CS794/CO673: Optimization for Data Science

Lec 12: Minimax

Yaoliang Yu

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS
DAVID R. CHERITON SCHOOL
OF COMPUTER SCIENCE

October 28, 2022

Problem

Minimax problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Two players $\mathbf{w}$ and $\mathbf{z}$, in $\mathbb{W} \subseteq \mathbb{R}^p$ and $\mathbb{Z} \subseteq \mathbb{R}^d$, respectively
- $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$, the payoff function
- $\mathbf{w}$-player aims to minimize the payoff f
- $\mathbf{z}$-player aims to maximize the payoff f, or equivalently to minimize $-f$

Problem

Minimax problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Two players $\mathbf{w}$ and $\mathbf{z}$, in $\mathbb{W} \subseteq \mathbb{R}^p$ and $\mathbb{Z} \subseteq \mathbb{R}^d$, respectively
- $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$, the payoff function
- $\mathbf{w}$-player aims to minimize the payoff f
- $\mathbf{z}$-player aims to maximize the payoff f, or equivalently to minimize $-f$

Problem

Minimax problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Two players $\mathbf{w}$ and $\mathbf{z}$, in $\mathbb{W} \subseteq \mathbb{R}^p$ and $\mathbb{Z} \subseteq \mathbb{R}^d$, respectively
- $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$, the payoff function
- $\mathbf{w}$-player aims to minimize the payoff f
- $\mathbf{z}$-player aims to maximize the payoff f, or equivalently to minimize $-f$

Problem

Minimax problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Two players $\mathbf{w}$ and $\mathbf{z}$, in $\mathbb{W} \subseteq \mathbb{R}^p$ and $\mathbb{Z} \subseteq \mathbb{R}^d$, respectively
- $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$, the payoff function
- $\mathbf{w}$-player aims to minimize the payoff f
- $\mathbf{z}$-player aims to maximize the payoff f, or equivalently to minimize $-f$

Problem

Minimax problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Two players $\mathbf{w}$ and $\mathbf{z}$, in $\mathbb{W} \subseteq \mathbb{R}^p$ and $\mathbb{Z} \subseteq \mathbb{R}^d$, respectively
- $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$, the payoff function
- $\mathbf{w}$-player aims to minimize the payoff f
- $\mathbf{z}$-player aims to maximize the payoff f, or equivalently to minimize $-f$

Understanding Minimax

- Introducing the upper and lower envelope functions:

$$\overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}), \qquad \underline{f}(\mathbf{z}) := \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})$$

- Minimax becomes the familiar minimization problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w})$$

- "Twin" (or dual) maximin problem:

$$\mathfrak{d}^\star = \left[\sup_{\mathbf{z}\in\mathbb{Z}} \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})\right] = \sup_{\mathbf{z}\in\mathbb{Z}} \underline{f}(\mathbf{z})$$

- Even for a smooth payoff f the envelopes $\underline{f}$ and $\overline{f}$ may still be nonsmooth!

Understanding Minimax

- Introducing the upper and lower envelope functions:

$$\overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}), \qquad \underline{f}(\mathbf{z}) := \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})$$

- Minimax becomes the familiar minimization problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w})$$

- "Twin" (or dual) maximin problem:

$$\mathfrak{d}^\star = \left[\sup_{\mathbf{z}\in\mathbb{Z}} \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})\right] = \sup_{\mathbf{z}\in\mathbb{Z}} \underline{f}(\mathbf{z})$$

- Even for a smooth payoff f the envelopes $\underline{f}$ and $\overline{f}$ may still be nonsmooth!

Understanding Minimax

- Introducing the upper and lower envelope functions:

$$\overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}), \qquad \underline{f}(\mathbf{z}) := \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})$$

- Minimax becomes the familiar minimization problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w})$$

- “Twin” (or dual) maximin problem:

$$\mathfrak{d}^\star = \left[\sup_{\mathbf{z}\in\mathbb{Z}} \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})\right] = \sup_{\mathbf{z}\in\mathbb{Z}} \underline{f}(\mathbf{z})$$

- Even for a smooth payoff f the envelopes $\underline{f}$ and $\overline{f}$ may still be nonsmooth!

Understanding Minimax

- Introducing the upper and lower envelope functions:

$$\overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}), \qquad \underline{f}(\mathbf{z}) := \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})$$

- Minimax becomes the familiar minimization problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \ \overline{f}(\mathbf{w})$$

- “Twin” (or dual) maximin problem:

$$\mathfrak{d}^\star = \left[\sup_{\mathbf{z}\in\mathbb{Z}} \ \inf_{\mathbf{w}\in\mathbb{W}} \ f(\mathbf{w},\mathbf{z})\right] = \sup_{\mathbf{z}\in\mathbb{Z}} \ \underline{f}(\mathbf{z})$$

- Even for a smooth payoff f the envelopes $\underline{f}$ and $\overline{f}$ may still be nonsmooth!

Understanding Minimax

- Introducing the upper and lower envelope functions:

$$\overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}), \qquad \underline{f}(\mathbf{z}) := \inf_{\mathbf{w}\in\mathbb{W}} f(\mathbf{w},\mathbf{z})$$

- Minimax becomes the familiar minimization problem:

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \ \overline{f}(\mathbf{w})$$

- "Twin" (or dual) maximin problem:

$$\mathfrak{d}^\star = \left[\sup_{\mathbf{z}\in\mathbb{Z}} \ \inf_{\mathbf{w}\in\mathbb{W}} \quad f(\mathbf{w},\mathbf{z})\right] = \sup_{\mathbf{z}\in\mathbb{Z}} \ \underline{f}(\mathbf{z})$$

- Even for a smooth payoff f the envelopes $\underline{f}$ and $\overline{f}$ may still be nonsmooth!

Example: $\mathfrak{p}_\star > \mathfrak{d}^\star$

Consider the simple bilinear problem:

$$\inf_{w \neq 0} \sup_{z \neq 0} wz.$$

- $\overline{f}(w) = \infty$ and $\underline{f}(z) = -\infty$
- $\mathfrak{p}_\star = \infty$ and $\mathfrak{d}^\star = -\infty$

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \sup_{z \in [-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$

Example: $\mathfrak{p}_\star > \mathfrak{d}^\star$

Consider the simple bilinear problem:

$$\inf_{w \neq 0} \sup_{z \neq 0} wz.$$

- $\overline{f}(w) = \infty$ and $\underline{f}(z) = -\infty$
- $\mathfrak{p}_\star = \infty$ and $\mathfrak{d}^\star = -\infty$

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \sup_{z \in [-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$

Example: $\mathfrak{p}_\star > \mathfrak{d}^\star$

Consider the simple bilinear problem:

$$\inf_{w \neq 0} \; \sup_{z \neq 0} \; wz.$$

- $\overline{f}(w) = \infty$ and $\underline{f}(z) = -\infty$
- $\mathfrak{p}_\star = \infty$ and $\mathfrak{d}^\star = -\infty$

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \; \sup_{z \in [-1,1]} \; wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$

Example: $\mathfrak{p}_\star > \mathfrak{d}^\star$

Consider the simple bilinear problem:

$$\inf_{w \neq 0} \sup_{z \neq 0} wz.$$

- $\overline{f}(w) = \infty$ and $\underline{f}(z) = -\infty$
- $\mathfrak{p}_\star = \infty$ and $\mathfrak{d}^\star = -\infty$

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \sup_{z \in [-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$

Example: $\mathfrak{p}_\star > \mathfrak{d}^\star$

Consider the simple bilinear problem:

$$\inf_{w \neq 0} \sup_{z \neq 0} wz.$$

- $\overline{f}(w) = \infty$ and $\underline{f}(z) = -\infty$
- $\mathfrak{p}_\star = \infty$ and $\mathfrak{d}^\star = -\infty$

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \sup_{z \in [-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$

Saddle Point

Definition: Saddle point

We call the pair $(\mathbf{w}_\star, \mathbf{z}^\star) \in \mathbb{W} \times \mathbb{Z}$ a saddle point of $f(\mathbf{w}, \mathbf{z})$ over $\mathbb{W} \times \mathbb{Z}$ if

$$\forall \mathbf{w} \in \mathbb{W},\ \forall \mathbf{z} \in \mathbb{Z},\quad f(\mathbf{w}_\star, \mathbf{z}) \leq f(\mathbf{w}_\star, \mathbf{z}^\star) \leq f(\mathbf{w}, \mathbf{z}^\star).$$

- Fixing $\mathbf{w}_\star$, $\quad \mathbf{z}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_\star, \mathbf{z})$, as can be seen from the left inequality
- Fixing $\mathbf{z}^\star$, $\quad \mathbf{w}_\star \in \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}^\star)$, as can be seen from the right inequality
- We will study algorithms that find a saddle point, i.e. solving the primal $\mathfrak{p}_\star$ and dual $\mathfrak{d}^\star$ *simultaneously*

Saddle Point

Definition: Saddle point

We call the pair $(\mathbf{w}_\star, \mathbf{z}^\star) \in \mathbb{W} \times \mathbb{Z}$ a saddle point of $f(\mathbf{w}, \mathbf{z})$ over $\mathbb{W} \times \mathbb{Z}$ if

$$\forall \mathbf{w} \in \mathbb{W},\ \forall \mathbf{z} \in \mathbb{Z}, \quad f(\mathbf{w}_\star, \mathbf{z}) \leq f(\mathbf{w}_\star, \mathbf{z}^\star) \leq f(\mathbf{w}, \mathbf{z}^\star).$$

- Fixing $\mathbf{w}_\star$, $\quad \mathbf{z}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_\star, \mathbf{z})$, as can be seen from the left inequality
- Fixing $\mathbf{z}^\star$, $\quad \mathbf{w}_\star \in \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}^\star)$, as can be seen from the right inequality
- We will study algorithms that find a saddle point, i.e. solving the primal $\mathfrak{p}_\star$ and dual $\mathfrak{d}^\star$ *simultaneously*

Saddle Point

Definition: Saddle point

We call the pair $(\mathbf{w}_\star, \mathbf{z}^\star) \in \mathbb{W} \times \mathbb{Z}$ a saddle point of $f(\mathbf{w}, \mathbf{z})$ over $\mathbb{W} \times \mathbb{Z}$ if

$$\forall \mathbf{w} \in \mathbb{W},\ \forall \mathbf{z} \in \mathbb{Z},\quad f(\mathbf{w}_\star, \mathbf{z}) \le f(\mathbf{w}_\star, \mathbf{z}^\star) \le f(\mathbf{w}, \mathbf{z}^\star).$$

- Fixing $\mathbf{w}_\star$, $\quad \mathbf{z}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_\star, \mathbf{z})$, as can be seen from the left inequality
- Fixing $\mathbf{z}^\star$, $\quad \mathbf{w}_\star \in \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}^\star)$, as can be seen from the right inequality
- We will study algorithms that find a saddle point, i.e. solving the primal $\mathfrak{p}_\star$ and dual $\mathfrak{d}^\star$ *simultaneously*

Saddle Point

Definition: Saddle point

We call the pair $(\mathbf{w}_\star, \mathbf{z}^\star) \in \mathbb{W} \times \mathbb{Z}$ a saddle point of $f(\mathbf{w}, \mathbf{z})$ over $\mathbb{W} \times \mathbb{Z}$ if

$$\forall \mathbf{w} \in \mathbb{W},\ \forall \mathbf{z} \in \mathbb{Z}, \quad f(\mathbf{w}_\star, \mathbf{z}) \leq f(\mathbf{w}_\star, \mathbf{z}^\star) \leq f(\mathbf{w}, \mathbf{z}^\star).$$

- Fixing $\mathbf{w}_\star$, $\mathbf{z}^\star \in \underset{\mathbf{z} \in \mathbb{Z}}{\operatorname{argmax}} f(\mathbf{w}_\star, \mathbf{z})$, as can be seen from the left inequality
- Fixing $\mathbf{z}^\star$, $\mathbf{w}_\star \in \underset{\mathbf{w} \in \mathbb{W}}{\operatorname{argmin}} f(\mathbf{w}, \mathbf{z}^\star)$, as can be seen from the right inequality
- We will study algorithms that find a saddle point, i.e. solving the primal $\mathfrak{p}_\star$ and dual $\mathfrak{d}^\star$ *simultaneously*

Weak and Strong Duality

Theorem: Weak duality

Weak duality, i.e. $\mathfrak{p}_\star \geq \mathfrak{d}^\star$, always holds.

- When equality holds we say strong duality holds

Definition: Optimal sets

$$\mathbb{W}_\star := \operatorname*{argmin}_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w}), \qquad \mathbb{Z}^\star := \operatorname*{argmax}_{\mathbf{z}\in\mathbb{Z}} \underline{f}(\mathbf{z}).$$

Theorem: Strong duality and saddle points

Assuming $\mathbb{W}_\star$ and $\mathbb{Z}^\star$ are nonempty. Then, strong duality holds iff there exists a saddle point, in which case $\mathbb{W}_\star \times \mathbb{Z}^\star$ is the set of all saddle points.

Weak and Strong Duality

Theorem: Weak duality

Weak duality, i.e. $\mathfrak{p}_\star \geq \mathfrak{d}^\star$, always holds.

- When equality holds we say strong duality holds

Definition: Optimal sets

$$\mathbb{W}_\star := \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} \overline{f}(\mathbf{w}), \qquad \mathbb{Z}^\star := \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} \underline{f}(\mathbf{z}).$$

Theorem: Strong duality and saddle points

Assuming $\mathbb{W}_\star$ and $\mathbb{Z}^\star$ are nonempty. Then, strong duality holds iff there exists a saddle point, in which case $\mathbb{W}_\star \times \mathbb{Z}^\star$ is the set of all saddle points.

Stability

Definition: More optimal sets

For $\mathbf{w} \in \mathbb{W}$ and $\mathbf{z} \in \mathbb{Z}$ we also define the sets

$$\mathbb{Z}^{\mathbf{w}} := \mathbb{Z}(\mathbf{w}) := \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}), \qquad \mathbb{W}_{\mathbf{z}} := \mathbb{W}(\mathbf{z}) := \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

- Let $(\mathbf{w}_\star, \mathbf{z}^\star)$ be a saddle point of f over $\mathbb{W} \times \mathbb{Z}$
- Clearly, $\mathbb{W}_\star \subseteq \mathbb{W}(\mathbf{z}^\star), \qquad \mathbb{Z}^\star \subseteq \mathbb{Z}(\mathbf{w}_\star)$
- The saddle point $(\mathbf{w}_\star, \mathbf{z}^\star)$ is stable if equality holds
- If both $(\mathbf{w}_\star, \mathbf{z}^\star)$ and $(\mathbf{u}_\star, \mathbf{v}^\star)$ are saddle points, then so are $(\mathbf{w}_\star, \mathbf{v}^\star)$ and $(\mathbf{u}_\star, \mathbf{z}^\star)$

Stability

Definition: More optimal sets

For $\mathbf{w} \in \mathbb{W}$ and $\mathbf{z} \in \mathbb{Z}$ we also define the sets

$$\mathbb{Z}^{\mathbf{w}} := \mathbb{Z}(\mathbf{w}) := \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}), \qquad \mathbb{W}_{\mathbf{z}} := \mathbb{W}(\mathbf{z}) := \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

- Let $(\mathbf{w}_\star, \mathbf{z}^\star)$ be a saddle point of f over $\mathbb{W} \times \mathbb{Z}$
- Clearly, $\mathbb{W}_\star \subseteq \mathbb{W}(\mathbf{z}^\star), \qquad \mathbb{Z}^\star \subseteq \mathbb{Z}(\mathbf{w}_\star)$
- The saddle point $(\mathbf{w}_\star, \mathbf{z}^\star)$ is stable if equality holds
- If both $(\mathbf{w}_\star, \mathbf{z}^\star)$ and $(\mathbf{u}_\star, \mathbf{v}^\star)$ are saddle points, then so are $(\mathbf{w}_\star, \mathbf{v}^\star)$ and $(\mathbf{u}_\star, \mathbf{z}^\star)$

Stability

Definition: More optimal sets

For $\mathbf{w} \in \mathbb{W}$ and $\mathbf{z} \in \mathbb{Z}$ we also define the sets

$$\mathbb{Z}^{\mathbf{w}} := \mathbb{Z}(\mathbf{w}) := \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}), \qquad \mathbb{W}_{\mathbf{z}} := \mathbb{W}(\mathbf{z}) := \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

- Let $(\mathbf{w}_\star, \mathbf{z}^\star)$ be a saddle point of f over $\mathbb{W} \times \mathbb{Z}$
- Clearly, $\mathbb{W}_\star \subseteq \mathbb{W}(\mathbf{z}^\star), \qquad \mathbb{Z}^\star \subseteq \mathbb{Z}(\mathbf{w}_\star)$
- The saddle point $(\mathbf{w}_\star, \mathbf{z}^\star)$ is stable if equality holds
- If both $(\mathbf{w}_\star, \mathbf{z}^\star)$ and $(\mathbf{u}_\star, \mathbf{v}^\star)$ are saddle points, then so are $(\mathbf{w}_\star, \mathbf{v}^\star)$ and $(\mathbf{u}_\star, \mathbf{z}^\star)$

