The duality between maximum separation and minimum distance

David Bremner[1]

UNB

June 19, 2008

[1] Joint work with Peter Gritzmann and Thomas Burger

Outline

Outline

Margin

Definition

Given $P, Q \subset \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ *classifies* P and Q with *margin* γ if

$$\gamma = \inf_{p \in P} f(p) - \sup_{q \in Q} f(q) = \inf_{p \in P, q \in Q} f(p) - f(q)$$

Margin

Definition

Given $P, Q \subset \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ *classifies* P and Q with *margin* γ if

$$\gamma = \inf_{p \in P} f(p) - \sup_{q \in Q} f(q) = \inf_{p \in P, q \in Q} f(p) - f(q)$$

Roughly speaking

We want to maximize γ/γ^*, where γ^* is a *best possible margin* for some family $\mathcal{F} \ni f$

Margin

Definition

Given $P, Q \subset \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ *classifies* P and Q with *margin* γ if

$$\gamma = \inf_{p \in P} f(p) - \sup_{q \in Q} f(q) = \inf_{p \in P, q \in Q} f(p) - f(q)$$

Toy example

- $P, Q \subset [0, 1]$
- $\mathcal{F} = \{x \to x, x \to -x\}$
- $\gamma^* = 1$

Margin

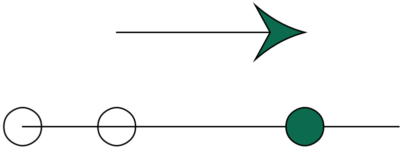
Definition

Given $P, Q \subset \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ *classifies* P and Q with *margin* γ if

$$\gamma = \inf_{p \in P} f(p) - \sup_{q \in Q} f(q) = \inf_{p \in P, q \in Q} f(p) - f(q)$$

Toy example

- $P, Q \subset [0,1]$
- $\mathcal{F} = \{ x \to x, x \to -x \}$
- $\gamma^* = 1$



$\gamma = 1/2$

Margin

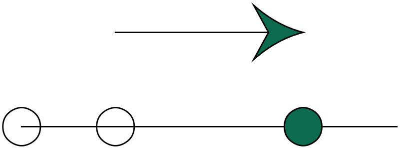
Definition

Given $P, Q \subset \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ *classifies* P and Q with *margin* γ if

$$\gamma = \inf_{p \in P} f(p) - \sup_{q \in Q} f(q) = \inf_{p \in P, q \in Q} f(p) - f(q)$$

Toy example

- $P, Q \subset [0, 1]$
- $\mathcal{F} = \{ x \to x, x \to -x \}$
- $\gamma^* = 1$



$\gamma = 1/2$

Margin

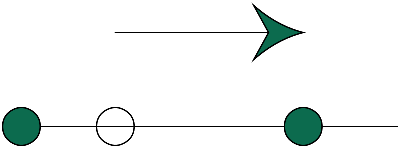
Definition

Given $P, Q \subset \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ *classifies* P and Q with *margin* γ if

$$\gamma = \inf_{p \in P} f(p) - \sup_{q \in Q} f(q) = \inf_{p \in P, q \in Q} f(p) - f(q)$$

Toy example

- $P, Q \subset [0, 1]$
- $\mathcal{F} = \{ x \to x, x \to -x \}$
- $\gamma^* = 1$



