A Lagrangian Relaxation Method for Approximating the Analytic Center of a Polytope*

Masakazu Kojima† Nimrod Megiddo‡ Shinji Mizuno§

September 1992

Abstract. The analytic center of a polytope $P_+ = \{\boldsymbol{x} \geq \boldsymbol{0} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},\ \boldsymbol{e}^T\boldsymbol{x} = 1\}$ is characterized by a saddle point condition

$$\min_{\boldsymbol{y}\in R^m} \max_{\boldsymbol{x}\in S_{++}} g(\boldsymbol{x},\boldsymbol{y}) = \max_{\boldsymbol{x}\in S_{++}} \min_{\boldsymbol{y}\in R^m} g(\boldsymbol{x},\boldsymbol{y}) \ ,$$

on the Lagrangian function

$$g(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \sum_{j=1}^{n} \log x_j$$

where $\boldsymbol{A} \in R^{m\times n}$, $\boldsymbol{b} \in R^m$, and $S_{++} = \{\boldsymbol{x} > \boldsymbol{0} : \sum_{j=1}^n x_j = 1\}$. This paper presents properties of the marginal function $f(\boldsymbol{y}) = \max\{g(\boldsymbol{x},\boldsymbol{y}) : \boldsymbol{x} \in S_{++}\}$ and explores the possibilities of a Lagrangian relaxation method for approximating the analytic center.

1. Introduction

The problem of approximating the analytic center [6] of a polytope underlies many interior-point algorithms (see [7] and the references therein) for linear programming. For a polytope P defined by

$$\boldsymbol{a}_i^T\boldsymbol{x} = \alpha_i \ (i = 1,2,\ldots,\ell) \quad \text{and} \quad \boldsymbol{b}_j^T\boldsymbol{x} \leq \beta_j \ (j = 1,2,\ldots,m),$$

*Research supported in part by ONR under contract N00014-91-C-0026. This research was done when M. Kojima and S. Mizuno visited at the IBM Almaden Research Center.

†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

the analytic center $\bar{\boldsymbol{x}}$ is the maximizer of

$$\sum_{j=1}^{m} \log(\beta_j - \boldsymbol{b}_j^T \boldsymbol{x})$$

subject to

$$\boldsymbol{a}_i^T \boldsymbol{x} = \alpha_i \ (i = 1, 2, \ldots, \ell) \ \text{ and } \ \boldsymbol{b}_j^T \boldsymbol{x} < \beta_j \ (j = 1, 2, \ldots, m).$$

Here $\boldsymbol{a}_i \in R^n$, $\alpha_i \in R$, $\boldsymbol{b}_j \in R^n$ and $\beta_j \in R$.

Consider the linear program: Minimize $\boldsymbol{c}^T \boldsymbol{x}$ subject to $\boldsymbol{x} \in P$, where $\boldsymbol{c} \in R^n$ and $P \subset R^n$ is a polyhedral set. We assume the set of minimal solutions of the latter is nonempty and bounded. Denote the minimal value by μ^*. For every $\mu > \mu^*$, the level set $P(\mu) = \{\boldsymbol{x} \in P : \boldsymbol{c}^T \boldsymbol{x} \leq \mu\}$ is nonempty and the set consisting of the analytic centers $\boldsymbol{x}(\mu)$ of the level set $P(\mu)$ is a smooth curve converging to a minimal solution of the linear program as $\mu \to \mu^*$. The set $\{\boldsymbol{x}(\mu) : \mu > \mu^*\}$ is called the central trajectory or the path of centers. Thus, tracing the central trajectory numerically, we obtain an approximate minimal solution of the linear program. Renegar [5] proposed a polynomial-time algorithm for linear programming based on successive approximations of the analytic center $\boldsymbol{x}(\mu)$ for an appropriate sequence $\{\mu^k > \mu^*\}$ converging to μ^*. Many other algorithms have since then been developed along similar lines (see, *e.g.*, the survey article [7] and its references).

In each iteration of Newton's method one needs to solve a linear system of equations induced by a local linearization of the problem, which generally requires $O(n^3)$ arithmetic operations, where n denotes the dimension of the linear system. Various numerical methods have been developed which require $O(n^2)$ arithmetic operations per iteration and often work on large scale nonlinear problems more efficiently than Newton's method.

This note presents a Lagrangian relaxation method which transforms the problem of approximating the analytic center of a polytope into an unconstrained minimization of a strictly convex C^∞ function $f : R^m \to R$ over R^m. This transformation itself is standard. It may not be powerful enough for developing a theoretically more efficient algorithm for linear programming. Practically, however, this transformation seems attractive in the sense that it makes possible to incorporate various unconstrained minimization methods such as conjugate gradients methods and quasi-Newton methods into the problem of finding the analytic center of a polytope.