Stability

Definition: More optimal sets

For $\mathbf{w} \in \mathbb{W}$ and $\mathbf{z} \in \mathbb{Z}$ we also define the sets

$$\mathbb{Z}^{\mathbf{w}} := \mathbb{Z}(\mathbf{w}) := \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}), \qquad \mathbb{W}_{\mathbf{z}} := \mathbb{W}(\mathbf{z}) := \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

- Let $(\mathbf{w}_\star, \mathbf{z}^\star)$ be a saddle point of f over $\mathbb{W} \times \mathbb{Z}$
- Clearly, $\mathbb{W}_\star \subseteq \mathbb{W}(\mathbf{z}^\star), \qquad \mathbb{Z}^\star \subseteq \mathbb{Z}(\mathbf{w}_\star)$
- The saddle point $(\mathbf{w}_\star, \mathbf{z}^\star)$ is stable if equality holds
- If both $(\mathbf{w}_\star, \mathbf{z}^\star)$ and $(\mathbf{u}_\star, \mathbf{v}^\star)$ are saddle points, then so are $(\mathbf{w}_\star, \mathbf{v}^\star)$ and $(\mathbf{u}_\star, \mathbf{z}^\star)$

Stability

Definition: More optimal sets

For $\mathbf{w} \in \mathbb{W}$ and $\mathbf{z} \in \mathbb{Z}$ we also define the sets

$$\mathbb{Z}^{\mathbf{w}} := \mathbb{Z}(\mathbf{w}) := \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}), \qquad \mathbb{W}_{\mathbf{z}} := \mathbb{W}(\mathbf{z}) := \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

- Let $(\mathbf{w}_\star, \mathbf{z}^\star)$ be a saddle point of f over $\mathbb{W} \times \mathbb{Z}$
- Clearly, $\mathbb{W}_\star \subseteq \mathbb{W}(\mathbf{z}^\star), \qquad \mathbb{Z}^\star \subseteq \mathbb{Z}(\mathbf{w}_\star)$
- The saddle point $(\mathbf{w}_\star, \mathbf{z}^\star)$ is stable if equality holds
- If both $(\mathbf{w}_\star, \mathbf{z}^\star)$ and $(\mathbf{u}_\star, \mathbf{v}^\star)$ are saddle points, then so are $(\mathbf{w}_\star, \mathbf{v}^\star)$ and $(\mathbf{u}_\star, \mathbf{z}^\star)$

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w\in[-1,1]} \sup_{z\in[-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w\in[-1,1]} \overline{f}(w)\right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z\in[-1,1]} \underline{f}(z)\right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w\in[-1,1]} 0\cdot w\right] = [-1,1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z\in[-1,1]} 0\cdot z\right] = [-1,1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0,0)$ is not stable

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w\in[-1,1]} \sup_{z\in[-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w\in[-1,1]} \overline{f}(w)\right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z\in[-1,1]} \underline{f}(z)\right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w\in[-1,1]} 0\cdot w\right] = [-1,1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z\in[-1,1]} 0\cdot z\right] = [-1,1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0,0)$ is not stable

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w\in[-1,1]} \sup_{z\in[-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w\in[-1,1]} \overline{f}(w)\right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z\in[-1,1]} \underline{f}(z)\right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w\in[-1,1]} 0\cdot w\right] = [-1,1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z\in[-1,1]} 0\cdot z\right] = [-1,1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0,0)$ is not stable

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \sup_{z \in [-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w \in [-1,1]} \overline{f}(w) \right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z \in [-1,1]} \underline{f}(z) \right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w \in [-1,1]} 0 \cdot w \right] = [-1, 1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z \in [-1,1]} 0 \cdot z \right] = [-1, 1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0, 0)$ is not stable

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \ \sup_{z \in [-1,1]} \ wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w \in [-1,1]} \overline{f}(w)\right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z \in [-1,1]} \underline{f}(z)\right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w \in [-1,1]} 0 \cdot w\right] = [-1, 1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z \in [-1,1]} 0 \cdot z\right] = [-1, 1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0, 0)$ is not stable

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w\in[-1,1]} \sup_{z\in[-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w\in[-1,1]} \overline{f}(w)\right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z\in[-1,1]} \underline{f}(z)\right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w\in[-1,1]} 0\cdot w\right] = [-1, 1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z\in[-1,1]} 0\cdot z\right] = [-1, 1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0, 0)$ is not stable

Example: $\mathfrak{p}_\star = \mathfrak{d}^\star$

Consider the simple constrained bilinear problem:

$$\inf_{w \in [-1,1]} \sup_{z \in [-1,1]} wz.$$

- $\overline{f}(w) = |w|$ and $\underline{f}(z) = -|z|$
- $\mathfrak{p}_\star = 0$ and $\mathfrak{d}^\star = 0$
- $\mathbb{W}_\star = \left[\operatorname*{argmin}_{w \in [-1,1]} \overline{f}(w)\right] = \{0\}$ and $\mathbb{Z}^\star = \left[\operatorname*{argmax}_{z \in [-1,1]} \underline{f}(z)\right] = \{0\}$
- $\mathbb{W}(\mathbf{z}_\star) = \left[\operatorname*{argmin}_{w \in [-1,1]} 0 \cdot w\right] = [-1, 1]$ and $\mathbb{Z}(\mathbf{w}_\star) = \left[\operatorname*{argmax}_{z \in [-1,1]} 0 \cdot z\right] = [-1, 1]$
- $\mathbb{W}_\star \subsetneq \mathbb{W}(\mathbf{z}_\star)$ and $\mathbb{Z}_\star \subsetneq \mathbb{Z}(\mathbf{w}_\star)$
- The unique saddle point $(\mathbf{w}_\star, \mathbf{z}^\star) = (0, 0)$ is not stable

Example: Robust optimization

Learn models that are robust against worst-case perturbations:

$$\inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \sup_{\|\mathbf{z}\|\le\epsilon} \ell(y, \langle \mathbf{x}+\mathbf{z}; \mathbf{w}\rangle) \equiv \inf_{\mathbf{w}} \sup_{\|\mathbf{z}(\cdot)\|\le\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- Minimizer as a defender that tries to learn a good model $\mathbf{w}$
- Maximizer as an attacker that tries to construct a difficult dataset through perturbations $\mathbf{z}$
- When the attacker acts first while the defender responds:

$$\sup_{\|\mathbf{z}(\cdot)\|\le\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- May perturbe distribution $\mathcal{D}$ under metric dist (distributionally robust opt):

$$\inf_{\mathbf{w}} \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\le\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle) \geq \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\le\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle)$$

A. Ben-Tal et al. "Robust Optimization". Princeton University Press, 2009.

Example: Robust optimization

Learn models that are robust against worst-case perturbations:

$$\inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \sup_{\|\mathbf{z}\|\le\epsilon} \ell(y, \langle \mathbf{x}+\mathbf{z}; \mathbf{w}\rangle) \equiv \inf_{\mathbf{w}} \sup_{\|\mathbf{z}(\cdot)\|\le\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- Minimizer as a defender that tries to learn a good model $\mathbf{w}$
- Maximizer as an attacker that tries to construct a difficult dataset through perturbations $\mathbf{z}$
- When the attacker acts first while the defender responds:

$$\sup_{\|\mathbf{z}(\cdot)\|\le\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- May perturbe distribution $\mathcal{D}$ under metric dist (distributionally robust opt):

$$\inf_{\mathbf{w}} \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\le\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle) \geq \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\le\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle)$$

A. Ben-Tal et al. "Robust Optimization". Princeton University Press, 2009.

Example: Robust optimization

Learn models that are robust against worst-case perturbations:

$$\inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \sup_{\|\mathbf{z}\|\leq\epsilon} \ell(y, \langle \mathbf{x}+\mathbf{z}; \mathbf{w}\rangle) \equiv \inf_{\mathbf{w}} \sup_{\|\mathbf{z}(\cdot)\|\leq\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- Minimizer as a defender that tries to learn a good model $\mathbf{w}$
- Maximizer as an attacker that tries to construct a difficult dataset through perturbations $\mathbf{z}$
- When the attacker acts first while the defender responds:

$$\sup_{\|\mathbf{z}(\cdot)\|\leq\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- May perturbe distribution $\mathcal{D}$ under metric dist (distributionally robust opt):

$$\inf_{\mathbf{w}} \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\leq\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle) \geq \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\leq\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle)$$

A. Ben-Tal et al. "Robust Optimization". Princeton University Press, 2009.

Example: Robust optimization

Learn models that are robust against worst-case perturbations:

$$\inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \sup_{\|\mathbf{z}\|\leq\epsilon} \ell(y, \langle \mathbf{x}+\mathbf{z}; \mathbf{w}\rangle) \equiv \inf_{\mathbf{w}} \sup_{\|\mathbf{z}(\cdot)\|\leq\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- Minimizer as a defender that tries to learn a good model $\mathbf{w}$
- Maximizer as an attacker that tries to construct a difficult dataset through perturbations $\mathbf{z}$
- When the attacker acts first while the defender responds:

$$\sup_{\|\mathbf{z}(\cdot)\|\leq\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- May perturb distribution $\mathcal{D}$ under metric dist (distributionally robust opt):

$$\inf_{\mathbf{w}} \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\leq\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle) \geq \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\leq\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle)$$

A. Ben-Tal et al. "Robust Optimization". Princeton University Press, 2009.

Example: Robust optimization

Learn models that are robust against worst-case perturbations:

$$\inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \sup_{\|\mathbf{z}\|\leq\epsilon} \ell(y, \langle \mathbf{x}+\mathbf{z}; \mathbf{w}\rangle) \equiv \inf_{\mathbf{w}} \sup_{\|\mathbf{z}(\cdot)\|\leq\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- Minimizer as a defender that tries to learn a good model $\mathbf{w}$
- Maximizer as an attacker that tries to construct a difficult dataset through perturbations $\mathbf{z}$
- When the attacker acts first while the defender responds:

$$\sup_{\|\mathbf{z}(\cdot)\|\leq\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}} \ell(y, \langle \mathbf{x}+\mathbf{z}(\mathbf{x},y); \mathbf{w}\rangle)$$

- May perturbe distribution $\mathcal{D}$ under metric dist (distributionally robust opt):

$$\inf_{\mathbf{w}} \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\leq\epsilon} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle) \geq \sup_{\mathrm{dist}(\tilde{\mathcal{D}},\mathcal{D})\leq\epsilon} \inf_{\mathbf{w}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\tilde{\mathcal{D}}} \ell(y, \langle \mathbf{x}; \mathbf{w}\rangle)$$

A. Ben-Tal et al. "Robust Optimization". Princeton University Press, 2009.

Example: Lasso revisited

Let us consider the familiar (square root) linear regression problem:

$$\inf_{\mathbf{w}} \ \|X\mathbf{w} - \mathbf{y}\|_2, \quad \text{where} \quad X = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top.$$

Now suppose we perturb each *feature*, i.e., columns in X, independently, arriving at the robust linear regression problem:

$$\inf_{\mathbf{w}} \ \sup_{\forall j, \|\mathbf{z}_j\|_2 \le \lambda} \ \|(X + Z)\mathbf{w} - \mathbf{y}\|_2,$$

where the perturbation matrix $Z = [\mathbf{z}_1, \ldots, \mathbf{z}_d]$.

Prove that robust linear regression is exactly equivalent to (square-root) Lasso (note the absence of the square on the ℓ_2 norm):

$$\inf_{\mathbf{w} \in \mathbb{R}^d} \ \|X\mathbf{w} - \mathbf{y}\|_2 + \lambda\|\mathbf{w}\|_1, \quad \text{where} \quad \|\mathbf{w}\|_1 = \sum_j |w_j|$$

Theorem: Minimax

Let $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$ be a real-valued function on convex sets $\mathbb{W}$ and $\mathbb{Z}$. Suppose

- $f(\mathbf{w}, \cdot) : \mathbb{Z} \to \mathbb{R}$ is continuous and concave on $\mathbb{Z}$ for each $\mathbf{w} \in \mathbb{W}$;
- $f(\cdot, \mathbf{z}) : \mathbb{W} \to \mathbb{R}$ is continuous and convex on $\mathbb{W}$ for each $\mathbf{z} \in \mathbb{Z}$;
- for some finite $F \subseteq \mathbb{Z}$, $\max_{\mathbf{z} \in F} f(\cdot, \mathbf{z})$ is inf-compact, i.e.

$$\bigcap_{\mathbf{z} \in F} \{\mathbf{w} \in \mathbb{W} : f(\mathbf{w}, \mathbf{z}) \leq \alpha\} \text{ is compact for all } \alpha \in \mathbb{R};$$

then strong duality holds and the minimum of the primal problem is attained:

$$\min_{\mathbf{w} \in \mathbb{W}} \sup_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}) = \sup_{\mathbf{z} \in \mathbb{Z}} \inf_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

A similar statement holds by swapping the role of $\mathbf{w}$ and $\mathbf{z}$.

- $\mathbb{W}$ is compact, which is the usual assumption; or
- $\mathbb{W}$ is closed and f is strongly convex in $\mathbf{w}$

Theorem: Minimax

Let $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$ be a real-valued function on convex sets $\mathbb{W}$ and $\mathbb{Z}$. Suppose

- $f(\mathbf{w}, \cdot) : \mathbb{Z} \to \mathbb{R}$ is continuous and concave on $\mathbb{Z}$ for each $\mathbf{w} \in \mathbb{W}$;
- $f(\cdot, \mathbf{z}) : \mathbb{W} \to \mathbb{R}$ is continuous and convex on $\mathbb{W}$ for each $\mathbf{z} \in \mathbb{Z}$;
- for some finite $F \subseteq \mathbb{Z}$, $\max_{\mathbf{z} \in F} f(\cdot, \mathbf{z})$ is inf-compact, i.e.

$$\bigcap_{\mathbf{z} \in F} \{\mathbf{w} \in \mathbb{W} : f(\mathbf{w}, \mathbf{z}) \leq \alpha\} \text{ is compact for all } \alpha \in \mathbb{R};$$

then strong duality holds and the minimum of the primal problem is attained:

$$\min_{\mathbf{w} \in \mathbb{W}} \sup_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}) = \sup_{\mathbf{z} \in \mathbb{Z}} \inf_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

A similar statement holds by swapping the role of $\mathbf{w}$ and $\mathbf{z}$.

- $\mathbb{W}$ is compact, which is the usual assumption; or
- $\mathbb{W}$ is closed and f is strongly convex in $\mathbf{w}$

Theorem: Minimax

Let $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$ be a real-valued function on convex sets $\mathbb{W}$ and $\mathbb{Z}$. Suppose

- $f(\mathbf{w}, \cdot) : \mathbb{Z} \to \mathbb{R}$ is continuous and concave on $\mathbb{Z}$ for each $\mathbf{w} \in \mathbb{W}$;
- $f(\cdot, \mathbf{z}) : \mathbb{W} \to \mathbb{R}$ is continuous and convex on $\mathbb{W}$ for each $\mathbf{z} \in \mathbb{Z}$;
- for some finite $F \subseteq \mathbb{Z}$, $\max_{\mathbf{z} \in F} f(\cdot, \mathbf{z})$ is inf-compact, i.e.

$$\bigcap_{\mathbf{z} \in F} \{\mathbf{w} \in \mathbb{W} : f(\mathbf{w}, \mathbf{z}) \leq \alpha\} \text{ is compact for all } \alpha \in \mathbb{R};$$

then strong duality holds and the minimum of the primal problem is attained:

$$\min_{\mathbf{w} \in \mathbb{W}} \sup_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}) = \sup_{\mathbf{z} \in \mathbb{Z}} \inf_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

A similar statement holds by swapping the role of $\mathbf{w}$ and $\mathbf{z}$.

- $\mathbb{W}$ is compact, which is the usual assumption; or
- $\mathbb{W}$ is closed and f is strongly convex in $\mathbf{w}$

Theorem: Minimax

Let $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$ be a real-valued function on convex sets $\mathbb{W}$ and $\mathbb{Z}$. Suppose

- $f(\mathbf{w}, \cdot) : \mathbb{Z} \to \mathbb{R}$ is continuous and concave on $\mathbb{Z}$ for each $\mathbf{w} \in \mathbb{W}$;
- $f(\cdot, \mathbf{z}) : \mathbb{W} \to \mathbb{R}$ is continuous and convex on $\mathbb{W}$ for each $\mathbf{z} \in \mathbb{Z}$;
- for some finite $F \subseteq \mathbb{Z}$, $\max_{\mathbf{z} \in F} f(\cdot, \mathbf{z})$ is inf-compact, i.e.

$$\bigcap_{\mathbf{z} \in F} \{\mathbf{w} \in \mathbb{W} : f(\mathbf{w}, \mathbf{z}) \leq \alpha\} \text{ is compact for all } \alpha \in \mathbb{R};$$

then strong duality holds and the minimum of the primal problem is attained:

$$\min_{\mathbf{w} \in \mathbb{W}} \sup_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}) = \sup_{\mathbf{z} \in \mathbb{Z}} \inf_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

A similar statement holds by swapping the role of $\mathbf{w}$ and $\mathbf{z}$.

