A DUAL AND INTERIOR POINT APPROACH TO SOLVE CONVEX MIN-MAX PROBLEMS

JOS F. STURM and SHUZHONG ZHANG
Econometric Institute
Erasmus University Rotterdam

Abstract. In this paper we propose an interior point method for solving the dual form of min-max type problems. The dual variables are updated by means of a scaling supergradient method. The boundary of the dual feasible region is avoided by the use of a logarithmic barrier function. A major difference with other interior point methods is the nonsmoothness of the objective function.

1. Introduction

Consider the following problem

$$(\mathrm{P}) \qquad \min_{x\in\mathcal{X}} \max_{1\le i\le m} f_i(x)$$

where we assume that the functions $f_i(x)$, $1 \le i \le m$, are real valued convex functions defined on a convex and compact subset $\mathcal{X}$ of $\Re^n$.

Clearly, we have

$$\min_{x\in\mathcal{X}} \max_{1\le i\le m} f_i(x) = \min_{x\in\mathcal{X}} \max_{y\in S} y^T f(x)$$

where S is m-dimensional unit simplex given by

$$S := \{y \in \Re^m : \sum_{i=1}^{m} y_i = 1 \text{ and } y_i \ge 0,\ 1 \le i \le m\}$$

and the m-dimensional vector function $f(x)$ is given by

$$f(x) := (f_1(x), f_2(x), \cdots, f_m(x))^T.$$

Since the function $y^T f(x)$ is convex in x for fixed $y \in S$, and is concave in y for fixed $x \in \mathcal{X}$, it follows that (see e.g. Sion [6])

$$\min_{x\in\mathcal{X}} \max_{y\in S} y^T f(x) = \max_{y\in S} \min_{x\in\mathcal{X}} y^T f(x). \tag{1}$$

From now on we shall concentrate on the dual problem of (P) given by

$$(\mathrm{D}) \qquad \max_{y\in S} h(y),$$

where the dual objective function is defined as

$$h(y) := \min_{x\in\mathcal{X}} y^T f(x).$$

Note that the domain of h is S. Clearly, $h(y)$ is a concave function.

In two recent papers by Barros, Frenk, Schaible and Zhang [1, 2], fast algorithms for solving generalized fractional programming were constructed on the basis of a similar duality relation. The dual problem (D) can be derived using the Lagrangian function. For a thorough discussion on the Lagrange duality theory for convex programming, we refer to the book of Hiriart-Urruty and Lemaréchal [5].

Observe that Problem (D) has a very simple constraint set. However, the function $h(y)$ is in general non-differentiable. Throughout this paper we shall use an oracle to get an optimal solution $\bar{x}$ of the following problem:

$$\min_{x\in\mathcal{X}} y^T f(x) \tag{2}$$

where $y \in S$. Using this oracle we not only know the function value $h(y) = y^T f(\bar{x})$, but also an element belonging to the supergradient set. More precisely,

$$f(\bar{x}) \in \partial h(y)$$

where $\partial h(y)$ denotes the supergradient set of h at point y.

The basic underlying idea is that we first introduce a logarithmic barrier for Problem (D), and then apply a scaling and projection supergradient method maximizing the barrier function. Due to lack of differentiability in $h(y)$, the convergence analysis differs in flavor from usual path-following algorithms. The advantage of our approach is that we do not require any knowledge on the functions f_i, $i = 1, 2, \cdots, m$, and the structure of the constraint set $\mathcal{X}$. Remark that for the cases where m is relatively large compared to the dimension n, and the constraint set $\mathcal{X}$ is simple, solving (2) is much easier than solving the original problem.

The notation we use is as follows. The superscript of a vector is used to denote the iteration number, e.g. in the k-th iteration we have $y^{(k)}$; the subscript will denote the coordinate, e.g. the i-th coordinate of $y^{(k)}$ is $y_i^{(k)}$; capitalization of a vector will denote the diagonal matrix taking the elements from the vector in the diagonal, e.g. $Y^{(k)} = \text{diag}(y_1^{(k)}, \cdots, y_m^{(k)})$. We denote the all-one vector by e, the Euclidean norm (the L_2 norm) simply by $\|\bullet\|$ and the L_∞ norm by $\|\bullet\|_\infty$.