$\gamma = -1/4$

Linear Classifiers and Support Functionals

Linear Classifiers

$$f(x) = \langle z, x \rangle := \sum_{i=1}^{n} z_i x_i$$

Support Functionals

For $K \subset \mathbb{R}^d, w \in (\mathbb{R}^d)^*$,

$$\sup(K; w) := \sup_{k \in K} \langle w, k \rangle \qquad \text{(upper)}$$

$$\inf(K; w) = \inf_{k \in K} \langle w, k \rangle \qquad \text{(lower)}$$

$$= -\sup(K; -w)$$

Linear Classifiers and Support Functionals

Linear Classifiers

$$f(x) = \langle z, x \rangle := \sum_{i=1}^{n} z_i x_i$$

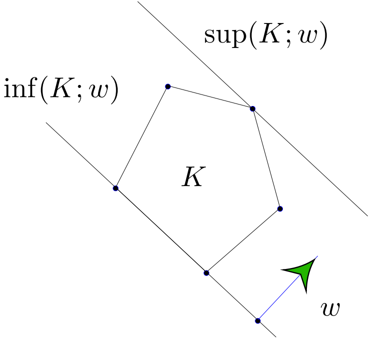
Support Functionals

For $K \subset \mathbb{R}^d, w \in (\mathbb{R}^d)^*$,

$$\sup(K; w) := \sup_{k \in K} \langle w, k \rangle \qquad \text{(upper)}$$

$$\inf(K; w) = \inf_{k \in K} \langle w, k \rangle \qquad \text{(lower)}$$

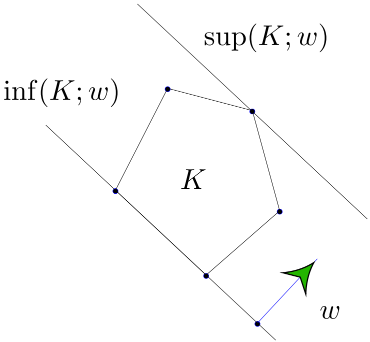
$$= -\sup(K; -w)$$



Linear Classifiers and Support Functionals

Linear Classifiers

$$f(x) = \langle z, x \rangle := \sum_{i=1}^{n} z_i x_i$$



Support Functionals

For $K \subset \mathbb{R}^d, w \in (\mathbb{R}^d)^*$,

$$\sup(K; w) := \sup_{k \in K} \langle w, k \rangle \qquad \text{(upper)}$$

$$\inf(K; w) = \inf_{k \in K} \langle w, k \rangle \qquad \text{(lower)}$$

$$= -\sup(K; -w)$$

W.l.o.g, K is convex

$$\sup(X; w) = \sup(\text{conv}(X); w)$$

Linear Margin

Margin optimization

$$\max \inf(P; z) - \sup(Q; z) \quad \text{such that}$$
$$z \in (\mathbb{R}^d)^*$$

Linear Margin

Margin optimization

$$\max \inf(P; z) - \sup(Q; z) \quad \text{such that}$$
$$z \in (\mathbb{R}^d)^*$$

If $\inf(P; z) - \sup(Q; z) = \gamma$ then for $\tau \geq 0$,

$$\inf(P; \tau z) - \sup(Q; \tau z) = \tau\gamma$$

Linear Margin

Margin optimization

$$\max \inf(P; z) - \sup(Q; z) \quad \text{such that}$$
$$z \in \quad \mathbb{B}^*$$

If $\inf(P; z) - \sup(Q; z) = \gamma$ then for $\tau \geq 0$,

$$\inf(P; \tau z) - \sup(Q; \tau z) = \tau\gamma$$

Scaling up

So that margin(P, Q) is not always $+\infty$, $z \in \mathbb{B}^*$, $\mathbb{B}^*$ bounded.

Linear Margin

Margin optimization

$$\max \inf(P; z) - \sup(Q; z) \quad \text{such that}$$
$$z \in \quad \mathbb{B}^* \setminus L$$

If $\inf(P; z) - \sup(Q; z) = \gamma$ then for $\tau \geq 0$,

$$\inf(P; \tau z) - \sup(Q; \tau z) = \tau\gamma$$

Scaling down

For some region $\vec{0} \in L \subset \mathbb{B}^*$, $z \notin L$

Linear Margin

Margin optimization

$$\max \inf(P; z) - \sup(Q; z) \quad \text{such that}$$
$$z \in \quad \mathbb{B}^* \setminus L$$

Without much loss of generality

- W.l.o.g. $\mathbb{B}^*$ is a *convex body*, i.e. convex, has interior.
- It seems natural then to choose $L = \operatorname{int} \mathbb{B}^*$.

Linear Margin

Margin optimization

$$\max \inf(P; z) - \sup(Q; z) \quad \text{such that}$$
$$z \in \operatorname{bd} \mathbb{B}^*$$

Without much loss of generality

- W.l.o.g. $\mathbb{B}^*$ is a *convex body*, i.e. convex, has interior.
- It seems natural then to choose $L = \operatorname{int} \mathbb{B}^*$.

Outline

Minkowski Norms

- $X \subset \mathbb{R}^d$ is *centrally symmetric* if $x \in X \Rightarrow -x \in X$.
- For any centrally symmetric convex body $\mathbb{B}$, the *Minkowski Norm*

$$\|x\|_{\mathbb{B}} := \inf_{\lambda \geq 0} \{ \lambda \geq 0 \mid x \in \lambda \mathbb{B} \}$$

- For any $K \subset \mathbb{R}^d$, the *polar* $K^* = \{ y \in \mathbb{R}^n \mid \sup(K; y) \leq 1 \}$.

