J Glob Optim (2009) 45:337–353
DOI 10.1007/s10898-008-9378-7

Global optimization for a class of fractional programming problems

Shu-Cherng Fang · David Y. Gao · Ruey-Lin Sheu · Wenxun Xing

Received: 9 October 2007 / Accepted: 11 November 2008 / Published online: 2 December 2008

Abstract This paper presents a canonical dual approach to minimizing the sum of a quadratic function and the ratio of two quadratic functions, which is a type of non-convex optimization problem subject to an elliptic constraint. We first relax the fractional structure by introducing a family of parametric subproblems. Under proper conditions on the "problem-defining" matrices associated with the three quadratic functions, we show that the canonical dual of each subproblem becomes a one-dimensional concave maximization problem that exhibits no duality gap. Since the infimum of the optima of the parameterized subproblems leads to a solution to the original problem, we then derive some optimality conditions and existence conditions for finding a global minimizer of the original problem. Some numerical results using the quasi-Newton and line search methods are presented to illustrate our approach.

Keywords Quadratic fractional programming · Sum-of-ratios · Global optimization · Canonical duality

S.-C. Fang (✉)
Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, USA
e-mail: fang@eos.ncsu.edu

D. Y. Gao
Department of Mathematics, Virginia Tech, Blacksburgh, VA, USA
e-mail: gao@vt.edu

R.-L. Sheu
Department of Mathematics, National Cheng Kung University, Tainan, Taiwan, ROC
e-mail: rsheu@mail.ncku.edu.tw

W. Xing
Department of Mathematical Sciences, Tsinghua University, Beijing, China
e-mail: wxing@math.tsinghua.edu.cn

1 Introduction

We study in this paper the following quadratic fractional programming problem:

$$(\mathcal{P}): \quad \min \left\{ P_0(\mathbf{x}) = f(\mathbf{x}) + \frac{g(\mathbf{x})}{h(\mathbf{x})} \quad : \quad \mathbf{x} \in \mathcal{X} \right\} \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ and

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x}, \;\; g(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T G\mathbf{x}, \quad h(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T H\mathbf{x} - \mathbf{b}^T\mathbf{x},$$

with $Q \in \mathbb{R}^{n\times n}$ being symmetric, $G \in \mathbb{R}^{n\times n}$ symmetric positive semi-definite, $H \in \mathbb{R}^{n\times n}$ symmetric negative definite and $\mathbf{f}, \mathbf{b} \in \mathbb{R}^n$. Assume that $\mu_0^{-1} = h(H^{-1}\mathbf{b}) > 0$ and $\delta \in (0, \mu_0^{-1}]$, then the feasible domain $\mathcal{X}$ is defined to be

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) \geq \delta > 0\},$$

which forms a constraint of elliptic type.

Problem $(\mathcal{P})$ belongs to a class of "sum-of-ratios" problems that have been actively studied for several decades. The ratios often stand for efficiency measures representing performance-to-cost, profit-to-revenue, return-to-risk, or signal-to-noise for numerous applications in economics, transportation science, finance, engineering, etc. [1,6,10,16,18,19,27,30]. Depending on the nature of each application, the functions f, g, h can be affine, convex, concave, or neither. However, even for the simplest case in which f, g, h are all affine functions, problem $(\mathcal{P})$ is still a global optimization problem that may have multiple local optima [5,26]. In particular, Freund and Jarre [12] showed that the sum-of-ratios problem $(\mathcal{P})$ is NP-complete when f, g are convex and h is concave (Our setting fits this category.). Due to computational complexity, most known algorithms work on the problems with linear-ratios using the branch-and-bound approach [2–4,17,21], although there do exist some different approaches [25,32]. Related work on nonlinear fractional programming can be referred to [22–24].

Due to the non-convexity involved in the fractional structure, the ordinary Lagrangean dual only provides a weak duality theorem that may bear a positive duality gap. Interestingly, Scott and Jefferson [29] proposed a signomial dual [9,28] for the sum-of-affine-ratios problems. In their approach, the strong duality theorem holds but the weak duality theorem is missing. Notice that the optimal solutions, when f, g, h are affine and linearly independent, always lie on the boundary of the feasible region (see Craven [7]). In this paper, we are motivated by this property to develop a canonical dual approach based on Gao and others' work [11,13,31] for solving problem $(\mathcal{P})$.

In Sect. 2, we first parameterize problem $(\mathcal{P})$ into a family of subprograms $\{(\mathcal{P}_\mu)\}$, in which each subproblem is a (possibly non-convex) quadratic program subject to one quadratic constraint. Similar parametric idea can be found in [12,20] Then, we show the infimum of the optima of the parameterized subproblems provides a solution to problem $(\mathcal{P})$. Since each subproblem $(\mathcal{P}_\mu)$ may be a non-convex problem, a canonical dual problem $(\mathcal{P}_\mu^d)$ is derived. We provide some sufficient conditions to establish both the weak and strong duality theorems (the so called *perfect duality*) for the pair of $(\mathcal{P}_\mu)$and$(\mathcal{P}_\mu^d)$. In Sect. 3, we study the topological properties of the feasible domain of the canonical dual problem and find the domain is a one-dimensional ray that could be open or closed and whose boundary can be characterized by the largest eigenvalue of a matrix composed of the problem-defining matrices Q, G and H. Then we develop some existence conditions under which a global optimizer of the original problem $(\mathcal{P})$ can indeed be identified by solving the corresponding canonical dual problems. In Sect. 4, we provide two numerical examples. The first example is a one-dimensional problem whose dual can be analytically solved to verify the correctness of the proposed approach. The second one is a two-dimensional example which sheds some lights on the numerical issues arising from the canonical dual. We show that a quasi-Newton method with line search using the Armijo's rule works efficiently.

2 Sufficiency for global optimality

In order to solve problem $(\mathcal{P})$, consider the following family of parameterized subproblem:

$$(\mathcal{P}_\mu): \quad \min \left\{ P_\mu(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \mu g(\mathbf{x}) \quad : \quad \mathbf{x} \in \mathcal{X}_\mu \right\}, \tag{2}$$

where $\mu \in [\mu_0, \delta^{-1}]$ and

$$\mathcal{X}_\mu = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) \geq \mu^{-1} \geq \delta > 0\}$$

is a convex set. We immediately have the following result:

Lemma 1 *Problem $(\mathcal{P})$ is equivalent to $(\mathcal{P}_\mu)$ in the sense that*

$$\inf_{\mathbf{x}\in\mathcal{X}} P_0(\mathbf{x}) = \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{\mathbf{x}\in\mathcal{X}_\mu} P_\mu(\mathbf{x}). \tag{3}$$

Proof It is easy to see that

$$\begin{aligned}
&\inf_{\mathbf{x}\in\mathcal{X}} P_0(\mathbf{x}) \\
&= \inf_{\mathbf{x}\in\mathcal{X}} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \frac{g(\mathbf{x})}{h(\mathbf{x})} \right\} \\
&= \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{h(\mathbf{x})=\mu^{-1}} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \frac{g(\mathbf{x})}{h(\mathbf{x})} \right\} \\
&= \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{h(\mathbf{x})=\mu^{-1}} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \mu g(\mathbf{x}) \right\} \\
&\geq \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{\mathbf{x}\in\mathcal{X}_\mu} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \mu g(\mathbf{x}) \right\} \\
&= \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{\mathbf{x}\in\mathcal{X}_\mu} P_\mu(\mathbf{x}).
\end{aligned}$$

Conversely,

$$\begin{aligned}
&\inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{\mathbf{x}\in\mathcal{X}_\mu} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \mu g(\mathbf{x}) \right\} \\
&= \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{h(\mathbf{x})\geq\mu^{-1}} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \mu g(\mathbf{x}) \right\} \\
&\geq \inf_{\mu\in[\mu_0,\delta^{-1}]} \inf_{h(\mathbf{x})\geq\mu^{-1}} \left\{ \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \frac{g(\mathbf{x})}{h(\mathbf{x})} \right\} \ (\text{since } g(\mathbf{x}) \geq 0) \\
&= \inf_{\mathbf{x}\in\mathcal{X}} P_0(\mathbf{x}).
\end{aligned}$$

This completes the proof of the lemma. □

Now, for any given $\mu \in [\mu_0, \delta^{-1}]$, we define

$$G_\mu(\sigma) = Q + \mu G - \sigma H, \quad \text{for } \sigma \geq 0, \tag{4}$$

$$\mathcal{S}_\mu^+ = \{\sigma \geq 0 \mid \; G_\mu(\sigma) \succ 0\}, \tag{5}$$

$$P_\mu^d(\sigma) = \frac{\sigma}{\mu} - \frac{1}{2}(\mathbf{f} - \sigma \mathbf{b})^T G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma \mathbf{b}), \tag{6}$$

where '$\succ$' means positive definiteness of a matrix. In the following we let $\partial \mathcal{S}_\mu^+$ denote the boundary of $\mathcal{S}_\mu^+$, $\det G_\mu(\sigma)$ be the determinant of the matrix $G_\mu(\sigma)$ and $\sigma_{\max}$ represent the maximum root of the equation defined by $\det G_\mu(\sigma) = 0$. Then we have the following topological properties of $\mathcal{S}_\mu^+$.