2. Main Results

Let $\boldsymbol{A} \in R^{m \times n}$, $\boldsymbol{b} \in R^m$ and $\boldsymbol{e} = (1, \ldots, 1)^T \in R^n$. Consider a polytope of the form

$$P_+ = \{\boldsymbol{x} \in R^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \ \boldsymbol{e}^T \boldsymbol{x} = 1, \ \boldsymbol{x} \geq \boldsymbol{0}\} \ .$$

Any nonempty bounded feasible region can be easily transformed into such a form. We assume P_+ satisfies the following condition:

Condition 2.1. (i) $P_{++} \equiv \{\boldsymbol{x} \in P_+ : \boldsymbol{x} > \boldsymbol{0}\}$ is nonempty.
(ii) The rows of $\begin{bmatrix} \boldsymbol{A} \\ \boldsymbol{e}^T \end{bmatrix} \in R^{(m+1)\times n}$ are linearly independent.

The analytic center of P_+ is defined to be the maximizer of $\sum_{j=1}^n \log x_j$ in P_{++}. Define

$$S_+ = \{\boldsymbol{x} \in R^n : \boldsymbol{e}^T\boldsymbol{x} = 1,\ \boldsymbol{x} \geq \boldsymbol{0}\} \quad , \quad S_{++} = \{\boldsymbol{x} \in S_+ : \boldsymbol{x} > \boldsymbol{0}\}$$
$$g(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) + \sum_{j=1}^n \log x_j \quad ((\boldsymbol{x}, \boldsymbol{y}) \in S_{++} \times R^m)\ . \tag{1}$$

Here $\boldsymbol{y} \in R^m$ is the Lagrange multiplier associated with $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. The function g is the *Lagrangian*. The vector $\bar{\boldsymbol{x}}$ is the analytic center of P_+ if and only if there exists a $\bar{\boldsymbol{y}} \in R^m$ such that the saddle point condition

$$g(\boldsymbol{x}, \bar{\boldsymbol{y}}) \leq g(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \leq g(\bar{\boldsymbol{x}}, \boldsymbol{y}) \quad \text{for every} \quad \boldsymbol{y} \in R^m \quad \text{and } \boldsymbol{x} \in S_{++} \tag{2}$$

holds. Define

$$\boldsymbol{\phi}(\boldsymbol{y}) = \arg\max\{g(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \in S_{++}\} \ \text{ and } f(\boldsymbol{y}) = \max\{g(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \in S_{++}\}.$$

Then

$$f(\boldsymbol{y}) = g(\boldsymbol{\phi}(\boldsymbol{y}), \boldsymbol{y}) = \boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{b}) + \sum_{j=1}^n \log \phi_j(\boldsymbol{y}) \quad \text{for every } \boldsymbol{y} \in R^m, \tag{3}$$
$$f(\bar{\boldsymbol{y}}) = g(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}) \leq g(\bar{\boldsymbol{x}}, \boldsymbol{y}) \leq f(\boldsymbol{y}) \ \text{ for every } \boldsymbol{y} \in R^m \ \text{ and } \ \bar{\boldsymbol{x}} = \boldsymbol{\phi}(\bar{\boldsymbol{y}}).$$

The *marginal function* f has some nice properties given in the following theorem:

Theorem 2.2. (i) *f is a C^∞.*
(ii) $\boldsymbol{\nabla} f(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{b}$.
(iii) *If $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{y})$ and $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{x})$ (i.e., $\boldsymbol{X}$ is the diagonal matrix with the components of $\boldsymbol{x}$ in its diagonal), then* $\boldsymbol{\nabla}^2 f(\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{X}\left(\boldsymbol{I} - \dfrac{\boldsymbol{x}\boldsymbol{x}^T}{\boldsymbol{x}^T\boldsymbol{x}}\right)\boldsymbol{X}\boldsymbol{A}^T$,
(iv) *$\boldsymbol{\nabla}^2 f(\boldsymbol{y})$ is positive-definite at every $\boldsymbol{y} \in R^m$, hence f is strictly convex.*

Proof: See Appendix. ∎

We now show how to compute $\boldsymbol{\phi}(\boldsymbol{y})$. For every $(\boldsymbol{x}, u, \boldsymbol{y}) \in R^n_{++} \times R^{1+m}$, define

$$\boldsymbol{F}(\boldsymbol{x}, u, \boldsymbol{y}) = \begin{pmatrix} \boldsymbol{\nabla}_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{y}) + u\boldsymbol{e} \\ \boldsymbol{e}^T\boldsymbol{x} - 1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^{-1}\boldsymbol{e} + \boldsymbol{A}^T\boldsymbol{y} + u\boldsymbol{e} \\ \boldsymbol{e}^T\boldsymbol{x} - 1 \end{pmatrix}. \tag{4}$$