- $\mathbb{W}$ is compact, which is the usual assumption; or
- $\mathbb{W}$ is closed and f is strongly convex in $\mathbf{w}$

Theorem: Minimax

Let $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$ be a real-valued function on convex sets $\mathbb{W}$ and $\mathbb{Z}$. Suppose

- $f(\mathbf{w}, \cdot) : \mathbb{Z} \to \mathbb{R}$ is continuous and concave on $\mathbb{Z}$ for each $\mathbf{w} \in \mathbb{W}$;
- $f(\cdot, \mathbf{z}) : \mathbb{W} \to \mathbb{R}$ is continuous and convex on $\mathbb{W}$ for each $\mathbf{z} \in \mathbb{Z}$;
- for some finite $F \subseteq \mathbb{Z}$, $\max_{\mathbf{z} \in F} f(\cdot, \mathbf{z})$ is inf-compact, i.e.

$$\bigcap_{\mathbf{z} \in F} \{\mathbf{w} \in \mathbb{W} : f(\mathbf{w}, \mathbf{z}) \leq \alpha\} \text{ is compact for all } \alpha \in \mathbb{R};$$

then strong duality holds and the minimum of the primal problem is attained:

$$\min_{\mathbf{w} \in \mathbb{W}} \sup_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}) = \sup_{\mathbf{z} \in \mathbb{Z}} \inf_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

A similar statement holds by swapping the role of $\mathbf{w}$ and $\mathbf{z}$.

- $\mathbb{W}$ is compact, which is the usual assumption; or
- $\mathbb{W}$ is closed and f is strongly convex in $\mathbf{w}$

Theorem: Minimax

Let $f : \mathbb{W} \times \mathbb{Z} \to \mathbb{R}$ be a real-valued function on convex sets $\mathbb{W}$ and $\mathbb{Z}$. Suppose

- $f(\mathbf{w}, \cdot) : \mathbb{Z} \to \mathbb{R}$ is continuous and concave on $\mathbb{Z}$ for each $\mathbf{w} \in \mathbb{W}$;
- $f(\cdot, \mathbf{z}) : \mathbb{W} \to \mathbb{R}$ is continuous and convex on $\mathbb{W}$ for each $\mathbf{z} \in \mathbb{Z}$;
- for some finite $F \subseteq \mathbb{Z}$, $\max_{\mathbf{z} \in F} f(\cdot, \mathbf{z})$ is inf-compact, i.e.

$$\bigcap_{\mathbf{z} \in F} \{\mathbf{w} \in \mathbb{W} : f(\mathbf{w}, \mathbf{z}) \leq \alpha\} \text{ is compact for all } \alpha \in \mathbb{R};$$

then strong duality holds and the minimum of the primal problem is attained:

$$\min_{\mathbf{w} \in \mathbb{W}} \sup_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}, \mathbf{z}) = \sup_{\mathbf{z} \in \mathbb{Z}} \inf_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}).$$

A similar statement holds by swapping the role of $\mathbf{w}$ and $\mathbf{z}$.

- $\mathbb{W}$ is compact, which is the usual assumption; or
- $\mathbb{W}$ is closed and f is strongly convex in $\mathbf{w}$

Example: Lagrangian duality and Slater's condition

For the generic constrained minimization problem

$$\inf_{\mathbf{w}} \ h(\mathbf{w}) \quad \text{s.t.} \quad \mathbf{g}(\mathbf{w}) \leq 0$$

we may construct the Lagrangian which implicitly removes the functional constraints:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \geq 0} \ \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} .$$

- h and $\mathbf{g}$ convex $\implies$ f convex in $\mathbf{w}$ and linear (hence concave) in $\mathbf{z}$
- Slater's condition: $\exists \mathbf{w}_0 \in \operatorname{dom} h$ such that $\mathbf{g}(\mathbf{w}_0) < 0 \implies f$ sup-compact in $\mathbf{z}$
- Applying the minimax theorem ($\mathbf{w}$ and $\mathbf{z}$ switched) we obtain strong duality:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \geq 0} \ \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} \qquad = \qquad \max_{\mathbf{z} \geq 0} \inf_{\mathbf{w}} \ \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})}$$

- For any $\mathbf{z}^\star$, $\mathbb{W}(\mathbf{z}^\star) \supseteq \mathbb{W}_\star$, whereas equality holds if (say) h is strictly convex

Example: Lagrangian duality and Slater's condition

For the generic constrained minimization problem

$$\inf_{\mathbf{w}} \; h(\mathbf{w}) \quad \text{s.t.} \quad \mathbf{g}(\mathbf{w}) \leq 0$$

we may construct the Lagrangian which implicitly removes the functional constraints:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \geq 0} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w}, \mathbf{z})}.$$

- h and $\mathbf{g}$ convex $\implies$ f convex in $\mathbf{w}$ and linear (hence concave) in $\mathbf{z}$
- Slater's condition: $\exists \mathbf{w}_0 \in \operatorname{dom} h$ such that $\mathbf{g}(\mathbf{w}_0) < 0 \implies f$ sup-compact in $\mathbf{z}$
- Applying the minimax theorem ($\mathbf{w}$ and $\mathbf{z}$ switched) we obtain strong duality:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \geq 0} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w}, \mathbf{z})} \qquad = \qquad \max_{\mathbf{z} \geq 0} \inf_{\mathbf{w}} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w}, \mathbf{z})}$$

- For any $\mathbf{z}^\star$, $\mathbb{W}(\mathbf{z}^\star) \supseteq \mathbb{W}_\star$, whereas equality holds if (say) h is strictly convex

Example: Lagrangian duality and Slater's condition

For the generic constrained minimization problem

$$\inf_{\mathbf{w}} \; h(\mathbf{w}) \quad \text{s.t.} \quad \mathbf{g}(\mathbf{w}) \le 0$$

we may construct the Lagrangian which implicitly removes the functional constraints:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \ge 0} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} .$$

- h and $\mathbf{g}$ convex $\implies$ f convex in $\mathbf{w}$ and linear (hence concave) in $\mathbf{z}$
- Slater's condition: $\exists \mathbf{w}_0 \in \operatorname{dom} h$ such that $\mathbf{g}(\mathbf{w}_0) < 0$ $\implies$ f sup-compact in $\mathbf{z}$
- Applying the minimax theorem ($\mathbf{w}$ and $\mathbf{z}$ switched) we obtain strong duality:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \ge 0} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} \qquad = \qquad \max_{\mathbf{z} \ge 0} \inf_{\mathbf{w}} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})}$$

- For any $\mathbf{z}^\star$, $\mathbb{W}(\mathbf{z}^\star) \supseteq \mathbb{W}_\star$, whereas equality holds if (say) h is strictly convex

Example: Lagrangian duality and Slater's condition

For the generic constrained minimization problem

$$\inf_{\mathbf{w}} \; h(\mathbf{w}) \quad \text{s.t.} \quad \mathbf{g}(\mathbf{w}) \leq 0$$

we may construct the Lagrangian which implicitly removes the functional constraints:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \geq 0} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} .$$

- h and $\mathbf{g}$ convex $\implies$ f convex in $\mathbf{w}$ and linear (hence concave) in $\mathbf{z}$
- Slater's condition: $\exists \mathbf{w}_0 \in \operatorname{dom} h$ such that $\mathbf{g}(\mathbf{w}_0) < 0 \implies f$ sup-compact in $\mathbf{z}$
- Applying the minimax theorem ($\mathbf{w}$ and $\mathbf{z}$ switched) we obtain strong duality:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \geq 0} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} \qquad = \qquad \max_{\mathbf{z} \geq 0} \inf_{\mathbf{w}} \; \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})}$$

- For any $\mathbf{z}^\star$, $\mathbb{W}(\mathbf{z}^\star) \supseteq \mathbb{W}_\star$, whereas equality holds if (say) h is strictly convex

Example: Lagrangian duality and Slater's condition

For the generic constrained minimization problem

$$\inf_{\mathbf{w}} \ h(\mathbf{w}) \quad \text{s.t.} \quad \mathbf{g}(\mathbf{w}) \le 0$$

we may construct the Lagrangian which implicitly removes the functional constraints:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \ge 0} \ \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} .$$

- h and $\mathbf{g}$ convex $\implies$ f convex in $\mathbf{w}$ and linear (hence concave) in $\mathbf{z}$
- Slater's condition: $\exists \mathbf{w}_0 \in \operatorname{dom} h$ such that $\mathbf{g}(\mathbf{w}_0) < 0 \implies f$ sup-compact in $\mathbf{z}$
- Applying the minimax theorem ($\mathbf{w}$ and $\mathbf{z}$ switched) we obtain strong duality:

$$\inf_{\mathbf{w}} \sup_{\mathbf{z} \ge 0} \ \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})} \quad = \quad \max_{\mathbf{z} \ge 0} \inf_{\mathbf{w}} \ \underbrace{h(\mathbf{w}) + \langle \mathbf{g}(\mathbf{w}), \mathbf{z} \rangle}_{f(\mathbf{w},\mathbf{z})}$$

- For any $\mathbf{z}^\star$, $\mathbb{W}(\mathbf{z}^\star) \supseteq \mathbb{W}_\star$, whereas equality holds if (say) h is strictly convex

Fenchel Conjugate

The Fenchel conjugate of a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is defined as:

$$f^*(\mathbf{w}^*) = \sup_{\mathbf{w}} \langle \mathbf{w}; \mathbf{w}^* \rangle - f(\mathbf{w}),$$

which is always closed and convex (even when f is not).

- Fenchel-Young inequality follows from the definition:

  $$f(\mathbf{w}) + f^*(\mathbf{w}^*) \geq \langle \mathbf{w}; \mathbf{w}^* \rangle ,$$

  with equality iff $\mathbf{w}^* = \partial f(\mathbf{w})$.

- $f^{**} = f$ iff f is (closed) convex

Fenchel Conjugate

The Fenchel conjugate of a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is defined as:

$$f^*(\mathbf{w}^*) = \sup_{\mathbf{w}} \langle \mathbf{w}; \mathbf{w}^* \rangle - f(\mathbf{w}),$$

which is always closed and convex (even when f is not).

- Fenchel-Young inequality follows from the definition:

  $$f(\mathbf{w}) + f^*(\mathbf{w}^*) \geq \langle \mathbf{w}; \mathbf{w}^* \rangle ,$$

  with equality iff $\mathbf{w}^* = \partial f(\mathbf{w})$.
- $f^{**} = f$ iff f is (closed) convex

Fenchel Conjugate

The Fenchel conjugate of a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is defined as:

$$f^*(\mathbf{w}^*) = \sup_{\mathbf{w}} \langle \mathbf{w}; \mathbf{w}^* \rangle - f(\mathbf{w}),$$

which is always closed and convex (even when f is not).

- Fenchel-Young inequality follows from the definition:

  $$f(\mathbf{w}) + f^*(\mathbf{w}^*) \geq \langle \mathbf{w}; \mathbf{w}^* \rangle,$$

  with equality iff $\mathbf{w}^* = \partial f(\mathbf{w})$.

- $f^{**} = f$ iff f is (closed) convex

Example: Fenchel-Rockafellar duality

$$\begin{aligned}\inf_{\mathbf{w}}\ g(A\mathbf{w}) + h(\mathbf{w}) &= \inf_{\mathbf{w}} \sup_{\mathbf{z}}\ \underbrace{\langle A\mathbf{w}; \mathbf{z}\rangle - g^*(\mathbf{z}) + h(\mathbf{w})}_{f(\mathbf{w},\mathbf{z})} \\ &\geq \sup_{\mathbf{z}} \inf_{\mathbf{w}} \langle A\mathbf{w}; \mathbf{z}\rangle - g^*(\mathbf{z}) + h(\mathbf{w}) \\ &= -\inf_{\mathbf{z}} \sup_{\mathbf{w}} \left\langle \mathbf{w}; -A^\top \mathbf{z}\right\rangle + g^*(\mathbf{z}) - h(\mathbf{w}) \\ &= -\inf_{\mathbf{z}} g^*(\mathbf{z}) + h^*(-A^\top \mathbf{z})\end{aligned}$$

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$, provided that g and h are both convex
- Conditions for strong duality include:
  - $\mathbf{0} \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} h)$, i.e. for any $\mathbf{d}$ there exists some $\lambda = \lambda(\mathbf{d}) > 0$ such that for any $t \in [0, \lambda]$, there exists $\mathbf{w} \in \operatorname{dom} h$ so that $A\mathbf{w} + t\mathbf{d} \in \operatorname{dom} g$
  - $A \operatorname{dom} h \cap \operatorname{cont}(g) \neq \emptyset$, where $\operatorname{cont}(g)$ is the set of points at which g is continuous

Example: Fenchel-Rockafellar duality

$$
\begin{aligned}
\inf_{\mathbf{w}} \; g(A\mathbf{w}) + h(\mathbf{w}) &= \inf_{\mathbf{w}} \sup_{\mathbf{z}} \; \underbrace{\langle A\mathbf{w}; \mathbf{z}\rangle - g^*(\mathbf{z}) + h(\mathbf{w})}_{f(\mathbf{w},\mathbf{z})} \\
&\geq \sup_{\mathbf{z}} \inf_{\mathbf{w}} \langle A\mathbf{w}; \mathbf{z}\rangle - g^*(\mathbf{z}) + h(\mathbf{w}) \\
&= -\inf_{\mathbf{z}} \sup_{\mathbf{w}} \left\langle \mathbf{w}; -A^\top \mathbf{z}\right\rangle + g^*(\mathbf{z}) - h(\mathbf{w}) \\
&= -\inf_{\mathbf{z}} g^*(\mathbf{z}) + h^*(-A^\top \mathbf{z})
\end{aligned}
$$

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$, provided that g and h are both convex
- Conditions for strong duality include:
  - $\mathbf{0} \in \mathrm{core}(\mathrm{dom}\, g - A\, \mathrm{dom}\, h)$, i.e. for any $\mathbf{d}$ there exists some $\lambda = \lambda(\mathbf{d}) > 0$ such that for any $t \in [0, \lambda]$, there exists $\mathbf{w} \in \mathrm{dom}\, h$ so that $A\mathbf{w} + t\mathbf{d} \in \mathrm{dom}\, g$
  - $A\, \mathrm{dom}\, h \cap \mathrm{cont}(g) \neq \emptyset$, where $\mathrm{cont}(g)$ is the set of points at which g is continuous

Example: Fenchel-Rockafellar duality

$$
\begin{aligned}
\inf_{\mathbf{w}} \; g(A\mathbf{w}) + h(\mathbf{w}) &= \inf_{\mathbf{w}} \sup_{\mathbf{z}} \; \underbrace{\langle A\mathbf{w}; \mathbf{z}\rangle - g^*(\mathbf{z}) + h(\mathbf{w})}_{f(\mathbf{w},\mathbf{z})} \\
&\geq \sup_{\mathbf{z}} \inf_{\mathbf{w}} \langle A\mathbf{w}; \mathbf{z}\rangle - g^*(\mathbf{z}) + h(\mathbf{w}) \\
&= -\inf_{\mathbf{z}} \sup_{\mathbf{w}} \left\langle \mathbf{w}; -A^\top \mathbf{z} \right\rangle + g^*(\mathbf{z}) - h(\mathbf{w}) \\
&= -\inf_{\mathbf{z}} g^*(\mathbf{z}) + h^*(-A^\top \mathbf{z})
\end{aligned}
$$

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$, provided that g and h are both convex
- Conditions for strong duality include:
  - $\mathbf{0} \in \operatorname{core}(\operatorname{dom} g - A \operatorname{dom} h)$, i.e. for any $\mathbf{d}$ there exists some $\lambda = \lambda(\mathbf{d}) > 0$ such that for any $t \in [0, \lambda]$, there exists $\mathbf{w} \in \operatorname{dom} h$ so that $A\mathbf{w} + t\mathbf{d} \in \operatorname{dom} g$
  - $A \operatorname{dom} h \cap \operatorname{cont}(g) \neq \emptyset$, where $\operatorname{cont}(g)$ is the set of points at which g is continuous

Example: Fenchel-Rockafellar duality

$$
\begin{aligned}
\inf_{\mathbf{w}} \; g(A\mathbf{w}) + h(\mathbf{w}) &= \inf_{\mathbf{w}} \sup_{\mathbf{z}} \; \underbrace{\langle A\mathbf{w}; \mathbf{z} \rangle - g^*(\mathbf{z}) + h(\mathbf{w})}_{f(\mathbf{w},\mathbf{z})} \\
&\geq \sup_{\mathbf{z}} \inf_{\mathbf{w}} \langle A\mathbf{w}; \mathbf{z} \rangle - g^*(\mathbf{z}) + h(\mathbf{w}) \\
&= -\inf_{\mathbf{z}} \sup_{\mathbf{w}} \left\langle \mathbf{w}; -A^\top \mathbf{z} \right\rangle + g^*(\mathbf{z}) - h(\mathbf{w}) \\
&= -\inf_{\mathbf{z}} g^*(\mathbf{z}) + h^*(-A^\top \mathbf{z})
\end{aligned}
$$