We organize the presentation in the following way. In Section 2, we will introduce the search direction and present the new algorithm. The convergence analysis of the algorithm is carried out in Section 3 and some remarks concluding the discussion are made in Section 4.

2. The scaling supergradient method

We introduce now the logarithmic barrier function

$$h_\mu(y) := h(y) + \mu \sum_{i=1}^{m} \log y_i.$$

Observe that $h_\mu(y)$ is a strictly concave function, for which the supergradient set is given by

$$\partial h_\mu(y) = \partial h(y) + \mu Y^{-1} e. \qquad (3)$$

The concept of logarithmic barrier was introduced by Frisch [4] to steer the iterates away from the boundary. The optimizer of the barrier function will be a nearly optimal solution to (D) if the multiple μ of the barrier term is small, as it is shown in the following lemma.

Lemma 1 *If $\bar{y} \in S$ is such that $h_\mu(\bar{y}) = \max_{y \in S} h_\mu(y)$ then*

$$h(\bar{y}) \geq \max_{y \in S} h(y) - m\mu.$$

PROOF.
From the concavity of h_μ, it follows that

$$0 \in \partial h_\mu(\bar{y}),$$

i.e., there exists $\eta \in \partial h(\bar{y})$ such that

$$\eta + \mu \bar{Y}^{-1} e = 0.$$

By the concavity of h, we have for $y^* \in \arg\max_{y \in S} h(y)$ that

$$\begin{aligned}
\max_{y \in S} h(y) &\leq h(\bar{y}) + \eta^T (y^* - \bar{y}) \\
&= h(\bar{y}) - \mu e^T \bar{Y}^{-1} (y^* - \bar{y}) \\
&= h(\bar{y}) + \mu (m - e^T \bar{Y}^{-1} y^*) \\
&\leq h(\bar{y}) + m\mu
\end{aligned}$$

where we used $\bar{y}, y^* \in S$.

□

In this paper we shall maximize h_μ over S for a prefixed parameter $\mu > 0$. We shall fix $0 < \mu < \epsilon/m$ if an ϵ-optimal solution is desired.

Assume that the current iterate $y^{(k)} \in \overset{o}{S}$, where $\overset{o}{S}$ denotes the relative interior of S. Calling Oracle (2) we obtain

$$x^{(k)} \in \arg\min_{x \in \mathcal{X}} (y^{(k)})^T f(x).$$

Let $g^{(k)} := f(x^{(k)}) + \mu (Y^{(k)})^{-1} e$. Hence, by (3) we know that

$$g^{(k)} \in \partial h_\mu(y^{(k)}).$$

As a search direction we propose a scaled supergradient direction, which coincides with the supergradient direction of the function $h_\mu(Y^{(k)} z)$ on the domain $\{z : (y^{(k)})^T z = 1\}$. The scaling transformation $z = (Y^{(k)})^{-1} y$ is based on the idea of

Dikin's affine scaling algorithm [3] for linear programming. Remark that this scaling maps the current iterate $y^{(k)}$ into the all-one vector e.

To simplify notations, we write

$$P_v := I_m - \frac{1}{\|v\|^2} v v^T$$

to denote the orthogonal projection matrix onto the kernel of a given vector $v \in R^m$.

The scaled supergradient direction we propse is $Y^{(k)} d^{(k)}$, where

$$d^{(k)} := \frac{1}{\left\| P_{y^{(k)}} Y^{(k)} g^{(k)} \right\|} P_{y^{(k)}} Y^{(k)} g^{(k)}.$$

Remark that

$$Y^{(k)} d^{(k)} = \arg \max_{e^T w = 0} \{ (g^{(k)})^T w : \left\| (Y^{(k)})^{-1} w \right\| \le 1 \}.$$

It is easily seen that $y^{(k)} + t_k Y^{(k)} d^{(k)} \in \overset{o}{S}$ if $| t_k | < 1$. In this paper, we require that

$$0 < t_k \le \alpha < 1 \text{ for } k = 0, 1, \cdots$$

along with the classical conditions of the supergradient step length (cf. Shor [7]), viz.

$$\lim_{k \to \infty} t_k = 0$$

$$\sum_{k=0}^{\infty} t_k = \infty.$$

For simplicity we let $\alpha := \frac{1}{2}$. As an example, one may choose $t_k = \frac{1}{k+2}$ for $k = 0, 1, \cdots$.