Then the maximizer $\boldsymbol{\phi}(\boldsymbol{y})$ of $g(\cdot, \boldsymbol{y})$ in S_{++} is characterized by the stationarity condition

$$\boldsymbol{F}(\boldsymbol{\phi}(\boldsymbol{y}), \psi(\boldsymbol{y}), \boldsymbol{y}) = \boldsymbol{0} \quad \text{and} \quad \boldsymbol{\phi}(\boldsymbol{y}) > \boldsymbol{0} \ , \tag{5}$$

where $\psi(\boldsymbol{y})$ denotes the Lagrange multiplier associated with $\boldsymbol{e}^T\boldsymbol{x} = 1$. The above is reduced to

$$\phi_j(\boldsymbol{y}) = \frac{1}{d_j(\boldsymbol{y}) - \psi(\boldsymbol{y})} \ (j = 1, 2, \ldots, n), \tag{6}$$

$$\sum_{j=1}^{n} \frac{1}{d_j(\boldsymbol{y}) - \psi(\boldsymbol{y})} = 1 \quad \text{and } d_j(\boldsymbol{y}) - \psi(\boldsymbol{y}) > 0 \ (j = 1, 2, \ldots, n) \tag{7}$$

where $\boldsymbol{d}(\boldsymbol{y}) = -\boldsymbol{A}^T\boldsymbol{y}$. The solution $\psi(\boldsymbol{y})$ of the equation in (7) lies in the interval $[d_{min}(\boldsymbol{y}) - n, \ d_{min}(\boldsymbol{y}) - 1]$, where $d_{min}(\boldsymbol{y}) = \min\{d_j(\boldsymbol{y}) : j = 1, 2, \ldots, n\}$, and can be computed by binary search over $[d_{min}(\boldsymbol{y}) - n, \ d_{min}(\boldsymbol{y}) - 1]$ (or by Newton method practically). Therefore, given $\boldsymbol{y} \in R^m$, it is easy to compute $\boldsymbol{\phi}(\boldsymbol{y})$. Both $f(\boldsymbol{y})$ and $\boldsymbol{\nabla} f(\boldsymbol{y})$ can be calculated in $O(n^2)$ arithmetic operations. The computation of the $\boldsymbol{\nabla}^2 f(\boldsymbol{y})$, however, requires $O(n^3)$ operations. Also, given $\boldsymbol{y} \in R^m$ and $\boldsymbol{\Delta y}$, the quadratic approximation $f(\boldsymbol{y}) + \theta \boldsymbol{\nabla} f(\boldsymbol{y})^T \boldsymbol{\Delta y} + \frac{1}{2}\theta^2 \boldsymbol{\Delta y}^T \boldsymbol{\nabla}^2 f(\boldsymbol{y}) \boldsymbol{\Delta y}$ of $f(\boldsymbol{y} + \theta \boldsymbol{\Delta y})$ in terms of the step size θ takes $O(n^2)$ arithmetic operations.

To summarize, the Lagrangian relaxation method transforms the problem of computing the analytic center into the unconstrained minimization of f over R^m, so that $\bar{\boldsymbol{x}}$ is the analytic center if and only if $\bar{\boldsymbol{y}}$ is the minimizer of f and $\bar{\boldsymbol{x}} = \boldsymbol{\phi}(\bar{\boldsymbol{y}})$ is the maximizer of $g(\boldsymbol{x}, \bar{\boldsymbol{y}})$ in S_{++}.

We now restrict attention to the steepest descent method for minimizing f.

Step 0: Let $k = 1$ and $\boldsymbol{y}^1 \in R^m$.
Step 1: Let $\boldsymbol{x}^k = \boldsymbol{\phi}(\boldsymbol{y}^k)$ and $\boldsymbol{d}^k = -\boldsymbol{\nabla} f(\boldsymbol{y}^k) = -(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{b})$.
Step 2: Choose $\theta^k > 0$. Set $\boldsymbol{y}^{k+1} = \boldsymbol{y}^k + \theta^k \boldsymbol{d}^k$, $k = k + 1$, and go to Step 1.

The following theorem indicates great flexibility with regard to step-lengths θ that ensure global convergence to $\bar{\boldsymbol{y}}$.

Theorem 2.3. *If the sequence $\{\theta^k > 0\}$ is bounded, then so is $\{\boldsymbol{y}^k\}$.*

Proof: See Appendix. ∎

We mention two ways of choosing θ^k that ensure global convergence.