Lemma 2 *Given any $\mu \in [\mu_0, \delta^{-1}]$, then*

(a) $G_\mu(\sigma) \succ 0$ *as σ becomes large enough*;
(b) $\mathcal{S}_\mu^+$ *is a ray in* $\mathbb{R}^1$;
(c) $\partial \mathcal{S}_\mu^+ = \{\sigma \in \mathbb{R}^1 | \sigma = \max\{0, \sigma_{max}\}\}$.

Proof (a) Since $-H$ is positive definite, there exists $\gamma > 0$ such that $-H - \gamma I \succ 0$ with I being an identity matrix. Notice that when σ is large enough, the matrix $Q + \mu G + \sigma \gamma I$ is diagonally dominant with positive diagonal elements. From [8], we know it is positive definite. Consequently, $G_\mu(\sigma)$ is positive definite as σ becomes large enough.
(b) Assume that $\bar{\sigma} \in \mathcal{S}_\mu^+$, then $Q + \mu G - \bar{\sigma} H \succ 0$. Hence for any $\sigma > \bar{\sigma} \geq 0$, $Q + \mu G - \sigma H = (Q + \mu G - \bar{\sigma} H) + (\sigma - \bar{\sigma})(-H) \succ 0$. This means $\mathcal{S}_\mu^+$ is a ray.
(c) Since H is symmetric negative definite, we can write $-H = LL^T$ such that L is lower triangular with positive diagonal elements. In this way, we have

$$G_\mu(\sigma) = L(B + \sigma I)L^T, \tag{7}$$

where $B = L^{-1}(Q + \mu G)(L^{-1})^T$. It is easy to see that

$$G_\mu \succ 0 \Longleftrightarrow \; B + \sigma I \succ 0 \tag{8}$$

and

$$\det G_\mu(\sigma) = 0 \Longleftrightarrow \det(B + \sigma I) = 0. \tag{9}$$

It follows from (9) that if $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ are the roots of $\det G_\mu(\sigma) = 0$, then they are the eigenvalues of the matrix $-B$. Therefore, $\sigma - \sigma_1 \leq \sigma - \sigma_2 \leq \cdots \leq \sigma - \sigma_n$ are the eigenvalues of the matrix $B + \sigma I$. By (8), we know that the smallest eigenvalue $\sigma - \sigma_1$ of $B + \sigma I$ is positive if and only if $G_\mu(\sigma)$ is positive definite. Consequently, when $\sigma_1 < 0$, the matrix $G_\mu(\sigma)$ must be positive definite for $\sigma \geq 0$. In this case, $\partial \mathcal{S}_\mu^+ = \{0\}$. On the other hand, if $\sigma_1 > 0$, then the maximum root $\sigma_{\max} = \sigma_1$ becomes the boundary point of $\mathcal{S}_\mu^+$. □

Lemma 3 *For any given $\mu \in [\mu_0, \delta^{-1}]$, the canonical dual function $P_\mu^d(\sigma)$ is concave and C^1 continuous with a decreasing derivative over $\mathcal{S}_\mu^+$.*

Proof By direct calculation, we have

$$\frac{d}{d\sigma} P_\mu^d(\sigma) = \frac{1}{\mu} - \mathbf{x}(\sigma)^T \left(\frac{1}{2} H \mathbf{x}(\sigma) - \mathbf{b} \right) \tag{10}$$

and

$$\frac{d^2}{d\sigma^2} P_\mu^d(\sigma) = -(H\mathbf{x}(\sigma)^T - \mathbf{b})^T G_\mu^{-1}(\sigma)(H\mathbf{x}(\sigma)^T - \mathbf{b}), \tag{11}$$

where $\mathbf{x}(\sigma) = G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b})$. It is obvious that $\frac{d^2}{d\sigma^2} P_\mu^d(\sigma) \le 0$ provided that $G_\mu(\sigma)$ is positive definite. This completes the proof. □

Given any $\mu \in [\mu_0, \delta^{-1}]$, consider the following canonical dual problem $(\mathcal{P}_\mu^d)$:

$$(\mathcal{P}_\mu^d) \quad \sup\left\{ P_\mu^d(\sigma) \quad : \quad \sigma \in \mathcal{S}_\mu^+ \right\}.$$

Theorem 1 (Weak Duality) *If there exists a global maximizer σ_μ of $P_\mu^d(\sigma)$ over $\mathcal{S}_\mu^+$, then the vector*

$$\mathbf{x}_\mu = G_\mu^{-1}(\sigma_\mu)(\mathbf{f} - \sigma_\mu \mathbf{b}) \tag{12}$$

is a global minimizer of $(\mathcal{P}_\mu)$ over $\mathcal{X}_\mu$ and

$$P_\mu^d(\sigma) \le P_\mu(\mathbf{x}), \quad \forall\, (\mathbf{x}, \sigma) \in \mathcal{X}_\mu \times \mathcal{S}_\mu^+. \tag{13}$$

Proof Let $\Lambda(\cdot) : \mathbb{R}^n \to \mathbb{R}$ be the so-called *geometrical transformation* (see [13–15]) defined by

$$y = \Lambda(\mathbf{x}) = \mu^{-1} + \mathbf{b}^T\mathbf{x} - \frac{1}{2}\mathbf{x}^T H\mathbf{x} \tag{14}$$

and let

$$V(y) = \begin{cases} 0 & \text{if } y \le 0; \\ +\infty & \text{otherwise,} \end{cases} \tag{15}$$

whose conjugate function is

$$V^\sharp(\sigma) = \begin{cases} 0 & \text{if } \sigma \ge 0; \\ +\infty & \text{otherwise.} \end{cases}$$

Then, Problem $(\mathcal{P}_\mu)$ in (2) can be written as the following unconstrained optimization problem

$$\min\left\{ P(\mathbf{x}) = V(\Lambda(\mathbf{x})) + \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \frac{\mu}{2}\mathbf{x}^T G\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \right\}. \tag{16}$$

Following [15], we define the so-called "total complementary function" as

$$\Xi(\mathbf{x}, \sigma) = \Lambda(\mathbf{x})^T\sigma - V^\sharp(\sigma) + \frac{1}{2}\mathbf{x}^T Q\mathbf{x} - \mathbf{f}^T\mathbf{x} + \frac{\mu}{2}\mathbf{x}^T G\mathbf{x} \tag{17}$$

for $\mathbf{x} \in \mathbb{R}^n$ and $\sigma \in \mathcal{S}_\mu^+$. Since $V^\sharp(\sigma) = 0$ when $\sigma \ge 0$ and $\Lambda(\mathbf{x})$ can be substituted by (14), the total complementary function can be simplified as

$$\Xi(\mathbf{x}, \sigma) = \frac{\sigma}{\mu} + \frac{1}{2}\mathbf{x}^T(G_\mu(\sigma))\mathbf{x} - (\mathbf{f} - \sigma\mathbf{b})^T\mathbf{x}, \tag{18}$$

where $G_\mu(\sigma)$ is defined in (4). Note that $\Xi(\mathbf{x}, \sigma)$ is convex in $\mathbf{x} \in \mathbb{R}^n$ for any given $\sigma \in \mathcal{S}_\mu^+$ and affine (hence concave) in σ for any given $\mathbf{x} \in \mathbb{R}^n$. Therefore, for each $\sigma \in \mathcal{S}_\mu^+$, the criticality condition