Minkowski Norms

- $X \subset \mathbb{R}^d$ is *centrally symmetric* if $x \in X \Rightarrow -x \in X$.
- For any centrally symmetric convex body $\mathbb{B}$, the *Minkowski Norm*

$$\|x\|_{\mathbb{B}} := \inf_{\lambda \geq 0} \{ \lambda \geq 0 \mid x \in \lambda\mathbb{B} \}$$

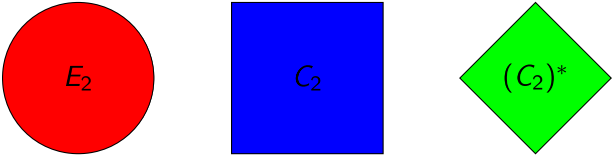
- For any $K \subset \mathbb{R}^d$, the *polar* $K^* = \{ y \in \mathbb{R}^n \mid \sup(K; y) \leq 1 \}$.

Minkowski Norms

- $X \subset \mathbb{R}^d$ is *centrally symmetric* if $x \in X \Rightarrow -x \in X$.
- For any centrally symmetric convex body $\mathbb{B}$, the *Minkowski Norm*

$$\|x\|_{\mathbb{B}} := \inf_{\lambda \geq 0} \{\, \lambda \geq 0 \mid x \in \lambda \mathbb{B} \,\}$$

- For any $K \subset \mathbb{R}^d$, the *polar* $K^* = \{\, y \in \mathbb{R}^n \mid \sup(K; y) \leq 1 \,\}$.



$\|\cdot\|_2 = \sqrt{x_1^2 + x_2^2}$ $\quad$ $\|\cdot\|_\infty = \max(|x_1|, |x_2|)$ $\quad$ $\|\cdot\|_1 = |x_1| + |x_2|$

Minkowski Norms

- $X \subset \mathbb{R}^d$ is *centrally symmetric* if $x \in X \Rightarrow -x \in X$.
- For any centrally symmetric convex body $\mathbb{B}$, the *Minkowski Norm*

$$\|x\|_{\mathbb{B}} := \inf_{\lambda \geq 0} \{ \lambda \geq 0 \mid x \in \lambda \mathbb{B} \}$$

- For any $K \subset \mathbb{R}^d$, the *polar* $K^* = \{ y \in \mathbb{R}^n \mid \sup(K; y) \leq 1 \}$.

$$\begin{aligned} \mathbb{S} &= \{ x \in \mathbb{B} : \|x\|_{\mathbb{B}} = 1 \} && = \operatorname{bd} \mathbb{B} \\ \mathbb{S}^* &= \{ z \in \mathbb{B}^* : \|z\|_{\mathbb{B}^*} = 1 \} && = \operatorname{bd} \mathbb{B}^* \end{aligned}$$

Minkowski Norms

- $X \subset \mathbb{R}^d$ is *centrally symmetric* if $x \in X \Rightarrow -x \in X$.
- For any centrally symmetric convex body $\mathbb{B}$, the *Minkowski Norm*

$$\|x\|_{\mathbb{B}} := \inf_{\lambda \geq 0} \{\, \lambda \geq 0 \mid x \in \lambda \mathbb{B} \,\}$$

- For any $K \subset \mathbb{R}^d$, the *polar* $K^* = \{\, y \in \mathbb{R}^n \mid \sup(K; y) \leq 1 \,\}$.

$$\begin{aligned} \mathbb{S} &= \{\, x \in \mathbb{B} : \|x\|_{\mathbb{B}} = 1 \,\} && = \operatorname{bd} \mathbb{B} \\ \mathbb{S}^* &= \{\, z \in \mathbb{B}^* : \|z\|_{\mathbb{B}^*} = 1 \,\} && = \operatorname{bd} \mathbb{B}^* \end{aligned}$$

- Choose $z \in \operatorname{bd} \mathbb{B}^*$

$$\begin{aligned} \max \inf(P; z) - \sup(Q; z) \\ z \in \mathbb{S}^* \end{aligned}$$

