Saddle Point: Background and Projected SM

R.D.C. Monteiro[1]

[1]School of Industrial and SystemsEngineering
Georgia Institute of Technology

January 24, 2024

Saddle point duality

Let $X \times Y \subset \mathbb{R}^n \times \mathbb{R}^m$ and $\Psi : X \times Y \to \mathbb{R}$ be given and define $p : X \to (-\infty, \infty]$ and $d : Y \to [-\infty, \infty)$ as

$$p(x) := \sup\{\Psi(x, y') : y' \in Y\} \quad \forall x \in X \tag{1}$$

$$d(y) := \inf\{\Psi(x', y) : x' \in X\} \quad \forall y \in Y \tag{2}$$

and let $Y(x)$ and $X(y)$ denote the set of optimal solutions of (1) and (2), respectively. Consider also the pair of min-max and max-min problems

$$p_* := \inf\{p(x) : x \in X\} = \inf_{x \in X} \sup_{y \in Y} \Psi(x, y), \tag{3}$$

$$d_* := \sup\{d(y) : y \in Y\} = \sup_{y \in Y} \inf_{x \in X} \Psi(x, y). \tag{4}$$

and let X_* and Y_* denote the set of optimal solutions of (3) and (4), respectively

Lemma

(Weak duality): For every $(x, y) \in X \times Y$,

$$p(x) \geq \Psi(x, y) \geq d(y)$$

As a consequence,

- *if* $(x_*, y_*) \in X_* \times Y_*$, *then*

$$p_* \geq \Psi(x_*, y_*) \geq d_* \tag{5}$$

- *for every* $(x, y) \in X \times Y$,

$$p(x) \geq p_* \geq d_* \geq d(y) \tag{6}$$

Proof: Due to the definition of $p(x)$ and $d(y)$ in (1) and (2), it follows that $p(x) \geq \Psi(x, y) \geq d(y)$ for every $(x, y) \in X \times Y$. The latter relation with $(x, y) = (x_*, y_*)$ together with the definition of X_* and Y_* then imply (5). The first and third inequalities in (6) follow from the definitions of p_* and d_*, respectively, and the second one follows immediately from the fact that $p(x) \geq d(y)$ for every $(x, y) \in X \times Y$

Proposition

For a given pair $(x_*, y_*) \in X \times Y$, the following conditions are equivalent:

(a) $p(x_*) = d(y_*) \in \mathbb{R}$

(b) $p_* = d_* \in \mathbb{R}$ *and* $(x_*, y_*) \in X_* \times Y_*$

(c) $\Psi(x, y_*) \geq \Psi(x_*, y_*) \geq \Psi(x_*, y)$ *for every* $(x, y) \in X \times Y$

(d) $(x_*, y_*) \in X(y_*) \times Y(x_*)$

As a consequence, if $\Psi(\cdot, y) - \Psi(x, \cdot)$ *is convex for every* $(x, y) \in X \times Y$*, then the above conditions are equivalent to*

$$(0, 0) \in \partial[\Psi(\cdot, y_*) - \Psi(x_*, \cdot)](x_*, y_*)$$

Def: A pair $(x_*, y_*) \in X \times Y$ as above is called a *saddle point*

Proof: By the weak duality lemma, it follows that (a) holds iff

$$p(x_*) = p_*, \quad d(y_*) = d_*, \quad p_* = d_*$$

or equivalently, (b) holds due to the fact that

$$X_* = \{x_* \in X : p(x_*) = p_*\}, \quad Y_* = \{y_* \in Y : d(y_*) = d_*\}.$$

Also, in view of the definitions of the functions p and d, (c) is equivalent to

$$d(y_*) = \Psi(x_*, y_*) = p(x_*)$$

and hence to (a) because of the weak duality lemma with $(x, y) = (x_*, y_*)$. Moreover, (c) and (d) are obviously equivalent

Remarks:

a) the set of saddle points is a "rectangle", i.e., of the form $X_* \times Y_*$

b) the value of Ψ is constant over the set of saddle points and is equal to $p_* = d_*$

Proposition

Assume that $X \subset \mathbb{R}^n$ is convex, $Y \subset \mathbb{R}^m$ is compact convex, Ψ is convex-concave, i.e., $\Psi(\cdot, y) - \Psi(x, \cdot)$ is closed convex for every $(x, y) \in X \times Y$ Then,

$$\inf_{x \in X} \sup_{y \in Y} \Psi(x, y) = \sup_{y \in Y} \inf_{x \in X} \Psi(x, y) < \infty$$