$$\frac{\partial \Xi}{\partial \mathbf{x}} = G_\mu(\sigma)\mathbf{x} - (\mathbf{f} - \sigma\mathbf{b}) = 0 \tag{19}$$

leads to the global minimizer $\mathbf{x}(\sigma) = G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b})$ of $\Xi(\mathbf{x}, \sigma)$ with respect to $\mathbf{x}$. Moreover,

$$\begin{aligned}\min_{\mathbf{x}\in\mathbb{R}^n} \Xi(\mathbf{x},\sigma) &= \Xi(\mathbf{x}(\sigma),\sigma)\\ &= \frac{\sigma}{\mu} + \frac{1}{2}\mathbf{x}(\sigma)^T(G_\mu(\sigma))\mathbf{x}(\sigma) - (\mathbf{f}-\sigma\mathbf{b})^T\mathbf{x}(\sigma)\\ &= \frac{\sigma}{\mu} + \frac{1}{2}\mathbf{x}(\sigma)^T(\mathbf{f}-\sigma\mathbf{b}) - (\mathbf{f}-\sigma\mathbf{b})^T\mathbf{x}(\sigma)\\ &= \frac{\sigma}{\mu} - \frac{1}{2}(\mathbf{f}-\sigma\mathbf{b})^T\mathbf{x}(\sigma)\\ &= \frac{\sigma}{\mu} - \frac{1}{2}(\mathbf{f}-\sigma\mathbf{b})^T(G_\mu(\sigma))^{-1}(\mathbf{f}-\sigma\mathbf{b})\\ &= P_\mu^d(\sigma).\end{aligned}$$

By the assumption, σ_μ is a global maximizer of $P_\mu^d(\sigma)$ over $\mathcal{S}_\mu^+$. If σ_μ is an interior of $\mathcal{S}_\mu^+$, then $\frac{d}{d\sigma}P_\mu^d(\sigma_\mu) = 0$. Otherwise, we have $\sigma_\mu = \max\{0, \sigma_{\max}\}$ and $\frac{d}{d\sigma}P_\mu^d(\sigma_\mu) \le 0$. In either case, if we denote $\mathbf{x}_\mu = \mathbf{x}(\sigma_\mu) = G_\mu^{-1}(\sigma_\mu)(\mathbf{f}-\sigma_\mu\mathbf{b})$, it follows from (10) that $\frac{1}{\mu} - \mathbf{x}_\mu^T(\frac{1}{2}H\mathbf{x}_\mu - \mathbf{b}) \le 0$. Namely, $\mathbf{x}_\mu \in \mathcal{X}_\mu$. Therefore, for any $\sigma \in \mathcal{S}_\mu^+$ we have

$$\begin{aligned}P_\mu^d(\sigma) &\le P_\mu^d(\sigma_\mu)\\ &= \min_{\mathbf{x}\in\mathbb{R}^n} \Xi(\mathbf{x},\sigma_\mu)\\ &= \Xi(\mathbf{x}_\mu,\sigma_\mu)\\ &= \min_{\mathbf{x}\in\mathcal{X}_\mu} \Xi(\mathbf{x},\sigma_\mu)\\ &\le \Lambda(\mathbf{x})^T\sigma - V^\sharp(\sigma) + \frac{1}{2}\mathbf{x}^TQ\mathbf{x} - \mathbf{f}^T\mathbf{x} + \frac{\mu}{2}\mathbf{x}^TG\mathbf{x},\ \forall\mathbf{x}\in\mathcal{X}_\mu\\ &\le P_\mu(\mathbf{x}),\ \forall\mathbf{x}\in\mathcal{X}_\mu.\end{aligned}$$

The last inequality comes from the fact that $\Lambda(\mathbf{x}) \le 0, \sigma \ge 0$, and $V^\sharp(\sigma) = 0$ when $(\mathbf{x},\sigma) \in \mathcal{X}_\mu \times \mathcal{S}_\mu^+$. This completes the proof. □

Theorem 2 (Strong Duality) *If σ_μ is a critical point of $P_\mu^d(\sigma)$ over $\mathcal{S}_\mu^+$, then $(\mathcal{P}_\mu^d)$ is perfectly dual to $(\mathcal{P}_\mu)$ in the sense that the vector*

$$\mathbf{x}_\mu = G_\mu^{-1}(\sigma_\mu)(\mathbf{f}-\sigma_\mu\mathbf{b}) \tag{20}$$

is a global minimizer of $(\mathcal{P}_\mu)$, σ_μ is a global maximizer of $(\mathcal{P}_\mu^d)$ and

$$\min_{\mathbf{x}\in\mathcal{X}_\mu} P_\mu(\mathbf{x}) = P_\mu(\mathbf{x}_\mu) = P_\mu^d(\sigma_\mu) = \max_{\sigma\in\mathcal{S}_\mu^+} P_\mu^d(\sigma). \tag{21}$$

Proof The proof basically follows that of the weak duality Theorem 1. The only difference lies in the assumption that σ_μ is a critical point of $P_\mu^d(\sigma)$ over $\mathcal{S}_\mu^+$. In this case, $\frac{d}{d\sigma}P_\mu^d(\sigma_\mu) = 0$. Consequently, $\mathbf{x}_\mu = \mathbf{x}(\sigma_\mu) = G_\mu^{-1}(\sigma_\mu)(\mathbf{f}-\sigma_\mu\mathbf{b})$ is on the boundary of $\mathcal{X}$, i.e., $\mathbf{x}_\mu \in \partial\mathcal{X}_\mu$. Hence $\Lambda(\mathbf{x}_\mu) = 0$ and it further implies that

$$P_\mu^d(\sigma_\mu) = \Xi(\mathbf{x}_\mu,\sigma_\mu) = P_\mu(\mathbf{x}_\mu).$$

Then Eq. 21 follows naturally. □

The above results immediately lead to the following sufficient condition for finding the global optimizer of problem $(\mathcal{P})$:

Corollary 1 (Sufficiency) *If there exists a critical point $\sigma_\mu \in \mathcal{S}_\mu^+$ for every $\mu \in [\mu_0, \delta^{-1}]$, then*

$$\min_{\mathbf{x}\in\mathcal{X}} P_0(\mathbf{x}) = \min_{\mu\in[\mu_0,\delta^{-1}]} P_\mu^d(\sigma_\mu). \tag{22}$$

3 Existence of global optimality

Recall that $\sigma_{\max}$ is the maximum root of $\det G_\mu(\sigma) = 0$. It can be found by using the power method or QR methods. Once it is calculated, we know $\mathcal{S}_\mu^+$ is a one dimensional (open or closed) ray starting from $\sigma_{\max}$ or 0 toward infinity. The next result provides an easy-to-check condition for the existence of a global optimal solution σ_μ to problem $(\mathcal{P}_\mu^d)$ over $\mathcal{S}_\mu^+$ with any given $\mu \in [\mu_0, \delta^{-1}]$.

Theorem 3 (Existence) *Given any $\mu \in [\mu_0, \delta^{-1}]$, if*

$$\lim_{\sigma(\in\mathcal{S}_\mu^+)\to\partial\mathcal{S}_\mu^+} \frac{dP_\mu^d(\sigma)}{d\sigma} > 0 \tag{23}$$

and

$$\lim_{\sigma\to\infty} \frac{dP_\mu^d(\sigma)}{d\sigma} < 0, \tag{24}$$

then the canonical dual problem $(\mathcal{P}_\mu^d)$ has at least one global optimal solution $\sigma_\mu \in \mathcal{S}_\mu^+$.