Our scaling supergradient algorithm generates the following sequence of dual variables belonging to $\overset{o}{S}$,

$$y^{(0)} = \frac{1}{m} e$$

and

$$y^{(k+1)} := y^{(k)} + t_k Y^{(k)} d^{(k)} \text{ for } k = 0, 1, 2, \cdots.$$

In the next section, it will be shown that

$$\limsup_{k \to \infty} h_\mu(y^{(k)}) = \max_{y \in S} h_\mu(y).$$

3. Convergence analysis

In the previous section, we have already seen that the sequence $\{y^{(k)}\}$ is contained in the relative interior of S. We shall now prove that our barrier method avoids the boundary so well that the sequence is actually contained in a closed subset of $\overset{o}{S}$.

By definition,

$$\left\| P_{y^{(k)}} Y^{(k)} g^{(k)} \right\| d^{(k)} = P_{y^{(k)}} Y^{(k)} f(x^{(k)}) + \mu P_{y^{(k)}} e.$$

Using $\min_{y \in S} \|y\|^2 = \frac{1}{m}$, it follows that

$$P_{y^{(k)}} e = e - \frac{y^{(k)}}{\left\| y^{(k)} \right\|^2} \geq e - m y^{(k)}.$$

Since $\mathcal{X}$ is convex and compact, all the convex functions f_i, $1 \leq i \leq m$, are uniformly bounded on $\mathcal{X}$. Letting $f_\infty := \max_{x \in \mathcal{X}} \|f(x)\|_\infty$, we have

$$\left\| P_{y^{(k)}} Y^{(k)} f(x) \right\| \leq f_\infty \left\| y^{(k)} \right\| \leq f_\infty$$

so that

$$\left\| P_{y^{(k)}} Y^{(k)} g^{(k)} \right\| d^{(k)} \geq \mu e - (f_\infty + m\mu) y^{(k)}.$$

This implies that

$$y_i^{(k+1)} \geq y_i^{(k)} \text{ for } i \text{ with } y_i^{(k)} \leq \frac{\mu}{f_\infty + m\mu}.$$

Since $0 < t_k \leq \frac{1}{2}$, we have

$$y^{(k+1)} \geq \frac{1}{2} y^{(k)}$$

for any k. Because $y^{(0)} = \frac{1}{m} e$, it follows that

$$\inf_k \min_{1 \leq i \leq m} y_i^{(k)} \geq c_1, \tag{4}$$

where $c_1 := \frac{1}{2} \frac{\mu}{f_\infty + m\mu}$.

Now we use (4) and the fact that all the limit points form a closed set contained in $\overset{o}{S}$ to conclude that there is one limit point, say $\bar{y}$, which attains the maximum function value in $h_\mu(y)$ among all the limit points. Let y^* be the maximum point of $h_\mu(y)$ in S. We shall now concentrate on proving $h_\mu(\bar{y}) = h_\mu(y^*)$.

The proof is done by contradiction. Suppose from now on that

$$h_\mu(\bar{y}) < h_\mu(y^*). \tag{5}$$

Let the upper level set of $h_\mu(y)$ at $\bar{y}$ be

$$L := \{ y \in S : \; h_\mu(y) \geq h_\mu(\bar{y}) \}.$$

By this construction, there will be no other limit point in $\overset{o}{L}$.

Clearly, $y^* \in \overset{o}{L}$. Moreover, there exists a positive number θ such that

$$B(y^*;\theta) \cap S \subseteq \overset{o}{L} \tag{6}$$

where $B(y^*;\theta)$ denotes a ball with center y^* and radius θ.