Rule 1: Fix $\theta^* > 0$ and $\alpha, \chi \in (0, 1)$. For every k, choose $\theta^k = \chi^p \theta^*$, where $p \geq 0$ is the smallest integer such that $f(\boldsymbol{y}^k + \chi^p \theta^* \boldsymbol{d}^k) \leq f(\boldsymbol{y}^k) - \alpha \chi^p \theta^* \|\boldsymbol{d}^k\|^2$.

Rule 2: Fix $\theta^1 > 0$ and α and χ as in Rule 1. For every $k \geq 2$ let

$$\theta^k = \begin{cases} \chi\theta^{k-1} & \text{if } f(\boldsymbol{x}^k) > f(\boldsymbol{x}^{k-1}) - \alpha\theta^k \|\boldsymbol{d}^k\|^2, \\ \theta^{k-1} & \text{otherwise,} \end{cases}$$

Rule 1 is Armijo's rule (see Luenberger [3]). Rule 2 is simpler since it requires only one evaluation of f per iteration. In both cases, $\{\theta^k\}$ is bounded. By Theorem 2.3, there exists a compact $D \subset R^m$ such that $y^k \in D$ for all k. Since f is continuous, there is a κ such that

$$f(\boldsymbol{y}^k) \geq \kappa \quad (k = 1, 2, \ldots) \ . \tag{8}$$

On the other hand, since f is C^∞, there is a positive number $\bar{\theta}$ such that

$$f(\boldsymbol{y} + \theta\boldsymbol{\nabla} f(\boldsymbol{y})) \leq f(\boldsymbol{y}) - \theta\alpha\|\boldsymbol{\nabla} f(\boldsymbol{y})\|^2 \text{ if } \boldsymbol{y} \in D \text{ and } 0 \leq \theta \leq \bar{\theta}.$$

Therefore, the inequalities

$$\theta^k \geq \chi\bar{\theta} \text{ and } f(\boldsymbol{y}^{k+1}) = f(\boldsymbol{y}^k + \theta^k\boldsymbol{d}^k) \leq f(\boldsymbol{y}^k) - \theta^k\alpha\|\boldsymbol{d}^k\|^2 \tag{9}$$

hold for all k in the case of Rule 1, and for every $k \geq k^* > 0$ in the case of Rule 2. Let $k^* = 1$ in the case of Rule 1. Suppose, on the contrary, that $\|\boldsymbol{d}^k\| = \|\boldsymbol{\nabla} f(\boldsymbol{y}^k)\|$ does not converge to zero. Then, we can find a $\delta > 0$ and an infinite subsequence K of $\{k = k^*, k^* + 1, \ldots\}$ such that $\|\boldsymbol{d}^k\| \geq \delta$ for every $k \in K$. Hence we have by (9) that

$$\begin{aligned} &f(\boldsymbol{y}^{k+1}) \leq f(\boldsymbol{y}^k) - \chi\bar{\theta}\alpha\delta^2 \text{ for every } k \in K, \\ &f(\boldsymbol{y}^{k+1}) \leq f(\boldsymbol{y}^k) \text{ for every } k \geq k^*. \end{aligned}$$

This implies $f(\boldsymbol{y}^k) \to -\infty$ as $k \to \infty$, which contradicts (8). Thus, $\|\boldsymbol{d}^k\| = \|\boldsymbol{\nabla} f(\boldsymbol{y}^k)\|$ converges to zero. Since f is strictly convex and C^∞, we conclude that $\{\boldsymbol{y}^k\}$ converges to the unique minimizer of f.

3. Concluding Remarks

We obtain $-\boldsymbol{\nabla}^2 f(\boldsymbol{y})^{-1}(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{b})$ as a search direction if we apply Newton's method for minimizing f. We can derive this direction from the stationarity condition for $\boldsymbol{x} \in R_{++}$ to be the analytic center, *i.e.*,

$$\boldsymbol{F}(\boldsymbol{x}, u, \boldsymbol{y}) = \boldsymbol{0}, \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \text{ and } \boldsymbol{x} > \boldsymbol{0} \tag{10}$$

If $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{y})$ then all the relations in (10), except perhaps $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, hold for some $u \in R$. Hence Newton's equations for approximating a solution of the equations in (10) at $(\boldsymbol{x}, u, \boldsymbol{y})$ are

$$\boldsymbol{A}^T\boldsymbol{\Delta}\boldsymbol{y} - \boldsymbol{X}^{-2}\boldsymbol{\Delta}\boldsymbol{x} + (du)\boldsymbol{e} = \boldsymbol{0}, \quad \boldsymbol{A}\boldsymbol{\Delta}\boldsymbol{x} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \quad \text{and} \quad \boldsymbol{e}^T\boldsymbol{\Delta}\boldsymbol{x} = 0 \ .$$

It can be easily verified that $\boldsymbol{\Delta y}$ coincides with the direction $-\boldsymbol{\nabla}^2 f(\boldsymbol{y})^{-1}(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{b})$. This direction is closely related to the one used in the primal-dual infeasible-interior-point algorithms of [4] and [1].