Proof Since $\frac{dP_\mu^d(\sigma)}{d\sigma}$ is continuous and decreasing over $\mathcal{S}_\mu^+$, it follows from (23) and (24) that there exists one $\sigma_\mu \in \mathcal{S}_\mu^+$ to be a critical point of $P_\mu^d(\sigma)$. By Theorem 2, we know σ_μ is a global optimal solution of $(\mathcal{P}_\mu^d)$. □

The existence condition in Theorem 3 actually implies that σ_μ lies in the interior of $\mathcal{S}_\mu^+$. We may extend our results to the situation that the point σ_μ lies on the boundary $\partial\mathcal{S}_\mu^+$, i.e.,

$$\lim_{\sigma(\in\mathcal{S}_\mu^+)\to\partial\mathcal{S}_\mu^+} \frac{dP_\mu^d(\sigma)}{d\sigma} = 0. \tag{25}$$

There are two possible cases for this to happen. Case 1: when $\sigma_{\max} < 0$, then $\partial\mathcal{S}_\mu^+ = \{0\}$ and $\det G_\mu(0) > 0$. In other words, $\sigma_\mu = 0 \in \mathcal{S}_\mu^+$ is a global optimal solution of $(\mathcal{P}_\mu^d)$. Case 2: when $\sigma_{\max} \geq 0$, then the set $\mathcal{S}_\mu^+$ may become open and condition (25) leads to a critical point that does not lie in $\mathcal{S}_\mu^+$. We thus would like to sharpen the result of Theorem 2 as follows.

Lemma 4 *Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s$, $i = 1, 2, \ldots, s$, be the distinct roots of* $\det G_\mu(\sigma) = 0$ *and I_i be the identical matrix with its dimensionality being the multiplicity of σ_i. Then, there exists a non-singular matrix N such that $NHN^T = -I$ and*

$$G_\mu(\sigma) = N^{-1}\Sigma(\sigma)(N^{-1})^T, \tag{26}$$

where $\Sigma(\sigma) = \sigma I + Diag\,(-\sigma_1 I_1, -\sigma_2 I_2, \ldots, -\sigma_s I_s)$.

Proof From (7), we know $G_\mu(\sigma)$ can be decomposed as

$$G_\mu(\sigma) = L(B + \sigma I)L^T,$$

where L is lower triangular with $-H = LL^T$ and $B = L^{-1}(Q + \mu G)(L^{-1})^T$. Since B is symmetric, we can further diagonalize B by

$$B = M^T(\text{Diag}\,(-\sigma_1 I_1, -\sigma_2 I_2, \ldots, -\sigma_s I_s))M,$$

where M is orthogonal. Simply choose $N = ML^{-1}$, the rest of the proof follows. □

Keep the notation used in Lemma 4 and assume that r_1 is the multiplicity of the root $\sigma_{\max} = \sigma_1$. Then we have the next result.

Theorem 4 *If* $\partial\mathcal{S}_\mu^+ = \{\sigma_1\}$ *with* $\sigma_1 \geq 0$ *and*

$$\lim_{\sigma(\in\mathcal{S}_\mu^+)\to\sigma_1^+} \frac{dP_\mu^d(\sigma)}{d\sigma} = 0, \tag{27}$$

then σ_1 *defines a global minimizer* $\bar{\mathbf{x}}$ *of problem* ($\mathcal{P}_\mu$) *such that*

$$\bar{\mathbf{x}} = \lim_{\sigma(\in\mathcal{S}_\mu^+)\to\sigma_1^+} G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b}). \tag{28}$$

Proof Denote

$$\mathbf{d}(\sigma) = (d_1(\sigma), d_2(\sigma), \ldots, d_n(\sigma))^T = N(\mathbf{f} - \sigma\mathbf{b}). \tag{29}$$

For $\sigma \in \mathcal{S}_\mu^+$, by (10) we know

$$\begin{aligned}
\frac{dP_\mu^d(\sigma)}{d\sigma} &= \frac{1}{\mu} - \frac{1}{2}(\mathbf{f} - \sigma\mathbf{b})^T G_\mu^{-1}(\sigma) H G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b}) + \mathbf{b}^T G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b}) \\
&= \frac{1}{\mu} - \frac{1}{2}(N(\mathbf{f}-\sigma\mathbf{b}))^T \Sigma(\sigma)^{-1} NHN^T \Sigma(\sigma)^{-1} N(\mathbf{f}-\sigma\mathbf{b}) + (N\mathbf{b})^T \Sigma(\sigma)^{-1} N(\mathbf{f}-\sigma\mathbf{b}) \\
&= \frac{1}{\mu} + \frac{1}{2}\mathbf{d}(\sigma)^T \Sigma(\sigma)^{-2}\mathbf{d}(\sigma) + (N\mathbf{b})^T \Sigma(\sigma)^{-1}\mathbf{d}(\sigma) \\
&= \frac{1}{\mu} + \frac{1}{2}\frac{d_1^2(\sigma) + d_2^2(\sigma) + \cdots + d_{r_1}^2(\sigma)}{(\sigma - \sigma_1)^2} + \frac{1}{2}\mathbf{d}_{n-r_1}(\sigma)^T \Sigma_{n-r_1}(\sigma)^{-2}\mathbf{d}_{n-r_1}(\sigma) \\
&\quad + \frac{t_1 d_1(\sigma) + t_2 d_2(\sigma) + \cdots + t_{r_1} d_{r_1}(\sigma)}{\sigma - \sigma_1} + \mathbf{t}_{n-r_1}^T \Sigma_{n-r_1}(\sigma)^{-1}\mathbf{d}_{n-r_1}(\sigma) \\
&= \frac{1}{\mu} + \frac{1}{2}\sum_{k=1}^{r_1}\left[\left(\frac{d_k(\sigma)}{\sigma - \sigma_1} + t_k\right)^2 - t_k^2\right] + \frac{1}{2}\mathbf{d}_{n-r_1}(\sigma)^T \Sigma_{n-r_1}(\sigma)^{-2}\mathbf{d}_{n-r_1}(\sigma) \\
&\quad + \mathbf{t}_{n-r_1}^T \Sigma_{n-r_1}(\sigma)^{-1}\mathbf{d}_{n-r_1}(\sigma)
\end{aligned}$$

where $\mathbf{t} = N\mathbf{b}$ and $\mathbf{d}_{n-r_1}(\sigma)$, $\Sigma_{n-r_1}(\sigma)$, $\mathbf{t}_{n-r_1}$ represent the last $n - r_1$ elements of $\mathbf{d}(\sigma)$, $\Sigma(\sigma)$, and $\mathbf{t}$, respectively. As $\sigma \to \sigma_1$, the vector $\mathbf{d}(\sigma)$ converges to $N(\mathbf{f} - \sigma_1\mathbf{b})$. However, condition (27) enforces that $d_1(\sigma), d_2(\sigma), \ldots, d_{r_1}(\sigma)$ converge to 0 no slower than $(\sigma - \sigma_1)$ does. Otherwise, $\frac{dP_\mu^d(\sigma)}{d\sigma}$ would tend to positive infinity. Consequently, each of

$$\bar{y}_i \triangleq \lim_{\sigma(\in\mathcal{S}_\mu^+)\to\sigma_1^+} \frac{d_i(\sigma)}{\sigma - \sigma_1}, \quad 1 \leq i \leq r_1$$

exists with a finite value.

A unique solution of $(\bar{y}_{r_1+1}, \ldots, \bar{y}_n)^T$ is then determined by

$$\begin{pmatrix} (\sigma_1 - \sigma_2)\, I_2 & \ldots & \ldots & 0 \\ \vdots & (\sigma_1 - \sigma_3)\, I_3 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & (\sigma_1 - \sigma_s)\, I_s \end{pmatrix} \begin{pmatrix} \bar{y}_{r_1+1} \\ \bar{y}_{r_1+2} \\ \vdots \\ \bar{y}_n \end{pmatrix} \begin{pmatrix} d_{r_1+1}(\sigma_1) \\ d_{r_1+2}(\sigma_1) \\ \vdots \\ d_n(\sigma_1). \end{pmatrix} \tag{30}$$

Now

$$\begin{aligned} &\lim_{\sigma(\in \mathcal{S}_\mu^+)\to\sigma_1^+} G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma \mathbf{b}) \\ &\quad = \mathbf{N^T}(\bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_{r_1}, \bar{\mathbf{y}}_{r_1+1}, \ldots, \bar{\mathbf{y}}_{\mathrm{n}})^{\mathbf{T}}, \end{aligned} \tag{31}$$

exists to be a finite vector. Thus we can define

$$\bar{\mathbf{x}} \triangleq \lim_{\sigma(\in \mathcal{S}_\mu^+)\to\sigma_1^+} G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma \mathbf{b}),$$

and rewrite (27) as

$$\frac{1}{\mu} - \frac{1}{2}\bar{\mathbf{x}}^T H \bar{\mathbf{x}} + \mathbf{b}^T \bar{\mathbf{x}} = 0.$$

This shows $\bar{\mathbf{x}}$ is a primal feasible solution that resides on the boundary of $\mathcal{X}_\mu$.

