Horizontal and Vertical Decomposition in Interior Point Methods for Linear Programs*

Masakazu Kojima† Nimrod Megiddo‡ Shinji Mizuno§ Susumu Shindoh¶

Abstract. Corresponding to the linear program:

$$\text{Maximize } \boldsymbol{c}^T\boldsymbol{x} \ \text{ subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{a},\ \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0},$$

we introduce two functions in the penalty parameter $t > 0$ and the Lagrange relaxation parameter vector $\boldsymbol{w}$,

$$\begin{aligned} \tilde{f}^p(t, \boldsymbol{w}) &= \max\{\boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{w}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{a}) + t\sum_{j=1}^{n} \ln x_j : \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} > \boldsymbol{0}\} \\ &\quad \text{(for horizontal decomposition)}, \\ \tilde{f}^d(t, \boldsymbol{w}) &= \min\{\boldsymbol{a}^T\boldsymbol{w} + \boldsymbol{b}^T\boldsymbol{y} - t\sum_{j=1}^{n} \ln z_j : \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{w},\ \boldsymbol{z} > \boldsymbol{0}\} \\ &\quad \text{(for vertical decomposition)}. \end{aligned}$$

For each $t > 0$, $\tilde{f}^p(t, \cdot)$ and $\tilde{f}^d(t, \cdot)$ are strictly convex C^∞ functions with a common minimizer $\hat{\boldsymbol{w}}(t)$, which converges to an optimal Lagrange multiplier vector $\boldsymbol{w}^*$ associated with the constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}$ as $t \to 0$, and enjoy the strong self-concordance property given by Nesterov and Nemirovsky. Based on these facts, we present conceptual algorithms with the use of Newton's method for tracing the trajectory $\{\hat{\boldsymbol{w}}(t) : t > 0\}$, and analyze their computational complexity.

1. Introduction.

This paper presents a theoretical framework for incorporating horizontal and vertical decomposition techniques into interior-point methods (see for example the survey paper [13])

*Research supported in part by ONR Contract N00014-94-C-0007

†Department of Information Sciences, Tokyo Institute of Technology, Oh-Okayama, Meguro-ku, Tokyo 152, Japan.

‡IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120, and School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel

§The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan.

¶Department of Mathematics and Physics, The National Defense Academy, Hashirimizu 1-10-20, Yokosuka, Kanagawa, 239, Japan.

for linear programs. Let $\boldsymbol{a} \in R^m$, $\boldsymbol{b} \in R^k$, $\boldsymbol{c} \in R^n$, $\boldsymbol{A} \in R^{m \times n}$ and $\boldsymbol{B} \in R^{k \times n}$. We consider the following equality form linear program $\mathcal{P}^*$ and its dual $\mathcal{D}^*$:

$$\begin{array}{llll} \mathcal{P}^* & \text{Maximize} & \boldsymbol{c}^T\boldsymbol{x} & \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{a},\ \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}. \\ \mathcal{D}^* & \text{Minimize} & \boldsymbol{a}^T\boldsymbol{w} + \boldsymbol{b}^T\boldsymbol{y} & \text{subject to} \quad \boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c},\ \boldsymbol{z} \geq \boldsymbol{0}. \end{array}$$

Let

$$\begin{array}{rcl} R_{++} & = & \{t \in R : t > 0\ \} \text{ (the set of positive numbers)}, \\ R^n_{++} & = & \{\boldsymbol{x} \in R^n : \boldsymbol{x} > \boldsymbol{0}\ \} \text{ (the positive orthant)}, \\ P_{++} & = & \{\boldsymbol{x} \in R^n_{++} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{a},\ \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b}\ \} \\ & & \text{(the interior of the feasible region of } \mathcal{P}^*), \\ D_{++} & = & \{(\boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) \in R^m \times R^k \times R^n_{++} : \boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c}\ \} \\ & & \text{(the interior of the feasible region of } \mathcal{D}^*), \\ Q_{++} & = & \{\boldsymbol{x} \in R^n_{++} : \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b}\ \} \\ D_{++}(\boldsymbol{w}) & = & \{(\boldsymbol{y}, \boldsymbol{z}) \in R^k \times R^n_{++} : \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{w}\ \}. \end{array}$$

Then we obviously see that

$$\begin{array}{rcl} P_{++} & = & \{\boldsymbol{x} \in R^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}\} \bigcap Q_{++}, \\ D_{++} & = & \bigcup_{w \in R^m} \{(\boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) \in R^{m+k+n} : (\boldsymbol{y}, \boldsymbol{z}) \in D_{++}(\boldsymbol{w})\}. \end{array}$$

Throughout the paper we impose the following assumptions on the constraint set of $\mathcal{P}^*$:

(A) P_{++} is nonempty and bounded.

(B) Q_{++} is nonempty and bounded.

(C) The $(m+k) \times n$ matrix $\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix}$ has full row rank.

For every $(t, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) \in R_{++} \times R^n_{++} \times R^m \times R^k \times R^n_{++}$, define

$$\begin{array}{rcl} f^p(t, \boldsymbol{w}, \boldsymbol{x}) & = & \boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{w}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{a}) + t\sum_{j=1}^n \ln x_j \\ f^d(t, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) & = & \boldsymbol{a}^T\boldsymbol{w} + \boldsymbol{b}^T\boldsymbol{y} - t\sum_{j=1}^n \ln z_j\ . \end{array}$$

Here $t \in R_{++}$ denotes the barrier parameter and $\boldsymbol{w} \in R^m$ the Lagrange multiplier vector associated with the constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}$.

We can regard the horizontal (or vertical) decomposition technique as a numerical tracing of the trajectory which consists of the solution $\hat{\boldsymbol{w}}(t)$ ($t \in R_{++}$) of the following parametric min-max (or min-min) problem:

$$\text{minimize}_{\,w \in R^m} \quad \text{maximize}\ \{f^p(t, \boldsymbol{w}, \boldsymbol{x}) : \boldsymbol{x} \in Q_{++}\}, \tag{1}$$

$$\text{minimize}_{\,w \in R^m} \quad \text{minimize}\ \{f^d(t, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) : (\boldsymbol{y}, \boldsymbol{z}) \in D_{++}(\boldsymbol{w})\}. \tag{2}$$

Under the assumptions (A), (B) and (C), the inner maximization problem in (1) (or the inner minimization problem in (2)) has a unique solution for every $(t, \boldsymbol{w}) \in R_{++} \times R^m$, so that we can consistently define the optimal value functions and the optimizers of the inner problems; for every $(t, \boldsymbol{w}) \in R_{++} \times R^m$, let

$$\begin{aligned}
\tilde{f}^p(t, \boldsymbol{w}) &= \max\{f^p(t, \boldsymbol{w}, \boldsymbol{x}) : \boldsymbol{x} \in Q_{++}\}, \\
\tilde{\boldsymbol{x}}(t, \boldsymbol{w}) &= \arg\max \{f^p(t, \boldsymbol{w}, \boldsymbol{x}) : \boldsymbol{x} \in Q_{++}\}, \\
\tilde{f}^d(t, \boldsymbol{w}) &= \min\{f^d(t, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) : (\boldsymbol{y}, \boldsymbol{z}) \in D_{++}(\boldsymbol{w})\}, \\
(\tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w})) &= \arg\min \{f^d(t, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) : (\boldsymbol{y}, \boldsymbol{z}) \in D_{++}(\boldsymbol{w})\}.
\end{aligned}$$

We are mainly concerned with theoretical aspects of the horizontal and vertical decomposition techniques. In particular we show the following features.

(a) For every $t \in R_{++}$, the function $\tilde{f}^p(t, \cdot) : R^m \to R$ is a strictly convex C^∞ function with the unique minimizer $\hat{\boldsymbol{w}}(t)$ over R^m.

(b) For every $(t, \boldsymbol{w}) \in R_{++} \times R^m$, we can use the primal-dual interior-point method ([5, 7, 8, 9, 11], etc.) to compute $(\tilde{\boldsymbol{x}}(t, \boldsymbol{w}), \tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$, $\tilde{f}^p(t, \boldsymbol{w})$, the gradient vector $\tilde{f}^p_w(t, \boldsymbol{w})$ w.r.t. $\boldsymbol{w}$ and the Hessian matrix $\tilde{f}^p_{ww}(t, \boldsymbol{w})$ w.r.t. $\boldsymbol{w}$.

(c) The set $\{(\tilde{\boldsymbol{x}}(t, \hat{\boldsymbol{w}}(t)), \hat{\boldsymbol{w}}(t), \tilde{\boldsymbol{y}}(t, \hat{\boldsymbol{w}}(t)), \tilde{\boldsymbol{z}}(t, \hat{\boldsymbol{w}}(t))) : t \in R_{++} \}$ forms the central trajectory [9] converging to the analytic center of the optimal solution set of the primal-dual pair of linear programs $\mathcal{P}^*$ and $\mathcal{D}^*$.

(d) $\{\tilde{f}^p(t, \cdot) : t \in R_{++} \}$ is a strongly self-concordant family [10] (with the parameter functions $\alpha(t) = t,\ \gamma(t) = \mu(t) = 1,\ \xi(t) = \sqrt{n}/t,\ \eta(t) = \sqrt{n}/(2t)$).

(e) $\tilde{f}^d(t, \boldsymbol{w}) - \tilde{f}^p(t, \boldsymbol{w}) = nt(1 - \ln t)$ for every $(t, \boldsymbol{w}) \in R_{++} \times R^m$. Note that the right hand side is independent of $\boldsymbol{w}$. Hence (a), (b), (c) and (d) above remain valid even if we replace $\tilde{f}^p$ by $\tilde{f}^d$.

In view of the features (a), (b), (c) and (e) above, we obtain approximate optimal solutions of $\mathcal{P}^*$ and $\mathcal{D}^*$ if we numerically trace the trajectory $\{\hat{\boldsymbol{w}}(t) : t > 0\}$ in the $\boldsymbol{w}$-space until the barrier parameter $t \in R_{++}$ gets sufficiently small. It should be emphasized that $\hat{\boldsymbol{w}}(t)$ is the common minimizer of strictly convex C^∞ functions $\tilde{f}^p(t, \cdot)$ and $\tilde{f}^d(t, \cdot)$ over the entire space R^m. Thus we can utilize various unconstrained minimization techniques such as Newton's, quasi-Newton and conjugate gradient methods to approximate $\hat{\boldsymbol{w}}(t)$.

The feature (d) is a major theoretical contribution of this paper, which makes it possible to effectively utilize Newton's method for tracing the trajectory $\{\hat{\boldsymbol{w}}(t) : t > 0\}$. The notion and the theory of self-concordance were given by Nesterov and Nemirovsky [10] for a wide class of polynomial-time interior-point methods for convex programming.

After listing in Section 2 symbols and notation used throughout the paper, we show in Section 3 that the inner optimization problems in (1) and (2) share a common necessary and sufficient optimality condition, and derive the feature (e) from the condition.

Section 4 describes several derivatives of the function $\tilde{f}^p : R_{++} \times R^m \rightarrow R$ including the gradient $\tilde{f}^p_w(t, \boldsymbol{w})$ and the positive definite Hessian matrix $\tilde{f}^p_{ww}(t, \boldsymbol{w})$ w.r.t. $\boldsymbol{w} \in R^m$. Details of calculation of derivatives are given in Appendix. Section 5 establishes that the family $\{\tilde{f}^p(t, \cdot) : t \in R_{++}\}$ is strongly self-concordant. In Section 6 we present conceptual decomposition algorithms with the use of Newton's method, and investigate their polynomial-time computational complexity.