The problem of minimizing the logarithmic barrier function $\boldsymbol{c}^T\boldsymbol{x} - \mu \sum_{j=1}^{n} \log x_j$ $(\boldsymbol{x} \in P_{++})$ over P_{++} is directly related with interior-point methods for linear programs. Here $\boldsymbol{c} \in R^n$ is a constant vector and $\mu > 0$ is a parameter. We can easily adapt the Lagrange relaxation method to this problem. In fact, if we replace the Lagrangian function g (see (1)) by

$$g(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{c}^T\boldsymbol{x} + \mu \sum_{j=1}^{n} \log x_j \quad ((\boldsymbol{x}, \boldsymbol{y}) \in S_{++} \times R^m_{++}),$$

all the results presented so far remain valid subject to minor modifications as follows.

- Replace the Hessian matrix of $\boldsymbol{f}$ by $\boldsymbol{\nabla}^2 f(\boldsymbol{y}) = \dfrac{1}{\mu}\boldsymbol{A}\boldsymbol{X}\left(\boldsymbol{I} - \dfrac{\boldsymbol{x}\boldsymbol{x}^T}{\boldsymbol{x}^T\boldsymbol{x}}\right)\boldsymbol{X}\boldsymbol{A}^T$.
- Replace $\boldsymbol{d}(\boldsymbol{y}) = -\boldsymbol{A}^T\boldsymbol{y}$ by $\boldsymbol{d}(\boldsymbol{y}) = \boldsymbol{c} - \boldsymbol{A}^T\boldsymbol{y}$, and (4), (6) and (7) by

$$\boldsymbol{F}(\boldsymbol{x}, u, \boldsymbol{y}) = \begin{pmatrix} \boldsymbol{\nabla}_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{y}) + u\boldsymbol{e} \\ \boldsymbol{e}^T\boldsymbol{x} - 1 \end{pmatrix} = \begin{pmatrix} -\boldsymbol{c} + \mu\boldsymbol{X}^{-1}\boldsymbol{e} + \boldsymbol{A}^T\boldsymbol{y} + u\boldsymbol{e} \\ \boldsymbol{e}^T\boldsymbol{x} - 1 \end{pmatrix},$$

$$\phi_j(\boldsymbol{y}) = \frac{\mu}{d_j(\boldsymbol{y}) - \psi(\boldsymbol{y})} \ (j = 1, 2, \ldots, n),$$

$$\sum_{j=1}^{n} \frac{\mu}{d_j(\boldsymbol{y}) - \psi(\boldsymbol{y})} = 1 \ \text{ and } d_j(\boldsymbol{y}) - \psi(\boldsymbol{y}) > 0 \ (j = 1, 2, \ldots, n)$$

respectively. The solution $\psi(\boldsymbol{y})$ of the last equation now lies in the interval $[d_{min}(\boldsymbol{y}) - n\mu,\ d_{min}(\boldsymbol{y}) - \mu]$.

The efficiency of the Lagrangian relaxation method is determined by how efficiently we can approximate $\bar{\boldsymbol{y}}$. We have shown the global convergence of the steepest descent method. In practice, however, the steepest descent method is not expected to work as efficiently as sophisticated methods such as conjugate directions and quasi-Newton methods. Thus, it would be very interesting to investigate how such techniques can be combined effectively with the Lagrangian relaxation method.

References

[1] M. Kojima, N. Megiddo, and S. Mizuno, "A primal-dual infeasible-interior-point algorithm for linear programming," *Mathematical Programming*, to appear.

[2] M. Kojima, N. Megiddo, and S. Mizuno, "A conjugate direction method for approximating the analytic center of a polytope," Research Report RJ 8540 IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120-6099, November 1991.

[3] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, Menlo Park, California, 1984.

[4] I. J. Lustig, "Feasibility issues in a primal-dual interior-point method for linear programming," *Mathematical Programming* **49** (1990/91) 145–162.

[5] J. Renegar, "A polynomial-time algorithm based on Newton's method for linear programming," *Mathematical Programming* **40** (1988) 59–94.

[6] G. Sonnevend, "An 'analytical centre' for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming," in: *Lecture Notes in Control and Information Sciences 84*, Springer, New York, 1985, pp. 866–876.

[7] M. J. Todd, "Recent developments and new directions in linear programming," in: N. Iri and K. Tanabe, eds., *Mathematical Programming, Recent Developments and Applications*, Kluwer Academic Publishers, London, 1989, pp. 109–157.