Let $\overline{\mathcal{S}}_\mu^+$ be the closure of $\mathcal{S}_\mu^+$ and define extensively the total complementarity function as

$$\widehat{\Xi}(\mathbf{x}, \sigma) = \frac{\sigma}{\mu} + \frac{1}{2}\mathbf{x}^T (G_\mu(\sigma))\mathbf{x} - (\mathbf{f} - \sigma \mathbf{b})^T \mathbf{x}, \tag{32}$$

for $\mathbf{x} \in \mathbb{R}^n$ and $\sigma \in \overline{\mathcal{S}}_\mu^+$. Since $G_\mu(\sigma_1)$ is positive semi-definite, $\widehat{\Xi}(\mathbf{x}, \sigma)$ must be convex in $\mathbf{x} \in \mathbb{R}^n$ for each $\sigma \in \overline{\mathcal{S}}_\mu^+$. Taking partial derivatives at $(\bar{\mathbf{x}}, \sigma_1)$, we have

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \widehat{\Xi}(\bar{\mathbf{x}}, \sigma_1) &= G_\mu(\sigma_1)\bar{\mathbf{x}} - (\mathbf{f} - \sigma_1 \mathbf{b}) \\ &= \lim_{\sigma(\in \mathcal{S}_\mu^+)\to\sigma_1^+} G_\mu(\sigma)\bar{\mathbf{x}} - (\mathbf{f} - \sigma_1 \mathbf{b}) \\ &= \lim_{\sigma(\in \mathcal{S}_\mu^+)\to\sigma_1^+} (\mathbf{f} - \sigma \mathbf{b}) - (\mathbf{f} - \sigma_1 \mathbf{b}) \\ &= 0. \end{aligned}$$

Consequently,

$$\begin{aligned} \widehat{\Xi}(\bar{\mathbf{x}}, \sigma_1) &= \min_{\mathbf{x}\in\mathbb{R}^n} \widehat{\Xi}(\mathbf{x}, \sigma_1) \\ &= \min_{\mathbf{x}\in\mathcal{X}_\mu} \widehat{\Xi}(\mathbf{x}, \sigma_1) \\ &\le \Lambda(\mathbf{x})^T \sigma_1 - V^\sharp(\sigma_1) + \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \mathbf{f}^T \mathbf{x} + \frac{\mu}{2}\mathbf{x}^T G \mathbf{x},\ \forall \mathbf{x} \in \mathcal{X}_\mu \\ &\le P_\mu(\mathbf{x}),\ \forall \mathbf{x} \in \mathcal{X}_\mu. \end{aligned}$$

Since $\bar{\mathbf{x}}$ is on the boundary of $\mathcal{X}_\mu$ and σ_1 is assumed to be non-negative, we have

$$\widehat{\Xi}(\bar{\mathbf{x}}, \sigma_1) = P_\mu(\bar{\mathbf{x}}) \le P_\mu(\mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{X}_\mu. \tag{33}$$

This proves the theorem. □

Theorem 5 *Assume that* $\partial\mathcal{S}_\mu^+ = \{\sigma_1\}$ *with* $\sigma_1 \geq 0$ *and*

$$\lim_{\sigma(\in\mathcal{S}_\mu^+)\to\sigma_1^+} \frac{dP_\mu^d(\sigma)}{d\sigma} = 0.$$

If we further define

$$P_\mu^d(\sigma_1) = \frac{\sigma_1}{\mu} - \frac{1}{2}(\mathbf{f} - \sigma_1\mathbf{b})^T\bar{\mathbf{x}}, \tag{34}$$

then $P_\mu^d(\sigma)$ *is right continuous at* σ_1 *and attains its maximum over* $\overline{\mathcal{S}}_\mu^+$ *at* σ_1. *In this case,*

$$P_\mu^d(\sigma) \leq P_\mu(\mathbf{x}),\ \forall\mathbf{x} \in \mathcal{X}_\mu \text{ and } \sigma \in \overline{\mathcal{S}}_\mu^+. \tag{35}$$

Proof From (6), for each $\sigma \in \mathcal{S}_\mu^+$, we have

$$P_\mu^d(\sigma) = \frac{\sigma}{\mu} - \frac{1}{2}(\mathbf{f} - \sigma\mathbf{b})^T G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b}).$$

Moreover,

$$\begin{aligned}\lim_{\sigma(\in\mathcal{S}_\mu^+)\to\sigma_1^+} P_\mu^d(\sigma) &= \frac{\sigma}{\mu} - \frac{1}{2}(\mathbf{f} - \sigma\mathbf{b})^T\bar{x}\\ &= P_\mu^d(\sigma_1).\end{aligned}$$

This shows that $P_\mu^d(\sigma_1)$ as defined in (34) makes the function $P_\mu^d(\sigma)$ right continuous at σ_1. Since $\frac{dP_\mu^d(\sigma)}{d\sigma}$ is decreasing (Lemma 3) and $\lim_{\sigma\to\sigma_1^+}\frac{dP_\mu^d(\sigma)}{d\sigma} = 0$, we know $\frac{dP_\mu^d(\sigma)}{d\sigma} \leq 0$. Hence, $P_\mu^d(\sigma)$ is decreasing over $\mathcal{S}_\mu^+$. The right continuity of $P_\mu^d(\sigma)$ at σ_1 further ensures that σ_1 must be the global maximum of $P_\mu^d(\sigma)$ over $\overline{\mathcal{S}}_\mu^+$.

Now, from (32), we have

$$\begin{aligned}\widehat{\Xi}(\bar{\mathbf{x}}, \sigma_1) &= \frac{\sigma_1}{\mu} + \frac{1}{2}\bar{\mathbf{x}}^T(G_\mu(\sigma_1))\bar{\mathbf{x}} - (\mathbf{f} - \sigma_1\mathbf{b})^T\bar{\mathbf{x}}\\ &= \frac{\sigma_1}{\mu} + \frac{1}{2}\bar{\mathbf{x}}^T \lim_{\sigma\to\sigma_1^+} G_\mu(\sigma) \lim_{\sigma\to\sigma_1^+} G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b}) - (\mathbf{f} - \sigma_1\mathbf{b})^T\bar{\mathbf{x}}\\ &= \frac{\sigma_1}{\mu} + \frac{1}{2}\bar{\mathbf{x}}^T(\mathbf{f} - \sigma_1\mathbf{b}) - (\mathbf{f} - \sigma_1\mathbf{b})^T\bar{\mathbf{x}}\\ &= P_\mu^d(\sigma_1)\\ &\leq P_\mu(\mathbf{x}),\ \forall\mathbf{x} \in \mathcal{X}_\mu. \qquad \text{(by (33))}\end{aligned}$$

This completes the proof. □

With Theorem 4 and 5, the existence condition of Theorem 3 can be generalized as

$$\lim_{\sigma(\in\mathcal{S}_\mu^+)\to\partial\mathcal{S}_\mu^+} \frac{dP_\mu^d(\sigma)}{d\sigma} \geq 0$$

and

$$\lim_{\sigma\to\infty} \frac{dP_\mu^d(\sigma)}{d\sigma} < 0.$$

In this case, we may find an optimal solution

$$\sigma_\mu \in \partial \mathcal{S}_\mu^+ \text{ with } \frac{dP_\mu^d(\sigma_\mu)}{d\sigma} = 0, \quad \det G_\mu(\sigma_\mu) = 0 \text{ and } G_\mu(\sigma_\mu) \succeq 0.$$

The following existence condition for global optimizers of the original problem then comes naturally:

Theorem 6 *If*

$$\lim_{\sigma(\in \mathcal{S}_\mu^+) \to \partial \mathcal{S}_\mu^+} \frac{dP_\mu^d(\sigma)}{d\sigma} \geq 0 \text{ and } \lim_{\sigma \to \infty} \frac{dP_\mu^d(\sigma)}{d\sigma} < 0 \tag{36}$$

hold for every $\mu \in [\mu_0, \delta^{-1}]$*, then*

$$\min_{\mathbf{x} \in \mathcal{X}} P_0(\mathbf{x}) = \min_{\mu \in [\mu_0, \delta^{-1}]} P_\mu^d(\sigma_\mu).$$

4 Numerical examples

Example 1 Let us begin with a simple one-dimensional example by taking $Q = -10$, $f = 1$, $G = 1$, $H = -1$, $b = 1$ and $\delta = 0.01$ to form a specific problem:

$$\min \ P_0(x) = -5x^2 - x + \frac{0.5x^2}{-0.5x^2 - x}$$

over the feasible domain

$$\chi = \{x \in R | -0.5x^2 - x \geq 0.01\} = \{-1.9899 \leq x \leq -0.01005\}.$$

The objective function $P_0(x)$ has a singularity at -2 and it is neither convex nor concave over χ. (See Fig. 1 for the graph of $P_0(x)$.)