Although we will state a horizontal decomposition for the general linear program $\mathcal{P}^*$ and a vertical decomposition mainly for its dual $\mathcal{D}^*$, it is interesting in practice to apply them to a special case where the matrix $\boldsymbol{B}$ has a block diagonal structure:

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_1 & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}_2 & \cdots & \boldsymbol{O} \\ & & \cdots & \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{B}_\ell \end{pmatrix}. \tag{3}$$

Our horizontal and vertical decomposition techniques applied to such special cases are roughly corresponding to the Dantzig-Wolfe and the Benders decomposition methods, respectively. See for example the book [6]. It should be noted that when the matrix $\boldsymbol{B}$ has a block diagonal structure as in (3), we can decompose the inner maximization problem in (1) (or the inner minimization problem in (2)) into ℓ smaller subproblems of the same type.

Many studies ([1, 2, 12], etc.) have been done on how we exploit special structures such as generalized upper bounds, block diagonal structures and staircase structures in interior-point methods. The conceptual algorithms given in Section 6 has a close relation with a compact-inverse implementation (see Section 9 of [2]) of the primal-dual interior-point method applied to $\mathcal{P}^*$ and $\mathcal{D}^*$. This is shown in Section 7 where we discuss some issues toward implementation of the horizontal and vertical decomposition techniques.

2. Notation.

$$\begin{aligned}
R_{++} &= \{t \in R : t > 0\}, \quad R^n_{++} = \{\boldsymbol{x} \in R^n : \boldsymbol{x} > \boldsymbol{0}\}, \\
P_{++} &= \{\boldsymbol{x} \in R^n_{++} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{a},\ \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b}\}, \\
D_{++} &= \{(\boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) \in R^m \times R^k \times R^n_{++} : \boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c}\}, \\
Q_{++} &= \{\boldsymbol{x} \in R^n_{++} : \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b}\}, \\
D_{++}(\boldsymbol{w}) &= \{(\boldsymbol{y}, \boldsymbol{z}) \in R^k \times R^n_{++} : \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{w}\}, \\
f^p(t, \boldsymbol{w}, \boldsymbol{x}) &= \boldsymbol{c}^T\boldsymbol{x} - \boldsymbol{w}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{a}) + t\sum_{j=1}^n \ln x_j, \\
f^d(t, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) &= \boldsymbol{a}^T\boldsymbol{w} + \boldsymbol{b}^T\boldsymbol{y} - t\sum_{j=1}^n \ln z_j,
\end{aligned}$$

$$
\begin{aligned}
\tilde{f}^p(t,\boldsymbol{w}) &= \max\{f^p(t,\boldsymbol{w},\boldsymbol{x}) : \boldsymbol{x} \in Q_{++}\},\\
\tilde{\boldsymbol{x}}(t,\boldsymbol{w}) &= \arg\max\ \{f^p(t,\boldsymbol{w},\boldsymbol{x}) : \boldsymbol{x} \in Q_{++}\},\\
\tilde{f}^d(t,\boldsymbol{w}) &= \min\{f^d(t,\boldsymbol{w},\boldsymbol{y},\boldsymbol{z}) : (\boldsymbol{y},\boldsymbol{z}) \in D_{++}(\boldsymbol{w})\},\\
(\tilde{\boldsymbol{y}}(t,\boldsymbol{w}),\tilde{\boldsymbol{z}}(t,\boldsymbol{w})) &= \arg\min\ \{f^d(t,\boldsymbol{w},\boldsymbol{y},\boldsymbol{z}) : (\boldsymbol{y},\boldsymbol{z}) \in D_{++}(\boldsymbol{w})\},\\
\tilde{\boldsymbol{X}} &= \tilde{\boldsymbol{X}}(t,\boldsymbol{w}) = \mathrm{diag}\ \tilde{\boldsymbol{x}}(t,\boldsymbol{w}), \quad \tilde{\boldsymbol{Z}} \ = \ \tilde{\boldsymbol{Z}}(t,\boldsymbol{w}) = \mathrm{diag}\ \tilde{\boldsymbol{z}}(t,\boldsymbol{w}),\\
\tilde{\Delta} &= \tilde{\boldsymbol{X}}^{1/2}\tilde{\boldsymbol{Z}}^{-1/2},\\
\hat{\boldsymbol{w}}(t) &= \arg\min\ \{\tilde{f}^p(t,\boldsymbol{w}) : \boldsymbol{w} \in R^m\ \},\\
\lambda_* &= 2-3^{1/2} = 0.2679\ldots\ , \quad \theta_* \ = \ 1/4 = 0.25\ .
\end{aligned}
$$

3. Duality between $\tilde{f}^p$ and $\tilde{f}^d$.

Given $t > 0$ and $\boldsymbol{w} \in R^m$, the term $\boldsymbol{a}^T\boldsymbol{w}$ involved in the objective functions $f^p(t,\boldsymbol{w},\cdot)$ of the inner problem in (1) and $f^d(t,\boldsymbol{w},\cdot,\cdot)$ of the inner problem in (2) is constant. Hence we may eliminate it from the objective functions when we are concerned with the inner problems in (1) and (2). Thus the pair of the inner problems is equivalent to the pair:

$$
\begin{array}{llll}
\mathcal{P}(t,\boldsymbol{w}) & \text{Maximize} & (\boldsymbol{c}-\boldsymbol{A}^T\boldsymbol{w})^T\boldsymbol{x} + t\sum_{j=1}^n \ln x_j & \text{subject to} \quad \boldsymbol{x} \in Q_{++}.\\
\mathcal{D}(t,\boldsymbol{w}) & \text{Minimize} & \boldsymbol{b}^T\boldsymbol{y} - t\sum_{j=1}^n \ln z_j & \text{subject to} \quad (\boldsymbol{y},\boldsymbol{z}) \in D_{++}(\boldsymbol{w}).
\end{array}
$$

This pair has exactly the same structure as the one often studied with the primal-dual interior-point method ([5, 7, 8, 9, 11], etc.). Therefore we can apply the primal-dual interior-point method to the pair of problems above. By using the well-known argument (see for example the paper [9]) we obtain the theorem below that states a common necessary and sufficient optimality condition for the inner maximization problem in (1) (*i.e.,* $\mathcal{P}(t,\boldsymbol{w})$) and the inner minimization problem in (2) (*i.e.,* $\mathcal{D}(t,\boldsymbol{w})$).

Theorem 3.1. *Let* $(t,\boldsymbol{w}) \in R_{++} \times R^m$. *Then* $\boldsymbol{x} \in R^n$ *is a maximizer of* $f^p(t,\boldsymbol{w},\cdot)$ *over* Q_{++} *and* $(\boldsymbol{y},\boldsymbol{z}) \in R^{k+n}$ *is a minimizer of* $f^d(t,\boldsymbol{w},\cdot,\cdot)$ *over* $D_{++}(\boldsymbol{w})$ *if and only if*

$$
\left.
\begin{array}{l}
\boldsymbol{Bx}-\boldsymbol{b}=\boldsymbol{0},\ \boldsymbol{x}>\boldsymbol{0},\\
\boldsymbol{A}^T\boldsymbol{w}+\boldsymbol{B}^T\boldsymbol{y}-\boldsymbol{z}-\boldsymbol{c}=\boldsymbol{0},\ \boldsymbol{z}>\boldsymbol{0},\\
\boldsymbol{Xz}-t\boldsymbol{e}=\boldsymbol{0}.
\end{array}
\right\} \tag{4}
$$

Here $\boldsymbol{X}$ *denotes the diagonal matrix of the coordinates of* $\boldsymbol{x} \in R^n$ *and* $\boldsymbol{e} = (1,\ldots,1)^T \in R^n$.

Let $(t,\boldsymbol{w}) \in R_{++} \times R^m$. Suppose that $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$ satisfies (4). Then $(\boldsymbol{w},\boldsymbol{y},\boldsymbol{z})$ is a feasible solution of the original dual linear program $\mathcal{D}^*$. Hence $\boldsymbol{a}^T\boldsymbol{w}+\boldsymbol{b}^T\boldsymbol{y}$ gives an upper bound of the maximal objective value of the original primal linear program $\mathcal{P}^*$ to be solved. On the other hand, $\boldsymbol{x}$ is not a feasible solution of $\mathcal{P}^*$ unless the additional equality $\boldsymbol{Ax}-\boldsymbol{a}=\boldsymbol{0}$

holds. It will be shown in Section 4 that the equality holds when and only when $\boldsymbol{w}$ is a minimizer of the outer problem in (1), *i.e.*, a minimizer of $\tilde{f}^p(t, \cdot)$ over R^m.

Recall that $\tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ and $(\tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$ denote the maximizer of the inner problem in (1) and the minimizer of the inner problem in (2), respectively. For every fixed $\boldsymbol{w} \in R^m$, the set $\{(\tilde{\boldsymbol{x}}(t, \boldsymbol{w}), \tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w})) : t \in R_{++}\}$ forms the central trajectory of the primal-dual pair of linear programs:

$$\begin{array}{llll} \mathcal{P}(\boldsymbol{w}) & \text{Maximize} & (\boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{w})^T\boldsymbol{x} & \text{subject to} \quad \boldsymbol{B}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0}. \\ \mathcal{D}(\boldsymbol{w}) & \text{Minimize} & \boldsymbol{b}^T\boldsymbol{y} & \text{subject to} \quad \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{w},\ \boldsymbol{z} \geq \boldsymbol{0}. \end{array}$$

Theorem 3.1 implies that the relations

$$\left.\begin{array}{l} \boldsymbol{B}\tilde{\boldsymbol{x}}(t, \boldsymbol{w}) - \boldsymbol{b} = \boldsymbol{0},\ \tilde{\boldsymbol{x}}(t, \boldsymbol{w}) > \boldsymbol{0}, \\ \boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\tilde{\boldsymbol{y}}(t, \boldsymbol{w}) - \tilde{\boldsymbol{z}}(t, \boldsymbol{w}) - \boldsymbol{c} = \boldsymbol{0},\ \tilde{\boldsymbol{z}}(t, \boldsymbol{w}) > \boldsymbol{0}, \\ \tilde{\boldsymbol{X}}(t, \boldsymbol{w})\tilde{\boldsymbol{z}}(t, \boldsymbol{w}) - t\boldsymbol{e} = \boldsymbol{0} \end{array}\right\} \tag{5}$$

hold for every $(t, \boldsymbol{w}) \in R_{++} \times R^m$. It follows that

$$\begin{aligned} & \tilde{f}^d(t, \boldsymbol{w}) - \tilde{f}^p(t, \boldsymbol{w}) \\ = \;& \boldsymbol{b}^T\tilde{\boldsymbol{y}}(t, \boldsymbol{w}) - \boldsymbol{c}^T\tilde{\boldsymbol{x}}(t, \boldsymbol{w}) + \boldsymbol{w}^T\boldsymbol{A}\tilde{\boldsymbol{x}}(t, \boldsymbol{w}) - t\sum_{j=1}^n \ln \tilde{x}_j(t, \boldsymbol{w})\tilde{z}_j(t, \boldsymbol{w}) \\ = \;& \left(\boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\tilde{\boldsymbol{y}}(t, \boldsymbol{w}) - \boldsymbol{c}\right)^T \tilde{\boldsymbol{x}}(\boldsymbol{t}, \boldsymbol{w}) - t\sum_{j=1}^n \ln t \\ = \;& \tilde{\boldsymbol{x}}(t, \boldsymbol{w})^T\tilde{\boldsymbol{z}}(t, \boldsymbol{w}) - nt\ln t \;=\; nt(1 - \ln t). \end{aligned}$$

Therefore we obtain:

Theorem 3.2. $\tilde{f}^d(t, \boldsymbol{w}) - \tilde{f}^p(t, \boldsymbol{w}) = nt(1 - \ln t)$ *for every* $(t, \boldsymbol{w}) \in R_{++} \times R^m$.

4. Derivatives.

In this section, we present various derivatives of $\tilde{f}^p : R_{++} \times R^m \to R$. Details of calculation of derivatives are omitted here but are given in Appendix.

Theorem 4.1.