Appendix

Proof of Theorem 2.2:

For the proof of (i), recall that $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{y})$, the unique maximizer of $g(\cdot, \boldsymbol{y})$ in $\boldsymbol{x} \in S_{++}$, satisfies the stationarity condition (5). The Jacobian matrix $\boldsymbol{DF}(\boldsymbol{x}, u, \boldsymbol{y})$ of $\boldsymbol{F}$ with respect to $(\boldsymbol{x}, u)$ is

$$\boldsymbol{DF}(\boldsymbol{x}, u, \boldsymbol{y}) = \begin{pmatrix} -\boldsymbol{X}^{-2} & \boldsymbol{e} \\ \boldsymbol{e}^T & \boldsymbol{0} \end{pmatrix},$$

which is nonsingular for $\boldsymbol{x} \in R^n_{++}$. Thus, by the implicit function theorem, $\boldsymbol{\phi}$ is C^∞. Since f can be written as in (3), and g is C^∞, it follows that f is C^∞. See, *e.g.*, Appendix A of [3].

(ii) Differentiating the equality in (3) with respect to y_i, we get

$$\begin{aligned} \frac{\partial f(\boldsymbol{y})}{\partial y_i} &= \sum_{j=1}^{n} \frac{\partial g(\boldsymbol{\phi}(\boldsymbol{y}), \boldsymbol{y})}{\partial x_j} \frac{\partial x_j(\boldsymbol{y})}{\partial y_i} + \frac{\partial g(\boldsymbol{\phi}(\boldsymbol{y}), \boldsymbol{y})}{\partial y_i} \\ &= \sum_{j=1}^{n} \frac{\partial g(\boldsymbol{\phi}(\boldsymbol{y}), \boldsymbol{y})}{\partial x_j} \frac{\partial x_j(\boldsymbol{y})}{\partial y_i} + \boldsymbol{A}_i \boldsymbol{\phi}(\boldsymbol{y}) - b_i, \end{aligned}$$

where $\boldsymbol{A}_i$ is the ith row of $\boldsymbol{A}$. On the other hand, it follows from (5) that

$$\left.\begin{array}{ll} \dfrac{\partial g(\boldsymbol{\phi}(\boldsymbol{y}),\boldsymbol{y})}{\partial x_j} + \psi(\boldsymbol{y}) = 0 & (j = 1,2,\ldots,n) \\ \boldsymbol{e}^T \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{y})}{\partial y_i} = 0 & (i = 1,2,\ldots,m) \end{array}\right\} \quad \text{for every } \boldsymbol{y} \in R^m.$$

Therefore,

$$\frac{\partial f(\boldsymbol{y})}{\partial y_i} \;=\; -\psi(\boldsymbol{y}) \sum_{j=1}^{n} \frac{\partial \phi_j(\boldsymbol{y})}{\partial y_i} + \boldsymbol{A}_i \boldsymbol{\phi}(\boldsymbol{y}) - b_i = \boldsymbol{A}_i \boldsymbol{\phi}(\boldsymbol{y}) - b_i \quad (i = 1,2,\ldots,m)$$

(iii) Differentiating the equality in (5) with respect to y_i

$$\frac{\partial \boldsymbol{F}(\boldsymbol{\phi}(\boldsymbol{y}),\psi(\boldsymbol{y}),\boldsymbol{y})}{\partial y_i} = \begin{pmatrix} -(\boldsymbol{\Phi}(\boldsymbol{y}))^{-2} \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{y})}{\partial y_i} + \dfrac{\partial \psi(\boldsymbol{y})}{\partial y_i}\boldsymbol{e} + \boldsymbol{A}_i^T \\ \boldsymbol{e}^T \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{y})}{\partial y_i} \end{pmatrix} = \boldsymbol{0}.$$

Here $\boldsymbol{\Phi}(\boldsymbol{y}) = \mathrm{diag}(\boldsymbol{\phi}(\boldsymbol{y}))$. By solving the linear equations for $\partial\psi(\boldsymbol{y})/\partial y_i$ and $\partial\boldsymbol{\phi}(\boldsymbol{y})/\partial y_i$, we obtain that

$$\left.\begin{array}{l} \dfrac{\partial \psi(\boldsymbol{y})}{\partial y_i} = -\dfrac{\boldsymbol{\phi}(\boldsymbol{y})^T}{\boldsymbol{\phi}(\boldsymbol{y})^T\boldsymbol{\phi}(\boldsymbol{y})}\, \boldsymbol{\Phi}(\boldsymbol{y})\boldsymbol{A}_i^T, \\ \dfrac{\partial \boldsymbol{\phi}(\boldsymbol{y})}{\partial y_i} = \boldsymbol{\Phi}(\boldsymbol{y}) \left(\boldsymbol{I} - \dfrac{\boldsymbol{\phi}(\boldsymbol{y})\boldsymbol{\phi}(\boldsymbol{y})^T}{\boldsymbol{\phi}(\boldsymbol{y})^T\boldsymbol{\phi}(\boldsymbol{y})} \right) \boldsymbol{\Phi}(\boldsymbol{y})\boldsymbol{A}_i^T \end{array}\right\} \quad (i = 1,2,\ldots,m)$$

