\ \Delta w_2 \\ \Delta s_F \\ \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} D_A(v - \pi^A) \\ D_1^Z(z_1 - \pi_1^Z) \\ D_2^Z(z_2 - \pi_2^Z) \\ D_1^W(w_1 - \pi_1^W) \\ D_2^W(w_2 - \pi_2^W) \\ L_F\big(y - w\big) \\ D_Y(y - \pi^Y) \\ E_F\big(g - J^T y - A^T v - z\big) \end{pmatrix}, \quad (10.11)$$

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

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

The diagonal matrix $L_F(L_L^T(D_1^W)^{-1}L_L + L_U^T(D_2^W)^{-1}L_U)L_F^T$ is nonsingular if every slack is either fixed or bounded above or below. If we define $D_W = \big(L_F(L_L^T(D_1^W)^{-1}L_L + L_U^T(D_2^W)^{-1}L_U)L_F^T\big)^{-1}$, then premultiplying the equations (10.11) by the matrix

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

gives the block upper-triangular system

$$
\begin{pmatrix}
D_A & 0 & 0 & 0 & 0 & 0 & A_F & 0\\
0 & D_1^Z & 0 & 0 & 0 & 0 & E_{LF} & 0\\
0 & 0 & D_2^Z & 0 & 0 & 0 & -E_{UF} & 0\\
0 & 0 & 0 & D_1^W & 0 & L_{LF} & 0 & 0\\
0 & 0 & 0 & 0 & D_2^W & -L_{UF} & 0 & 0\\
0 & 0 & 0 & 0 & 0 & D_W^{-1} & 0 & L_F\\
0 & 0 & 0 & 0 & 0 & 0 & J_F & \bar{D}_Y\\
0 & 0 & 0 & 0 & 0 & 0 & \bar{H}_F & -J_F^T
\end{pmatrix}
\begin{pmatrix}
\Delta v\\ \Delta z_1\\ \Delta z_2\\ \Delta w_1\\ \Delta w_2\\ \Delta s_F\\ \Delta x_F\\ \Delta y
\end{pmatrix}
= -
\begin{pmatrix}
D_A(v-\pi^A)\\
D_1^Z(z_1-\pi_1^Z)\\
D_2^Z(z_2-\pi_2^Z)\\
D_1^W(w_1-\pi_1^W)\\
D_2^W(w_2-\pi_2^W)\\
L_F(y-\pi^W)\\
L_F^T D_W L_F\big(y-\pi^W\big) + D_Y\big(y-\pi^Y\big)\\
E_F\big(g - J^T y - A^T\pi^A - \pi_F^Z\big)
\end{pmatrix},
$$

where $\bar{H}_F = E_F(H + A^T D_A^{-1} A + E_L^T(D_1^Z)^{-1}E_L + E_U^T(D_2^Z)^{-1}E_U)E_F^T$, $\bar{D}_Y = D_Y + L_F^T D_W L_F$, $\pi^W = L_L^T\pi_1^W - L_U^T\pi_2^W$ and $\pi_F^Z = E_L^T\pi_1^Z - E_U^T\pi_2^Z$. Using block back substitution, Δx_F and Δy can be computed by solving the equations

$$
\begin{pmatrix} H_F & -J_F^T\\ J_F & \bar{D}_Y \end{pmatrix}
\begin{pmatrix} \Delta x_F\\ \Delta y \end{pmatrix}
= -
\begin{pmatrix}
E_F\big(g - J^T y - A^T\pi^A - \pi_F^Z\big)\\
L_F^T D_W L_F\big(y-\pi^W\big) + D_Y\big(y-\pi^Y\big)
\end{pmatrix}. \tag{10.12}
$$

The full vector Δx is then 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

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

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 (10.10) 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 (10.10) 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 fourth and fifth block of equations of the Newton equations for a zero of (10.8) to solve for Δv gives $\Delta v = -(v - \widehat{\pi}^A)$, with $\widehat{\pi}^A = v^E - \frac{1}{\mu^A}\big(A(x+\Delta x) - b\big)$. Finally, the vectors Δw_X and Δz_X are recovered as $\Delta w_X = [\, y + \Delta y - w\,]_X$ and $\Delta z_X = [\, g + H\Delta x - J^T(y+\Delta y) - z\,]_X$.

10.6. Summary: computations associated with the general problem

The results of the preceding section implies that the solution of the path-following equations $F'(v)\Delta v = -F(v)$ with F and F' given by (10.8) and (10.9) 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 $X_1^\mu = \operatorname{diag}(E_L x - \ell^X + \mu^B e)$, $X_2^\mu = \operatorname{diag}(u^X - E_U x + \mu^B e)$, $Z_1 = \operatorname{diag}(z_1)$, $Z_2 = \operatorname{diag}(z_2)$, $W_1 = \operatorname{diag}([w_1]_i)$, $W_2 = \operatorname{diag}([w_2]_i)$, $S_1^\mu = \operatorname{diag}(L_L s - \ell^S + \mu^B e)$ and $S_2^\mu = \operatorname{diag}(u^S - L_U s + \mu^B e)$. Given the quantities