By Lemma 1, we have $\min_{\mathbf{x} \in \mathcal{X}} P_0(\mathbf{x}) = \min_{\mu \in [2,100]} \min_{\mathbf{x} \in \mathcal{X}_\mu} P_\mu(\mathbf{x})$ with

$$\begin{aligned} P_\mu(x) &= \frac{1}{2} x^T Q x - x^T f + \mu g(x) \\ &= (-5 + 0.5\mu)x^2 - x \end{aligned}$$

and $\chi_\mu = \{-1 - \sqrt{1 - 2\mu^{-1}} \leq x \leq -1 + \sqrt{1 - 2\mu^{-1}}\}$. Directly minimizing $P_\mu(\mathbf{x})$ could be difficult. For example, if $\mu \in [2, 10)$, $P_\mu(\mathbf{x})$ is concave so that we are facing a family of less desirable concave minimization problems. Contrarily, each canonical dual functional $P_\mu^d(\sigma)$ of $P_\mu(\mathbf{x})$ is concave over $\mathcal{S}_\mu^+$. Therefore, for each $\mu \in [2, 100]$, we are to maximize

$$\begin{aligned} P_\mu^d(\sigma) &= \frac{\sigma}{\mu} - \frac{1}{2}(f - \sigma b)^t G_\mu^{-1}(\sigma)(f - \sigma b) \\ &= \frac{\sigma}{\mu} - \frac{(1 - \sigma)^2}{2(-10 + \mu + \sigma)} \end{aligned}$$

over $\mathcal{S}_\mu^+ = \{\sigma \mid \sigma > \max\{0, 10 - \mu\}\}$. The derivative of $P_\mu^d(\sigma)$ becomes

$$\frac{d}{d\sigma} P_\mu^d(\sigma) = \frac{1}{\mu} + \frac{1}{2}\left(\frac{1 - \sigma}{-10 + \mu + \sigma}\right)^2 + \frac{1 - \sigma}{-10 + \mu + \sigma}.$$

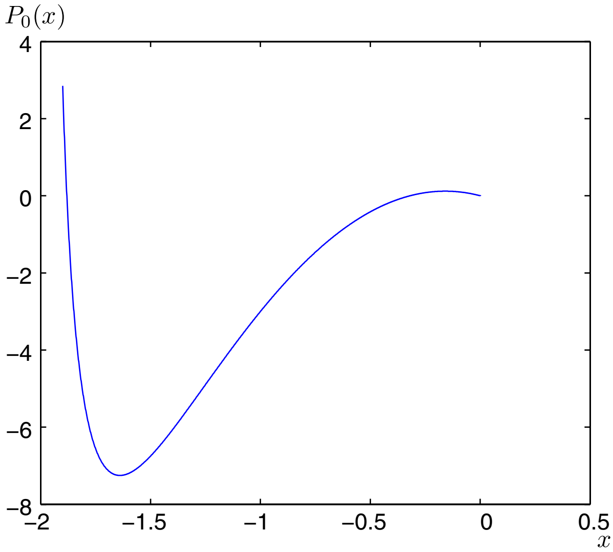

Fig. 1 Graph of $P_0(x)$ for Example 1

It is obvious that

$$\lim_{\sigma\to\infty} \frac{d}{d\sigma} P_\mu^d(\sigma) = \frac{1}{\mu} + \frac{1}{2} - 1 = \frac{1}{\mu} - \frac{1}{2} < 0 \text{ for } \mu > 2.$$

Hence the condition (24) is satisfied for $\mu > 2$. On the other hand,

$$\lim_{\sigma\to\partial\mathcal{S}_\mu^+} \frac{d}{d\sigma} P_\mu^d(\sigma) = \begin{cases} \infty, & \text{if } \mu \in [2,9),\ \partial\mathcal{S}_\mu^+ = \{10-\mu\}; \\ \frac{-7}{18}, & \text{if } \mu = 9,\ \partial\mathcal{S}_\mu^+ = \{1\}; \\ \infty, & \text{if } \mu \in (9,10],\ \partial\mathcal{S}_\mu^+ = \{10-\mu\}; \\ \frac{1}{\mu} + \frac{1}{2}(\frac{1}{\mu-10})^2 + \frac{1}{\mu-10} > 0, & \text{if } \mu \in (10,100],\ \partial\mathcal{S}_\mu^+ = \{0\}. \end{cases}$$

This implies that condition (23) holds except for $\mu = 9$. By Theorem 3, the maximizer σ_μ of $P_\mu^d(\sigma)$ for each $\mu \in [2,100] \setminus \{2,9\}$ exists. We can then implement a simple Newton method to locate σ_μ and define $P^d(\mu) = P_\mu^d(\sigma_\mu)$. Figure 2 shows the graph of $P_\mu^d(\sigma_\mu)$ in terms of μ. It has three local minima at $\mu = 2$, $\mu = 3.3825$ and $\mu = 100$, respectively, and one global maximum at $\mu = 9$ which is a cusp of $P^d(\mu)$. Finally, we minimize $P^d(\mu)$ over $\mu \in [2,100]$ using the line search with Armijo's rule (to be further described in Example 2 below) to obtain the global minimum at $\mu = 3.3825$ whose corresponding primal solution is $x = -1.6393$.

Example 2 We consider a two-dimensional problem with

$$Q = \begin{bmatrix} -1 & 6 \\ 6 & 5 \end{bmatrix},\ G = \begin{bmatrix} 5 & 1 \\ 1 & 2 \end{bmatrix},\ H = \begin{bmatrix} -7 & 3 \\ 3 & -2 \end{bmatrix},\ \mathbf{f} = \begin{bmatrix} -8 \\ 2 \end{bmatrix},\ \mathbf{b} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

The constraint set $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) \geq \delta = 0.01\}$ is an ellipse together with all its interior. This example is a non-convex global optimization problem subject to an elliptic constraint. Since $\mu_0^{-1} = h(H^{-1}b) = 20.3$, we consider the parametric programs $P_\mu(\mathbf{x})$ and its canonical dual functional $p_\mu^d(\sigma)$ for $\mu \in [0.04926, 100]$.