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

$$\frac{\partial f(\boldsymbol{\phi}(\boldsymbol{y}))}{\partial y_i} = \boldsymbol{A}_i \boldsymbol{\phi}(\boldsymbol{y}) - b_i \quad (i = 1,2,\ldots,m).$$

Hence, for every i $(i = 1,2,\ldots,m)$ and k $(k = 1,2,\ldots,m)$,

$$\frac{\partial^2 f(\boldsymbol{\phi}(\boldsymbol{y}))}{\partial y_i \partial y_k} = \boldsymbol{A}_i \frac{\partial \boldsymbol{\phi}(\boldsymbol{y})}{\partial y_k} = \boldsymbol{A}_i \boldsymbol{\Phi}(\boldsymbol{y}) \left(\boldsymbol{I} - \frac{\boldsymbol{\phi}(\boldsymbol{y})\boldsymbol{\phi}(\boldsymbol{y})^T}{\boldsymbol{\phi}(\boldsymbol{y})^T\boldsymbol{\phi}(\boldsymbol{y})} \right) \boldsymbol{\Phi}(\boldsymbol{y})\boldsymbol{A}_k^T.$$

This implies (iii).

(iv) Let $\boldsymbol{y} \in R^m$ and $\boldsymbol{x} = \boldsymbol{\phi}(\boldsymbol{y})$. Obviously, $\boldsymbol{\nabla}^2 f(\boldsymbol{y})$ is symmetric and positive semi-definite. It suffices to show that there is no $\boldsymbol{0} \neq \boldsymbol{v} \in R^m$ for which

$$\boldsymbol{0} = \left(\boldsymbol{I} - \frac{\boldsymbol{x}\boldsymbol{x}^T}{\boldsymbol{x}^T\boldsymbol{x}} \right) \boldsymbol{X}\boldsymbol{A}^T\boldsymbol{v} = \boldsymbol{X} \left(\boldsymbol{A}^T\boldsymbol{v} - \frac{\boldsymbol{x}^T\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{v}}{\boldsymbol{x}^T\boldsymbol{x}}\boldsymbol{e} \right).$$

Assume, on the contrary, that such a $\boldsymbol{v}$ exists. Since $\boldsymbol{X} = \mathrm{diag}(\boldsymbol{x})$,

$$\boldsymbol{A}^T\boldsymbol{v} - \boldsymbol{e}\frac{\boldsymbol{x}^T\boldsymbol{X}\boldsymbol{A}^T\boldsymbol{v}}{\boldsymbol{x}^T\boldsymbol{x}} = \boldsymbol{0}\ .$$

This contradicts (ii) of Condition 2.1.

Proof of Theorem 2.3:

For $\boldsymbol{y} \in B \equiv \{\boldsymbol{y} \in R^m : \|\boldsymbol{y}\| = 1\}$, let $\boldsymbol{\xi}(\boldsymbol{y})$ be the minimizer of $\|\boldsymbol{\xi}\|^2$ subject to $\boldsymbol{A}\boldsymbol{\xi} = \boldsymbol{y}$ and $\boldsymbol{e}^T\boldsymbol{\xi} = 0$. It follows that $\boldsymbol{\xi} : B \to R^n$ is continuous. So there is $\gamma > 0$ such that

$$\bar{\boldsymbol{x}} + \gamma\boldsymbol{\xi}(\boldsymbol{y}) > \boldsymbol{0} \text{ for every } \boldsymbol{y} \in B. \tag{11}$$

Here $\bar{\boldsymbol{x}}$ denotes the analytic center of P_+. Since $\bar{\boldsymbol{x}} \in P_{++}$,

$$\boldsymbol{A}(\bar{\boldsymbol{x}} + \gamma\boldsymbol{\xi}(\boldsymbol{y})) - \boldsymbol{b} = \gamma\boldsymbol{y} \quad \text{and} \quad \boldsymbol{e}^T(\bar{\boldsymbol{x}} + \gamma\boldsymbol{\xi}(\boldsymbol{y})) = 1 \quad \text{for every } \boldsymbol{y} \in B. \tag{12}$$