(i) $\tilde{f}^p_w(t, \boldsymbol{w}) = \boldsymbol{a} - \boldsymbol{A}\tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ *for every* $(t, \boldsymbol{w}) \in R_{++} \times R^m$.

(ii) $$\begin{aligned} \tilde{f}^p_{ww}(t, \boldsymbol{w}) &= \boldsymbol{A}\tilde{\Delta}\left(\boldsymbol{I} - \tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)\tilde{\Delta}\boldsymbol{A}^T \\ &= \frac{1}{t}\boldsymbol{A}\tilde{\boldsymbol{X}}\left(\boldsymbol{I} - \tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T \end{aligned}$$
for every $(t, \boldsymbol{w}) \in R_{++} \times R^m$.

Here $\tilde{\Delta} = \tilde{\boldsymbol{X}}^{1/2}\tilde{\boldsymbol{Z}}^{-1/2}$, $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{X}}(t, \boldsymbol{w}) =$ *diag* $\tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ *and* $\tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{Z}}(t, \boldsymbol{w}) =$ *diag* $\tilde{\boldsymbol{z}}(t, \boldsymbol{w})$.

As a corollary we obtain:

Theorem 4.2. *Let* $t \in R_{++}$ *be fixed arbitrarily.*

(i) *The Hessian matrix* $\tilde{f}^p_{ww}(t, \boldsymbol{w})$ *is positive definite for every* $\boldsymbol{w} \in R^m$.

(ii) *The function* $\tilde{f}^p(t, \cdot) : R^m \to R$ *is a strictly convex* C^∞ *function with the unique minimizer over* R^m.

Proof: (i) Obviously the Hessian matrix

$$\tilde{f}^p_{ww}(t, \boldsymbol{w}) = \boldsymbol{A}\tilde{\Delta}\left(\boldsymbol{I} - \tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)\tilde{\Delta}\boldsymbol{A}^T$$

is positive semi-definite. Hence it suffices to show that the matrix is nonsingular. Assume on the contrary that

$$\boldsymbol{A}\tilde{\Delta}\left(\boldsymbol{I} - \tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)\tilde{\Delta}\boldsymbol{A}^T\boldsymbol{u} = \boldsymbol{0}$$

for some nonzero $\boldsymbol{u} \in R^m$. Since $\left(\boldsymbol{I} - \tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)$ is a projection matrix, we have

$$\left(\boldsymbol{I} - \tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)\tilde{\Delta}\boldsymbol{A}^T\boldsymbol{u} = \boldsymbol{0},$$

which implies that

$$\boldsymbol{A}^T\boldsymbol{u} - \boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{A}^T\boldsymbol{u} = \boldsymbol{0}.$$

This contradicts the assumption (C).

(ii) The strict convexity of $\tilde{f}^p(t, \cdot)$ follows directly from (i). We have observed that the mappings $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$, $\tilde{\boldsymbol{z}}$ on $R_{++} \times R^m$ are defined implicitly through the equalities in (4) (see (5)). Since the Jacobian matrix of the left hand side of the equalities with respect to $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is nonsingular at every $(t, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in R_{++} \times R^m \times R^n_{++} \times R^k \times R^n_{++}$ and the left hand side is C^∞ differentiable with respect to $(t, \boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, we see by the implicit function theorem (for example [3]) that the mappings $\tilde{\boldsymbol{x}}$, $\tilde{\boldsymbol{y}}$, $\tilde{\boldsymbol{z}}$ are C^∞ differentiable with respect to $(t, \boldsymbol{w})$. It is well-known [9] that under the assumptions (A)and (C), there exists a unique $(\boldsymbol{w}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ for which

$$\left.\begin{array}{l} \boldsymbol{A}\boldsymbol{x} - \boldsymbol{a} = \boldsymbol{0},\ \boldsymbol{B}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0},\ \boldsymbol{x} > \boldsymbol{0}, \\ \boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} - \boldsymbol{c} = \boldsymbol{0},\ \boldsymbol{z} > \boldsymbol{0}, \\ \boldsymbol{X}\boldsymbol{z} - t\boldsymbol{e} = \boldsymbol{0} \end{array}\right\} \tag{6}$$

hold. We know by Theorems 3.1 and 4.1 that (6) is a necessary and sufficient condition for $\boldsymbol{w} \in R^m$ to be a minimizer of $\tilde{f}^p(t, \cdot)$ over R^m. This ensures the existence and the unique minimizer of $\tilde{f}^p(t, \cdot)$. ∎

For every $t \in R_{++}$, define $\hat{\boldsymbol{w}}(t)$ to be the minimizer of $\tilde{f}^p(t,\cdot)$ over R^m whose existence and uniqueness have been shown above; *i.e.*, $\hat{\boldsymbol{w}}(t) = \arg\min \{\tilde{f}^p(t,\boldsymbol{w}) : \boldsymbol{w} \in R^m\}$. In view of Theorem 3.2, we see that $\hat{\boldsymbol{w}}(t) = \arg\min \{\tilde{f}^d(t,\boldsymbol{w}) : \boldsymbol{w} \in R^m\}$. Furthermore, for every $t \in R_{++}$, $(\boldsymbol{x},\boldsymbol{w},\boldsymbol{y},\boldsymbol{z}) = (\tilde{\boldsymbol{x}}(t,\hat{\boldsymbol{w}}(t)),\hat{\boldsymbol{w}}(t),\tilde{\boldsymbol{y}}(t,\hat{\boldsymbol{w}}(t)),\tilde{\boldsymbol{z}}(t,\hat{\boldsymbol{w}}(t)))$ satisfies (6). Thus the set $\{(\tilde{\boldsymbol{x}}(t,\hat{\boldsymbol{w}}(t)),\hat{\boldsymbol{w}}(t),\tilde{\boldsymbol{y}}(t,\hat{\boldsymbol{w}}(t)),\tilde{\boldsymbol{z}}(t,\hat{\boldsymbol{w}}(t))) : t \in R_{++}\}$ forms the central trajectory converging to the analytic center of the optimal solution set of the primal-dual pair of linear programs $\mathcal{P}^*$ and $\mathcal{D}^*$.

The theorem below will be utilized in Section 5 where we discuss the self-concordant property of the family $\{\tilde{f}^p(t,\cdot) : t \in R_{++}\}$ of functions.

Theorem 4.3. *Let $\boldsymbol{w} \in R^m$ and $\boldsymbol{h} \in R^m$ be fixed arbitrarily. For every $(t,s) \in R_{++} \times R$, let*

$$\begin{aligned}
\boldsymbol{X}(t,s) &= \tilde{\boldsymbol{X}}(t,\boldsymbol{w}+s\boldsymbol{h}),\\
\boldsymbol{u}(t,s) &= \left(\boldsymbol{I} - \boldsymbol{X}(t,s)\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{X}(t,s)^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\boldsymbol{X}(t,s)\right)\boldsymbol{X}(t,s)\boldsymbol{A}^T\boldsymbol{h},\\
\boldsymbol{v}(t,s) &= \left(\boldsymbol{I} - \boldsymbol{X}(t,s)\boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{X}(t,s)^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\boldsymbol{X}(t,s)\right)\boldsymbol{e},
\end{aligned}$$

Then the following (i) – (v) hold for every $s \in R$.

(i) $\dfrac{d\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds} = (\boldsymbol{a} - \boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w}+s\boldsymbol{h}))^T\boldsymbol{h}$

(ii) $\dfrac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^2} = \dfrac{1}{t}\displaystyle\sum_{j=1}^n u_j(t,s)^2.$

(iii) $\dfrac{d^3\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^3} = -\dfrac{2}{t^2}\displaystyle\sum_{j=1}^n u_j(t,s)^3.$

(iv) $\dfrac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{dtds} = -\dfrac{\boldsymbol{u}(t,s)^T\boldsymbol{v}(t,s)}{t}.$

(v) $\dfrac{d^3\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{dtds^2} = \dfrac{\boldsymbol{u}(t,s)^T\left(2\,diag\ \boldsymbol{v}(t,s) - \boldsymbol{I}\right)\boldsymbol{u}(t,s)}{t^2}.$

5. Self-Concordance.

In this section, we apply the notion and the theory of self-concordance to the family $\{\tilde{f}^p(t,\cdot) : t \in R_{++}\}$ of functions according to the paper [10] by Nesterov and Nemirovsky. To avoid complicated notation and discussion, however, we employ a simplified version of the self-concordance. Our definition of the "global self-concordance" and related properties shown below are less general than those of the original strong self-concordance given in the paper

[10], but can be adapted directly to the family $\{\tilde{f}^p(t,\cdot) : t \in R_{++}\}$. The readers who are familiar to the strong self-concordance can easily see that if a family $\{F(t,\cdot) : t \in R_{++}\}$ of functions on R^m is globally self-concordant then $\mathcal{F} = \{R^m, F(t,\cdot), R^m\}_{t \in R_{++}}$ is strongly self-concordant.

Definition 5.1. A function F on R^m is globally self-concordant with the parameter value $\alpha > 0$ if it satisfies:

(a) F is a strictly convex C^3 function on R^m with a positive definite Hessian matrix $F_{ww}(\boldsymbol{w})$ at every $\boldsymbol{w} \in R^m$.
(b) There is a unique global minimizer of F over R^m.
(c) For every $\boldsymbol{w} \in R^m$ and $\boldsymbol{h} \in R^m$,

$$\left| \left. \frac{d^3 F(\boldsymbol{w} + s\boldsymbol{h})}{ds^3} \right|_{s=0} \right| \leq 2\alpha^{-1/2} \left(\left. \frac{d^2 F(\boldsymbol{w} + s\boldsymbol{h})}{ds^2} \right|_{s=0} \right)^{3/2}$$

Definition 5.2. A family $\{F(t,\cdot) : t \in R_{++}\}$ of functions on R^m is globally self-concordant with the positive parameter functions α, ξ and η if it satisfies:

(d) $\alpha,\ \xi,\ \eta : R_{++} \to R_{++}$ are continuous functions.
(e) For every $t \in R_{++}$, $F(t,\cdot)$ is a globally self-concordant function on R^m with the parameter value $\alpha(t)$.
(f) $F(t,\boldsymbol{w})$, $F_w(t,\boldsymbol{w})$ and $F_{ww}(t,\boldsymbol{w})$ are differentiable in $t \in R_{++}$ and the resultant derivatives in $t \in R_{++}$ are continuous in $(t,\boldsymbol{w}) \in R_{++} \times R^m$.
(g) For every $(t,\boldsymbol{w}) \in R_{++} \times R^m$ and $\boldsymbol{h} \in R^m$,

$$\left| \left. \frac{d^2 F(t, \boldsymbol{w} + s\boldsymbol{h})}{dt ds} \right|_{s=0} \right| \quad \leq \quad \xi(t)\alpha(t)^{1/2} \left(\left. \frac{d^2 F(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^2} \right|_{s=0} \right)^{1/2}.$$

(h) For every $(t,\boldsymbol{w}) \in R_{++} \times R^m$ and $\boldsymbol{h} \in R^m$,

$$\left| \left. \frac{d^3 F(t, \boldsymbol{w} + s\boldsymbol{h})}{dt ds^2} \right|_{s=0} \right| \quad \leq \quad 2\eta(t) \left(\left. \frac{d^2 F(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^2} \right|_{s=0} \right).$$

Let $\{F(t,\cdot) : t \in R_{++}\}$ be a globally self-concordant family of functions on R^m with the positive parameter functions $\alpha,\ \xi,\ \eta : R_{++} \to R_{++}$. For every $t \in R_{++}$, define Newton's decrement of the function $F(t,\cdot)$ at $\boldsymbol{w} \in R^m$ by

$$\lambda(t,\boldsymbol{w}) = \inf\{\lambda : |F_w(t,\boldsymbol{w})^T \boldsymbol{h}| \leq \lambda\alpha(t)^{1/2} \left(\boldsymbol{h}^T F_{ww}(t,\boldsymbol{w})\boldsymbol{h}\right)^{1/2} \text{ for every } \boldsymbol{h} \in R^m\}.$$