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$, provided that g and h are both convex
- Conditions for strong duality include:
  - $\mathbf{0} \in \mathrm{core}(\mathrm{dom}\, g - A\, \mathrm{dom}\, h)$, i.e. for any $\mathbf{d}$ there exists some $\lambda = \lambda(\mathbf{d}) > 0$ such that for any $t \in [0, \lambda]$, there exists $\mathbf{w} \in \mathrm{dom}\, h$ so that $A\mathbf{w} + t\mathbf{d} \in \mathrm{dom}\, g$
  - $A\, \mathrm{dom}\, h \cap \mathrm{cont}(g) \neq \emptyset$, where $\mathrm{cont}(g)$ is the set of points at which g is continuous

Example: Fenchel-Rockafellar duality

$$\begin{aligned} \inf_{\mathbf{w}} \ g(A\mathbf{w}) + h(\mathbf{w}) &= \inf_{\mathbf{w}} \sup_{\mathbf{z}} \ \underbrace{\langle A\mathbf{w}; \mathbf{z} \rangle - g^*(\mathbf{z}) + h(\mathbf{w})}_{f(\mathbf{w},\mathbf{z})} \\ &\geq \sup_{\mathbf{z}} \inf_{\mathbf{w}} \langle A\mathbf{w}; \mathbf{z} \rangle - g^*(\mathbf{z}) + h(\mathbf{w}) \\ &= -\inf_{\mathbf{z}} \sup_{\mathbf{w}} \left\langle \mathbf{w}; -A^\top \mathbf{z} \right\rangle + g^*(\mathbf{z}) - h(\mathbf{w}) \\ &= -\inf_{\mathbf{z}} g^*(\mathbf{z}) + h^*(-A^\top \mathbf{z}) \end{aligned}$$

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$, provided that g and h are both convex
- Conditions for strong duality include:
  - $\mathbf{0} \in \mathrm{core}(\mathrm{dom}\, g - A\, \mathrm{dom}\, h)$, i.e. for any $\mathbf{d}$ there exists some $\lambda = \lambda(\mathbf{d}) > 0$ such that for any $t \in [0, \lambda]$, there exists $\mathbf{w} \in \mathrm{dom}\, h$ so that $A\mathbf{w} + t\mathbf{d} \in \mathrm{dom}\, g$
  - $A\, \mathrm{dom}\, h \cap \mathrm{cont}(g) \neq \emptyset$, where $\mathrm{cont}(g)$ is the set of points at which g is continuous

Alternating

- The saddle point definition suggests the following natural alternating algorithm:

Algorithm 1: Alternating Minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathbb{W} \times \mathbb{Z} \cap \operatorname{dom} f$
for $t = 0, 1, 2, \ldots$ do
    $\mathbf{w}_{t+1} \leftarrow \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}_t)$
    $\mathbf{z}_{t+1} \leftarrow \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_{t+1}, \mathbf{z})$    // or $\mathbf{z}_{t+1} \leftarrow \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_t, \mathbf{z})$
```

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$ $\implies$ each step is a convex Problem

Alternating

- The saddle point definition suggests the following natural alternating algorithm:

```
Algorithm 2: Alternating Minimax
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathbb{W} \times \mathbb{Z} \cap \operatorname{dom} f$
for $t = 0, 1, 2, \ldots$ do
    $\mathbf{w}_{t+1} \leftarrow \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}_t)$
    $\mathbf{z}_{t+1} \leftarrow \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_{t+1}, \mathbf{z})$    // or $\mathbf{z}_{t+1} \leftarrow \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_t, \mathbf{z})$
```

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$ $\implies$ each step is a convex Problem

Alternating

- The saddle point definition suggests the following natural alternating algorithm:

```
Algorithm 3: Alternating Minimax
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathbb{W} \times \mathbb{Z} \cap \operatorname{dom} f$
for $t = 0, 1, 2, \ldots$ do
    $\mathbf{w}_{t+1} \leftarrow \operatorname*{argmin}_{\mathbf{w} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}_t)$
    $\mathbf{z}_{t+1} \leftarrow \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_{t+1}, \mathbf{z})$    // or $\mathbf{z}_{t+1} \leftarrow \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}_t, \mathbf{z})$
```

- f is convex in $\mathbf{w}$ and concave in $\mathbf{z}$ $\implies$ each step is a convex Problem

Example: Alternating does not work!

$$\min_{w\in[-1,1]} \max_{z\in[-1,1]} wz.$$

It is easy to see that strong duality holds and

$$\overline{f}(w) = |w|, \qquad \underline{f}(z) = -|z|,$$

so that we have a unique saddle point $(w_\star, z^\star) = (0, 0)$, which is not stable:

$$\mathbb{W}(0) = [-1, 1] \supsetneq \mathbb{W}_\star = \{0\} \text{ and similarly } \mathbb{Z}(0) = [-1, 1] \supsetneq \mathbb{Z}^\star = \{0\}.$$

Applying the alternating algorithm with any $z_0 \neq 0$ we obtain

$$\begin{aligned} z_0 \neq 0 \implies w_1 = z_1 &= -\operatorname{sign}(z_0) \\ \implies w_2 = z_2 &= \operatorname{sign}(z_0) \\ \implies w_3 = z_3 &= -\operatorname{sign}(z_0), \qquad \text{oscillating!} \end{aligned}$$

Example: Alternating does not work?

$$\min_{w\in[-1,1]} \max_{z\in[-1,1]} z\exp(w).$$

It is easy to see that strong duality holds and

$$\overline{f}(w) = \exp(w), \qquad \underline{f}(z) = z\exp(-\operatorname{sign}(z)),$$

so that we have a unique saddle point $(w_\star, z^\star) = (-1, 1)$ which is now stable.

Applying the alternating algorithm with any z_0 we obtain

$$w_1 = -\operatorname{sign}(z_0), z_1 = 1 \implies w_2 = -1, z_2 = 1 \implies w_3 = -1, z_3 = 1 \implies \cdots,$$

which converges to the unique saddle point in two iterations!

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}) = \inf_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w}), \quad \text{where} \quad \overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Apply the subgradient algorithm to minimize $\overline{f}(\mathbf{w})$
- $\partial\overline{f}(\mathbf{w}) = \partial f(\mathbf{w},\mathbf{z}^\star)$ where $\mathbf{z}^\star \in \mathrm{argmax}_{\mathbf{z}\in\mathbb{Z}}\ f(\mathbf{w},\mathbf{z})$

Algorithm 4: Uzawa's algorithm for minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathbb{W} \times \mathbb{Z} \cap \mathrm{dom}\, f$
for $t = 0, 1, \ldots$ do
    $\mathbf{z}_t = \mathrm{argmax}_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w}_t, \mathbf{z})$          // solve inner maximization exactly
    compute subgradient $\mathbf{g}_t = \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)$     // treating $\mathbf{z}_t$ as constant
    choose step size $\eta_t$                                              // see ??
    optional: $\mathbf{g}_t \leftarrow \mathbf{g}_t / \|\mathbf{g}_t\|$                          // normalization
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \mathbf{g}_t]$                 // subgrad on outer minimization
```

H. Uzawa. "Iterative methods for concave programming". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 154–165, J. M. Danskin. "The theory of max-min and its application to weapons allocation problems". Springer, 1967, V. F. Dem'yanov. "On the minimax problem". *Soviet Mathematics Doklady*, vol. 187, no. 2 (1969), pp. 255–258.

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}) = \inf_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w}), \quad \text{where} \quad \overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Apply the subgradient algorithm to minimize $\overline{f}(\mathbf{w})$
- $\partial\overline{f}(\mathbf{w}) = \partial f(\mathbf{w},\mathbf{z}^\star)$ where $\mathbf{z}^\star \in \mathrm{argmax}_{\mathbf{z}\in\mathbb{Z}} \ f(\mathbf{w},\mathbf{z})$

Algorithm 5: Uzawa's algorithm for minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathbb{W} \times \mathbb{Z} \cap \mathrm{dom}\, f$
for $t = 0, 1, \ldots$ do
    $\mathbf{z}_t = \mathrm{argmax}_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w}_t, \mathbf{z})$          // solve inner maximization exactly
    compute subgradient $\mathbf{g}_t = \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)$          // treating $\mathbf{z}_t$ as constant
    choose step size $\eta_t$          // see ??
    optional: $\mathbf{g}_t \leftarrow \mathbf{g}_t / \|\mathbf{g}_t\|$          // normalization
    $\mathbf{w}_{t+1} = P_{\mathbb{W}}[\mathbf{w}_t - \eta_t \mathbf{g}_t]$          // subgrad on outer minimization
```

H. Uzawa. "Iterative methods for concave programming". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 154–165, J. M. Danskin. "The theory of max-min and its application to weapons allocation problems". Springer, 1967, V. F. Dem'yanov. "On the minimax problem". *Soviet Mathematics Doklady*, vol. 187, no. 2 (1969), pp. 255–258.

$$\mathfrak{p}_\star = \inf_{\mathbf{w}\in\mathbb{W}} \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z}) = \inf_{\mathbf{w}\in\mathbb{W}} \overline{f}(\mathbf{w}), \quad \text{where} \quad \overline{f}(\mathbf{w}) := \sup_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w},\mathbf{z})$$

- Apply the subgradient algorithm to minimize $\overline{f}(\mathbf{w})$
- $\partial\overline{f}(\mathbf{w}) = \partial f(\mathbf{w},\mathbf{z}^\star)$ where $\mathbf{z}^\star \in \mathrm{argmax}_{\mathbf{z}\in\mathbb{Z}}\ f(\mathbf{w},\mathbf{z})$

Algorithm 6: Uzawa's algorithm for minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathbb{W} \times \mathbb{Z} \cap \mathrm{dom}\, f$
for $t = 0, 1, \ldots$ do
    $\mathbf{z}_t = \mathrm{argmax}_{\mathbf{z}\in\mathbb{Z}} f(\mathbf{w}_t, \mathbf{z})$          // solve inner maximization exactly
    compute subgradient $\mathbf{g}_t = \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)$   // treating $\mathbf{z}_t$ as constant
    choose step size $\eta_t$                                         // see ??
    optional: $\mathbf{g}_t \leftarrow \mathbf{g}_t / \|\mathbf{g}_t\|$                  // normalization
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \mathbf{g}_t]$                 // subgrad on outer minimization
```

H. Uzawa. "Iterative methods for concave programming". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 154–165, J. M. Danskin. "The theory of max-min and its application to weapons allocation problems". Springer, 1967, V. F. Dem'yanov. "On the minimax problem". *Soviet Mathematics Doklady*, vol. 187, no. 2 (1969), pp. 255–258.

- Replace exact inner maximization in Uzawa with a single gradient ascent step

```
Algorithm 7: Gradient Descent Ascent (GDA) for Minimax
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathrm{dom}\, f \cap \mathbb{W} \times \mathbb{Z}$
for $t = 0, 1, \ldots$ do
    choose step size $\eta_t > 0$
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)]$        // GD on minimization
    $\mathbf{z}_{t+1} = \mathrm{P}_{\mathbb{Z}}[\mathbf{z}_t - \eta_t \partial_{\mathbf{z}} - f(\mathbf{w}_t, \mathbf{z}_t)]$        // GA on maximization
```

- Use different step sizes on $\mathbf{w}$ and $\mathbf{z}$
- Use $\mathbf{w}_{t+1}$ in the update on $\mathbf{z}$ (or vice versa)
- Use stochastic gradients in both steps (more on this later)
- After every update in $\mathbf{w}$, perform k updates in $\mathbf{z}$ (or vice versa)

G. W. Brown and J. v. Neumann. "Solutions of Games by Differential Equations". In: *Contributions to the Theory of Games I*. ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1950, pp. 73–79, K. J. Arrow and L. Hurwicz. "Gradient method for concave programming I: Local results". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 117–126.

- Replace exact inner maximization in Uzawa with a single gradient ascent step

Algorithm 8: Gradient Descent Ascent (GDA) for Minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathrm{dom}\, f \cap \mathbb{W} \times \mathbb{Z}$
for $t = 0, 1, \ldots$ do
    choose step size $\eta_t > 0$
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)]$        // GD on minimization
    $\mathbf{z}_{t+1} = \mathrm{P}_{\mathbb{Z}}[\mathbf{z}_t - \eta_t \partial_{\mathbf{z}} - f(\mathbf{w}_t, \mathbf{z}_t)]$        // GA on maximization
```

- Use different step sizes on $\mathbf{w}$ and $\mathbf{z}$
- Use $\mathbf{w}_{t+1}$ in the update on $\mathbf{z}$ (or vice versa)
- Use stochastic gradients in both steps (more on this later)
- After every update in $\mathbf{w}$, perform k updates in $\mathbf{z}$ (or vice versa)

G. W. Brown and J. v. Neumann. "Solutions of Games by Differential Equations". In: *Contributions to the Theory of Games I*. ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1950, pp. 73–79, K. J. Arrow and L. Hurwicz. "Gradient method for concave programming I: Local results". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 117–126.

- Replace exact inner maximization in Uzawa with a single gradient ascent step

Algorithm 9: Gradient Descent Ascent (GDA) for Minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathrm{dom}\, f \cap \mathbb{W} \times \mathbb{Z}$
for $t = 0, 1, \ldots$ do
    choose step size $\eta_t > 0$
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)]$    // GD on minimization
    $\mathbf{z}_{t+1} = \mathrm{P}_{\mathbb{Z}}[\mathbf{z}_t - \eta_t \partial_{\mathbf{z}} - f(\mathbf{w}_t, \mathbf{z}_t)]$    // GA on maximization
```

- Use different step sizes on $\mathbf{w}$ and $\mathbf{z}$
- Use $\mathbf{w}_{t+1}$ in the update on $\mathbf{z}$ (or vice versa)
- Use stochastic gradients in both steps (more on this later)
- After every update in $\mathbf{w}$, perform k updates in $\mathbf{z}$ (or vice versa)

G. W. Brown and J. v. Neumann. "Solutions of Games by Differential Equations". In: *Contributions to the Theory of Games I*. ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1950, pp. 73–79, K. J. Arrow and L. Hurwicz. "Gradient method for concave programming I: Local results". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 117–126.

- Replace exact inner maximization in Uzawa with a single gradient ascent step

Algorithm 10: Gradient Descent Ascent (GDA) for Minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \operatorname{dom} f \cap \mathbb{W} \times \mathbb{Z}$
for $t = 0, 1, \ldots$ do
    choose step size $\eta_t > 0$
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)]$    // GD on minimization
    $\mathbf{z}_{t+1} = \mathrm{P}_{\mathbb{Z}}[\mathbf{z}_t - \eta_t \partial_{\mathbf{z}}-f(\mathbf{w}_t, \mathbf{z}_t)]$    // GA on maximization
```

- Use different step sizes on $\mathbf{w}$ and $\mathbf{z}$
- Use $\mathbf{w}_{t+1}$ in the update on $\mathbf{z}$ (or vice versa)
- Use stochastic gradients in both steps (more on this later)
- After every update in $\mathbf{w}$, perform k updates in $\mathbf{z}$ (or vice versa)

G. W. Brown and J. v. Neumann. "Solutions of Games by Differential Equations". In: *Contributions to the Theory of Games I*. ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1950, pp. 73–79, K. J. Arrow and L. Hurwicz. "Gradient method for concave programming I: Local results". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 117–126.

- Replace exact inner maximization in Uzawa with a single gradient ascent step

Algorithm 11: Gradient Descent Ascent (GDA) for Minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathrm{dom}\, f \cap \mathbb{W} \times \mathbb{Z}$
for $t = 0, 1, \ldots$ do
    choose step size $\eta_t > 0$
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)]$    // GD on minimization
    $\mathbf{z}_{t+1} = \mathrm{P}_{\mathbb{Z}}[\mathbf{z}_t - \eta_t \partial_{\mathbf{z}} - f(\mathbf{w}_t, \mathbf{z}_t)]$    // GA on maximization
```

- Use different step sizes on $\mathbf{w}$ and $\mathbf{z}$
- Use $\mathbf{w}_{t+1}$ in the update on $\mathbf{z}$ (or vice versa)
- Use stochastic gradients in both steps (more on this later)
- After every update in $\mathbf{w}$, perform k updates in $\mathbf{z}$ (or vice versa)

G. W. Brown and J. v. Neumann. "Solutions of Games by Differential Equations". In: *Contributions to the Theory of Games I*. ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1950, pp. 73–79, K. J. Arrow and L. Hurwicz. "Gradient method for concave programming I: Local results". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 117–126.