More, if the above supremum is finite then $Y_* \neq \emptyset$ (i.e., it is achieved)

More generally, the above result also holds under the condition that Y is closed convex and there exists $x_0 \in X$ such that

$$\{y \in Y : -\Psi(x_0, y) \leq \gamma\}$$

is bounded for every $\gamma \in \mathbb{R}$

Corollary

In addition to the assumptions of the above proposition, if X is also compact, then the value of the above saddle point problem is finite and $X_* \times Y_*$ is nonempty and compact

Projected SM for Saddle Point

Assume

- $X \times Y \subset \mathbb{R}^n \times \mathbb{R}^m$ is nonempty closed convex
- $\Psi : X \times Y \to \mathbb{R}$ is a differentiable closed convex-concave function on $X \times Y$. i.e., $\Psi(\cdot, y) - \Psi(x, \cdot)$ is closed convex for every $(x, y) \in X \times Y$
- for every $y \in Y$,

$$X(y) := \mathrm{Argmin}_{x \in X} \Psi(x, y) \neq \emptyset$$

- there holds

$$\inf_{x \in X} \sup_{y \in Y} \Psi(x, y) = \sup_{y \in Y} \min_{x \in X} \Psi(x, y) = \sup_{y \in Y} d(y)$$

is finite

Idea: Apply projected SM to the dual problem

$$\sup_{y \in Y} d(y)$$

or equivalently,

$$\inf_{y \in Y} (-d)(y)$$

Lemma

Have

a) $-d(\cdot) \in \overline{\text{Conv}}\,(Y)$

b) *if* $x_y \in X(y)$ *then* $-\nabla_y \Psi(x_y, y) \in \partial(-d)(y)$

Proof: a) Since $-d(\cdot) = \sup_{x \in X} -\Psi(x, \cdot)$ and $-\Psi(x, \cdot) \in \overline{\text{Conv}}\,(Y)$, it follows that $-d \in \overline{\text{Conv}}\,(Y)$

b) Assume that $x_y \in X(y)$, and hence that $\Psi(x_y, y) = d(y)$. Then, for any $y' \in Y$, have

$$\begin{aligned}
(-d)(y') &= \sup_{x \in X} -\Psi(x, y') \geq -\Psi(x_y, y') \\
&\geq -\Psi(x_y, y) - \langle \nabla_y \Psi(x_y, y), y' - y \rangle \\
&= -d(y) - \langle \nabla_y \Psi(x_y, y), y' - y \rangle
\end{aligned}$$

and hence b) follows

Projected SM

0) $y_0 \in Y$ is given
1) For $k = 0, 1, 2, \ldots,$ do
   - compute $x_k \in X(y_k) := \mathrm{Argmin}_{x \in X} \Psi(x, y_k)$
   - choose stepsize $\lambda_k > 0$ and set $s_k = -\nabla_y \Psi(x_k, y_k)$
   - $y_{k+1} = Proj_Y(y_k - \lambda_k s_k)$

Assumption: Ther exists $M > 0$ such that

$$\|\nabla_y \Psi(x, y)\| \leq M$$

for every $y \in Y$ and $x \in X(y)$

This assumption implies that $\|s_k\| \leq M$ for every $k \geq 0$

Analysis

For every $y \in Y$, have

$$\begin{aligned}
\|y_{k+1} - y\|^2 &= \|P_Y(y_k - \lambda_k s_k) - P_Y(y)\|^2 \\
&\leq \|(y_k - \lambda_k s_k) - y\|^2 \\
&= \|y_k - y\|^2 + \|\lambda_k s_k\|^2 - 2\lambda_k \langle s_k, y_k - y \rangle
\end{aligned}$$

So, for every $y \in Y$,

$$\begin{aligned}
&\|y_k - y\|^2 - \|y_{k+1} - y\|^2 + \|\lambda_k s_k\|^2 \\
&\qquad \geq 2\lambda_k \langle s_k, y_k - y \rangle \\
\text{(due to def of } s_k) &\qquad = 2\lambda_k \langle \nabla_y(-\Psi)(x_k, y_k), y_k - y \rangle \\
(\text{since } -\Psi(x_k, \cdot) \text{ is convex}) &\qquad \geq 2\lambda_k [(-\Psi)(x_k, y_k) - (-\Psi)(x_k, y)] \\
&\qquad \geq 2\lambda_k [\Psi(x_k, y) - \Psi(x_k, y_k)] \\
(\text{since } x_k \in X(y_k)) &\qquad \geq 2\lambda_k [\Psi(x_k, y) - \Psi(x, y_k)] \quad \forall x \in X
\end{aligned}$$