Now we turn to consider an iterative point $y^{(k)}$. Let the upper level set at $y^{(k)}$ be

$$L_k := \{y \in S : \ h_\mu(y) \geq h_\mu(y^{(k)})\}.$$

Due to the concavity of h_μ, the projected supergradient direction $P_e g^{(k)}$ provides a normal direction in S of a supporting hyperplane for L_k at $y^{(k)}$.

Let $y \in L_k$. The distance from y to the hyperplane is given by

$$\frac{(g^{(k)})^T(y - y^{(k)})}{\left\| P_e g^{(k)} \right\|}. \tag{7}$$

Let $\hat{y} \in \overset{o}{S}$. Define

$$\delta^{(k)} := (d^{(k)})^T (Y^{(k)})^{-1} (\hat{y} - y^{(k)}). \tag{8}$$

Lemma 2 *Let $r > 0$. If $B(\hat{y};r) \cap S \subseteq L_k$ then there exists some constant c_2 such that*

$$\delta^{(k)} \geq r c_2 (h_\mu(y^*) - h_\mu(y^{(k)})).$$

PROOF.
Consider the following supporting hyperplane of L_k,

$$\{z : \ (g^{(k)})^T P_e (z - y^{(k)}) = 0\}.$$

The distance from $\hat{y}$ towards this supporting hyperplane is

$$(g^{(k)})^T(\hat{y} - y^{(k)}) / \left\| P_e g^{(k)} \right\|.$$

As $B(\hat{y};r) \cap S \subseteq L_k$ this implies

$$(g^{(k)})^T(\hat{y} - y^{(k)}) \geq r \left\| P_e g^{(k)} \right\|.$$

Therefore,

$$\begin{aligned} \delta^{(k)} &= (d^{(k)})^T (Y^{(k)})^{-1} (\hat{y} - y^{(k)}) \\ &= (g^{(k)})^T(\hat{y} - y^{(k)}) / \left\| P_{y^{(k)}} Y^{(k)} g^{(k)} \right\| \\ &\geq r \frac{\left\| P_e g^{(k)} \right\|}{\left\| P_{y^{(k)}} Y^{(k)} g^{(k)} \right\|}. \end{aligned} \tag{9}$$

Using the Cauchy-Schwartz inequality and the supergradient inequality, we have

$$\left\|P_e g^{(k)}\right\| \geq (g^{(k)})^T(y^* - y^{(k)}) / \left\|y^* - y^{(k)}\right\| \geq \frac{h(y^*) - h(y^{(k)})}{\left\|y^* - y^{(k)}\right\|}. \tag{10}$$

As $y^{(k)}$ and y^* both belong to the unit simplex, it follows $\left\|y^* - y^{(k)}\right\| \leq \sqrt{2}$. Moreover, there holds

$$\left\|P_{y^{(k)}} Y^{(k)} g^{(k)}\right\| \leq \left\|Y^{(k)} g^{(k)}\right\| \leq \left\|Y^{(k)} f(x^{(k)})\right\| + \mu \left\|e\right\| \leq f_\infty + \sqrt{m}\mu. \tag{11}$$

From (9)-(11) it follows that

$$\delta^{(k)} \geq r c_2 (h_\mu(y^*) - h_\mu(y^{(k)}))$$

for $c_2 = \frac{1}{\sqrt{2}(f_\infty + \sqrt{m}\mu)}$. □

Define

$$\rho^{(k)} := \left\|\hat{Y}^{-1}(\hat{y} - y^{(k)})\right\|. \tag{12}$$

We have the following relation:

Lemma 3 *There holds*

$$(\rho^{(k+1)})^2 \leq (\rho^{(k)})^2 - 2t_k[\delta^{(k)} - (\rho^{(k)})^2 - (1 + \rho^{(k)})^2 t_k/2].$$