- Replace exact inner maximization in Uzawa with a single gradient ascent step

Algorithm 12: Gradient Descent Ascent (GDA) for Minimax

```
Input: $(\mathbf{w}_0, \mathbf{z}_0) \in \mathrm{dom}\, f \cap \mathbb{W} \times \mathbb{Z}$
for $t = 0, 1, \ldots$ do
    choose step size $\eta_t > 0$
    $\mathbf{w}_{t+1} = \mathrm{P}_{\mathbb{W}}[\mathbf{w}_t - \eta_t \partial_{\mathbf{w}} f(\mathbf{w}_t, \mathbf{z}_t)]$        // GD on minimization
    $\mathbf{z}_{t+1} = \mathrm{P}_{\mathbb{Z}}[\mathbf{z}_t - \eta_t \partial_{\mathbf{z}} {-} f(\mathbf{w}_t, \mathbf{z}_t)]$        // GA on maximization
```

- Use different step sizes on $\mathbf{w}$ and $\mathbf{z}$
- Use $\mathbf{w}_{t+1}$ in the update on $\mathbf{z}$ (or vice versa)
- Use stochastic gradients in both steps (more on this later)
- After every update in $\mathbf{w}$, perform k updates in $\mathbf{z}$ (or vice versa)

G. W. Brown and J. v. Neumann. "Solutions of Games by Differential Equations". In: *Contributions to the Theory of Games I*. ed. by H. W. Kuhn and A. W. Tucker. Princeton University Press, 1950, pp. 73–79, K. J. Arrow and L. Hurwicz. "Gradient method for concave programming I: Local results". In: *Studies in linear and non-linear programming*. Ed. by K. J. Arrow et al. Standford University Press, 1958, pp. 117–126.

Example: Vanilla GDA may never converge for any step size

$$\min_{w\in[-1,1]} \max_{z\in[-1,1]} wz \quad \equiv \quad \max_{z\in[-1,1]} \min_{w\in[-1,1]} wz,$$

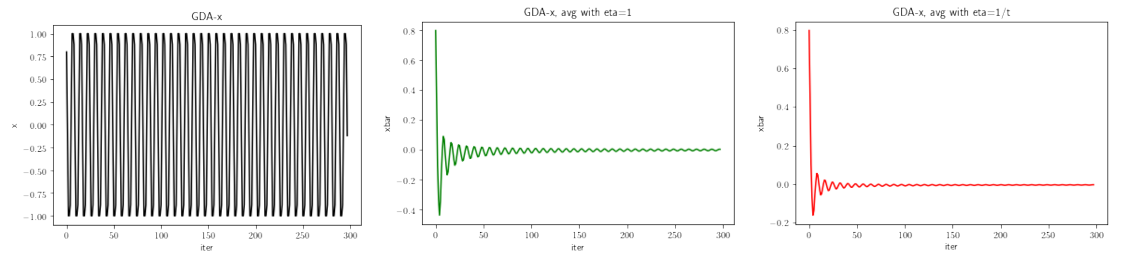
which has a unique (non-stable) saddle-point at $(w_\star, z^\star) = (0, 0)$.

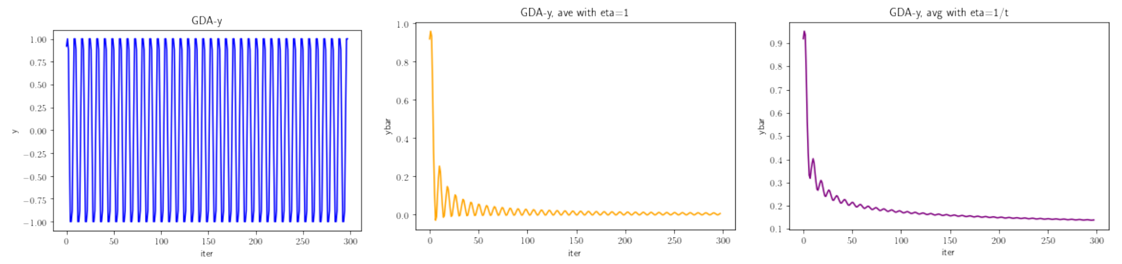
If we run vanilla (projected) GDA with step size $\eta_t \geq 0$, then

$$\begin{aligned} w_{t+1} &= [w_t - \eta_t z_t]_{-1}^{1}, \qquad z_{t+1} = [z_t + \eta_t w_t]_{-1}^{1} \\ w_{t+1}^2 + z_{t+1}^2 &\geq 1 \wedge [(w_t - \eta_t z_t)^2 + (z_t + \eta_t w_t)^2] \\ &= 1 \wedge [(1 + \eta_t^2)(w_t^2 + z_t^2)] \\ &\geq 1 \wedge (w_t^2 + z_t^2). \end{aligned}$$

Therefore, if we do *not* initialize at the saddle point $(w_\star, z^\star) = (0, 0)$, then

$$\|(w_t, z_t)\| \geq 1 \wedge \|(w_0, z_0)\| > 0 = \|(w_\star, z^\star)\|.$$


GDA-x
x
iter
GDA-x, avg with eta=1
xbar
iter
GDA-x, avg with eta=1/t
xbar
iter


GDA-y
y
iter
GDA-y, ave with eta=1
ybar
iter
GDA-y, avg with eta=1/t
ybar
iter

Example: Fenchel conjugate of Jensen-Shannon divergence

$$f(t) = t\log t - (t+1)\log(t+1) + \log 4.$$

We derive its Fenchel conjugate:

$$f^*(s) = \sup_t st - f(t) = \sup_t st - t\log t + (t+1)\log(t+1) - \log 4.$$

Taking derivative w.r.t. t we obtain

$$s - \log t - 1 + \log(t+1) + 1 = 0 \iff t = \tfrac{1}{\exp(-s)-1},$$

and plugging it back we get

$$\begin{aligned} f^*(s) &= \tfrac{s}{\exp(-s)-1} - \tfrac{1}{\exp(-s)-1}\log\tfrac{1}{\exp(-s)-1} + \tfrac{\exp(-s)}{\exp(-s)-1}\log\tfrac{\exp(-s)}{\exp(-s)-1} - \log 4 \\ &= \tfrac{s}{\exp(-s)-1} - \tfrac{1}{\exp(-s)-1}\log\tfrac{1}{\exp(-s)-1} + \tfrac{\exp(-s)}{\exp(-s)-1}\log\tfrac{1}{\exp(-s)-1} - \tfrac{s\exp(-s)}{\exp(-s)-1} - \log 4 \\ &= -s - \log(\exp(-s)-1) - \log 4 = -\log(1-\exp(s)) - \log 4. \end{aligned}$$

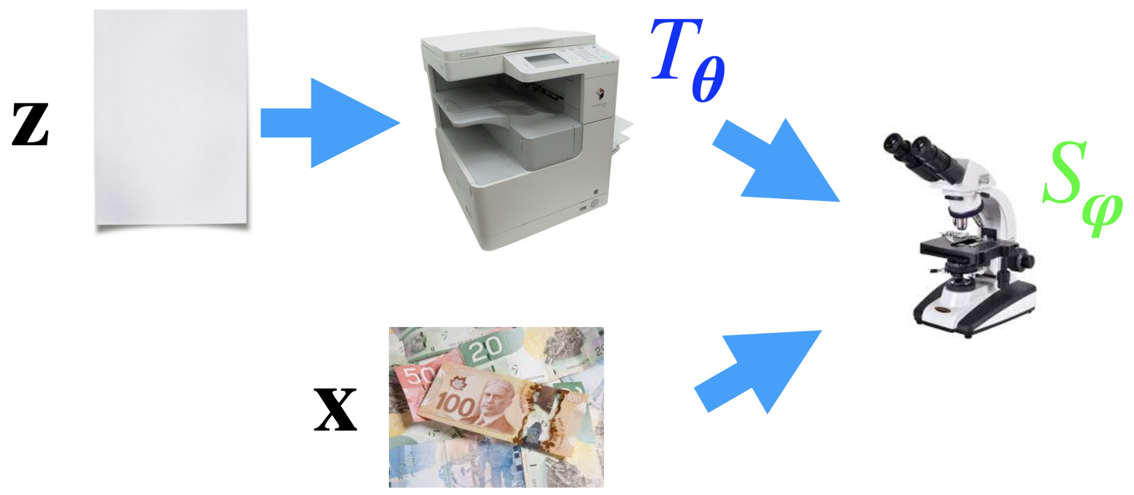
Definition: Generative adversarial networks (GAN)

$$\inf_{\boldsymbol{\theta}} \ \mathsf{JS}(\mathbf{X} \| \mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})), \quad \text{where} \quad \mathsf{JS}(\mathsf{p}\|\mathsf{q}) = \mathsf{D}_f(\mathsf{p}\|\mathsf{q}) = \mathsf{KL}(\mathsf{p}\|\tfrac{\mathsf{p}+\mathsf{q}}{2}) + \mathsf{KL}(\mathsf{p}\|\tfrac{\mathsf{p}+\mathsf{q}}{2})$$

To circumvent the lack of the density $\mathsf{q}(\mathbf{x})$ of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$, we expand using duality:

$$\begin{aligned} \mathsf{JS}(\mathbf{X}\|\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})) &= \int_{\mathbf{x}} f\big(\mathsf{p}(\mathbf{x})/\mathsf{q}(\mathbf{x})\big)\mathsf{q}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathbf{x}} [\sup_s s\mathsf{p}(\mathbf{x})/\mathsf{q}(\mathbf{x}) - f^*(s)]\mathsf{q}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ &= \int_{\mathbf{x}} [\sup_s s\mathsf{p}(\mathbf{x}) - f^*(s)\mathsf{q}(\mathbf{x})]\,\mathrm{d}\mathbf{x} \\ &= \sup_{\mathsf{S}:\mathbb{R}^d\to\mathbb{R}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\mathbf{x}} f^*(\mathsf{S}(\mathbf{x}))\mathsf{q}(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ &= \sup_{\mathsf{S}:\mathbb{R}^d\to\mathbb{R}} \mathbb{E}\mathsf{S}(\mathbf{X}) - \mathbb{E}f^*(\mathsf{S}(\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z}))) \\ &\geq \inf_{\boldsymbol{\theta}} \sup_{\boldsymbol{\phi}} \mathbb{E}\mathsf{S}_{\boldsymbol{\phi}}(\mathbf{X}) - \mathbb{E}f^*(\mathsf{S}_{\boldsymbol{\phi}}(\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z}))) \end{aligned}$$

I. Goodfellow et al. "Generative Adversarial Nets". In: *NIPS*. 2014.


$\mathbf{z}$
T_θ
S_φ
$\mathbf{x}$

$\mathbf{z}$ T_{θ^*} $\mathbf{x}$

Example: Catch me if you can

Let us consider the game between the generator $\mathsf{q}(\mathbf{x})$ (the implicit density of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$) and the discriminator $\mathsf{S}(\mathbf{x})$:

$$\inf_{\mathsf{q}} \sup_{\mathsf{S}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\, \mathrm{d}\mathbf{x} + \int_{\mathbf{x}} \log\big(1 - \exp(\mathsf{S}(\mathbf{x}))\big)\mathsf{q}(\mathbf{x})\, \mathrm{d}\mathbf{x} + \log 4.$$

- Fixing the generator q, what is the optimal discriminator S?
- Plugging the optimal discriminator S back in, what is the optimal generator?
- Fixing the discriminator S, what is the optimal generator q?
- Plugging the optimal generator q back in, what is the optimal discriminator?
- Does strong duality hold? Stability?

Example: Catch me if you can

Let us consider the game between the generator $\mathsf{q}(\mathbf{x})$ (the implicit density of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$) and the discriminator $\mathsf{S}(\mathbf{x})$:

$$\inf_{\mathsf{q}} \sup_{\mathsf{S}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\, \mathrm{d}\mathbf{x} + \int_{\mathbf{x}} \log\big(1 - \exp(\mathsf{S}(\mathbf{x}))\big)\mathsf{q}(\mathbf{x})\, \mathrm{d}\mathbf{x} + \log 4.$$

- Fixing the generator q, what is the optimal discriminator S?
- Plugging the optimal discriminator S back in, what is the optimal generator?
- Fixing the discriminator S, what is the optimal generator q?
- Plugging the optimal generator q back in, what is the optimal discriminator?
- Does strong duality hold? Stability?

Example: Catch me if you can

Let us consider the game between the generator $\mathsf{q}(\mathbf{x})$ (the implicit density of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$) and the discriminator $\mathsf{S}(\mathbf{x})$:

$$\inf_{\mathsf{q}} \sup_{\mathsf{S}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbf{x}} \log\left(1 - \exp(\mathsf{S}(\mathbf{x}))\right)\mathsf{q}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \log 4.$$

- Fixing the generator q, what is the optimal discriminator S?
- Plugging the optimal discriminator S back in, what is the optimal generator?
- Fixing the discriminator S, what is the optimal generator q?
- Plugging the optimal generator q back in, what is the optimal discriminator?
- Does strong duality hold? Stability?

Example: Catch me if you can

Let us consider the game between the generator $\mathsf{q}(\mathbf{x})$ (the implicit density of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$) and the discriminator $\mathsf{S}(\mathbf{x})$:

$$\inf_{\mathsf{q}} \sup_{\mathsf{S}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbf{x}} \log\big(1 - \exp(\mathsf{S}(\mathbf{x}))\big)\mathsf{q}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \log 4.$$

- Fixing the generator q, what is the optimal discriminator S?
- Plugging the optimal discriminator S back in, what is the optimal generator?
- Fixing the discriminator S, what is the optimal generator q?
- Plugging the optimal generator q back in, what is the optimal discriminator?
- Does strong duality hold? Stability?

Example: Catch me if you can

Let us consider the game between the generator $\mathsf{q}(\mathbf{x})$ (the implicit density of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$) and the discriminator $\mathsf{S}(\mathbf{x})$:

$$\inf_{\mathsf{q}} \sup_{\mathsf{S}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{\mathbf{x}} \log\big(1 - \exp(\mathsf{S}(\mathbf{x}))\big)\mathsf{q}(\mathbf{x})\,\mathrm{d}\mathbf{x} + \log 4.$$

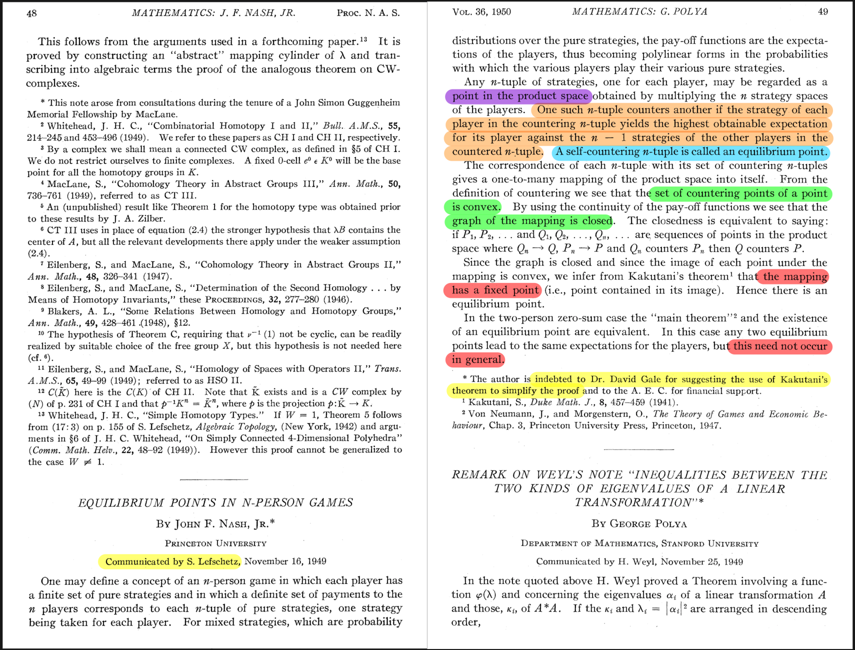
- Fixing the generator q, what is the optimal discriminator S?
- Plugging the optimal discriminator S back in, what is the optimal generator?
- Fixing the discriminator S, what is the optimal generator q?
- Plugging the optimal generator q back in, what is the optimal discriminator?
- Does strong duality hold? Stability?

Example: Catch me if you can

Let us consider the game between the generator $\mathsf{q}(\mathbf{x})$ (the implicit density of $\mathsf{T}_{\boldsymbol{\theta}}(\mathbf{Z})$) and the discriminator $\mathsf{S}(\mathbf{x})$:

$$\inf_{\mathsf{q}} \sup_{\mathsf{S}} \int_{\mathbf{x}} \mathsf{S}(\mathbf{x})\mathsf{p}(\mathbf{x})\, \mathrm{d}\mathbf{x} + \int_{\mathbf{x}} \log\big(1 - \exp(\mathsf{S}(\mathbf{x}))\big)\mathsf{q}(\mathbf{x})\, \mathrm{d}\mathbf{x} + \log 4.$$

- Fixing the generator q, what is the optimal discriminator S?
- Plugging the optimal discriminator S back in, what is the optimal generator?
- Fixing the discriminator S, what is the optimal generator q?
- Plugging the optimal generator q back in, what is the optimal discriminator?
- Does strong duality hold? Stability?



This follows from the arguments used in a forthcoming paper.[13] It is proved by constructing an "abstract" mapping cylinder of λ and transcribing into algebraic terms the proof of the analogous theorem on CW-complexes.

[*] This note arose from consultations during the tenure of a John Simon Guggenheim Memorial Fellowship by MacLane.

[2] Whitehead, J. H. C., "Combinatorial Homotopy I and II," *Bull. A.M.S.*, **55,** 214–245 and 453–496 (1949). We refer to these papers as CH I and CH II, respectively.