Define positive numbers θ^*, α, β and λ^* such that

$$\theta^* = \sup\{\theta^k : k = 1, 2, \dots\}, \tag{13}$$

$$\alpha = \min\left\{\sum_{j=1}^{n} \log(\bar{x}_j + \gamma\xi_j(\boldsymbol{y})) : \boldsymbol{y} \in B\right\}, \tag{14}$$

$$\beta = \max\{\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\| : \boldsymbol{x} \in S_+\}, \tag{15}$$

$$\lambda^* = \max\left\{\frac{(\theta^*\beta^2 - \alpha - n\log n)}{\gamma}, \quad \|\boldsymbol{y}^1\|\right\}. \tag{16}$$

We show that

$$\boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}) - \boldsymbol{b}) \geq \theta^*\beta^2 \quad \text{if } \|\boldsymbol{y}\| \geq \lambda^*. \tag{17}$$

Suppose $\boldsymbol{y} \in B$ and $\lambda \geq \lambda^*$; hence $\|\lambda\boldsymbol{y}\| \geq \lambda^*$. Since $\boldsymbol{\phi}(\lambda\boldsymbol{y}) \in S_{++}$ is by definition the maximizer of $g(\boldsymbol{x}, \lambda\boldsymbol{y})$ in S_{++}, and $\bar{\boldsymbol{x}} + \gamma\boldsymbol{\xi}(\boldsymbol{y}) \in S_{++}$ (see (11) and (12), we get

$$\begin{aligned}
&\lambda\boldsymbol{y}^T(\boldsymbol{A}(\boldsymbol{\phi}(\lambda\boldsymbol{y}) - \boldsymbol{b}) \\
&\geq \quad -\sum_{j=1}^{n} \log\phi_j(\lambda\boldsymbol{y}) + \lambda\boldsymbol{y}^T(\boldsymbol{A}(\bar{\boldsymbol{x}} + \gamma\boldsymbol{\xi}(\boldsymbol{y})) - \boldsymbol{b}) + \sum_{j=1}^{n} \log(\bar{x}_j + \gamma\xi_j(\boldsymbol{y})) \\
&\geq \quad -\sum_{j=1}^{n} \log(1/n) + \lambda\boldsymbol{y}^T(\gamma\boldsymbol{y}) + \alpha \quad (\text{by } \boldsymbol{\phi}(\lambda\boldsymbol{y}) \in S_{++}, \text{ (12) and (14)}) \\
&\geq \quad n\log n + \lambda^*\gamma + \alpha \quad (\text{by } \lambda \geq \lambda^* \text{ and } \boldsymbol{y} \in B) \\
&\geq \quad \theta^*\beta. \quad (\text{by (16)})
\end{aligned}$$

Thus we have proven (17).

Now we prove that

$$\|\boldsymbol{y}^k\| \leq \lambda^* + \theta^*\beta \quad \text{for every } k \ (k = 1, 2, \dots) \ . \tag{18}$$

By the definition (16) of λ^* and $\theta^*\beta \geq 0$, we know that $\|\boldsymbol{y}^1\| \leq \lambda^* + \theta^*\beta$. Assuming that $\|\boldsymbol{y}^k\| \leq \lambda^* + \theta^*\beta$, we will show that $\|\boldsymbol{y}^{k+1}\| \leq \lambda^* + \theta^*\beta$. By definition,

$$\boldsymbol{y}^{k+1} = \boldsymbol{y}^k - \theta^k(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}^k) - \boldsymbol{b}).$$

If $\|\boldsymbol{y}^k\| \leq \lambda^*$, we see from (13) and (15) that

$$\|\boldsymbol{y}^{k+1}\| \leq \|\boldsymbol{y}^k\| + \theta^k\|\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}^k) - \boldsymbol{b}\| \leq \lambda^* + \theta^*\beta.$$

If $\lambda^* < \|\boldsymbol{y}^k\| \leq \lambda^* + \theta^*\beta$, then

$$\begin{aligned}
\|\boldsymbol{y}^{k+1}\|^2 &= \left\|\boldsymbol{y}^k - \theta^k(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}^k) - \boldsymbol{b})\right\|^2 \\
&\leq \|\boldsymbol{y}^k\|^2 - 2\theta^k(\boldsymbol{y}^k)^T(\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}^k) - \boldsymbol{b}) + (\theta^k\|\boldsymbol{A}\boldsymbol{\phi}(\boldsymbol{y}^k) - \boldsymbol{b}\|)^2 \\
&\leq \|\boldsymbol{y}^k\|^2 - 2\theta^k\theta^*\beta^2 + (\theta^k\beta)^2 \quad \text{(by (15) and (17))} \\
&\leq \|\boldsymbol{y}^k\|^2 \quad \text{(by (13))} \\
&\leq \lambda^* + \theta^*\beta.
\end{aligned}$$

Thus we have proven (18). This completes the proof of Theorem 2.3.