Norm duality

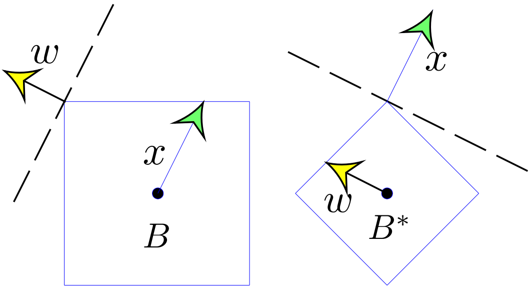
Polarity of unit balls connects norms:

$$\|x\|_{\mathbb{B}} = \sup(\mathbb{B}^*; x) \tag{ND}$$

Norm duality

Polarity of unit balls connects norms:

$$\|x\|_{\mathbb{B}} = \sup(\mathbb{B}^*; x) \tag{ND}$$



Norms define metrics

- Define

$$\mathrm{dist}_{\mathbb{B}}(x, y) := \|x - y\|_{\mathbb{B}}$$
$$\mathrm{dist}_{\mathbb{B}}(X, Y) := \inf_{x \in X, y \in Y} \|x - y\|_{\mathbb{B}}$$
$$H(z, \varphi) = \{\, x \in \mathbb{R}^n \mid \langle\, z, x \,\rangle = \varphi \,\}$$

- Where $\mathbb{B}$ is clear from context, we write $\|\cdot\|$, and $\mathrm{dist}(\cdot, \cdot)$.
- From (ND), we have for all $z \in \mathbb{S}^*$,

$$\mathrm{dist}(H(z, \mu_1), H(z, \mu_2)) = |\mu_1 - \mu_2| \tag{1}$$

Norms define metrics

- Define

$$\begin{aligned} \mathrm{dist}_{\mathbb{B}}(x, y) &:= \|x - y\|_{\mathbb{B}} \\ \mathrm{dist}_{\mathbb{B}}(X, Y) &:= \inf_{x \in X, y \in Y} \|x - y\|_{\mathbb{B}} \\ H(z, \varphi) &= \{\, x \in \mathbb{R}^n \mid \langle\, z, x \,\rangle = \varphi \,\} \end{aligned}$$

- Where $\mathbb{B}$ is clear from context, we write $\| \cdot \|$, and $\mathrm{dist}(\cdot, \cdot)$.
- From (ND), we have for all $z \in \mathbb{S}^*$,

$$\mathrm{dist}(H(z, \mu_1), H(z, \mu_2)) = |\mu_1 - \mu_2| \tag{1}$$

Norms define metrics

- Define

$$\begin{aligned} \operatorname{dist}_{\mathbb{B}}(x, y) &:= \|x - y\|_{\mathbb{B}} \\ \operatorname{dist}_{\mathbb{B}}(X, Y) &:= \inf_{x \in X, y \in Y} \|x - y\|_{\mathbb{B}} \\ H(z, \varphi) &= \{\, x \in \mathbb{R}^n \mid \langle\, z, x \,\rangle = \varphi \,\} \end{aligned}$$

- Where $\mathbb{B}$ is clear from context, we write $\|\cdot\|$, and $\operatorname{dist}(\cdot, \cdot)$.
- From (ND), we have for all $z \in \mathbb{S}^*$,

$$\operatorname{dist}(H(z, \mu_1), H(z, \mu_2)) = |\mu_1 - \mu_2| \tag{1}$$

Separability

Separable

$P, Q \subset \mathbb{R}^d$ are *separable* if $\exists z \neq \vec{0}$ such that $\sup(Q; z) \leq \inf(P; z)$.

Theorem (The Separating Hyperplane Theorem)

Given convex sets P and Q with $\operatorname{aff}(P \cup Q) = \mathbb{R}^d$, $\operatorname{relint} S \cap \operatorname{relint} T \neq \emptyset$ *if and only if P and Q are not separable.*

Separability

Separable

$P, Q \subset \mathbb{R}^d$ are *separable* if $\exists z \neq \vec{0}$ such that $\sup(Q; z) \leq \inf(P; z)$.

Theorem (The Separating Hyperplane Theorem)

Given convex sets P and Q with $\operatorname{aff}(P \cup Q) = \mathbb{R}^d$, $\operatorname{relint} S \cap \operatorname{relint} T \neq \emptyset$ if and only if P and Q are not separable.

Separability

Separable

$P, Q \subset \mathbb{R}^d$ are *separable* if $\exists z \neq \vec{0}$ such that $\sup(Q; z) \leq \inf(P; z)$.