PROOF.
Since $y^{(k+1)} = y^{(k)} + t_k Y^{(k)} d^{(k)}$ we have

$$\begin{aligned}(\rho^{(k+1)})^2 &= \left\|\hat{Y}^{-1}(\hat{y} - y^{(k)} - t_k Y^{(k)} d^{(k)})\right\|^2 \\ &= (\rho^{(k)})^2 - 2t_k\delta^{(k)} + 2t_k(\hat{y} - y^{(k)})^T \hat{Y}^{-1}(I - \hat{Y}^{-1}Y^{(k)})d^{(k)} \\ &\quad + t_k^2 \left\|\hat{Y}^{-1}Y^{(k)}d^{(k)}\right\|^2. \end{aligned} \tag{13}$$

Notice that

$$\left\|\hat{Y}^{-1}y^{(k)} - e\right\|_\infty \leq \left\|\hat{Y}^{-1}y^{(k)} - e\right\| = \rho^{(k)}. \tag{14}$$

Therefore, using $\left\|d^{(k)}\right\| = 1$ it follows

$$\left\|\hat{Y}^{-1}Y^{(k)}d^{(k)}\right\| \leq \left\|\hat{Y}^{-1}y^{(k)}\right\|_\infty \left\|d^{(k)}\right\| \leq 1 + \rho^{(k)}. \tag{15}$$

Similarly, using $\left\|d^{(k)}\right\| = 1$ and the Cauchy-Schwartz inequality, we have

$$\begin{aligned} \left| (\hat{y} - y^{(k)})^T \hat{Y}^{-1}(I - \hat{Y}^{-1}Y^{(k)})d^{(k)} \right| &\leq \left\|(I - \hat{Y}^{-1}Y^{(k)})\hat{Y}^{-1}(\hat{y} - y^{(k)})\right\| \\ &\leq \left\|(I - \hat{Y}^{-1}Y^{(k)})e\right\|_\infty \left\|\hat{Y}^{-1}(\hat{y} - y^{(k)})\right\| \\ &= \left\|\hat{Y}^{-1}y^{(k)} - e\right\|_\infty \rho^{(k)} \\ &\leq (\rho^{(k)})^2, \end{aligned}$$

where the last inequality follows from (14).

Substituting the above inequality and the inequality (15) into (13) yields the desired result. □

Define

$$\hat{y}_\lambda := (1-\lambda)\bar{y} + \lambda y^*,$$

where $0 < \lambda < 1$. Let $\hat{y}$ be $\hat{y}_\lambda$. By (6) there exists $\tilde{h} > h_\mu(\bar{y})$ such that the ball $B(y^*;\theta) \cap S$ will be contained in the upper level set

$$\{y \in S:\ h_\mu(y) \geq \tilde{h}\}.$$

Using the concavity of h_μ, this implies

$$B(\hat{y}_\lambda;\lambda\theta) \cap S \subset \{y \in S:\ h_\mu(y) \geq (1-\lambda)h_\mu(\bar{y}) + \lambda\tilde{h}\}.$$

Since

$$\limsup_{k\to\infty} h_\mu(y^{(k)}) = h_\mu(\bar{y}) < \tilde{h} \tag{16}$$

we obtain from Lemma 2 and (16) that for given $0 < \lambda < 1$ there must exist k_1 such that for all $k \geq k_1$,

$$\delta^{(k)} \geq c_2\lambda\theta(h_\mu(y^*) - \tilde{h}). \tag{17}$$

On the other hand, by (12) we have

$$\begin{aligned} \rho^{(k)} &= \left\|\hat{Y}^{-1}(\hat{y} - y^{(k)})\right\| \\ &\leq \left\|\hat{Y}^{-1}(\hat{y} - \bar{y})\right\| + \left\|\hat{Y}^{-1}(\bar{y} - y^{(k)})\right\| \\ &= \lambda\sqrt{\sum_{i=1}^m \left(\frac{\bar{y}_i - y_i^*}{(1-\lambda)\bar{y}_i + \lambda y_i^*}\right)^2} + \left\|\hat{Y}^{-1}(\bar{y} - y^{(k)})\right\|. \end{aligned} \tag{18}$$

As $\bar{y} \neq y^*$ is a limit point, there is an unbounded set $K(\lambda)$ of integers such that

$$\left\|\hat{Y}^{-1}(\bar{y} - y^{(k)})\right\| \leq \lambda\sqrt{\sum_{i=1}^m \left(\frac{\bar{y}_i - y_i^*}{(1-\lambda)\bar{y}_i + \lambda y_i^*}\right)^2} \tag{19}$$

for all $k \in K(\lambda)$.