Or alternatively, $\lambda(t, \boldsymbol{w})$ is defined by

$$\begin{aligned}\lambda(t, \boldsymbol{w})^2 &= 2\left(\frac{F(t, \boldsymbol{w}) - \inf\{\Phi(t, \boldsymbol{u}) : \boldsymbol{u} \in R^m\}}{\alpha(t)}\right) \\ &= \frac{1}{\alpha(t)} F_w(t, \boldsymbol{w})^T F_{ww}(t, \boldsymbol{w})^{-1} F_w(t, \boldsymbol{w}),\end{aligned}$$

where $\Phi(t, \cdot)$ is a quadratic approximation of the function $F(t, \cdot)$ at $\boldsymbol{w} \in R^m$;

$$\begin{aligned}\Phi(t, \boldsymbol{u}) &= F(t, \boldsymbol{w}) + F_w(t, \boldsymbol{w})^T(\boldsymbol{u} - \boldsymbol{w}) + \frac{1}{2}(\boldsymbol{u} - \boldsymbol{w})^T F_{ww}(t, \boldsymbol{w})(\boldsymbol{u} - \boldsymbol{w}) \\ &\quad \text{for every } (t, \boldsymbol{u}) \in R_{++} \times R^m.\end{aligned}$$

Newton's decrement is a continuous function from $R_{++} \times R^m$ into the set of nonnegative numbers such that for each $t \in R_{++}$, $\lambda(t, \cdot)$ takes the minimal value 0 over R^m at $\boldsymbol{w} \in R^m$ if and only if $\boldsymbol{w}$ is the unique global minimizer of $F(t, \cdot)$ over R^m. Thus, for every $\theta \in R_{++}$, the set

$$N(\theta) = \{(t, \boldsymbol{w}) \in R_{++} \times R^m : \lambda(t, \boldsymbol{w}) \leq \theta\}$$

forms a closed neighborhood of the trajectory

$$\begin{aligned}&\{(t, \boldsymbol{w}) \in R_{++} \times R^m : \lambda(t, \boldsymbol{w}) = 0\} \\ &= \{(t, \boldsymbol{w}) \in R_{++} \times R^m : \boldsymbol{w} = \arg\min \{F(t, \boldsymbol{w}) : \boldsymbol{w} \in R^m\}\},\end{aligned}$$

which we want to trace numerically. Let $\theta \in R_{++}$. Define the metric ρ_θ on R_{++} by

$$\begin{aligned}\rho_\theta(t'', t') &= \frac{1}{2}\max\{|\ln(\alpha(\tau')/\alpha(\tau''))| : \tau'', \tau' \in [t'', t']\} + \theta^{-1}\left|\int_{t''}^{t'} \xi(s)ds\right| + \left|\int_{t''}^{t'} \eta(s)ds\right| \\ &\quad \text{for every } t',\ t'' \in R_{++}.\end{aligned}$$

Let $\lambda_* = 2 - 3^{1/2} = 0.2679\cdots$. Define

$$\begin{aligned}\sigma(\lambda) &= \begin{cases}(1+\lambda)^{-1} & \text{if } \lambda > \lambda_* , \\ 1 & \text{if } \lambda \leq \lambda_* ,\end{cases} \\ \omega(\lambda) &= 1 - (1 - 3\lambda)^{1/3}.\end{aligned}$$

Theorem 5.3. *Assume that $\{F(t, \cdot) : t \in R_{++}\}$ is a globally self-concordant family of functions on R^m with the positive parameter functions $\alpha,\ \xi,\ \eta : R_{++} \rightarrow R_{++}$. Let $t' \in R_{++}$, $\boldsymbol{w}' \in R^m$ and $\lambda' = \lambda(t', \boldsymbol{w}')$. Let $\bar{\boldsymbol{w}} \in R^m$ be the Newton iterate at $\boldsymbol{w}'$ with the step length $\sigma(\lambda')$;*

$$\bar{\boldsymbol{w}} = \boldsymbol{w}' - \sigma(\lambda') F_{ww}(t', \boldsymbol{w}')^{-1} F_w(t', \boldsymbol{w}').$$

(i) *If* $\lambda' > \lambda_*$ *then*

$$F(t', \bar{\boldsymbol{w}}) - F(t', \boldsymbol{w}') \leq -\alpha(t')\left(\lambda_* - \ln(1+\lambda_*)\right) \leq -0.03\alpha(t').$$

(ii) *If* $\lambda' \leq \lambda_*$ *then*

$$\lambda(t', \bar{\boldsymbol{w}}) \leq \frac{(\lambda')^2}{(1-\lambda')^2} \begin{cases} = & \dfrac{\lambda'}{2} \quad \text{if } \lambda' = \lambda_*, \\ < & \dfrac{\lambda'}{2} \quad \text{if } \lambda' < \lambda_* \end{cases}$$

$$F(t', \boldsymbol{w}') - \min\{F(t', \boldsymbol{w}) : \boldsymbol{w} \in R^m\} \leq \frac{1}{2}\alpha(t')\omega(\lambda')^2 \frac{1+\omega(\lambda')}{1-\omega(\lambda')} \leq 4\lambda'\alpha(t').$$

(iii) *If* $\lambda' < \theta < \lambda_*$ *and* $\rho_\theta(t'', t') \leq \theta^{-1}(\theta - \lambda')$ *then* $\lambda(t'', \boldsymbol{w}') \leq \theta$.

Proof: See Section 1 of the paper [10] ∎

We have already seen that:

(a)' For every $t \in R_{++}$, the function $\tilde{f}^p(t, \cdot) : R^m \to R$ is a strictly convex C^∞ function with a positive definite Hessian matrix $\tilde{f}^p_{ww}(t, \boldsymbol{w})$ at every $\boldsymbol{w} \in R^m$.

(b)' For every $t \in R_{++}$, there is a unique global minimizer of $\tilde{f}^p(t, \cdot)$ over R^m.

(f)' $\tilde{f}^p(\cdot, \cdot)$ is C^∞ differentiable on $R_{++} \times R^m$.

Hence the following theorem is sufficient to establish that the family $\{\tilde{f}^p(t, \cdot) : t \in R_{++}\}$ of functions on R^m is globally self-concordant.

Theorem 5.4.

(d)' Define $\alpha(t) = t$, $\xi(t) = \dfrac{\sqrt{n}}{t}$ *and* $\eta(t) = \dfrac{\sqrt{n}}{2t}$ *for every* $t \in R_{++}$.

Then the following relations hold for every $(t, \boldsymbol{w}) \in R_{++} \times R^m$.

$$\text{(c)'} \quad \left| \left. \frac{d^3 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^3} \right|_{s=0} \right| \leq 2\alpha(t)^{-1/2} \left(\left. \frac{d^2 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^2} \right|_{s=0} \right)^{3/2}.$$

$$\text{(g)'} \quad \left| \left. \frac{d^2 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{dt\,ds} \right|_{s=0} \right| \leq \xi(t)\alpha(t)^{1/2} \left(\left. \frac{d^2 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^2} \right|_{s=0} \right)^{1/2}.$$

$$\text{(h)'} \quad \left| \left. \frac{d^3 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{dt\,ds^2} \right|_{s=0} \right| \leq 2\eta(t) \left(\left. \frac{d^2 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^2} \right|_{s=0} \right).$$

Proof: (c)' By Theorem 4.3, we have that

$$\begin{aligned}
\left(\frac{d^3\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^3}\right)^2 &= \left(\frac{2}{t^2}\sum_{j=1}^{n} u_j(t,s)^3\right)^2 \\
&= \frac{4}{t^4}\left(\sum_{j=1}^{n} u_j(t,s)^3\right)^2 \\
&\leq \frac{4}{t^4}\left(\sum_{j=1}^{n} u_j(t,s)^2\right)^3 \\
&= \frac{4}{t}\left(\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^2}\right)^3.
\end{aligned}$$

Thus (c)' follows.

(g)' It follows from the definition of $\boldsymbol{v}(t,s)$ that $|\boldsymbol{v}(t,s)| \leq \sqrt{n}$. Hence we see by Theorem 4.3 that

$$\begin{aligned}
\left|\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{dtds}\right| &= \left|\frac{\boldsymbol{u}(t,s)^T\boldsymbol{v}(t,s)}{t}\right| \\
&\leq \frac{\sqrt{n}\|\boldsymbol{u}(t,s)\|}{t} \\
&= \xi(t)\alpha(t)^{1/2}\left(\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^2}\right)^{1/2}.
\end{aligned}$$

Thus we have shown (g)'.

(h)' It is easily verified that

$$-\sqrt{n} \leq 2v_i(t,\boldsymbol{w}) - 1 \leq \sqrt{n} \text{ for } i = 1,2,\ldots,n. \tag{7}$$

Hence we see by Theorem 4.3 that

$$\begin{aligned}
\left|\frac{d^3\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{dtds^2}\right| &\leq \frac{\sqrt{n}}{t^2}\boldsymbol{u}(t,s)^T\boldsymbol{u}(t,s) \\
&= \frac{\sqrt{n}}{t}\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^2} \\
&= 2\eta(t)\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^2}.
\end{aligned}$$

Thus we have shown (h)'. ∎

Now we are ready to apply Theorem 5.4 to the family $\{\tilde{f}^p(t,\cdot) : t \in R_{++}\}$. Newton's decrement turns out to be

$\lambda(t,\boldsymbol{w})$

$$
\begin{aligned}
&= \inf\{\lambda : \left|\frac{d\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds}\right|_{s=s}\right| \le \lambda\alpha(t)^{1/2}\left(\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^2}\Big|_{s=s}\right)^{1/2},\ \forall \boldsymbol{h}\in R^m\ \} \\
&= \inf\{\lambda : \left|(\boldsymbol{a}-\boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w}))^T\boldsymbol{h}\right| \\
&\qquad \le \lambda\left(\boldsymbol{h}^T\boldsymbol{A}\tilde{\boldsymbol{X}}\left(\boldsymbol{I}-\tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T\boldsymbol{h}\right)^{1/2},\ \forall \boldsymbol{h}\in R^m\ \} \\
&= \inf\{\lambda : \left|(\boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w})-\boldsymbol{a})^T\boldsymbol{h}\right| \\
&\qquad \le \lambda t^{1/2}\left(\boldsymbol{h}^T\boldsymbol{A}\tilde{\Delta}\left(\boldsymbol{I}-\tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)\tilde{\Delta}\boldsymbol{A}^T\boldsymbol{h}\right)^{1/2},\ \forall \boldsymbol{h}\in R^m\ \}.
\end{aligned}
$$

Here $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{X}}(t,\boldsymbol{w})$ and $\tilde{\Delta} = \tilde{\boldsymbol{X}}(t,\boldsymbol{w})^{1/2}\tilde{\boldsymbol{Z}}(t,\boldsymbol{w})^{-1/2}$. Let $0<\theta$ and $0<t''<t'$. Then

$$
\begin{aligned}
\rho_\theta(t'',t') &= \frac{1}{2}\max\{|\ln(\alpha(\tau'')/\alpha(\tau'))| : \tau'',\ \tau'\in[t'',t']\ \} \\
&\quad +\theta^{-1}\left|\int_{t''}^{t'}\xi(s)ds\right| + \left|\int_{t''}^{t'}\eta(s)ds\right| \\
&= \frac{1}{2}\max\{|\ln(\tau''/\tau')| : \tau'',\ \tau'\in[t'',t']\ \} \\
&\quad +\theta^{-1}\left|\int_{t''}^{t'}\frac{\sqrt{n}}{s}ds\right| + \left|\int_{t''}^{t'}\frac{\sqrt{n}}{2s}ds\right| \\
&= \frac{1}{2}\ln(t'/t'') + \frac{\sqrt{n}}{\theta}(\ln t' - \ln t'') + \frac{\sqrt{n}}{2}(\ln t' - \ln t'') \\
&= \left(\frac{1}{2}+\frac{\sqrt{n}}{\theta}+\frac{\sqrt{n}}{2}\right)\ln(t'/t'').
\end{aligned}
$$