Analysis (continued)

Summing the above ineq from $k = 0$ to $k = K \geq 1$, we conclude that for every $(x, y) \in X \times Y$

$$\begin{aligned} \|y_0 - y\|^2 - \|y_K - y\|^2 + \sum_{k=1}^{K-1} \|\lambda_k s_k\|^2 \\ &\geq 2 \sum_{k=0}^{K-1} \lambda_k [\Psi(x_k, y) - \Psi(x, y_k)] \\ &\geq 2\Lambda_K [\Psi(\bar{x}_K, y) - \Psi(x, \bar{y}_K)] \end{aligned}$$

where $\Lambda_K := \sum_{k=0}^{K-1} \lambda_k$ and

$$\bar{x}_K := \frac{1}{\Lambda_k} \sum_{k=0}^{K-1} \lambda_k x_k \qquad \bar{y}_K := \frac{1}{\Lambda_k} \sum_{k=0}^{K-1} \lambda_k y_k$$

Analysis (continued)

Proposition

For every $K \geq 1$ and $y \in Y$, have

$$\frac{\|y_0 - y\|^2 + \sum_{k=0}^{K-1} \|\lambda_k s_k\|^2}{2\Lambda_K} \geq \Psi(\bar{x}_K, y) - d(\bar{y}_K)$$

If Y is compact, then we can maximize both sides with respect to $y \in Y$ to obtain

$$\frac{D_Y^2 + \sum_{k=0}^{K-1} \|\lambda_k s_k\|^2}{2\Lambda_K} \geq p(\bar{x}_K) - d(\bar{y}_K)$$

where D_Y is the diameter of Y

Application

$$(P) \quad p_* = \inf\{f(x) : g(x) \leq 0, x \in X\}$$

where

- X is nonempty closed convex
- $f : \mathbb{R}^n \to \mathbb{R}$ is convex
- $g = (g_1, \ldots, g_m)$ and each $g_i : \mathbb{R}^n \to \mathbb{R}$ is convex

Define $Y = \mathbb{R}^m_+$ and

$$\Psi(x, y) = f(x) + \langle y, g(x) \rangle \quad \forall y \in Y$$

Have

$$\sup_{y \in Y} \Psi(x, y) = \begin{cases} f(x) & \text{if } g(x) \leq 0 \\ +\infty & \text{otherwise} \end{cases}$$

and hence (P) is equivalent to

$$\inf_{x \in X} \sup_{y \in Y} \Psi(x, y)$$

Application (continued)

For any $y \in Y$ and $x \in X(y) = \mathrm{Argmin}\,\{\Psi(x, y) : x \in X\}$, have

$$-g(x) = -\nabla_y \Psi(x, y) \in \partial(-d)(y)$$

Assumption: There exists $M > 0$ such that

$$\|g(x)\| \leq M \quad \forall x \in X$$

Remark: Given $y_k \geq 0$, the SM iteration becomes

$$x_k \in \mathrm{Argmin}\,\{f(x) + \langle y_k, g(x)\rangle : x \in X\}$$

$$y_{k+1} = [y_k + \lambda_k g(x_k)]^+$$

Application (continued)

Proposition

For every $K \geq 1$, have

$$\|g^+(\bar{x}_K)\| + f(\bar{x}_K) - p_* \leq \frac{2(\|y_0\|^2 + 1) + \sum_{k=0}^{K-1} \lambda_k^2 M^2}{2\Lambda_K}$$

where $g^+(\cdot) = \max\{g(\cdot), 0\}$

Proof: For every $y \in Y = \mathbb{R}_+^m$, have

$$\Psi(\bar{x}_k, y) - p_* \leq \Psi(\bar{x}_k, y) - d(\bar{y}_k) \leq \frac{\|y_0 - y\|^2 + \sum_{k=0}^{K-1} \|\lambda_k s_k\|^2}{2\Lambda_K}$$

$$\leq \frac{\|y_0 - y\|^2 + \sum_{k=0}^{K-1} \lambda_k^2 M^2}{2\Lambda_K}$$

Now take $y = g^+(\bar{x}_K)/\|g^+(\bar{x}_K)\| \geq 0$ in the above ineq and use the fact that

$$\|y - y_0\|^2 \leq 2\left(\|y_0\|^2 + \|y\|^2\right) \leq 2\left(\|y_0\|^2 + 1\right)$$

$$\Psi(\bar{x}_k, y) = f(\bar{x}_K) + \langle y, g(\bar{x}_K)\rangle = f(\bar{x}_K) + \|g^+(\bar{x}_K)\|$$