Theorem (The Separating Hyperplane Theorem)

Given convex sets P and Q with $\operatorname{aff}(P \cup Q) = \mathbb{R}^d$, $\operatorname{relint} S \cap \operatorname{relint} T \neq \emptyset$ if and only if P and Q are not separable.

$$\operatorname{sep}(P, Q) := \begin{cases} 1 & P \text{ and } Q \text{ separable} \\ -1 & \text{otherwise} \end{cases}$$

Separability

Theorem (The Separating Hyperplane Theorem)

Given convex sets P and Q with $\operatorname{aff}(P \cup Q) = \mathbb{R}^d$, $\operatorname{relint} S \cap \operatorname{relint} T \neq \emptyset$ *if and only if P and Q are not separable.*

$$\operatorname{sep}(P, Q) := \begin{cases} 1 & P \text{ and } Q \text{ separable} \\ -1 & \text{otherwise} \end{cases}$$

Corollary

For P, Q convex polyhedra, $\operatorname{sep}(P, Q)$ *is computable in polynomial time.*

Margin as distance between parallel hyperplanes

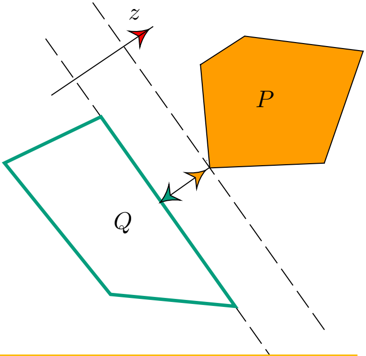
Dual Unit Ball Formulation

$$\max \inf(P; z) - \sup(Q; z)$$
$$z \in \mathbb{S}^*$$

Parallel Hyperplane Distance

$$\max \operatorname{sep}(P, Q) \cdot \operatorname{dist}(H(z, \inf(P; z)), H(z, \sup(Q; z)))$$
$$z \neq \vec{0}$$

Margin as distance between parallel hyperplanes



Dual Unit Ball Formulation

$$\max \inf(P; z) - \sup(Q; z)$$
$$z \in \mathbb{S}^*$$

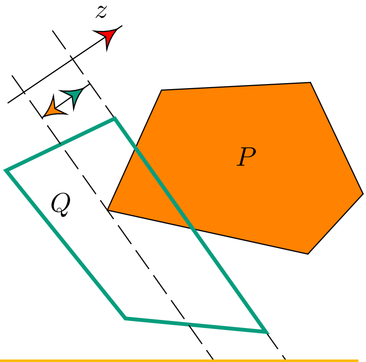
Parallel Hyperplane Distance

$$\max \operatorname{sep}(P, Q) \cdot \operatorname{dist}(H(z, \inf(P; z)), H(z, \sup(Q; z)))$$
$$z \neq \vec{0}$$

Margin as distance between parallel hyperplanes

Dual Unit Ball Formulation

$$\max \inf(P; z) - \sup(Q; z)$$
$$z \in \mathbb{S}^*$$



Parallel Hyperplane Distance

$$\max \operatorname{sep}(P, Q) \cdot \operatorname{dist}(H(z, \inf(P; z)), H(z, \sup(Q; z)))$$
$$z \neq \vec{0}$$

Outline

Minkowski Sums

The *Minkowski sum* and *Minkowski difference* of $P, Q \subset \mathbb{R}^n$ is defined as

$$P + Q = \{\, x + y \mid x \in P \text{ and } y \in Q \,\}$$
$$P - Q = P + (-Q) = \{\, x - y \mid x \in P \text{ and } y \in Q \,\}$$

Support is Minkowski-additive

$$\sup(P + Q; w) = \sup(P; w) + \sup(Q; w) \tag{2}$$

Define $\operatorname{margin}(P, Q)$ as the optimal value of

$$\max \inf(P - Q; w) \quad w \in \mathbb{S}^* \tag{MARGIN}$$

Minkowski Sums

The *Minkowski sum* and *Minkowski difference* of $P, Q \subset \mathbb{R}^n$ is defined as

$$P + Q = \{\, x + y \mid x \in P \text{ and } y \in Q \,\}$$
$$P - Q = P + (-Q) = \{\, x - y \mid x \in P \text{ and } y \in Q \,\}$$

Support is Minkowski-additive

$$\sup(P + Q; w) = \sup(P; w) + \sup(Q; w) \tag{2}$$

Define margin(P, Q) as the optimal value of

$$\max \inf(P - Q; w)$$
$$w \in \mathbb{S}^* \tag{MARGIN}$$

Minkowski Sums

The *Minkowski sum* and *Minkowski difference* of $P, Q \subset \mathbb{R}^n$ is defined as

$$P + Q = \{\, x + y \mid x \in P \text{ and } y \in Q \,\}$$
$$P - Q = P + (-Q) = \{\, x - y \mid x \in P \text{ and } y \in Q \,\}$$