Thus we obtain

$$
\rho_\theta(t'',t') = \left(\frac{1}{2}+\frac{\sqrt{n}}{\theta}+\frac{\sqrt{n}}{2}\right)\ln(t'/t''). \tag{8}
$$

Theorem 5.5. *Let $\theta_* = 1/4 < \lambda_* = 2-3^{1/2} = 0.2679\ldots$. Let $(t',\boldsymbol{w}')\in R_{++}\times R^m$ and $\lambda' = \lambda(t',\boldsymbol{w}')$. Let $\bar{\boldsymbol{w}}\in R^m$ be the Newton iterate at $\boldsymbol{w}'$ with the step length $\sigma(\lambda')$;*

$$
\bar{\boldsymbol{w}} = \boldsymbol{w}' - \sigma(\lambda')\tilde{f}^p_{ww}(t',\boldsymbol{w}')^{-1}\tilde{f}^p_w(t',\boldsymbol{w}').
$$

(i) *If $\lambda' > \lambda_*$ then*

$$
\tilde{f}^p(t',\bar{\boldsymbol{w}}) - \tilde{f}^p(t',\boldsymbol{w}') \le -t'(\lambda_* - \ln(1+\lambda_*)) \le -0.03t'.
$$

(ii) *If $\lambda' \le \lambda_*$ then*

$$
\lambda(t',\bar{\boldsymbol{w}}) \le \frac{(\lambda')^2}{(1-\lambda')^2}\begin{cases} = \dfrac{\lambda'}{2} & \text{if } \lambda' = \lambda_*, \\[2ex] < \dfrac{\lambda'}{2} & \text{if } \lambda' < \lambda_*, \end{cases}
$$

$$
\tilde{f}^p(t',\boldsymbol{w}') - \min\{\tilde{f}^p(t',\boldsymbol{w}) : \boldsymbol{w}\in R^m\} \le \frac{1}{2}t'\omega(\lambda')^2\frac{1+\omega(\lambda')}{1-\omega(\lambda')} \le 4\lambda't'.
$$

(iii) *If* $\lambda' \leq \theta_*/2$ *and* $(1-1/(11\sqrt{n}))t' \leq t'' \leq t'$ *then* $\lambda(t'', \boldsymbol{w}') \leq \theta_*$.

Proof: The assertions (i) and (ii) follow directly from Theorems 5.3 and 5.4, so that it suffices to show (iii). Assume that $\lambda' \leq \theta_*/2$ and $(1-1/(11\sqrt{n}))t' \leq t'' \leq t'$. Then we see by (8) that

$$\begin{aligned} \rho_{\theta_*}(t'', t') &= \left(\frac{1}{2} + \frac{\sqrt{n}}{\theta_*} + \frac{\sqrt{n}}{2}\right) \ln(t'/t'') \\ &\leq 5\sqrt{n} \ln \frac{1}{1-1/(11\sqrt{n})} \\ &\leq 5\sqrt{n}\frac{1}{10\sqrt{n}} \;=\; \frac{1}{2} \;\leq\; \theta_*^{-1}(\theta_* - \lambda'). \end{aligned}$$

Thus the desired result follows from (iii) of Theorem 5.3. ∎

6. Conceptual Algorithms.

In addition to the assumptions (A), (B) and (C) stated in the Introduction, we assume:

(D) For any $(t, \boldsymbol{w}) \in R_{++} \times R^m$, we can compute the exact maximizer $\tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ of $\tilde{f}^p(t, \boldsymbol{w}, \cdot)$ over Q_{++} and the exact minimizer $(\tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$ of $\tilde{f}^d(t, \boldsymbol{w}, \cdot, \cdot)$ over $D_{++}(\boldsymbol{w})$.

We describe two algorithms based on Theorem 5.5. For each fixed $t \in R_{++}$, Algorithm 6.1 approximates the minimizer $\hat{\boldsymbol{w}}(t)$ of $\tilde{f}^p(t, \boldsymbol{w})$ over R^m, while Algorithm 6.2 numerically traces the trajectory $\{(t, \hat{\boldsymbol{w}}(t)) : t \in R_{++}\}$ from a given initial point $(t^0, \boldsymbol{w}^0)$ in the neighborhood $N(\theta_*/2) = \{(t, \boldsymbol{w}) : \lambda(t, \boldsymbol{w}) \leq \theta_*/2\}$ of the trajectory $\{(t, \hat{\boldsymbol{w}}(t)) : t \in R_{++}\}$, where $\theta_* = 1/4$.

Let $(t, \boldsymbol{w}) \in R_{++} \times R^m$.

Algorithm 6.1.

Step 0: Let $q = 0$ and $\boldsymbol{w}^q = \boldsymbol{w}$.
Step 1: Compute $(\boldsymbol{x}^q, \boldsymbol{y}^q, \boldsymbol{z}^q) = (\tilde{\boldsymbol{x}}(t, \boldsymbol{w}^q), \tilde{\boldsymbol{y}}(t, \boldsymbol{w}^q), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}^q))$. Let $\lambda^q = \lambda(t, \boldsymbol{w}^q)$.
Step 2: Compute $\boldsymbol{w}^{q+1} = \boldsymbol{w}^q - \sigma(\lambda^q)\tilde{f}^p_{ww}(t, \boldsymbol{w}^q)^{-1}\tilde{f}^p_w(t, \boldsymbol{w}^q)$.
Step 3: Replace q by $q+1$, and go to Step 1.

Suppose $(t^0, \boldsymbol{w}^0) \in N(\theta_*/2)$. where $\theta_* = 1/4$.

Algorithm 6.2.

Step 0: Let $\delta = 1/(11\sqrt{n})$ and $r = 0$.
Step 1: Compute $(\boldsymbol{x}^r, \boldsymbol{y}^r, \boldsymbol{z}^r) = (\tilde{\boldsymbol{x}}(t^r, \boldsymbol{w}^r), \tilde{\boldsymbol{y}}(t^r, \boldsymbol{w}^r), \tilde{\boldsymbol{z}}(t^r, \boldsymbol{w}^r))$. Let $\lambda^r = \lambda(t^r, \boldsymbol{w}^r)$.
Step 2: Let $t^{r+1} = (1-\delta)t^r$. Compute $\boldsymbol{w}^{r+1} = \boldsymbol{w}^r - \tilde{f}^p_{ww}(t^{r+1}, \boldsymbol{w}^r)^{-1}\tilde{f}^p_w(t^{r+1}, \boldsymbol{w}^r)$.

Step 3: Replace r by $r+1$, and go to Step 1.

In view of Theorem 5.5, we see that:

(i) The sequence $\{(t, \boldsymbol{w}^q)\}$ generated by Algorithm 6.1 moves into the neighborhood

$$N(\theta_*/2) = \{(t, \boldsymbol{w}) : \lambda(t, \boldsymbol{w}) \leq \theta_*/2\}$$

of the trajectory $\{(t, \hat{\boldsymbol{w}}(t)) : t \in R_{++}\}$ in a finite number of iterations, and eventually converges to $\hat{\boldsymbol{w}}(t)$. Hence Algorithm 6.1 provides us with an initial point of Algorithm 6.2.

(ii) The sequence $\{(t^r, \boldsymbol{w}^r)\}$ generated by Algorithm 6.2 runs in the neighborhood

$$N(\theta_*/2) = \{(t, \boldsymbol{w}) : \lambda(t, \boldsymbol{w}) \leq \theta_*/2\}$$

and satisfies that

$$t^{r+1} = \left(1 - \frac{1}{11\sqrt{n}}\right) t^r \quad \text{for every } r = 0, 1, 2, \dots \; ; \tag{9}$$

hence $\lim_{r\to\infty} t^r = 0$. The sequence $\{(\boldsymbol{x}^r, \boldsymbol{w}^r, \boldsymbol{y}^r, \boldsymbol{z}^r)\}$ lies in a neighborhood of the central trajectory of the primal-dual pair of linear programs $\mathcal{P}^*$ and $\mathcal{D}^*$, and if $(\boldsymbol{x}^*, \boldsymbol{w}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ is a limiting point of the sequence then $\boldsymbol{x}^*$ and $(\boldsymbol{w}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ are optimal solutions of $\mathcal{P}^*$ and $\mathcal{D}^*$, respectively.

We now assume that all elements of the data $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ are integers to analyze computational complexity of Algorithms 6.1 and 6.2 in detail. Let L denote the input size of $\mathcal{P}^*$.

Lemma 6.3. *Let $t > 0$.*

(i) $\tilde{f}^p(t, \mathbf{0}) \leq n2^{2L} + tnL$.

(ii) $\min_{w \in R^m} \tilde{f}^p(t, \boldsymbol{w}) \geq -n2^{2L} - tnL$.

Proof: By (A) and (B) assumed in the Introduction and by the definition of L, we know that

$$\begin{array}{l} x_i \leq 2^L \leq \exp L \quad \text{for every } i = 1, 2, \dots, n \text{ and every } \boldsymbol{x} \in Q_{++}, \\ \bar{x}_i \geq 2^{-L} \geq \exp(-L) \quad \text{for every } i = 1, 2, \dots, n \text{ and some } \bar{\boldsymbol{x}} \in P_{++}. \end{array}$$

It follows that

$$f^p(t, \mathbf{0}, \boldsymbol{x}) \;\; = \;\; \boldsymbol{c}^T \boldsymbol{x} + t \sum_{j=1}^{n} \ln x_j$$

$$
\begin{aligned}
&\le n(2^L)(2^L)+tnL \\
&= n2^{2L}+tnL \quad \text{for every } \boldsymbol{x}\in Q_{++}, \\
f^p(t,\boldsymbol{w},\bar{\boldsymbol{x}}) &= \boldsymbol{c}^T\bar{\boldsymbol{x}}+t\sum_{j=1}^{n}\ln \bar{x}_j \\
&\ge -n(2^L)(2^L)+tn(-L) \\
&= -n2^{2L}-tnL \quad \text{for every } \boldsymbol{w}\in R^m.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\tilde{f}^p(t,\boldsymbol{0}) &= \max\{f^p(t,\boldsymbol{0},\boldsymbol{x}):\boldsymbol{x}\in Q_{++}\} \;\le\; n2^{2L}+tnL, \\
\min_{w\in R^m}\tilde{f}^p(t,\boldsymbol{w}) &= \min_{w\in R^m}\max\{f^p(t,\boldsymbol{w},\boldsymbol{x}):\boldsymbol{x}\in Q_{++}\} \\
&\ge \min_{w\in R^m} f^p(t,\boldsymbol{w},\bar{\boldsymbol{x}}) \\
&\ge -n2^{2L}-tnL.
\end{aligned}
$$

∎

Theorem 6.4.

(i) *Let* $t=2^{2L}$ *and* $\boldsymbol{w}^0=\boldsymbol{0}$. *Then Algorithm 6.1 generates in* $O(nL)$ *iterations a* $(t,\boldsymbol{w}^q)\in N(\theta_*/2)$.

(ii) *Let* $\epsilon>0$. *Then Algorithm 6.2 generates in* $O(\sqrt{n}\ln t^0/\epsilon)$ *iterations a* $(t^r,\boldsymbol{w}^r)\in N(\theta_*/2)$ *such that* $0<t^r<\epsilon$.