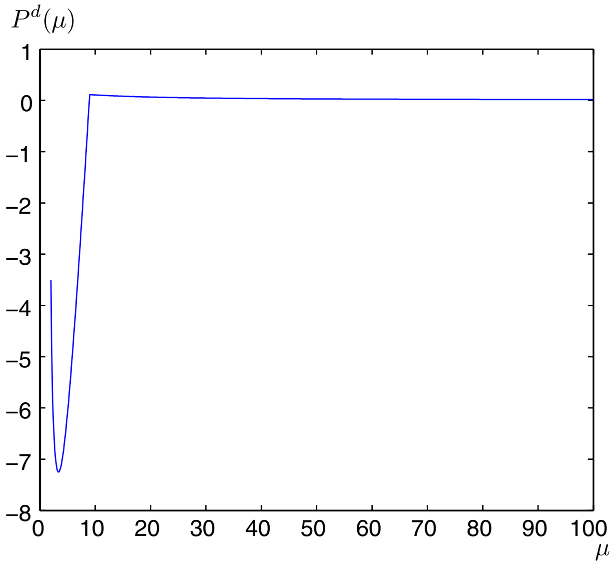

Fig. 2 Graph of $P^d(\mu)$ for Example 1

To verify Theorem 3, we first factorize $-H = LL^T$ with

$$L = \begin{bmatrix} \sqrt{7} & 0 \\ -3/\sqrt{7} & \sqrt{5}/\sqrt{7} \end{bmatrix}.$$

Then we use Matlab symbolic calculation to find the largest eigenvalue of $-L^{-1}(Q + \mu G)(L^{-1})^T$ as

$$-6.9 - 3\mu + 0.1\sqrt{5581 + 3920\mu + 720\mu^2} \quad \text{for} \quad \mu \in [0.04926, 100],$$

from which we can compute the boundary of $\mathcal{S}_\mu^+$ as

$$\partial\mathcal{S}_\mu^+ = \begin{cases} \{-6.9 - 3\mu + 0.1\sqrt{5581 + 3920\mu + 720\mu^2}\}, & \text{if } \mu \in [0.04926, 1.609); \\ \{0\}, & \text{if } \mu \in [1.609, 100]. \end{cases}$$

and

$$\lim_{\sigma \to \partial\mathcal{S}_\mu^+} \frac{d}{d\sigma} P_\mu^d(\sigma) = \begin{cases} \infty, & \text{if } \mu \in [0.04926, 1.609); \\ \geq 0, & \text{if } \mu \in [1.609, 10.659]; \\ < 0, & \text{if } \mu \in (10.659, 100]. \end{cases}$$

To check condition (24), we recall from the beginning of the proof for Theorem 4 that

$$\frac{dP_\mu^d(\sigma)}{d\sigma} = \frac{1}{\mu} + \frac{1}{2}\mathbf{d}(\sigma)^T \Sigma(\sigma)^{-2}\mathbf{d}(\sigma) + (N\mathbf{b})^T \Sigma(\sigma)^{-1}\mathbf{d}(\sigma)$$

and hence

$$\lim_{\sigma \to \infty} \frac{dP_\mu^d(\sigma)}{d\sigma} = \frac{1}{\mu} + \frac{1}{2}(N\mathbf{b})^T(N\mathbf{b}) - (N\mathbf{b})^T(N\mathbf{b})$$
$$= \frac{1}{\mu} - \frac{1}{2}\|N\mathbf{b}\|^2.$$

We can verify that

$$-20.29 \leq \frac{1}{\mu} - \frac{1}{2}\|N\mathbf{b}\|^2 \leq -13.6 \quad \text{for } \mu \in [0.04926, 100]$$

and thus condition (24) is met. From the above calculation, we have the strong duality on $\mu \in [0.04926, 10.659]$ since there is a critical point inside $\mathcal{S}_\mu^+$. We have, however, only the weak duality on $\mu \in (10.659, 100]$ since the global maximizer of $P_\mu^d(\sigma)$ occurs at $\sigma_\mu = 0$ but $\lim_{\sigma\to 0^+} \frac{d}{d\sigma} P_\mu^d(\sigma) < 0$. From Theorem 1 and 2, if we define

$$P^d(\mu) = \begin{cases} P_\mu^d(\sigma_\mu), & \text{for } \mu \in [0.04926, 10.659]; \\ P_\mu^d(0), & \text{for } \mu \in (10.659, 100], \end{cases}$$

then $P^d(\mu) \leq P_\mu(\mathbf{x}_\mu)$ with the equality sign being valid only on $\mu \in [0.04926, 10.659]$ (See Fig. 3 for the graph of $P^d(\mu)$.). What we can do is to minimize with respect to μ the lower bound function $P^d(\mu)$ of $P_\mu(\mathbf{x}_\mu)$. If the minimizer μ^* resides luckily in [0.04926, 10.659] (which is the case for this example), then we solve the master problem $(\mathcal{P})$. Otherwise, we only obtain a lower bound value of $(\mathcal{P})$. Nevertheless, to numerically minimize $P^d(\mu)$, we need to address a few issues as follows.

Although $P_\mu^d(\sigma)$ is concave for each $\mu \in [0.04926, 100]$, some of them have a very large "flat" region so that the Newton method may fail to converge. For example, when $\mu = 0.9593$, $\mathcal{S}_\mu^+ = \{\sigma \geq 0 \mid G_\mu(\sigma) \succ 0\} = \{\sigma > 0.2242\}$. The optimal solution for maximizing $P_{0.9593}^d(\sigma)$ over $\{\sigma > 0.2242\}$ is $\sigma_\mu = 1.47$ (See Fig. 4 for the graph of $P_{0.9593}^d(\sigma)$.). Such a function could cause numerical difficulty for the pure Newton method, should we not start from an initial solution close to $\sigma_\mu = 1.47$.

We can, however, introduce a quadratic penalty term to the function $P_\mu^d(\sigma)$ as follows:

$$\hat{P}_\mu^d(\sigma) = \sigma/\mu - \frac{1}{2}(\mathbf{f} - \sigma\mathbf{b})^T G_\mu^{-1}(\sigma)(\mathbf{f} - \sigma\mathbf{b}) - \frac{\gamma}{2}\sigma^2$$

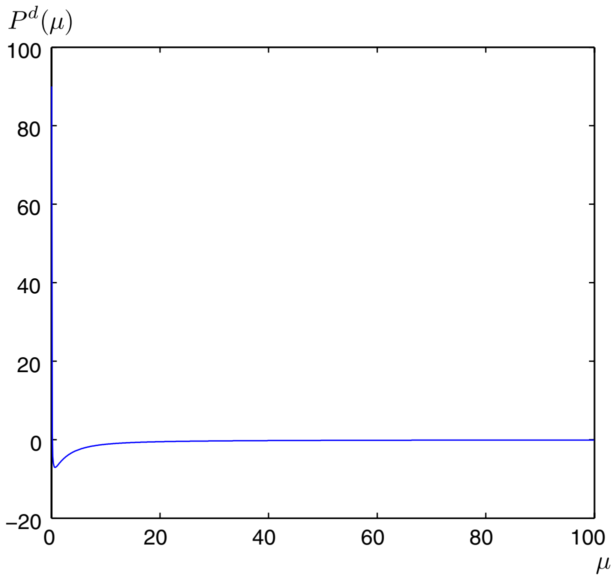

Fig. 3 Graph of $P^d(\mu)$ for Example 2

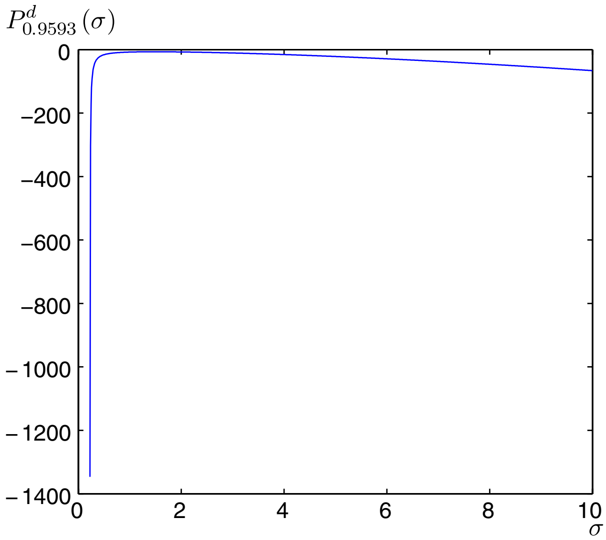

Fig. 4 Graph of $P^d_{0.9593}(\sigma)$ for Example 2

where $\gamma > 0$ is the penalty parameter to be reduced to 0 gradually. This will bend the flat region of $P^d_\mu(\sigma)$ for easier maximization numerically.