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

Solve the KKT system

$$\begin{pmatrix} H_F(x,y) + A_F^T D_A^{-1} A_F + D_Z^{-1} & -J_F(x)^T \\ J_F(x) & D_Y + L_F^T D_W L_F \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} E_F\big(g(x) - J(x)^T y - A^T \pi^A - \pi^Z\big) \\ L_F^T D_W L_F\big(y - \pi^W\big) + D_Y(y - \pi^Y)\big) \end{pmatrix}. \tag{10.13}$$

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

As $(x,s) \to (x^*, s^*)$ it holds that $\|D_Z^{-1}\|$ is 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 this system 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 $\widehat{D}_Z$ and $\widehat{D}_W$ denote diagonal matrices such that $\widehat{D}_Z^2 = (A_F^T D_A^{-1} A_F)^{-1}$ and $\widehat{D}_W^2 = (L_F^T D_W L_F)^{-1}$, then $\|\widehat{D}_Z\|$ and $\|\widehat{D}_W\|$ are bounded as $(x,s) \to (x^*, s^*)$. The equations (10.13) may be written in the form

$$\begin{pmatrix} \widehat{D}_Z H_F(x,y) \widehat{D}_Z + \widehat{D}_Z^2 D_Z^{-1} + I & -(\widehat{D}_W J_F(x) \widehat{D}_Z)^T \\ \widehat{D}_W J_F(x) \widehat{D}_Z & D_Y + L_F^T D_W L_F \end{pmatrix} \begin{pmatrix} \Delta x_F \\ \Delta y \end{pmatrix} = - \begin{pmatrix} \widehat{D}_Z E_F\big(g(x) - J(x)^T y - A^T \pi^A - \pi^Z\big) \\ \widehat{D}_W\big(L_F^T D_W L_F\big(y - \pi^W\big) + D_Y(y - \pi^Y)\big) \end{pmatrix}, \tag{10.14}$$

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

The associated line-search merit function (10.7) can be written as

$$
\begin{aligned}
f(x) - \big(c(x)-s\big)^T y^{\scriptscriptstyle E} + \frac{1}{2\mu^{\scriptscriptstyle P}}\|c(x)-s\|^2 + \frac{1}{2\mu^{\scriptscriptstyle P}}\|c(x)-s+\mu^{\scriptscriptstyle P}(y-y^{\scriptscriptstyle E})\|^2 \\
- (Ax-b)^T v^{\scriptscriptstyle E} + \frac{1}{2\mu^{\scriptscriptstyle A}}\|Ax-b\|^2 + \frac{1}{2\mu^{\scriptscriptstyle A}}\|Ax-b+\mu^{\scriptscriptstyle A}(v-v^{\scriptscriptstyle E})\|^2 \\
- \sum_{j=1}^{n_L}\Big\{\mu^{\scriptscriptstyle B}[\,z_1^{\scriptscriptstyle E}\,]_j \ln\big([\,z_1\,]_j[\,E_{\scriptscriptstyle L}x - \ell^{\scriptscriptstyle X} + \mu^{\scriptscriptstyle B}e\,]_j^2\big) - [\,z_1 \cdot (E_{\scriptscriptstyle L}x - \ell^{\scriptscriptstyle X} + \mu^{\scriptscriptstyle B}e)\,]_j\Big\} \\
- \sum_{j=1}^{n_U}\Big\{\mu^{\scriptscriptstyle B}[\,z_2^{\scriptscriptstyle E}\,]_j \ln\big([\,z_2\,]_j[\,u^{\scriptscriptstyle X} - E_{\scriptscriptstyle U}x + \mu^{\scriptscriptstyle B}e\,]_j^2\big) - [\,z_2 \cdot (u^{\scriptscriptstyle X} - E_{\scriptscriptstyle U}x + \mu^{\scriptscriptstyle B}e)\,]_j\Big\} \\
- \sum_{i=1}^{m_L}\Big\{\mu^{\scriptscriptstyle B}[\,w_1^{\scriptscriptstyle E}\,]_i \ln\big([\,w_1\,]_i[\,L_{\scriptscriptstyle L}s - \ell^{\scriptscriptstyle S} + \mu^{\scriptscriptstyle B}e\,]_i^2\big) - [\,w_1 \cdot (L_{\scriptscriptstyle L}s - \ell^{\scriptscriptstyle S} + \mu^{\scriptscriptstyle B}e)\,]_i\Big\} \\
- \sum_{i=1}^{m_U}\Big\{\mu^{\scriptscriptstyle B}[\,w_2^{\scriptscriptstyle E}\,]_i \ln\big([\,w_2\,]_i[\,u^{\scriptscriptstyle S} - L_{\scriptscriptstyle U}s + \mu^{\scriptscriptstyle B}e\,]_i^2\big) - [\,w_2 \cdot (u^{\scriptscriptstyle S} - L_{\scriptscriptstyle U}s + \mu^{\scriptscriptstyle B}e)\,]_i\Big\}.
\end{aligned} \tag{10.15}
$$

References

[1] P. E. Gill, V. Kungurtsev, and D. P. Robinson. A shifted primal-dual penalty-barrier method for nonlinear optimization. Center for Computational Mathematics Report CCoM 19-03, University of California, San Diego, 2019. 2