Proof: Let $q^*=\lceil 4nL/0.03\rceil+1$. Assume on the contrary that

$$(t,\boldsymbol{w}^q)\notin N(\lambda^*) \quad \text{for } q=0,1,2,\ldots,q^*.$$

By Theorems 6.3 and 5.5 we then see that

$$
\begin{aligned}
-n2^{2L}-tnL &\le \tilde{f}^p(t,\boldsymbol{w}^{q^*}) \\
&\le \tilde{f}^p(t,\boldsymbol{w}^0)-0.03t\times q^* \\
&\le n2^{2L}+tnL-4nLt \\
&\le n2^{2L}-2nL\times 2^{2L}-tnL \;<\; -n2^{2L}-tnL.
\end{aligned}
$$

This is a contradiction. Hence $(t,\boldsymbol{w}^q)\in N(\lambda^*)$ for some $q\le q^*$. By Theorem 5.5, we see that $(t,\boldsymbol{w}^{q+2})\in N(\theta_*/2)$. Thus we have shown (i). The assertion (ii) follows directly from the inequality (9). ∎

7. Toward Practical Implementation.

Among the assumptions imposed in our discussions so far, the most impractical one seems to be the assumption:

(D) For any $(t, \boldsymbol{w}) \in R_{++} \times R^m$, we can compute the exact maximizer $\tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ of $\tilde{f}^p(t, \boldsymbol{w}, \cdot)$ over Q_{++} and the exact minimizer $(\tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$ of $\tilde{f}^d(t, \boldsymbol{w}, \cdot, \cdot)$ over $D_{++}(\boldsymbol{w})$.

For every $(t, \boldsymbol{w}) \in R_{++} \times R^m$, the Newton direction $(\boldsymbol{dx}, \boldsymbol{dy}, \boldsymbol{dz})$ toward a point on the central trajectory of the primal-dual pair of $\mathcal{P}(\boldsymbol{w})$ and $\mathcal{D}(\boldsymbol{w})$ gives a search direction in the $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$-space, which was utilized in many primal-dual feasible- and infeasible-interior-point methods ([5, 7, 8, 9, 11], etc.); $(\boldsymbol{dx}, \boldsymbol{dy}, \boldsymbol{dz})$ is given by

$$\left.\begin{array}{l} \boldsymbol{B}\boldsymbol{dx} = -(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{b}), \\ \boldsymbol{B}^T\boldsymbol{dy} - \boldsymbol{dz} = -(\boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} - \boldsymbol{c}), \\ \boldsymbol{Z}\boldsymbol{dx} + \boldsymbol{X}\boldsymbol{dz} = -(\boldsymbol{X}\boldsymbol{z} - t\boldsymbol{e}). \end{array}\right\} \tag{10}$$

In our framework, we have paid little attention to such search directions but assumed that the point $(\tilde{\boldsymbol{x}}(t, \boldsymbol{w}), \tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$ into which all the flow induced by the search directions runs is available.

To develop practical computational methods, we need to weaken the assumption (D); at least we need to replace "exact" by "approximate" in (D). In Algorithms 6.1 and 6.2, the exact optimizers $\boldsymbol{x} = \tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ and $(\boldsymbol{y}, \boldsymbol{z}) = (\tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$ are necessary to compute the gradient $\tilde{f}^p_w(t, \boldsymbol{w}) = \boldsymbol{a} - \boldsymbol{A}\boldsymbol{x}$ and the Hessian matrix

$$\tilde{f}^p_{ww}(t, \boldsymbol{w}) = \boldsymbol{A}\Delta\left(\boldsymbol{I} - \Delta\boldsymbol{B}^T(\boldsymbol{B}\Delta^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\Delta\right)\Delta\boldsymbol{A}^T$$

with which we perform one Newton iteration to generate a new point in the $\boldsymbol{w}$-space;

$$\boldsymbol{w} + \sigma(\lambda(t, \boldsymbol{w}))\boldsymbol{dw},$$

where

$$\begin{array}{rcl} \boldsymbol{H}\boldsymbol{dw} & = & \boldsymbol{A}\boldsymbol{x} - \boldsymbol{a}, \\ \boldsymbol{H} & = & \boldsymbol{A}\Delta\left(\boldsymbol{I} - \Delta\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\Delta\right)\Delta\boldsymbol{A}^T, \\ \boldsymbol{G} & = & \boldsymbol{B}\Delta^2\boldsymbol{B}^T, \\ \Delta & = & (\mathrm{diag}\ \boldsymbol{x})^{1/2}(\mathrm{diag}\ \boldsymbol{z})^{-1/2}. \end{array} \tag{11}$$

It should be noted, however, that even when the exact optimizers are unavailable, the search direction $\boldsymbol{dw}$ above is well-defined for every pair of $\boldsymbol{x} > \boldsymbol{0}$ and $\boldsymbol{z} > \boldsymbol{0}$. Thus, for every $(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in R_{++} \times R^n_{++} \times R^k \times R^n_{++}$, the correspondence $\boldsymbol{w} \to \boldsymbol{dw}$ defines a search direction in the $\boldsymbol{w}$-space.

Another critical question to be addressed is whether solving the system (11) to compute a search direction $\boldsymbol{dw}$ in the $\boldsymbol{w}$-space is essentially easier than one iteration of the primal-dual interior-point method applied to the original pair of $\mathcal{P}^*$ and $\mathcal{D}^*$ without using the decomposition techniques. To compute $\boldsymbol{dw}$ by solving (11), we need

(i) a construction of the matrix $\boldsymbol{H} = \boldsymbol{A}\Delta\left(\boldsymbol{I} - \Delta\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\Delta\right)\Delta\boldsymbol{A}^T$, which requires an (implicit) inversion of the matrix $\boldsymbol{G} = \boldsymbol{B}\Delta^2\boldsymbol{B}^T$,

(ii) a solution $\boldsymbol{u}$ of a system of equations $\boldsymbol{H}\boldsymbol{dw} = \boldsymbol{s}$ for some $\boldsymbol{s} \in R^m$.

On the other hand, given a current iterate $(\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{y}, \boldsymbol{z}) \in R^n_{++} \times R^m \times R^k \times R^n_{++}$, the standard primal-dual interior-point method solves the system of equations in the search direction $(\boldsymbol{dx}, \boldsymbol{dw}, \boldsymbol{dy}, \boldsymbol{dz})$:

$$\left.\begin{array}{l} \boldsymbol{A}\boldsymbol{dx} = -(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{a}), \quad \boldsymbol{B}\boldsymbol{dx} = -(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{b}), \\ \boldsymbol{A}^T\boldsymbol{dw} + \boldsymbol{B}^T\boldsymbol{dy} - \boldsymbol{dz} = -(\boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} - \boldsymbol{c}), \\ \boldsymbol{X}\boldsymbol{dz} + \boldsymbol{Z}\boldsymbol{dx} = -(\boldsymbol{X}\boldsymbol{z} - t\boldsymbol{e}), \end{array}\right\} \tag{12}$$

where $t \geq 0$. We can easily verify that the search direction $(\boldsymbol{dx}, \boldsymbol{dw}, \boldsymbol{dy}, \boldsymbol{dz})$ can be represented as

$$\left.\begin{array}{rcl} \boldsymbol{H}\boldsymbol{dw} &=& \boldsymbol{A}\boldsymbol{x} - \boldsymbol{a} - \boldsymbol{A}\Delta^2\boldsymbol{B}^T\boldsymbol{G}^{-1}(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{b}) \\ && -\boldsymbol{A}\Delta\left(\boldsymbol{I} - \Delta\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\Delta\right)(\boldsymbol{r} + \boldsymbol{s}), \\ \boldsymbol{G}\boldsymbol{dy} &=& \boldsymbol{B}\boldsymbol{x} - \boldsymbol{b} - \boldsymbol{B}\Delta\left(\Delta\boldsymbol{A}^T\boldsymbol{dw} + \boldsymbol{r} + \boldsymbol{s}\right), \\ \Delta\boldsymbol{dz} &=& \Delta\boldsymbol{A}^T\boldsymbol{dw} + \Delta\boldsymbol{B}^T\boldsymbol{dy} + \boldsymbol{r}, \\ \Delta^{-1}\boldsymbol{dx} &=& -\Delta\boldsymbol{dz} - \boldsymbol{s}, \end{array}\right\} \tag{13}$$

where

$$\boldsymbol{r} \;=\; \Delta(\boldsymbol{A}^T\boldsymbol{w} + \boldsymbol{B}^T\boldsymbol{y} - \boldsymbol{z} - \boldsymbol{c}) \;\text{ and }\; \boldsymbol{s} = \boldsymbol{X}^{-1/2}\boldsymbol{Z}^{-1/2}(\boldsymbol{X}\boldsymbol{z} - t\boldsymbol{e}).$$

This representation (13) of the solution $(\boldsymbol{dx}, \boldsymbol{dw}, \boldsymbol{dy}, \boldsymbol{dz})$ of (12) is called a compact-inverse implementation in the literature [2]. Hence if we utilize the compact-inverse implementation (13), the majority of the computation of $(\boldsymbol{dx}, \boldsymbol{dw}, \boldsymbol{dy}, \boldsymbol{dz})$ are also spent for (i) and (ii).

Therefore we are required almost the same amount of computational work in solving (11) to compute a search direction $\boldsymbol{dw}$ as in one iteration of the compact-inverse implementation of the primal-dual interior-point method applied to the original pair of $\mathcal{P}^*$ and $\mathcal{D}^*$ without using the decomposition techniques. This observation is ironic to our nice theoretical results with the use of Newton's method presented so far, but sheds a new light on the compact-inverse implementation. In particular, we observe that when $\boldsymbol{x} = \tilde{\boldsymbol{x}}(t, \boldsymbol{w})$, $\boldsymbol{y} = \tilde{\boldsymbol{y}}(t, \boldsymbol{w})$ and $\boldsymbol{z} = \tilde{\boldsymbol{z}}(t, \boldsymbol{w})$, (13) turns out to be

$$\left.\begin{array}{l} \boldsymbol{H}\boldsymbol{dw} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{a}, \\ \boldsymbol{G}\boldsymbol{dy} = -\boldsymbol{B}\Delta^2\boldsymbol{A}^T\boldsymbol{dw}, \\ \Delta\boldsymbol{dz} = \Delta\boldsymbol{A}^T\boldsymbol{dw} + \Delta\boldsymbol{B}^T\boldsymbol{dy}, \\ \Delta^{-1}\boldsymbol{dx} = -\Delta\boldsymbol{dz}; \end{array}\right\}$$

hence $\boldsymbol{dw}$ determined by the compact-inverse implementation (13) is exactly the same as the search direction $\boldsymbol{dw}$ in the $\boldsymbol{w}$-space determined by (11).

We conclude the paper by a general idea of computational methods consisting of the following three steps:

Step xyz: Given $(t, \boldsymbol{w}) \in R_{++} \times R^m$ and $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in R^n_{++} \times R^k \times R^n_{++}$, compute a new iterate $(\boldsymbol{x}', \boldsymbol{y}', \boldsymbol{z}')$ by $\boldsymbol{x}' = \boldsymbol{x} + \alpha_p \boldsymbol{dx}$ and $(\boldsymbol{y}', \boldsymbol{z}') = (\boldsymbol{y}, \boldsymbol{z}) + \alpha_d(\boldsymbol{dy}, \boldsymbol{dz})$, where $(\boldsymbol{dx}, \boldsymbol{dy}, \boldsymbol{dz})$ is given by (10) and $\alpha_p,\ \alpha_d \geq 0$ are step lengths.

Step w: Given $t > 0$, $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in R^n_{++} \times R^k \times R^n_{++}$ and $\boldsymbol{w} \in R^m$, choose a search direction $\boldsymbol{dw}$ in the $\boldsymbol{w}$-space and generate a new iterate $\boldsymbol{w}'$ by $\boldsymbol{w}' = \boldsymbol{w} + \sigma \boldsymbol{dw}$, where $\sigma > 0$ is a step length.

Step t: Decrease t.