Support is Minkowski-additive

$$\sup(P + Q; w) = \sup(P; w) + \sup(Q; w) \tag{2}$$

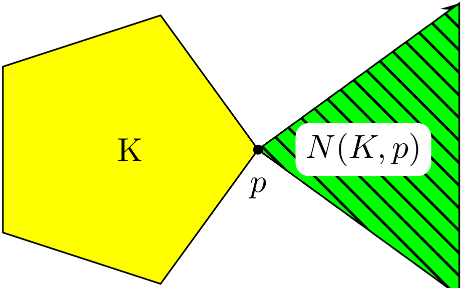
Define $\text{margin}(P, Q)$ as the optimal value of

$$\begin{aligned} \max \inf(P - Q; w) \\ w \in \mathbb{S}^* \end{aligned} \qquad \text{(MARGIN)}$$

Outer Normal Cones

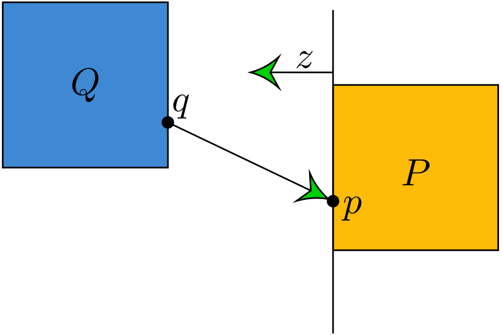
For a convex set $K \subset \mathbb{R}^n$ and a point $x \in K$ the *outer normal cone* of K at x is defined as

$$N(K, x) := \{ z \in (\mathbb{R}^d)^* \mid \langle z, x \rangle = \sup(K; z) \}.$$



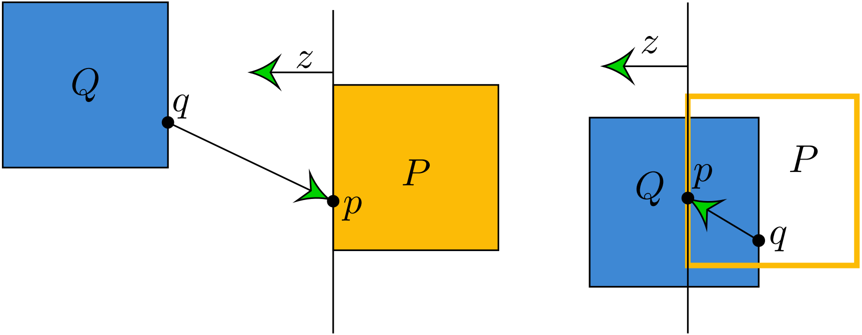
Margin by translation

$$\begin{aligned} \min \quad & \|p - q\| \\ \sup(P; z) = & \inf(Q + (p - q); z) \\ z \neq \vec{0}, \quad & q \in Q, \quad p \in P \end{aligned} \tag{3}$$



Margin by translation

$$
\begin{aligned}
\min \quad & \|p - q\| \\
\sup(P; z) = & \inf(Q + (p - q); z) \\
z \neq \vec{0}, \quad & q \in Q, \quad p \in P
\end{aligned}
\tag{3}
$$



Margin by translation

$$
\begin{aligned}
\min \quad & \|p - q\| \\
\sup(P; z) = & \inf(Q + (p - q); z) \\
z \neq \vec{0}, \quad & q \in Q, \quad p \in P
\end{aligned} \tag{3}
$$

- Minkowski additivity of support, twice.
- $\vec{0} \neq z \in N(K, r)$ iff $r \in \operatorname{bd} K$.

Margin by translation

$$\begin{aligned} \min \quad & \|p - q\| \\ \sup(P - Q; z) = & \langle z, p - q \rangle \\ z \neq \vec{0}, \quad & q \in Q, \quad p \in P \end{aligned} \tag{3}$$

- Minkowski additivity of support, twice.
- $\vec{0} \neq z \in N(K, r)$ iff $r \in \operatorname{bd} K$.