[3] By a complex we shall mean a connected CW complex, as defined in §5 of CH I. We do not restrict ourselves to finite complexes. A fixed 0-cell $e^0 \epsilon K^0$ will be the base point for all the homotopy groups in K.

[4] MacLane, S., "Cohomology Theory in Abstract Groups III," *Ann. Math.*, **50,** 736–761 (1949), referred to as CT III.

[5] An (unpublished) result like Theorem 1 for the homotopy type was obtained prior to these results by J. A. Zilber.

[6] CT III uses in place of equation (2.4) the stronger hypothesis that λB contains the center of A, but all the relevant developments there apply under the weaker assumption (2.4).

[7] Eilenberg, S., and MacLane, S., "Cohomology Theory in Abstract Groups II," *Ann. Math.*, **48,** 326–341 (1947).

[8] Eilenberg, S., and MacLane, S., "Determination of the Second Homology . . . by Means of Homotopy Invariants," these PROCEEDINGS, **32,** 277–280 (1946).

[9] Blakers, A. L., "Some Relations Between Homology and Homotopy Groups," *Ann. Math.*, **49,** 428–461 .(1948), §12.

[10] The hypothesis of Theorem C, requiring that ν^{-1} (1) not be cyclic, can be readily realized by suitable choice of the free group X, but this hypothesis is not needed here (cf. [6]).

[11] Eilenberg, S., and MacLane, S., "Homology of Spaces with Operators II," *Trans. A.M.S.*, **65,** 49–99 (1949); referred to as HSO II.

[12] $C(\tilde{K})$ here is the $C(K)$ of CH II. Note that $\tilde{K}$ exists and is a CW complex by (N) of p. 231 of CH I and that $p^{-1}K^n = \tilde{K}^n$, where p is the projection $p:\tilde{K} \to K$.

[13] Whitehead, J. H. C., "Simple Homotopy Types." If $W = 1$, Theorem 5 follows from (17:3) on p. 155 of S. Lefschetz, *Algebraic Topology*, (New York, 1942) and arguments in §6 of J. H. C. Whitehead, "On Simply Connected 4-Dimensional Polyhedra" (*Comm. Math. Helv.*, **22,** 48–92 (1949)). However this proof cannot be generalized to the case $W \neq 1$.

EQUILIBRIUM POINTS IN N-PERSON GAMES

BY JOHN F. NASH, JR.[*]

PRINCETON UNIVERSITY

Communicated by S. Lefschetz, November 16, 1949

One may define a concept of an n-person game in which each player has a finite set of pure strategies and in which a definite set of payments to the n players corresponds to each n-tuple of pure strategies, one strategy being taken for each player. For mixed strategies, which are probability distributions over the pure strategies, the pay-off functions are the expectations of the players, thus becoming polylinear forms in the probabilities with which the various players play their various pure strategies.

Any n-tuple of strategies, one for each player, may be regarded as a point in the product space obtained by multiplying the n strategy spaces of the players. One such n-tuple counters another if the strategy of each player in the countering n-tuple yields the highest obtainable expectation for its player against the $n - 1$ strategies of the other players in the countered n-tuple. A self-countering n-tuple is called an equilibrium point.

The correspondence of each n-tuple with its set of countering n-tuples gives a one-to-many mapping of the product space into itself. From the definition of countering we see that the set of countering points of a point is convex. By using the continuity of the pay-off functions we see that the graph of the mapping is closed. The closedness is equivalent to saying: if $P_1, P_2, \ldots$ and $Q_1, Q_2, \ldots, Q_n, \ldots$ are sequences of points in the product space where $Q_n \to Q$, $P_n \to P$ and Q_n counters P_n then Q counters P.

Since the graph is closed and since the image of each point under the mapping is convex, we infer from Kakutani's theorem[1] that the mapping has a fixed point (i.e., point contained in its image). Hence there is an equilibrium point.

In the two-person zero-sum case the "main theorem"[2] and the existence of an equilibrium point are equivalent. In this case any two equilibrium points lead to the same expectations for the players, but this need not occur in general.

[*] The author is indebted to Dr. David Gale for suggesting the use of Kakutani's theorem to simplify the proof and to the A. E. C. for financial support.

[1] Kakutani, S., *Duke Math. J.*, **8,** 457–459 (1941).

[2] Von Neumann, J., and Morgenstern, O., *The Theory of Games and Economic Behaviour*, Chap. 3, Princeton University Press, Princeton, 1947.

REMARK ON WEYL'S NOTE "INEQUALITIES BETWEEN THE TWO KINDS OF EIGENVALUES OF A LINEAR TRANSFORMATION"*

BY GEORGE POLYA

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY

Communicated by H. Weyl, November 25, 1949

In the note quoted above H. Weyl proved a Theorem involving a function $\varphi(\lambda)$ and concerning the eigenvalues α_i of a linear transformation A and those, κ_i, of A^*A. If the κ_i and $\lambda_i = |\alpha_i|^2$ are arranged in descending order,

Definition: Equilibrium in game theory

Suppose we have n players participating in a game, where the players act simultaneously by each choosing a strategy w_i and then receiving payoff

$$f_i(\mathbf{w}) = f_i(w_1, \ldots, w_n), \quad \mathbf{w} \in \mathbb{W}, \quad i = 1, \ldots, n.$$

Nash considered the product space $\mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$, and proved the existence of an equilibrium $\mathbf{w}^*$ where

$$\forall i, \quad w_i^* \in \underset{w_i \in \mathbb{W}_i}{\operatorname{argmax}} \; f_i(w_1^*, \ldots, w_{i-1}^*, w_i, w_{i+1}^*, \ldots, w_n^*),$$

under the assumption that each f_i is continuous in $\mathbf{w}$ and concave in w_i.

- No player unilaterally has motive to deviate from its strategy in equilibrium $\mathbf{w}^*$
- Striking similarity to alternating minimization where $f_i \equiv -f$
- Basically alternating minimax with $-f_1 = f = f_2$

J. F. Nash. "Equilibrium Points in N-Person Games". *Proceedings of the National Academy of Sciences*, vol. 36, no. 1 (1950), pp. 48–49.

Definition: Equilibrium in game theory

Suppose we have n players participating in a game, where the players act simultaneously by each choosing a strategy w_i and then receiving payoff

$$f_i(\mathbf{w}) = f_i(w_1, \ldots, w_n), \quad \mathbf{w} \in \mathbb{W}, \quad i = 1, \ldots, n.$$

Nash considered the product space $\mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$, and proved the existence of an equilibrium $\mathbf{w}^*$ where

$$\forall i, \quad w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \ f_i(w_1^*, \ldots, w_{i-1}^*, w_i, w_{i+1}^*, \ldots, w_n^*),$$

under the assumption that each f_i is continuous in $\mathbf{w}$ and concave in w_i.

- No player unilaterally has motive to deviate from its strategy in equilibrium $\mathbf{w}^*$
- Striking similarity to alternating minimization where $f_i \equiv -f$
- Basically alternating minimax with $-f_1 = f = f_2$

J. F. Nash. "Equilibrium Points in N-Person Games". *Proceedings of the National Academy of Sciences*, vol. 36, no. 1 (1950), pp. 48–49.

Definition: Equilibrium in game theory

Suppose we have n players participating in a game, where the players act simultaneously by each choosing a strategy w_i and then receiving payoff

$$f_i(\mathbf{w}) = f_i(w_1, \dots, w_n), \quad \mathbf{w} \in \mathbb{W}, \quad i = 1, \dots, n.$$

Nash considered the product space $\mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$, and proved the existence of an equilibrium $\mathbf{w}^*$ where

$$\forall i, \quad w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \; f_i(w_1^*, \dots, w_{i-1}^*, w_i, w_{i+1}^*, \dots, w_n^*),$$

under the assumption that each f_i is continuous in $\mathbf{w}$ and concave in w_i.

- No player unilaterally has motive to deviate from its strategy in equilibrium $\mathbf{w}^*$
- Striking similarity to alternating minimization where $f_i \equiv -f$
- Basically alternating minimax with $-f_1 = f = f_2$

J. F. Nash. "Equilibrium Points in N-Person Games". *Proceedings of the National Academy of Sciences*, vol. 36, no. 1 (1950), pp. 48–49.

Definition: Equilibrium in game theory

Suppose we have n players participating in a game, where the players act simultaneously by each choosing a strategy w_i and then receiving payoff

$$f_i(\mathbf{w}) = f_i(w_1, \ldots, w_n), \quad \mathbf{w} \in \mathbb{W}, \quad i = 1, \ldots, n.$$

Nash considered the product space $\mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$, and proved the existence of an equilibrium $\mathbf{w}^*$ where

$$\forall i, \quad w_i^* \in \underset{w_i \in \mathbb{W}_i}{\operatorname{argmax}} \; f_i(w_1^*, \ldots, w_{i-1}^*, w_i, w_{i+1}^*, \ldots, w_n^*),$$

under the assumption that each f_i is continuous in $\mathbf{w}$ and concave in w_i.

- No player unilaterally has motive to deviate from its strategy in equilibrium $\mathbf{w}^*$
- Striking similarity to alternating minimization where $f_i \equiv -f$
- Basically alternating minimax with $-f_1 = f = f_2$

J. F. Nash. "Equilibrium Points in N-Person Games". *Proceedings of the National Academy of Sciences*, vol. 36, no. 1 (1950), pp. 48–49.

Algorithm 13: Alternating algorithm for Nash equilibrium

```
Input: $\mathbf{w}_0 \in \mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$
1 for $t = 0, 1, \ldots$ do
2     for $i = 1, \ldots, n$ do
3         $w_{i,t+1} \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} f_i(w_{1,t}, \ldots, w_{i-1,t}, w_i, w_{i+1,t}, \ldots, w_{n,t})$    // simultaneously
```

- Line 3 as a multi-valued mapping $\mathsf{T} : \mathbb{W} \rightrightarrows \mathbb{W}$ such that $\mathbf{w}_{t+1} \in \mathsf{T}(\mathbf{w}_t)$
- T is compact convex valued and upper semicontinuous
- According to Kakutani's fixed point theorem there exists a fixed point $\mathbf{w}^* \in \mathsf{T}(\mathbf{w}^*)$
- Nash did not prove alternating will necessarily converge to any fixed point!
- $w \mapsto 2w$ admits a unique fixed point $w^* = 0$ but will never converge to it

---

H. W. Kuhn. "Simplicial approximation of fixed points". *Proceedings of the National Academy of Sciences*, vol. 61, no. 4 (1968), pp. 1238–1242, C. Daskalakis et al. "The Complexity of Computing a Nash Equilibrium". *SIAM Journal on Computing*, vol. 39, no. 1 (2009), pp. 195–259, X. Chen et al. "Settling the complexity of computing two-player Nash equilibria". *Journal of the ACM*, vol. 56, no. 3 (2009), p. 14, K. Etessami and M. Yannakakis. "On the Complexity of Nash Equilibria and Other Fixed Points". *SIAM Journal on Computing*, vol. 39, no. 6 (2010), pp. 2531–2597.

Algorithm 14: Alternating algorithm for Nash equilibrium

```
Input: $\mathbf{w}_0 \in \mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$
1 for $t = 0, 1, \ldots$ do
2     for $i = 1, \ldots, n$ do
3         $w_{i,t+1} \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} f_i(w_{1,t}, \ldots, w_{i-1,t}, w_i, w_{i+1,t}, \ldots, w_{n,t})$    // simultaneously
```

- Line 3 as a multi-valued mapping $\mathsf{T} : \mathbb{W} \rightrightarrows \mathbb{W}$ such that $\mathbf{w}_{t+1} \in \mathsf{T}(\mathbf{w}_t)$
- T is compact convex valued and upper semicontinuous
- According to Kakutani's fixed point theorem there exists a fixed point $\mathbf{w}^* \in \mathsf{T}(\mathbf{w}^*)$
- Nash did not prove alternating will necessarily converge to any fixed point!
- $w \mapsto 2w$ admits a unique fixed point $w^* = 0$ but will never converge to it

---

H. W. Kuhn. "Simplicial approximation of fixed points". *Proceedings of the National Academy of Sciences*, vol. 61, no. 4 (1968), pp. 1238–1242, C. Daskalakis et al. "The Complexity of Computing a Nash Equilibrium". *SIAM Journal on Computing*, vol. 39, no. 1 (2009), pp. 195–259, X. Chen et al. "Settling the complexity of computing two-player Nash equilibria". *Journal of the ACM*, vol. 56, no. 3 (2009), p. 14, K. Etessami and M. Yannakakis. "On the Complexity of Nash Equilibria and Other Fixed Points". *SIAM Journal on Computing*, vol. 39, no. 6 (2010), pp. 2531–2597.

Algorithm 15: Alternating algorithm for Nash equilibrium

```
Input: $\mathbf{w}_0 \in \mathbb{W} = \prod_{i=1}^n \mathbb{W}_i$
1 for $t = 0, 1, \ldots$ do
2     for $i = 1, \ldots, n$ do
3         $w_{i,t+1} \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} f_i(w_{1,t}, \ldots, w_{i-1,t}, w_i, w_{i+1,t}, \ldots, w_{n,t})$   // simultaneously
```

- Line 3 as a multi-valued mapping $\mathsf{T} : \mathbb{W} \rightrightarrows \mathbb{W}$ such that $\mathbf{w}_{t+1} \in \mathsf{T}(\mathbf{w}_t)$
- T is compact convex valued and upper semicontinuous
- According to Kakutani's fixed point theorem there exists a fixed point $\mathbf{w}^* \in \mathsf{T}(\mathbf{w}^*)$
- Nash did not prove alternating will necessarily converge to any fixed point!
- $w \mapsto 2w$ admits a unique fixed point $w^* = 0$ but will never converge to it

---

H. W. Kuhn. "Simplicial approximation of fixed points". *Proceedings of the National Academy of Sciences*, vol. 61, no. 4 (1968), pp. 1238–1242, C. Daskalakis et al. "The Complexity of Computing a Nash Equilibrium". *SIAM Journal on Computing*, vol. 39, no. 1 (2009), pp. 195–259, X. Chen et al. "Settling the complexity of computing two-player Nash equilibria". *Journal of the ACM*, vol. 56, no. 3 (2009), p. 14, K. Etessami and M. Yannakakis. "On the Complexity of Nash Equilibria and Other Fixed Points". *SIAM Journal on Computing*, vol. 39, no. 6 (2010), pp. 2531–2597.

Algorithm 16: Alternating algorithm for Nash equilibrium

```
Input: $\mathbf{w}_0 \in \mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$
1 for $t = 0, 1, \ldots$ do
2     for $i = 1, \ldots, n$ do
3         $w_{i,t+1} \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} f_i(w_{1,t}, \ldots, w_{i-1,t}, w_i, w_{i+1,t}, \ldots, w_{n,t})$    // simultaneously
```

- Line 3 as a multi-valued mapping $\mathsf{T} : \mathbb{W} \rightrightarrows \mathbb{W}$ such that $\mathbf{w}_{t+1} \in \mathsf{T}(\mathbf{w}_t)$
- T is compact convex valued and upper semicontinuous
- According to Kakutani's fixed point theorem there exists a fixed point $\mathbf{w}^* \in \mathsf{T}(\mathbf{w}^*)$
- Nash did not prove alternating will necessarily converge to any fixed point!
- $w \mapsto 2w$ admits a unique fixed point $w^* = 0$ but will never converge to it

---

H. W. Kuhn. "Simplicial approximation of fixed points". *Proceedings of the National Academy of Sciences*, vol. 61, no. 4 (1968), pp. 1238–1242, C. Daskalakis et al. "The Complexity of Computing a Nash Equilibrium". *SIAM Journal on Computing*, vol. 39, no. 1 (2009), pp. 195–259, X. Chen et al. "Settling the complexity of computing two-player Nash equilibria". *Journal of the ACM*, vol. 56, no. 3 (2009), p. 14, K. Etessami and M. Yannakakis. "On the Complexity of Nash Equilibria and Other Fixed Points". *SIAM Journal on Computing*, vol. 39, no. 6 (2010), pp. 2531–2597.

Algorithm 17: Alternating algorithm for Nash equilibrium

```
Input: $\mathbf{w}_0 \in \mathbb{W} = \prod_{i=1}^n \mathbb{W}_i$
1 for $t = 0, 1, \ldots$ do
2     for $i = 1, \ldots, n$ do
3         $w_{i,t+1} \in \underset{w_i \in \mathbb{W}_i}{\operatorname{argmax}} f_i(w_{1,t}, \ldots, w_{i-1,t}, w_i, w_{i+1,t}, \ldots, w_{n,t})$   // simultaneously
```

- Line 3 as a multi-valued mapping $\mathsf{T} : \mathbb{W} \rightrightarrows \mathbb{W}$ such that $\mathbf{w}_{t+1} \in \mathsf{T}(\mathbf{w}_t)$
- T is compact convex valued and upper semicontinuous
- According to Kakutani's fixed point theorem there exists a fixed point $\mathbf{w}^* \in \mathsf{T}(\mathbf{w}^*)$
- Nash did not prove alternating will necessarily converge to any fixed point!
- $w \mapsto 2w$ admits a unique fixed point $w^* = 0$ but will never converge to it

---

H. W. Kuhn. "Simplicial approximation of fixed points". *Proceedings of the National Academy of Sciences*, vol. 61, no. 4 (1968), pp. 1238–1242, C. Daskalakis et al. "The Complexity of Computing a Nash Equilibrium". *SIAM Journal on Computing*, vol. 39, no. 1 (2009), pp. 195–259, X. Chen et al. "Settling the complexity of computing two-player Nash equilibria". *Journal of the ACM*, vol. 56, no. 3 (2009), p. 14, K. Etessami and M. Yannakakis. "On the Complexity of Nash Equilibria and Other Fixed Points". *SIAM Journal on Computing*, vol. 39, no. 6 (2010), pp. 2531–2597.