Algorithms 6.1 and 6.2 may be regarded as ideal cases of such methods in which we always perform infinitely many iterations of Step xyz to compute the exact optimizers $\tilde{\boldsymbol{x}}(t, \boldsymbol{w})$ of $f^p(t, \boldsymbol{w}, \cdot)$ and $(\tilde{\boldsymbol{y}}(t, \boldsymbol{w}), \tilde{\boldsymbol{z}}(t, \boldsymbol{w}))$ of $f^d(t, \boldsymbol{w}, \cdot, \cdot)$ before we perform Step w with the search direction $\boldsymbol{dw}$ determined by (11).

Acknowledgment. We would like to thank an anonymous referee of the paper [4] which had been submitted to *Operations Research Letters*. Although the paper was not accepted for publication, the referee argued some interesting and significant points. In particular, the referee suggested us to see whether a real valued function

$$f(\boldsymbol{w}) = \max\{\boldsymbol{w}^T(\boldsymbol{B}\boldsymbol{x} - \boldsymbol{b}) + \sum_{j=1}^{n} \ln x_j : \boldsymbol{e}^T\boldsymbol{x} = 1,\ \boldsymbol{x} > \boldsymbol{0}\}$$

on R^m is self-concordant. This question motivated the study of the subject of the current paper.

References

[1] J. R. Birge and L. Qi, "Computing block-angular Karmarkar projections with applications to stochastic programming," *Management Science* 34 (1988) 1472–1479.

[2] I. C. Choi and F. Goldfarb, "Exploiting special structure in a primal-dual path-following algorithm," *Mathematical Programming* 58 (1993) 33–52.

[3] M. W. Hirsch and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra,* (Academic Press, New York, 1974).

[4] M. Kojima, N. Megiddo and S. Mizuno, "A Lagrangian Relaxation Method for Approximating the Analytic Center of a Polytope," Technical Report, IBM Almaden Research Center, San Jose, CA 95120–6099, USA, 1992.

[5] M. Kojima, S. Mizuno and A. Yoshise, "A primal-dual interior point algorithm for linear programming," In N. Megiddo, ed., *Progress in Mathematical Programming, Interior-Point and Related Methods* (Springer-Verlag, New York, 1989) 29–47.
[6] L. S. Lasdon, *Optimization Theory for Large Systems*, (Macmillan, New York, 1970).
[7] I. J. Lustig, "Feasibility issues in a primal-dual interior-point method for linear programming," *Mathematical Programming* **49** (1990/91) 145–162.
[8] R. Marsten, R. Subramanian, M. Saltzman, I. J. Lustig and D. Shanno, "Interior point methods for linear programming: Just call Newton, Lagrange, and Fiacco and McCormick!," *Interfaces* **20** (1990) 105–116.
[9] N. Megiddo, "Pathways to the optimal set in linear programming," In N. Megiddo, ed., *Progress in Mathematical Programming, Interior-Point and Related Methods* (Springer-Verlag, New York, 1989) 131–158.
[10] Ju. E. Nesterov and A. S. Nemirovsky, "Self-concordant functions and polynomial-time methods in convex programming," Report, Central Economical and Mathematical Institute, USSR Acad. Sci. (Moscow, USSR, 1989).
[11] K. Tanabe, "Centered Newton method for mathematical programming," In M. Iri and K. Yajima, eds., *System Modeling and Optimization* (Springer-Verlag, New York, 1988) 197–206.
[12] M. J .Todd, "Exploiting special structure in Karmarkar's linear programming algorithm," *Mathematical Programming* 41 (1988) 97–113.
[13] M. J .Todd, "Recent developments and new directions in linear programming," In M. Iri and K. Tanabe, eds., *Recent Developments and Applications* (Kluwer Academic Publishers, London, 1989) 109–157.

Appendix. Calculation of Derivatives of $\tilde{f}^p$.

Differentiating the identities in (5) by $\boldsymbol{w} \in R^m$, we first observe

$$\left.\begin{array}{l} \boldsymbol{B}\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) = \boldsymbol{O}, \quad \boldsymbol{A}^T + \boldsymbol{B}^T\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w}) - \tilde{\boldsymbol{z}}_w(t,\boldsymbol{w}) = \boldsymbol{O}, \\ \tilde{\boldsymbol{Z}}(t,\boldsymbol{w})\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) + \tilde{\boldsymbol{X}}(t,\boldsymbol{w})\tilde{\boldsymbol{z}}_w(t,\boldsymbol{w}) = \boldsymbol{O} \end{array}\right\} \tag{14}$$

for every $(t,\boldsymbol{w}) \in R_{++} \times R^m$. Differentiating the identities in (5) by $t \in R_{++}$, we also see

$$\left.\begin{array}{l} \boldsymbol{B}\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}) = \boldsymbol{O}, \quad \boldsymbol{B}^T\tilde{\boldsymbol{y}}_t(t,\boldsymbol{w}) - \tilde{\boldsymbol{z}}_t(t,\boldsymbol{w}) = \boldsymbol{O}, \\ \tilde{\boldsymbol{Z}}(t,\boldsymbol{w})\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}) + \tilde{\boldsymbol{X}}(t,\boldsymbol{w})\tilde{\boldsymbol{z}}_t(t,\boldsymbol{w}) = \boldsymbol{e} \end{array}\right\} \tag{15}$$

for every $(t,\boldsymbol{w}) \in R_{++} \times R^m$. By solving (14) in $\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w})$, $\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w})$, $\tilde{\boldsymbol{z}}_w(t,\boldsymbol{w})$ and (15) in $\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w})$, $\tilde{\boldsymbol{y}}_t(t,\boldsymbol{w})$, $\tilde{\boldsymbol{z}}_t(t,\boldsymbol{w})$, we obtain the following lemma.

Lemma A.

(i) *For every* $(t,\boldsymbol{w}) \in R_{++} \times R^m$,

$$\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) \;\; = \;\; -\frac{1}{t}\tilde{\boldsymbol{X}}\left(\boldsymbol{I} - \tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T,$$

$$\begin{aligned}
\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w}) &= -(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{A}^T, \\
\tilde{\boldsymbol{z}}_w(t,\boldsymbol{w}) &= \tilde{\boldsymbol{X}}^{-1}\left(\boldsymbol{I}-\tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T.
\end{aligned}$$

(ii) *For every* $(t,\boldsymbol{w}) \in R_{++} \times R^m$,

$$\begin{aligned}
\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}) &= \frac{1}{t}\tilde{\boldsymbol{X}}\left(\boldsymbol{I}-\tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\boldsymbol{e}, \\
\tilde{\boldsymbol{y}}_t(t,\boldsymbol{w}) &= (\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\boldsymbol{e}, \\
\tilde{\boldsymbol{z}}_t(t,\boldsymbol{w}) &= \boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\boldsymbol{e}.
\end{aligned}$$

Here $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{X}}(t,\boldsymbol{w})$ *and* $\tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{Z}}(t,\boldsymbol{w})$.

Proof: (i) It follows from the second and third identities of (14) that

$$\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) = -\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\tilde{\boldsymbol{z}}_w(t,\boldsymbol{w}) = -\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\left(\boldsymbol{A}^T+\boldsymbol{B}^T\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w})\right)$$

Hence, by the first identity of (14),

$$\boldsymbol{O} = \boldsymbol{B}\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) = -\boldsymbol{B}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\left(\boldsymbol{A}^T+\boldsymbol{B}^T\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w})\right).$$

Therefore we obtain that

$$\begin{aligned}
\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w}) &= -(\boldsymbol{B}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{A}^T \\
&= -(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{A}^T \text{ (since } \tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}} = t\boldsymbol{I}), \\
\tilde{\boldsymbol{z}}_w(t,\boldsymbol{w}) &= \boldsymbol{A}^T+\boldsymbol{B}^T\tilde{\boldsymbol{y}}_w(t,\boldsymbol{w}) \\
&= \boldsymbol{A}^T-\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{A}^T \\
&= \tilde{\boldsymbol{X}}^{-1}\left(\boldsymbol{I}-\tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T, \\
\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) &= -\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\tilde{\boldsymbol{z}}_w(t,\boldsymbol{w}) \\
&= -\frac{1}{t}\tilde{\boldsymbol{X}}\left(\boldsymbol{I}-\tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T.
\end{aligned}$$

(ii) In view of the the second and third identities of (15), we see that

$$\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}) = \tilde{\boldsymbol{Z}}^{-1}\boldsymbol{e}-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\tilde{\boldsymbol{z}}_t(t,\boldsymbol{w}) = \tilde{\boldsymbol{Z}}^{-1}\boldsymbol{e}-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{B}^T\tilde{\boldsymbol{y}}_t(t,\boldsymbol{w}).$$

Hence, by the first identity of (15),

$$\boldsymbol{O} = \boldsymbol{B}\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}) = \boldsymbol{B}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{e}-\boldsymbol{B}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{B}^T\tilde{\boldsymbol{y}}_t(t,\boldsymbol{w}).$$

Therefore we obtain

$$\begin{aligned}
\tilde{\boldsymbol{y}}_t(t,\boldsymbol{w}) &= (\boldsymbol{B}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{Z}}^{-1}\boldsymbol{e} = (\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\boldsymbol{e}, \\
\tilde{\boldsymbol{z}}_t(t,\boldsymbol{w}) &= \boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\boldsymbol{e}, \\
\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}) &= \tilde{\boldsymbol{Z}}^{-1}\boldsymbol{e}-\tilde{\boldsymbol{X}}\tilde{\boldsymbol{Z}}^{-1}\tilde{\boldsymbol{z}}_t(t,\boldsymbol{w}) = \frac{1}{t}\tilde{\boldsymbol{X}}\left(\boldsymbol{I}-\tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\boldsymbol{e}.
\end{aligned}$$

∎

Proof of Theorem 4.1:

By the definition,

$$\begin{aligned}
\tilde{f}^p_w(t,\boldsymbol{w}) &= \tilde{\boldsymbol{x}}_w(t,\boldsymbol{w})^T f^p_x(t,\boldsymbol{w},\tilde{\boldsymbol{x}}(t,\boldsymbol{w})) - (\boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w}) - \boldsymbol{a}) \\
&= \tilde{\boldsymbol{x}}_w(t,\boldsymbol{w})^T (\boldsymbol{B}^T \tilde{\boldsymbol{y}}(t,\boldsymbol{w})) - (\boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w}) - \boldsymbol{a}) \\
&= (\boldsymbol{B}\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}))^T \tilde{\boldsymbol{y}}(t,\boldsymbol{w}) - (\boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w}) - \boldsymbol{a}) \\
&\quad \text{(since } \boldsymbol{B}\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}) = \boldsymbol{0} \text{ by (14))} \\
&= \boldsymbol{a} - \boldsymbol{A}\tilde{\boldsymbol{x}}(t,\boldsymbol{w}).
\end{aligned}$$

Thus we have shown (i). By the definition and (i), we see that

$$\tilde{f}^p_{ww}(t,\boldsymbol{w}) = -\boldsymbol{A}\tilde{\boldsymbol{x}}_w(t,\boldsymbol{w}).$$

Hence we obtain by (i) of Lemma A and $\tilde{\boldsymbol{X}} = \sqrt{t}\tilde{\Delta}$ that

$$\begin{aligned}
\tilde{f}^p_{ww}(t,\boldsymbol{w}) &= \frac{1}{t}\boldsymbol{A}\tilde{\boldsymbol{X}}\left(\boldsymbol{I} - \tilde{\boldsymbol{X}}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\boldsymbol{X}}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\boldsymbol{X}}\right)\tilde{\boldsymbol{X}}\boldsymbol{A}^T \\
&= \boldsymbol{A}\tilde{\Delta}\left(\boldsymbol{I} - \tilde{\Delta}\boldsymbol{B}^T(\boldsymbol{B}\tilde{\Delta}^2\boldsymbol{B}^T)^{-1}\boldsymbol{B}\tilde{\Delta}\right)\tilde{\Delta}\boldsymbol{A}^T.
\end{aligned}$$

This completes the proof of Theorem 4.1.