Define shift(P, Q) as the solution to the following

$$\begin{aligned} & \min \|r\| \\ & r \in \operatorname{bd} P - Q \end{aligned} \tag{NORM}$$

Margin by translation

$$
\begin{aligned}
\min \quad & \|p - q\| \\
& z \in N(P - Q; p - q) \\
z \neq \vec{0}, \quad & q \in Q, \quad p \in P
\end{aligned}
\tag{3}
$$

Define $\text{shift}(P, Q)$ as the solution to the following

$$
\begin{aligned}
& \min \|r\| \\
& r \in \text{bd}\, P - Q
\end{aligned}
\tag{NORM}
$$

Proposition

For convex polyhedra P and Q, $\text{shift}(P, Q)$ is computable in polynomial time given a polynomial sized facet representation for $P - Q$

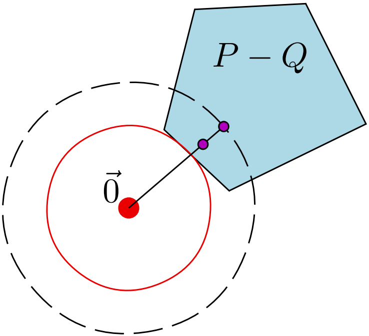
Outline

Finding the minimum translation in the separable case

Proposition

Let P, Q be convex bodies. If $\vec{0} \notin \mathrm{int}(P - Q)$, then

$$\mathrm{shift}(P, Q) = \min_{r \in \mathrm{bd}(P-Q)} \|k\| = \min_{r \in (P-Q)} \|k\|$$



Relaxing (MARGIN)

Proposition

Let P and Q be separable convex bodies.

$$\sup_{w \in \mathbb{S}^*} \inf(P - Q, w) = \sup_{w \in \mathbb{B}^*} \inf(P - Q, w)$$

Proof.

Let $w' := \operatorname{argmax}_{w \in \mathbb{B}^*} \inf(P - Q; w)$.

$$0 \leq \inf(P - Q; w')$$

$$\inf(P - Q; w') \leq \inf(P - Q; w'/\|w'\|_{\mathbb{B}^*})$$

□

Relaxing (MARGIN)

Proposition

Let P and Q be separable convex bodies.

$$\sup_{w \in \mathbb{S}^*} \inf(P - Q, w) = \sup_{w \in \mathbb{B}^*} \inf(P - Q, w)$$

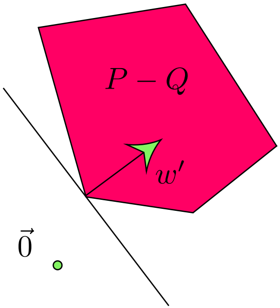
Proof.

Let $w' := \operatorname{argmax}_{w \in \mathbb{B}^*} \inf(P - Q; w)$.

$$0 \leq \inf(P - Q; w')$$

$$\inf(P - Q; w') \leq \inf(P - Q; w'/\|w'\|_{\mathbb{B}^*})$$

□



Relaxing (MARGIN)

Proposition

Let P and Q be separable convex bodies.

$$\sup_{w \in \mathbb{S}^*} \inf(P - Q, w) = \sup_{w \in \mathbb{B}^*} \inf(P - Q, w)$$

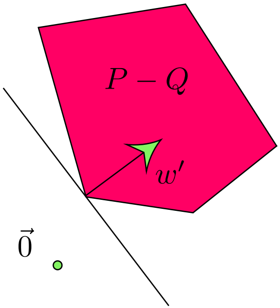
Proof.

Let $w' := \operatorname{argmax}_{w \in \mathbb{B}^*} \inf(P - Q; w)$.

$$0 \leq \inf(P - Q; w')$$

$$\inf(P - Q; w') \leq \inf(P - Q; w'/\|w'\|_{\mathbb{B}^*})$$

□



Weak Duality

Proposition

Let $r \in P - Q$, $z \in \mathbb{B}^*$

$$\|r\| \geq \inf(P - Q; z) \tag{4}$$

Proof.

Since $z \in \mathbb{B}^*$, from (ND),

$$\|r\| = \sup(\mathbb{B}^*; r) \geq \langle z, r \rangle \geq \inf(P - Q; z)$$

□

Weak Duality

Proposition

Let $r \in P - Q$, $z \in \mathbb{B}^*$

$$\|r\| \geq \inf(P - Q; z) \tag{4}$$

Proof.

Since $z \in \mathbb{B}^*$, from (ND),

$$\|r\| = \sup(\mathbb{B}^*; r) \geq \langle z, r \rangle \geq \inf(P - Q; z)$$

□

Weak Duality

Proposition

Let $r \in P - Q$, $z \in \mathbb{B}^*$

$$\|r\| \geq \inf(P - Q; z) \tag{4}$$

Proof.