Algorithm 18: Alternating algorithm for Nash equilibrium

```
Input: $\mathbf{w}_0 \in \mathbb{W} = \prod_{i=1}^{n} \mathbb{W}_i$
1 for $t = 0, 1, \ldots$ do
2     for $i = 1, \ldots, n$ do
3         $w_{i,t+1} \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} f_i(w_{1,t}, \ldots, w_{i-1,t}, w_i, w_{i+1,t}, \ldots, w_{n,t})$    // simultaneously
```

- Line 3 as a multi-valued mapping $\mathsf{T} : \mathbb{W} \rightrightarrows \mathbb{W}$ such that $\mathbf{w}_{t+1} \in \mathsf{T}(\mathbf{w}_t)$
- T is compact convex valued and upper semicontinuous
- According to Kakutani's fixed point theorem there exists a fixed point $\mathbf{w}^* \in \mathsf{T}(\mathbf{w}^*)$
- Nash did not prove alternating will necessarily converge to any fixed point!
- $w \mapsto 2w$ admits a unique fixed point $w^* = 0$ but will never converge to it

---

H. W. Kuhn. "Simplicial approximation of fixed points". *Proceedings of the National Academy of Sciences*, vol. 61, no. 4 (1968), pp. 1238–1242, C. Daskalakis et al. "The Complexity of Computing a Nash Equilibrium". *SIAM Journal on Computing*, vol. 39, no. 1 (2009), pp. 195–259, X. Chen et al. "Settling the complexity of computing two-player Nash equilibria". *Journal of the ACM*, vol. 56, no. 3 (2009), p. 14, K. Etessami and M. Yannakakis. "On the Complexity of Nash Equilibria and Other Fixed Points". *SIAM Journal on Computing*, vol. 39, no. 6 (2010), pp. 2531–2597.

Definition: Reducing n-person game to minimax

Quite remarkably, Nikaidô and Isoda proved the existence of a normalized equilibrium

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z}\in\mathbb{Z}} \ f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := \sum_i f_i(w_1, \ldots, w_{i-1}, z_i, w_{i+1}, \ldots, w_n)$$

is defined on the product space $\mathbb{W} \times \mathbb{Z}$ with $\mathbb{Z} = \mathbb{W}$.

Any normalized equilibrium is an equilibrium while the converse may not hold.

We can now formulate the (normalized) Nash equilibrium in n-person non-cooperative game as the minimax problem:

$$0 = \min_{\mathbf{w}\in\mathbb{W}} \max_{\mathbf{z}\in\mathbb{Z}} \ f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w}),$$

which is concave in $\mathbf{z} \in \mathbb{Z} = \mathbb{W}$ if each f_i is concave in in its i-th input.

H. Nikaidô and K. Isoda. "Note on non-cooperative convex games". *Pacific Journal of Mathematics*, vol. 5, no. 1 (1955), pp. 807–815.

- Zero-sum: two players (i.e. $n = 2$) with opposing payoff functions $f_1 + f_2 = 0$
- Saddle point is exactly Nash's equilibrium
- Payoff of either player at any equilibrium remains the same (i.e. $\pm[\mathfrak{p}_\star = \mathfrak{d}^\star]$)
- Strong duality implies it does not matter which player moves first
- Set of Nash equilibria enjoys the product/interchangeable structure

Example: Saddle point as Nash equilibrium

Let $f_1 = -g$ and $f_2 = g$ and consider normalized Nash equilibrium:

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z}\in\mathbb{Z}} \; f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := g(w_1, z_2) - g(z_1, w_2).$$

Or using the formulation of Nikaidô and Isoda:

$$0 \quad = \quad \left[\min_{\mathbf{w}\in\mathbb{W}} \; \max_{\mathbf{z}\in\mathbb{W}} \; f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w})\right] \quad = \quad \left[\min_{\mathbf{w}\in\mathbb{W}} \; \max_{\mathbf{z}\in\mathbb{W}} \; g(w_1, z_2) - g(z_1, w_2)\right],$$

which is a convex problem if g is convex-concave!

- Zero-sum: two players (i.e. $n = 2$) with opposing payoff functions $f_1 + f_2 = 0$
- Saddle point is exactly Nash's equilibrium
- Payoff of either player at any equilibrium remains the same (i.e. $\pm[\mathfrak{p}_\star = \mathfrak{d}^\star]$)
- Strong duality implies it does not matter which player moves first
- Set of Nash equilibria enjoys the product/interchangeable structure

Example: Saddle point as Nash equilibrium

Let $f_1 = -g$ and $f_2 = g$ and consider normalized Nash equilibrium:

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} \ f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := g(w_1, z_2) - g(z_1, w_2).$$

Or using the formulation of Nikaidô and Isoda:

$$0 \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \ \max_{\mathbf{z} \in \mathbb{W}} \ f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w})\right] \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \ \max_{\mathbf{z} \in \mathbb{W}} \ g(w_1, z_2) - g(z_1, w_2)\right],$$

which is a convex problem if g is convex-concave!

- Zero-sum: two players (i.e. $n = 2$) with opposing payoff functions $f_1 + f_2 = 0$
- Saddle point is exactly Nash's equilibrium
- Payoff of either player at any equilibrium remains the same (i.e. $\pm[\mathfrak{p}_\star = \mathfrak{d}^\star]$)
- Strong duality implies it does not matter which player moves first
- Set of Nash equilibria enjoys the product/interchangeable structure

Example: Saddle point as Nash equilibrium

Let $f_1 = -g$ and $f_2 = g$ and consider normalized Nash equilibrium:

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} \ f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := g(w_1, z_2) - g(z_1, w_2).$$

Or using the formulation of Nikaidô and Isoda:

$$0 \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \ \max_{\mathbf{z} \in \mathbb{W}} \ f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w})\right] \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \ \max_{\mathbf{z} \in \mathbb{W}} \ g(w_1, z_2) - g(z_1, w_2)\right],$$

which is a convex problem if g is convex-concave!

- Zero-sum: two players (i.e. $n = 2$) with opposing payoff functions $f_1 + f_2 = 0$
- Saddle point is exactly Nash's equilibrium
- Payoff of either player at any equilibrium remains the same (i.e. $\pm[\mathfrak{p}_\star = \mathfrak{d}^\star]$)
- Strong duality implies it does not matter which player moves first
- Set of Nash equilibria enjoys the product/interchangeable structure

Example: Saddle point as Nash equilibrium

Let $f_1 = -g$ and $f_2 = g$ and consider normalized Nash equilibrium:

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} \; f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := g(w_1, z_2) - g(z_1, w_2).$$

Or using the formulation of Nikaidô and Isoda:

$$0 \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \; \max_{\mathbf{z} \in \mathbb{W}} \; f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w}) \right] \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \; \max_{\mathbf{z} \in \mathbb{W}} \; g(w_1, z_2) - g(z_1, w_2) \right],$$

which is a convex problem if g is convex-concave!

- Zero-sum: two players (i.e. $n = 2$) with opposing payoff functions $f_1 + f_2 = 0$
- Saddle point is exactly Nash's equilibrium
- Payoff of either player at any equilibrium remains the same (i.e. $\pm[\mathfrak{p}_\star = \mathfrak{d}^\star]$)
- Strong duality implies it does not matter which player moves first
- Set of Nash equilibria enjoys the product/interchangeable structure

Example: Saddle point as Nash equilibrium

Let $f_1 = -g$ and $f_2 = g$ and consider normalized Nash equilibrium:

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z}\in\mathbb{Z}} \ f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := g(w_1, z_2) - g(z_1, w_2).$$

Or using the formulation of Nikaidô and Isoda:

$$0 \quad = \quad \left[\min_{\mathbf{w}\in\mathbb{W}} \ \max_{\mathbf{z}\in\mathbb{W}} \ f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w})\right] \quad = \quad \left[\min_{\mathbf{w}\in\mathbb{W}} \ \max_{\mathbf{z}\in\mathbb{W}} \ g(w_1, z_2) - g(z_1, w_2)\right],$$

which is a convex problem if g is convex-concave!

- Zero-sum: two players (i.e. $n = 2$) with opposing payoff functions $f_1 + f_2 = 0$
- Saddle point is exactly Nash's equilibrium
- Payoff of either player at any equilibrium remains the same (i.e. $\pm[\mathfrak{p}_\star = \mathfrak{d}^\star]$)
- Strong duality implies it does not matter which player moves first
- Set of Nash equilibria enjoys the product/interchangeable structure

Example: Saddle point as Nash equilibrium

Let $f_1 = -g$ and $f_2 = g$ and consider normalized Nash equilibrium:

$$\mathbf{w}^\star \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} f(\mathbf{w}^\star, \mathbf{z}), \quad \text{where} \quad f(\mathbf{w}, \mathbf{z}) := g(w_1, z_2) - g(z_1, w_2).$$

Or using the formulation of Nikaidô and Isoda:

$$0 \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \max_{\mathbf{z} \in \mathbb{W}} f(\mathbf{w}, \mathbf{z}) - f(\mathbf{w}, \mathbf{w}) \right] \quad = \quad \left[\min_{\mathbf{w} \in \mathbb{W}} \max_{\mathbf{z} \in \mathbb{W}} g(w_1, z_2) - g(z_1, w_2) \right],$$

which is a convex problem if g is convex-concave!

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

$$w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

  i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

$$w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

$$w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

$$w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

$$w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

$$w_i^* \in \operatorname*{argmax}_{w_i \in \mathbb{W}_i} \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

  i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.

Definition: Equilibrium in general sum games

- General sum: $f_1 + f_2 \neq c$ for any c or with $n \geq 3$ players
- Minimax: we call $\mathbf{w}^* \in \mathbb{W}$ a minimax equilibrium if

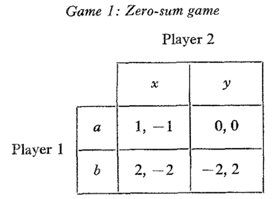
$$w_i^* \in \underset{w_i \in \mathbb{W}_i}{\operatorname{argmax}} \ \underline{f_i}(w_i), \quad \underline{f_i}(w_i) = \min_{\{w_j \in \mathbb{W}_j\}_{j \neq i}} f_i(w_1, \ldots, w_{i-1}, w_i, w_{i+1}, \ldots, w_n),$$

  - n two-player games: each player i plays against all other players
  - minimax equilibrium always exists (under mild conditions)
  - different from (normalized) Nash equilibrium, even for zero-sum two-player games, coincides with Nash equilibrium only when the latter exists
- Pareto: $\mathbf{w}^* \in \mathbb{W}$ a Pareto equilibrium if for any $\mathbf{w} \in \mathbb{W}$,

$$\mathbf{f}(\mathbf{w}) \leq \mathbf{f}(\mathbf{w}^*) \implies \mathbf{f}(\mathbf{w}) = \mathbf{f}(\mathbf{w}^*),$$

i.e., it is not possible to strictly improve any player's payoff without degrading some other player's.

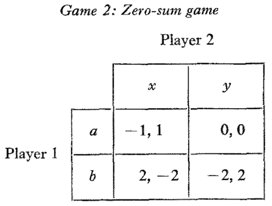
A. W. Starr and Y.-C. Ho. "Nonzero-sum differential games". *Journal of optimization theory and applications*, vol. 3, no. 3 (1969), pp. 184–206.



Game 1: Zero-sum game

Player 2

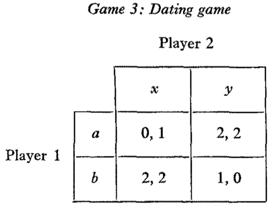
		x	y
Player 1	a	1, −1	0, 0
	b	2, −2	−2, 2



Game 2: Zero-sum game

Player 2

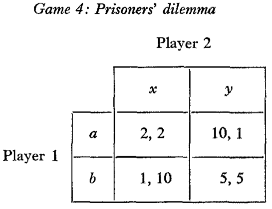
		x	y
Player 1	a	−1, 1	0, 0
	b	2, −2	−2, 2



Game 3: Dating game

Player 2

		x	y
Player 1	a	0, 1	2, 2
	b	2, 2	1, 0



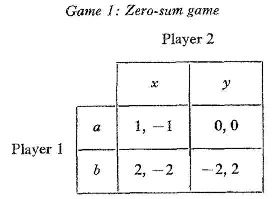
Game 4: Prisoners' dilemma

Player 2

		x	y
Player 1	a	2, 2	10, 1
	b	1, 10	5, 5

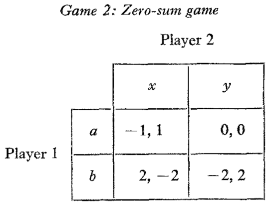
Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to "force" player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a "forces" player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!



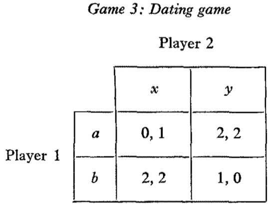
Game 1: Zero-sum game

		Player 2	
		x	y
Player 1	a	1, −1	0, 0
	b	2, −2	−2, 2



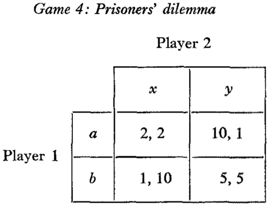
Game 2: Zero-sum game

		Player 2	
		x	y
Player 1	a	−1, 1	0, 0
	b	2, −2	−2, 2



Game 3: Dating game

		Player 2	
		x	y
Player 1	a	0, 1	2, 2
	b	2, 2	1, 0

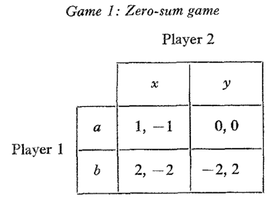


Game 4: Prisoners' dilemma

		Player 2	
		x	y
Player 1	a	2, 2	10, 1
	b	1, 10	5, 5

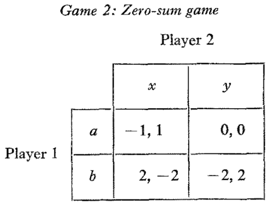
Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to "force" player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a "forces" player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!



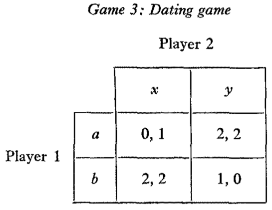
Game 1: Zero-sum game

		Player 2	
		x	y
Player 1	a	1, −1	0, 0
	b	2, −2	−2, 2



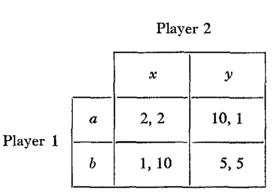
Game 2: Zero-sum game

		Player 2	
		x	y
Player 1	a	−1, 1	0, 0
	b	2, −2	−2, 2



Game 3: Dating game

		Player 2	
		x	y
Player 1	a	0, 1	2, 2
	b	2, 2	1, 0

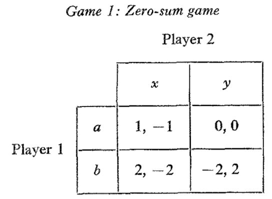


Game 4: Prisoners' dilemma

		Player 2	
		x	y
Player 1	a	2, 2	10, 1
	b	1, 10	5, 5

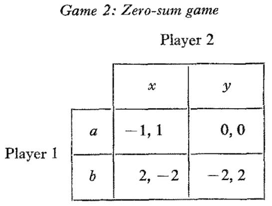
Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to "force" player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a "forces" player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!



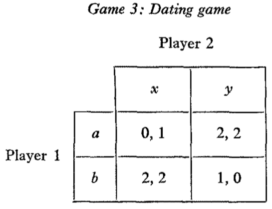
Game 1: Zero-sum game

		Player 2	
		x	*y*
Player 1	*a*	1, −1	0, 0
	b	2, −2	−2, 2



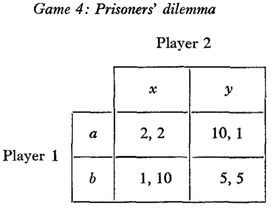
Game 2: Zero-sum game

		Player 2	
		x	*y*
Player 1	*a*	−1, 1	0, 0
	b	2, −2	−2, 2



Game 3: Dating game

		Player 2	
		x	*y*
Player 1	*a*	0, 1	2, 2
	b	2, 2	1, 0

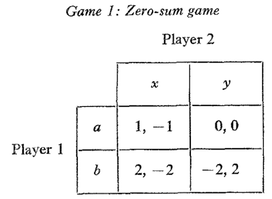


Game 4: Prisoners' dilemma

		Player 2	
		x	*y*
Player 1	*a*	2, 2	10, 1
	b	1, 10	5, 5

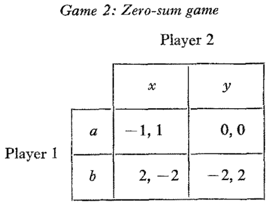
Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to "force" player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a "forces" player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!