Proof of Theorem 4.3:

For simplicity of notation, we use

$$\begin{array}{rll}
\boldsymbol{x}(t,s) & \text{to denote} & \tilde{\boldsymbol{x}}(t,\boldsymbol{w}+s\boldsymbol{h}), \\
\boldsymbol{X} & \text{to denote} & \boldsymbol{X}(t,s) = \tilde{\boldsymbol{X}}(t,\boldsymbol{w}+s\boldsymbol{h}), \\
\boldsymbol{G} & \text{to denote} & \boldsymbol{B}\boldsymbol{X}^2\boldsymbol{B}^T, \\
\boldsymbol{P} & \text{to denote} & \boldsymbol{I} - \boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X},
\end{array}$$

respectively.

(i) By (i) of Theorem 4.1,

$$\frac{d\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds} = \tilde{f}^p_w(t,\boldsymbol{w}+s\boldsymbol{h})^T\boldsymbol{h} = (\boldsymbol{a} - \boldsymbol{A}\boldsymbol{x}(t,s))^T\boldsymbol{h}.$$

Thus we have shown (i).

(ii) By (ii) of Theorem 4.1,

$$\begin{aligned}
\frac{d^2 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^2} &= \boldsymbol{h}^T \tilde{f}^p_{ww}(t, \boldsymbol{w} + s\boldsymbol{h})\boldsymbol{h} \\
&= \frac{1}{t}\boldsymbol{h}^T \boldsymbol{A}\boldsymbol{X}\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \;=\; \frac{1}{t}\boldsymbol{u}(t,s)^T\boldsymbol{u}(t,s).
\end{aligned}$$

Thus we have shown (ii).

(iii) We first observe that

$$\begin{aligned}
\frac{d\boldsymbol{x}(t,s)}{ds} &= \frac{d\tilde{\boldsymbol{x}}(t, \boldsymbol{w} + s\boldsymbol{h})}{ds} \\
&= \tilde{\boldsymbol{x}}_w(t, \boldsymbol{w} + s\boldsymbol{h})\boldsymbol{h} \\
&= -\frac{1}{t}\boldsymbol{X}\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \;=\; -\frac{1}{t}\boldsymbol{X}\boldsymbol{u}(t,s).
\end{aligned}$$

On the other hand, we see by (ii) that

$$\frac{d^3 \tilde{f}^p(t, \boldsymbol{w} + s\boldsymbol{h})}{ds^3} = \frac{2}{t}\boldsymbol{u}(t,s)^T \frac{d\boldsymbol{u}(t,s)}{ds}.$$

To evaluate $\frac{d\boldsymbol{u}(t,s)}{ds}$, let

$$\boldsymbol{q}(t,s) = \boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \;\text{ and }\; \boldsymbol{r}(t,s) = \boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}^2\boldsymbol{A}^T\boldsymbol{h}.$$

Then

$$\begin{aligned}
&\boldsymbol{G}\boldsymbol{r}(t,s) = \boldsymbol{B}\boldsymbol{X}^2\boldsymbol{A}^T\boldsymbol{h}, \\
&\boldsymbol{u}(t,s) = \boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} = \boldsymbol{q}(t,s) - \boldsymbol{X}\boldsymbol{B}^T\boldsymbol{r}(t,s).
\end{aligned} \tag{16}$$

It follows that

$$\begin{aligned}
&\frac{d\boldsymbol{q}(t,s)}{ds} = \boldsymbol{X}_s\boldsymbol{A}^T\boldsymbol{h}, \\
&2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_s\boldsymbol{B}^T\boldsymbol{r}(t,s) + \boldsymbol{B}\boldsymbol{X}^2\boldsymbol{B}^T\frac{d\boldsymbol{r}(t,s)}{ds} = 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_s\boldsymbol{A}^T\boldsymbol{h}.
\end{aligned}$$

Here $\boldsymbol{X}_s = \text{diag}\,\frac{d\boldsymbol{x}(t,s)}{ds}$. Hence

$$\begin{aligned}
\boldsymbol{G}\frac{d\boldsymbol{r}(t,s)}{ds} &= 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_s\boldsymbol{A}^T\boldsymbol{h} - 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_s\boldsymbol{B}^T\boldsymbol{r}(t,s) \\
&= 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_s\boldsymbol{A}^T\boldsymbol{h} - 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_s\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}^2\boldsymbol{A}^T\boldsymbol{h} \\
&= 2\boldsymbol{B}\boldsymbol{X}_s\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h},
\end{aligned}$$

$$
\begin{aligned}
\frac{d\boldsymbol{r}(t,s)}{ds} &= 2\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}_s\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}, \\
\frac{d\boldsymbol{u}(t,s)}{ds} &= \frac{d\boldsymbol{q}(t,s)}{ds} - \boldsymbol{X}_s\boldsymbol{B}^T\boldsymbol{r}(t,s) - \boldsymbol{X}\boldsymbol{B}^T\frac{d\boldsymbol{r}(t,s)}{ds} \\
&= \boldsymbol{X}_s\boldsymbol{A}^T\boldsymbol{h} - \boldsymbol{X}_s\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}^2\boldsymbol{A}^T\boldsymbol{h} - 2\boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}_s\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \\
&= \left(\boldsymbol{I} - 2\boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}\right)\boldsymbol{X}^{-1}\boldsymbol{X}_s\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
t\frac{d^3\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{ds^3} &= 2\boldsymbol{u}(t,s)^T\frac{d\boldsymbol{u}(t,s)}{ds} \\
&= 2\boldsymbol{h}^T\boldsymbol{A}\boldsymbol{X}\boldsymbol{P}\left(\boldsymbol{I} - 2\boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}\right)\boldsymbol{X}^{-1}\boldsymbol{X}_s\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \\
&= 2\boldsymbol{h}^T\boldsymbol{A}\boldsymbol{X}\boldsymbol{P}\boldsymbol{X}^{-1}\boldsymbol{X}_s\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \\
&= -2\boldsymbol{u}(t,s)^T\left(\frac{1}{t}\mathrm{diag}\ \boldsymbol{u}(t,s)\right)\boldsymbol{u}(t,s) \\
&= -\frac{2}{t}\sum_{j=1}^{n}u_j(t,s)^3.
\end{aligned}
$$

Thus we have shown (iii).

(iv) By (i), we see that

$$
\begin{aligned}
\frac{d^2\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{dtds} &= \frac{d((\boldsymbol{a}-\boldsymbol{A}\boldsymbol{x}(t,s))^T\boldsymbol{h})}{dt} \\
&= -\boldsymbol{h}^T\boldsymbol{A}\tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}+s\boldsymbol{h}) \\
&= -\frac{1}{t}\boldsymbol{h}^T\boldsymbol{A}\boldsymbol{X}\boldsymbol{P}\boldsymbol{e} \\
&= -\frac{1}{t}\boldsymbol{u}(t,s)^T\boldsymbol{v}(t,s).
\end{aligned}
$$

(v) By (ii), we see that

$$
\begin{aligned}
\frac{d^3\tilde{f}^p(t,\boldsymbol{w}+s\boldsymbol{h})}{dtds^2} &= \frac{d}{dt}\left\{\frac{\boldsymbol{u}(t,s)^T\boldsymbol{u}(t,s)}{t}\right\} \\
&= \frac{2}{t}\boldsymbol{u}(t,s)^T\frac{d\boldsymbol{u}(t,s)}{dt} - \frac{\boldsymbol{u}(t,s)^T\boldsymbol{u}(t,s)}{t^2}.
\end{aligned}
$$

Using $\boldsymbol{q}(t,s)$ and $\boldsymbol{r}(t,s)$, we represent $\boldsymbol{u}(t,s)$ as in (16) to evaluate $\dfrac{d\boldsymbol{u}(t,s)}{dt}$. By the definition of $\boldsymbol{q}(t,s)$ and $\boldsymbol{r}(t,s)$, we see that

$$
\begin{aligned}
&\frac{d\boldsymbol{q}(t,s)}{dt} = \boldsymbol{X}_t\boldsymbol{A}^T\boldsymbol{h}, \\
&2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{B}^T\boldsymbol{r}(t,s) + \boldsymbol{B}\boldsymbol{X}^2\boldsymbol{B}^T\frac{d\boldsymbol{r}(t,s)}{dt} = 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{A}^T\boldsymbol{h}.
\end{aligned}
$$

Here $\boldsymbol{X}_t = \text{ diag } \dfrac{d\boldsymbol{x}(t,s)}{dt}$. Hence

$$\begin{aligned}
\boldsymbol{G}\frac{d\boldsymbol{r}(t,s)}{dt} &= 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{A}^T\boldsymbol{h} - 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{B}^T\boldsymbol{r}(t,s) \\
&= 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{A}^T\boldsymbol{h} - 2\boldsymbol{B}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}^2\boldsymbol{A}^T\boldsymbol{h} \\
&= 2\boldsymbol{B}\boldsymbol{X}_t\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}, \\
\frac{d\boldsymbol{r}(t,s)}{dt} &= 2\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}_t\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}, \\
\frac{d\boldsymbol{u}(t,s)}{dt} &= \frac{d\boldsymbol{q}(t,s)}{dt} - \boldsymbol{X}_t\boldsymbol{B}^T\boldsymbol{r}(t,s) - \boldsymbol{X}\boldsymbol{B}^T\frac{d\boldsymbol{r}(t,s)}{dt} \\
&= \boldsymbol{X}_t\boldsymbol{A}^T\boldsymbol{h} - \boldsymbol{X}_t\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}^2\boldsymbol{A}^T\boldsymbol{h} - 2\boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}_t\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \\
&= \left(\boldsymbol{I} - 2\boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}\right)\boldsymbol{X}^{-1}\boldsymbol{X}_t\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}.
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
\boldsymbol{u}(t,s)^T\frac{d\boldsymbol{u}(t,s)}{dt} &= \boldsymbol{h}^T\boldsymbol{A}\boldsymbol{X}\boldsymbol{P}\left(\boldsymbol{I} - 2\boldsymbol{X}\boldsymbol{B}^T\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{X}\right)\boldsymbol{X}^{-1}\boldsymbol{X}_t\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h} \\
&= \boldsymbol{h}^T\boldsymbol{A}\boldsymbol{X}\boldsymbol{X}_t\boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}.
\end{aligned}$$

By the definition of $\boldsymbol{u}(t,s)$, $\boldsymbol{x}(t,s)$, $\boldsymbol{X}_t$ and Lemma A, we know that

$$\begin{aligned}
\boldsymbol{u}(t,s) &= \boldsymbol{P}\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{h}, \\
\frac{d\boldsymbol{x}(t,s)}{dt} &= \tilde{\boldsymbol{x}}_t(t,\boldsymbol{w}+s\boldsymbol{h}) = \frac{1}{t}\boldsymbol{X}\boldsymbol{v}(t,s), \\
\boldsymbol{X}^{-1}\boldsymbol{X}_t &= \boldsymbol{X}^{-1}\left(\text{ diag } \frac{d\boldsymbol{x}(t,s)}{dt}\right) = \frac{1}{t}\left(\text{ diag } \boldsymbol{v}(t,s)\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{d^3\tilde{f}^p}{ds^2dt} &= \frac{2}{t}\boldsymbol{u}(t,s)^T\frac{d\boldsymbol{u}(t,s)}{dt} - \frac{\boldsymbol{u}(t,s)^T\boldsymbol{u}(t,s)}{t^2} \\
&= \frac{2\boldsymbol{u}(t,s)^T\left(\text{diag } \boldsymbol{v}(t,s)\right)\boldsymbol{u}(t,s) - \boldsymbol{u}(t,s)^T\boldsymbol{u}(t,s)}{t^2} \\
&= \frac{\boldsymbol{u}(t,s)^T\left(2\text{diag } \boldsymbol{v}(t,s) - \boldsymbol{I}\right)\boldsymbol{u}(t,s)}{t^2}.
\end{aligned}$$

This completes the proof of Theorem 4.3.