Based on (17) and (18), there exists a sufficiently small constant $\lambda_0 > 0$ such that when $\lambda = \lambda_0$ and $k \in K(\lambda_0)$, then

$$(\rho^{(k)})^2 < \min\{\frac{1}{3}c_2\lambda_0\theta(h_\mu(y^*) - \tilde{h}), 1\}. \tag{20}$$

Let k_1 be chosen according to (17) for $\lambda = \lambda_0$.

Because $\lim_{k\to\infty} t_k = 0$, there is $k_2 \in K(\lambda_0)$ with $k_2 \geq k_1$, such that for all $k \geq k_2$ we have

$$2t_k < \frac{1}{3}c_2\lambda_0\theta(h_\mu(y^*) - \tilde{h}).$$

In particular, for $k \geq k_2$ and if (20) holds, then we have

$$\frac{(1+\rho^{(k)})^2}{2}t_k < 2t_k < \frac{1}{3}c_2\lambda_0\theta(h_\mu(y^*) - \tilde{h}). \tag{21}$$

Using (17), (20), (21) and applying Lemma 3, it follows that

$$\begin{aligned}(\rho^{(k+1)})^2 \ &\leq \ (\rho^{(k)})^2 - 2t_k(1 - \frac{1}{3} - \frac{1}{3})c_2\lambda_0\theta(h_\mu(y^*) - \tilde{h}) \\ &= \ (\rho^{(k)})^2 - \frac{2}{3}t_k c_2\lambda_0\theta(h_\mu(y^*) - \tilde{h}) \end{aligned} \tag{22}$$

for $k = k_2$. This implies that $\rho^{(k_2+1)} < \rho^{(k_2)}$ and so (20) and (21) hold for $k := k_2+1$, and consequently (22) also holds for $k := k_2 + 1$. Recursively applying (22) yields a contradiction since $\sum_{j=k_2}^{\infty} t_j = +\infty$. This shows that inequality (5) cannot be true, which, in turn, proves the desired convergence result. To summarize, we present the following main theorem of this paper.

Theorem 1 *There holds*

$$\limsup_{k\to\infty} h_\mu(y^{(k)}) = \max_{y\in S} h_\mu(y).$$

4. Concluding remarks

We have presented in this article an interior point method for solving a dual form of min-max type problems. An important question left is how to recover the primal solutions using approximately optimal dual variables and an approximately optimal objective value. We regard this as a topic for future research.

In a forthcoming paper, the authors will investigate a path-following scheme, extending the current results. Finally, we remark that our convergence proof fails for $\mu = 0$, in which case the method becomes comparable to the affine scaling algorithm for linear programming. It remains an open question whether the convergence still holds in that case.

References

1. A.I. Barros, J.B.G. Frenk, S. Schaible and S. Zhang, A new algorithm for generalized fractional programming, *Technical Report TI94-23*, Tinbergen Institute Rotterdam, 1994.
2. A.I. Barros, J.B.G. Frenk, S. Schaible and S. Zhang, How duality can be used to solve generalized fractional programming problems, 1994, submitted for publication.
3. I.I. Dikin, Iterative solutions of problems of linear and quadratic programming, *Soviet Mathematics Doklady* 8 (1967) 674-675.
4. K.R. Frisch, The logarithmic potential method for convex programming, Institute of Economics, University of Oslo, Oslo, Norway, 1955.

5. J.-B. Hiriart-Urruty and C. Lemaréchal, Convex analysis and minimization algorithms (vol. 1), Springer-Verlag, Berlin, 1993.
6. M. Sion, On general minimax theorems, *Pacific Journal of Mathematics* 8 (1958) 171-176.
7. N.Z. Shor, Minimization methods for non-differentiable functions, Springer-Verlag, Berlin, 1985.