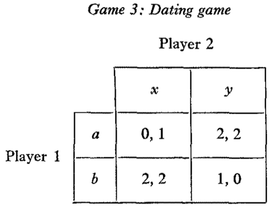
Game 1: Zero-sum game

		Player 2	
		x	y
Player 1	a	1, −1	0, 0
	b	2, −2	−2, 2



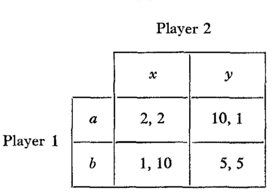
Game 2: Zero-sum game

		Player 2	
		x	y
Player 1	a	−1, 1	0, 0
	b	2, −2	−2, 2



Game 3: Dating game

		Player 2	
		x	y
Player 1	a	0, 1	2, 2
	b	2, 2	1, 0

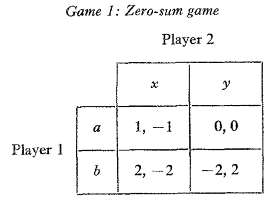


Game 4: Prisoners' dilemma

		Player 2	
		x	y
Player 1	a	2, 2	10, 1
	b	1, 10	5, 5

Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to "force" player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a "forces" player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!

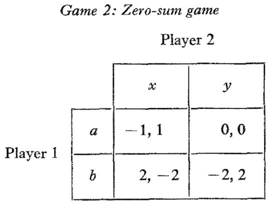


Game 1: Zero-sum game

Player 2

Player 1

	x	y
a	1, −1	0, 0
b	2, −2	−2, 2

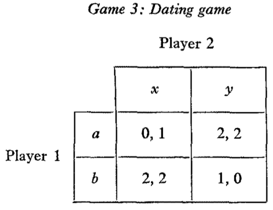


Game 2: Zero-sum game

Player 2

Player 1

	x	y
a	−1, 1	0, 0
b	2, −2	−2, 2

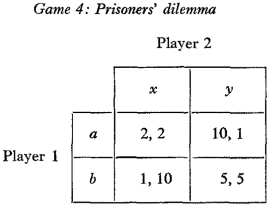


Game 3: Dating game

Player 2

Player 1

	x	y
a	0, 1	2, 2
b	2, 2	1, 0



Game 4: Prisoners' dilemma

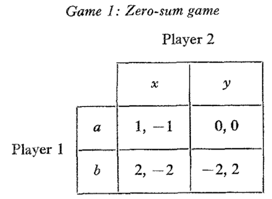
Player 2

Player 1

	x	y
a	2, 2	10, 1
b	1, 10	5, 5

Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to “force” player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a “forces” player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!

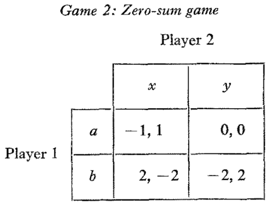


Game 1: Zero-sum game

Player 2

Player 1

	x	y
a	1, −1	0, 0
b	2, −2	−2, 2

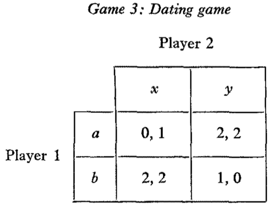


Game 2: Zero-sum game

Player 2

Player 1

	x	y
a	−1, 1	0, 0
b	2, −2	−2, 2

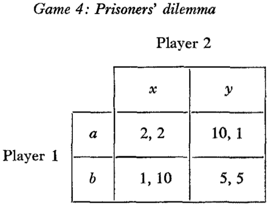


Game 3: Dating game

Player 2

Player 1

	x	y
a	0, 1	2, 2
b	2, 2	1, 0



Game 4: Prisoners' dilemma

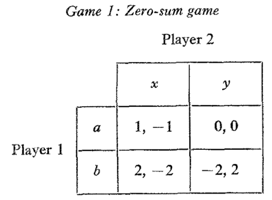
Player 2

Player 1

	x	y
a	2, 2	10, 1
b	1, 10	5, 5

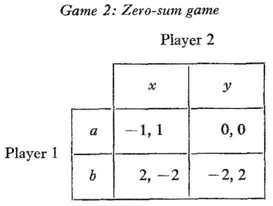
Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to “force” player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a “forces” player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!



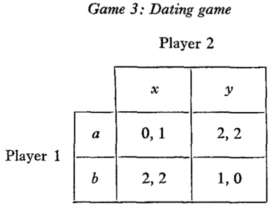
Game 1: Zero-sum game

		Player 2	
		x	y
Player 1	a	1, −1	0, 0
	b	2, −2	−2, 2



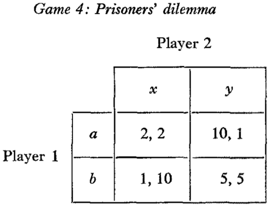
Game 2: Zero-sum game

		Player 2	
		x	y
Player 1	a	−1, 1	0, 0
	b	2, −2	−2, 2



Game 3: Dating game

		Player 2	
		x	y
Player 1	a	0, 1	2, 2
	b	2, 2	1, 0

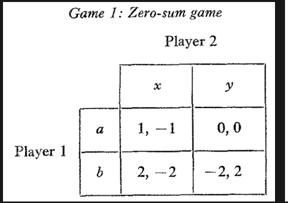


Game 4: Prisoners' dilemma

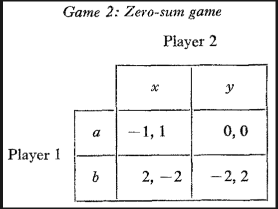
		Player 2	
		x	y
Player 1	a	2, 2	10, 1
	b	1, 10	5, 5

Example: Two-player general sum

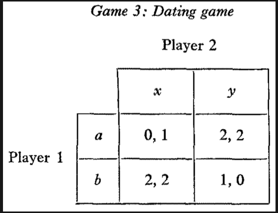
- Game 1: (a,x) is a NE; player 2 chooses x to “force” player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a “forces” player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!


Game 1: Zero-sum game

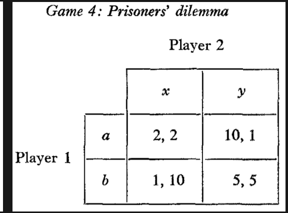
		Player 2	
		x	y
Player 1	a	1, −1	0, 0
	b	2, −2	−2, 2


Game 2: Zero-sum game

		Player 2	
		x	y
Player 1	a	−1, 1	0, 0
	b	2, −2	−2, 2


Game 3: Dating game

		Player 2	
		x	y
Player 1	a	0, 1	2, 2
	b	2, 2	1, 0


Game 4: Prisoners' dilemma

		Player 2	
		x	y
Player 1	a	2, 2	10, 1
	b	1, 10	5, 5

Example: Two-player general sum

- Game 1: (a,x) is a NE; player 2 chooses x to “force” player 1 to choose a
- Game 2: no NE but (a,x) is still a minimax equilibrium
- Game 3: NE (a,x) and (b,y), with different costs due to non-zero sum
  - player 1 committing to a “forces” player 2 to play x
  - in general sum, whoever moves first may gain a significant advantage!
- Game 4: unique NE (b,y) and yet (a,x) gives lower costs to both players!
  - (a,x) (also (a,y) and (b,x)) is a Pareto equilibrium
  - NE (b,y) is not!

Definition: Stackelberg equilibrium

Another interesting notion of equilibrium of two players, due to Stackelberg:

$$\mathbf{w}_* \in \underset{\mathbf{w}\in\mathbb{W}}{\operatorname{argmax}} f_1(\mathbf{w}, \mathbf{z}(\mathbf{w})), \quad \mathbf{z}(\mathbf{w}) \in \underset{\mathbf{z}\in\mathbb{Z}}{\operatorname{argmax}}\ f_2(\mathbf{w}, \mathbf{z}), \quad \mathbf{z}^* = \mathbf{z}(\mathbf{w}_*).$$

- Player $\mathbf{w}$ is the (big market) leader who acts first
- Player $\mathbf{z}$ is the follower (e.g. small competitor) who responds
- By acting first the leader has some advantage while the follower could threaten the leader to make trouble for both players!

H. von Stackelberg. "Market structure and equilibrium". Springer, 1934.

Definition: Stackelberg equilibrium

Another interesting notion of equilibrium of two players, due to Stackelberg:

$$\mathbf{w}_* \in \operatorname*{argmax}_{\mathbf{w}\in\mathbb{W}} f_1(\mathbf{w}, \mathbf{z}(\mathbf{w})), \quad \mathbf{z}(\mathbf{w}) \in \operatorname*{argmax}_{\mathbf{z}\in\mathbb{Z}} \ f_2(\mathbf{w}, \mathbf{z}), \quad \mathbf{z}^* = \mathbf{z}(\mathbf{w}_*).$$

- Player $\mathbf{w}$ is the (big market) leader who acts first
- Player $\mathbf{z}$ is the follower (e.g. small competitor) who responds
- By acting first the leader has some advantage while the follower could threaten the leader to make trouble for both players!

H. von Stackelberg. "Market structure and equilibrium". Springer, 1934.

Definition: Stackelberg equilibrium

Another interesting notion of equilibrium of two players, due to Stackelberg:

$$\mathbf{w}_* \in \operatorname*{argmax}_{\mathbf{w}\in\mathbb{W}} f_1(\mathbf{w}, \mathbf{z}(\mathbf{w})), \quad \mathbf{z}(\mathbf{w}) \in \operatorname*{argmax}_{\mathbf{z}\in\mathbb{Z}} \; f_2(\mathbf{w}, \mathbf{z}), \quad \mathbf{z}^* = \mathbf{z}(\mathbf{w}_*).$$

- Player $\mathbf{w}$ is the (big market) leader who acts first
- Player $\mathbf{z}$ is the follower (e.g. small competitor) who responds
- By acting first the leader has some advantage while the follower could threaten the leader to make trouble for both players!

H. von Stackelberg. "Market structure and equilibrium". Springer, 1934.

Definition: Stackelberg equilibrium

Another interesting notion of equilibrium of two players, due to Stackelberg:

$$\mathbf{w}_* \in \operatorname*{argmax}_{\mathbf{w} \in \mathbb{W}} f_1(\mathbf{w}, \mathbf{z}(\mathbf{w})), \quad \mathbf{z}(\mathbf{w}) \in \operatorname*{argmax}_{\mathbf{z} \in \mathbb{Z}} \ f_2(\mathbf{w}, \mathbf{z}), \quad \mathbf{z}^* = \mathbf{z}(\mathbf{w}_*).$$

- Player $\mathbf{w}$ is the (big market) leader who acts first
- Player $\mathbf{z}$ is the follower (e.g. small competitor) who responds
- By acting first the leader has some advantage while the follower could threaten the leader to make trouble for both players!

H. von Stackelberg. "Market structure and equilibrium". Springer, 1934.

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
    - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
    - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader “rips” off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader “rips” off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
    - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
    - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Example: Market price

- Leader and follower produce same product with quantity q_1 and q_2 at no cost
- The payoff for each player is $f_i(q_1, q_2) = q_i(4 - q_1 - q_2)_+, \quad i = 1, 2,$ where $p := (4 - q_1 - q_2)_+$ is say the market price for the product
- Given q_1, the optimal choice for the follower is $q_2 = \frac{4-q_1}{2}$, which in turn yields the optimal choice for the leader $q_1^* = 2$ hence $q_2^* = 1$
- By merely acting first the leader gets payoff 2 while the follower gets payoff 1
- Had the two players acted simultaneously, the Nash equilibrium is easily seen to be $q_1 = q_2 = \frac{4}{3}$ with payoff $\frac{16}{9}$ for both players
  - leader "rips" off follower!
- However, the follower can threaten the leader by intentionally deviating from its optimal response, which will hurt both players!
  - setting $q_2 = 4$ leads to 0 payoff for both players, which is clearly irrational but perhaps not uncommon in reality...

Data Poisoning

- Follower F (i.e., the defender) aims at minimizing $f = \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$:

$$\mathbf{w}_* = \mathbf{w}_*(\mathcal{D}_p) \in \underset{\mathbf{w}}{\operatorname{argmin}} \; \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$$

- Leader L (i.e., the attacker) aims at maximizing a different loss function $\ell = \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$ on the validation set $\mathcal{D}_v$:

$$\mathcal{D}_{p_*} \in \underset{\mathcal{D}_p}{\operatorname{argmax}} \, \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$$

- Stackelberg formulation (a.k.a. bilevel optimization):

$$\max_{\mathcal{D}_p} \; \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*), \;\; \text{s.t.} \;\; \mathbf{w}_* \in \underset{\mathbf{w}}{\operatorname{argmin}} \, \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w}).$$

Data Poisoning

- Follower F (i.e., the defender) aims at minimizing $f = \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$:

$$\mathbf{w}_* = \mathbf{w}_*(\mathcal{D}_p) \in \underset{\mathbf{w}}{\operatorname{argmin}} \; \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$$

- Leader L (i.e., the attacker) aims at maximizing a different loss function $\ell = \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$ on the validation set $\mathcal{D}_v$:

$$\mathcal{D}_{p*} \in \underset{\mathcal{D}_p}{\operatorname{argmax}} \, \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$$

- Stackelberg formulation (a.k.a. bilevel optimization):

$$\max_{\mathcal{D}_p} \; \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*), \;\; \text{s.t.} \;\; \mathbf{w}_* \in \underset{\mathbf{w}}{\operatorname{argmin}} \, \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w}).$$

Data Poisoning

- Follower F (i.e., the defender) aims at minimizing $f = \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$:

$$\mathbf{w}_* = \mathbf{w}_*(\mathcal{D}_p) \in \underset{\mathbf{w}}{\operatorname{argmin}} \; \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$$

- Leader L (i.e., the attacker) aims at maximizing a different loss function $\ell = \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$ on the validation set $\mathcal{D}_v$:

$$\mathcal{D}_{p*} \in \underset{\mathcal{D}_p}{\operatorname{argmax}} \, \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$$

- Stackelberg formulation (a.k.a. bilevel optimization):

$$\max_{\mathcal{D}_p} \; \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*), \; \text{ s.t. } \; \mathbf{w}_* \in \underset{\mathbf{w}}{\operatorname{argmin}} \, \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w}).$$

Data Poisoning

- Follower F (i.e., the defender) aims at minimizing $f = \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$:

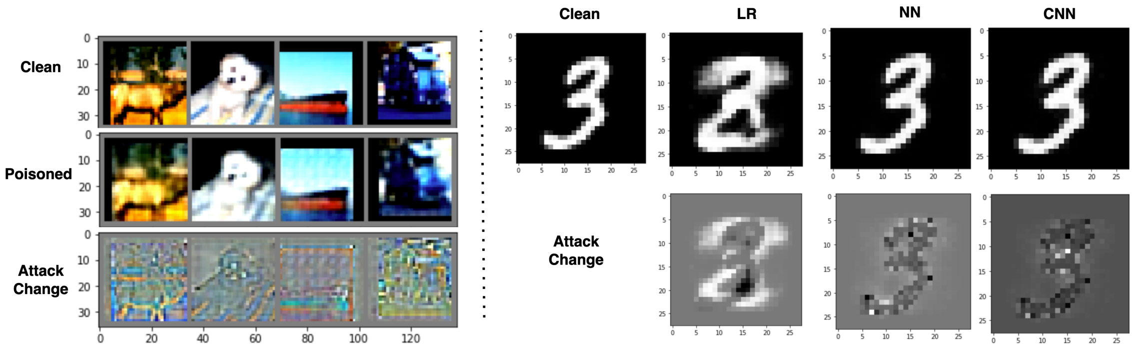
$$\mathbf{w}_* = \mathbf{w}_*(\mathcal{D}_p) \in \underset{\mathbf{w}}{\operatorname{argmin}} \; \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w})$$

- Leader L (i.e., the attacker) aims at maximizing a different loss function $\ell = \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$ on the validation set $\mathcal{D}_v$:

$$\mathcal{D}_{p*} \in \underset{\mathcal{D}_p}{\operatorname{argmax}} \, \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*)$$

- Stackelberg formulation (a.k.a. bilevel optimization):

$$\max_{\mathcal{D}_p} \; \mathcal{L}(\mathcal{D}_v, \mathbf{w}_*), \;\; \text{s.t.} \;\; \mathbf{w}_* \in \underset{\mathbf{w}}{\operatorname{argmin}} \, \mathcal{L}(\mathcal{D}_{tr} \cup \mathcal{D}_p, \mathbf{w}).$$


Clean
Poisoned
Attack
Change
0
10
20
30
0
20
40
60
80
100
120
Clean
LR
NN
CNN
0
5
10
15
20
25
Attack
Change