To minimize $P^d(\mu)$ over $\mu \in [\mu_0, \delta^{-1}]$, we use the line search with the Armijo's rule and thus it is not necessary to solve σ_μ for each μ. Suppose the current iterate is at $\mu_k \in [\mu_0, \delta^{-1}]$, we may approximate the derivative of $P^d(\mu)$ at $\mu = \mu_k$ by

$$d_k = \frac{d}{d\mu} P^d(\mu)|_{\mu=\mu_k} \doteq \frac{P^d(\mu_k + \varepsilon) - P^d(\mu_k)}{\varepsilon}$$

where $\varepsilon > 0$ is a selected parameter and the two terms in the numerator can be evaluated by the quasi-Newton method. If $d_k > 0$, the full step size s can be taken as the distance from the left boundary to μ_k, i.e., $s = \mu_0 - \mu_k$. Otherwise, we take $s = \delta^{-1} - \mu_k$ from the other end. Then, we select two parameters such that parameter α is close to 0 for scaling the slope d_k and parameter $\beta \in (0, 1)$ for scaling the full step size s. Let the test point t_n be defined as

$$t_n = \mu_k + (\beta^n s)(\alpha d_k)$$

and choose m to be the first non-negative integer such that

$$m = \min\{n \geq 0 \mid P^d(t_n) < P^d(\mu_k) + (\beta^n s)(\alpha d_k)\}.$$

Then, we can update

$$\mu_{k+1} = \mu_k + (\beta^m s)(\alpha d_k)$$

and repeat until d_k is nearly 0.

In our example, we use $\beta = 2/3$, $\alpha = 0.001$ and $\varepsilon = 0.01$. It took only 6 times of line search to reach the global minimum of $P^d_\mu(\sigma_\mu)$ at $\mu = 0.69076$ with a value of -7.0766. A total of 22 function evaluations including those used to find the search direction d_k and by the Armijo's rule are required. As a result, it took only 1.0615 cpu seconds to reach the optimal solution for this example, compared to 82.84 seconds taken by applying the grid method (with a grid size of 0.01).

5 Concluding remarks

The sum-of-ratios problems are considered to be difficult. In this paper, we study a special class of basic sum-of-ration problems in quadratic form and hope the results can lead to better understanding of fractional programming. We first parameterize such a problem into a family of subproblems. Then we develop a corresponding canonical duality theory, both in weak and strong duality form, to handle each subproblem. Based on the properties of the subproblems, we provide not only the extremality conditions for global optimality of the original problem, but also some easily checkable existence conditions to assure that the global optimal solutions of a subclass of quadratic sum-of-rations problems can indeed be found by solving a sequence of concave maximization problems.

Acknowledgements The First two Authors, namely, Shu-Cherng Fang and David Y. Gao, acknowledge with thanks for the support and sponsorship received under Grant Nos. DMI-0553310 and CCF-0514768, respectively, from the US National Science Foundation. Ruey-Lin Sheu has received sponsorship support from the Taiwan National Science Council, under Grant No. NSC 96-2115-M-006-014, and Wenxun Xing from Tsinghua Basic Research Foundation Grant No. 052201070, Chinese Ministry of Education Key Project No. 108005 and Chinese NSFC Grant No. 10801087, for which they thank the respective institutions.

References

1. Almogy, Y., Levin, O.: A class of fractional programming problems. Oper. Res. **19**, 57–67 (1971)
2. Benson, H.P.: Global optimization algorithm for the nonlinear sum of ratios problem. J. Optim. Theory Appl. **112**, 1–29 (2002)
3. Benson, H.P.: Using concave envelopes to globally solve the nonlinear sum of ratios problems. J. Glob. Optim. **22**, 343–364 (2002)
4. Benson, H.P.: On the global optimization of sum of linear fractional functions over a convex set. J. Optim. Theory Appl. **121**, 19–39 (2004)
5. Cambini, A., Crouzeix, J-P., Martein, L.: On the pseudoconvexity of a quadratic fractional function. Optimization **51**, 677–687 (2002)
6. Colantoni, C.S., Manes, R.P., Whinston, A.: Programming, profit rates, and pricing decisions. Account. Rev. **44**, 467–481 (1969)
7. Craven, B.D.: Fractional Programming, Sigma Series in Applied Mathematics, vol. 4. Heldermann Verlag, Berlin, Germany (1988)
8. Dahl, G.: A note on diagonally dominant matrices. Linear Algebra Appl. **317**, 217–224 (2000)
9. Duffin, R.J., Peterson, E.L.: Geometric programming with signomials. J. Optim. Theory Appl. **11**, 3–35 (1973)
10. Falk, J.E., Palocsay, S.W.: Optimizaing the sum of linear fractional functions. In: Floudas , C.A., Pardalos, P.M. Recent Advances in Global Optimization, pp. 221–258. Princeton University Press, Princeton, NJ (1992)
11. Fang, S.-C., Gao, D.Y., Sheu, R.-L., Wu, S.-Y.: Canonical dual approach for solving quadratic interger programming problems. J. Ind. Manag. Optim. **4**, 125–142 (2008)
12. Freund, R.W., Jarre, F.: Solving the sum-of-ratios problem by an interior-point method. J. Glob. Optim. **19**, 83–102 (2001)
13. Gao, D.Y.: Duality Principles in Nonconvex Systems: Theory, Methods and Applications, 454 pp. Kluwer Academic Publishers, Dordrecht/Boston/London (2000)
14. Gao, D.Y., Sherali, H.D.: Canonical duality theory: connections between nonconvex mechanics and global optimization. In: Gao, D.Y., Sherali, H.D. (eds.) Advances in Mechanics and Mathematics, vol. III, pp. 249–316. Springer (2006)
15. Gao, D.Y., Strang, G.: Geometric nonlinearity: potential energy, complementary energy, and the gap function. Quart. Appl. Math. **47**(3), 487–504 (1989)
16. Kanchan, P.K., Holland, A.S.B., Sahney, B.N.: Transportation techniques in linear-plus-fractional programming. Cahiers du CERO **23**, 153–157 (1981)
17. Konno, H., Fukaishi, K.: A branch and bound algorithm for solving low rank linear multiplicative and fractional programming problems. J. Glob. Optim. **18**, 283–299 (2000)

18. Konno, H., Inori, M.: Bond portfolio optimization by bilinear fractional programming. J. Oper. Res. Soc. Jpn. **32**, 143–158 (1989)
19. Konno, H., Watanabe, H.: Bond portfolio optimization problems and their application to index tracking: a partial optimiztion approach. J. Operations Res. Soc. Jpn. **39**, 295–306 (1996)
20. Konno, H., Yajima, Y., Matsui, T.: Parametric simplex algorithm for solving a special class on nonconvex minimization problems. J. Glob. Optim. **1**, 65–81 (1991)
21. Kuno, T.: A branch-and-bound algorithm for maximizing the sum of several linear ratios. J. Glob. Optim. **22**, 155–174 (2002)
22. Liang, Z.A., Huang, H.X., Pardalos, P.M.: Optimality conditions and duality for a class of nonlinear fractional programming problems. J. Optim. Theory Appl. **110**, 611–619 (2001)
23. Pardalos, P.M.: An algorithm for a class of nonlinear fractional problems using ranking of the vertices. BIT Numer. Math. **26**, 392–395 (1986)
24. Pardalos, P.M., Phillips, A.: Global optimization of fractional programs. J. Glob. Optim. **1**, 173–182 (1991)
25. Phuong, N.T.H., Tuy, H.: A unified monotonic approach to generalized linear fractional programming. J. Glob. Optim. **26**, 229–259 (2003)
26. Schaible, S.: A note on the sum of a linear and linear-fractional function. Nav. Res. Logist. Q. **24**, 691–693 (1977)
27. Schaible, S.: Fractional programming. In: Horst R., Pardalos P.M. (eds.) Handbook of Global Optimization, pp. 495–608. Kluwer Academic Publishers (1995)
28. Scott, C., Fang S.-C.: Geometric programming. Ann. Oper. Res. (Kluwer Academic Publishers), **105**, pp. 226 (2001)
29. Scott, C.H., Jefferson, T.R.: Duality of a nonconvex sum of ratios. J. Optim. Theory Appl. **98**, 151–159 (1998)
30. Stancu-Minasian, I.M.: Applications of the fractional programming. Econ. Comput. Econ. Cybern. Stud. Res. **1**, 69–86 (1980)
31. Wang, Z., Fang, S.-C., Gao, D.Y., Xing, W.: Global extremal conditions for multi-integer quadratic programming. J. Ind. Manag. Optim. **4**, 213–225 (2008)
32. Wu, W.-Y., Sheu, R.-L., Birbil, S.I.: Solving the sum-of-ratios problem by a stochastic search algorithm. J. Glob. Optim. **42**, 91–109 (2008)