Since $z \in \mathbb{B}^*$, from (ND),

$$\|r\| = \sup(\mathbb{B}^*; r) \geq \langle z, r \rangle \geq \inf(P - Q; z)$$

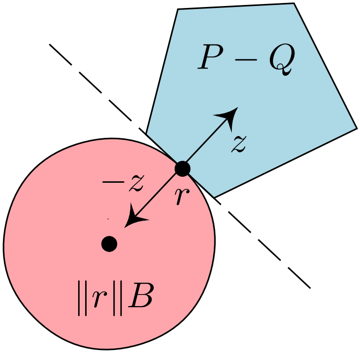
□

Duality of convex relaxations

Theorem

Let P and Q be separable convex bodies. Vectors $r \in P - Q$ and $z \in \mathbb{B}^*$ form a primal–dual optimal pair for (MARGIN) *and (NORM) iff*

$$\|r\|_{\mathbb{B}} = \langle z, r \rangle \qquad -z \in N(P - Q, r)$$



Skip proof

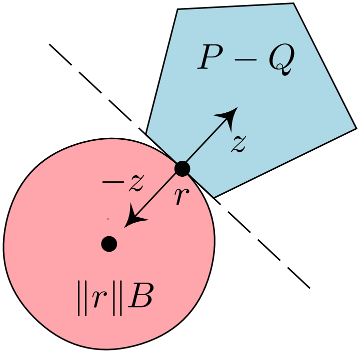
Necessity

Duality of convex relaxations

Theorem

Vectors $r \in P - Q$ *and* $z \in \mathbb{B}^*$ *form a primal–dual optimal pair iff*

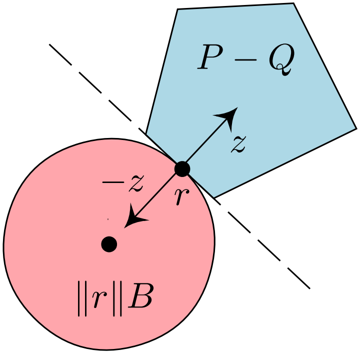
$$\|r\|_{\mathbb{B}} = \langle z, r \rangle \qquad -z \in N(P - Q, r)$$



Necessity

- $r = \operatorname{argmin}_{x \in P-Q} \|x\|$,
  $z = \operatorname{argmax}_{w \in \mathbb{B}^*} \inf(P - Q; w)$

Duality of convex relaxations



Theorem

Vectors $r \in P - Q$ *and* $z \in \mathbb{B}^*$ *form a primal–dual optimal pair iff*

$$\|r\|_{\mathbb{B}} = \langle z, r \rangle \qquad -z \in N(P - Q, r)$$

Necessity

- $r = \operatorname{argmin}_{x \in P-Q} \|x\|$,
  $z = \operatorname{argmax}_{w \in \mathbb{B}^*} \inf(P - Q; w)$
- We argue $\|r\|_{\mathbb{B}} = \langle z, r \rangle$.

Duality of convex relaxations

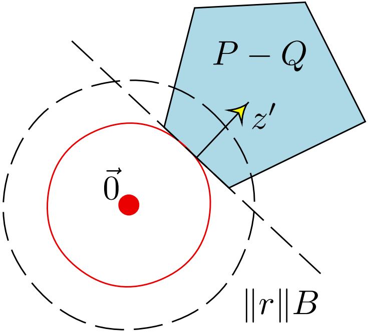
Theorem

Vectors $r \in P - Q$ and $z \in \mathbb{B}^*$ form a primal–dual optimal pair iff

$$\|r\|_{\mathbb{B}} = \langle z, r \rangle \qquad -z \in N(P - Q, r)$$

Necessity

- We argue $\|r\|_{\mathbb{B}} = \langle z, r \rangle$.
- int $\|r\|\mathbb{B} \cap (P - Q) = \emptyset$



Duality of convex relaxations

Theorem

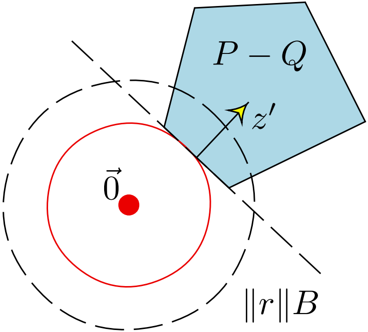
Vectors $r \in P - Q$ and $z \in \mathbb{B}^*$ form a primal–dual optimal pair iff

$$\|r\|_{\mathbb{B}} = \langle z, r \rangle \qquad -z \in N(P - Q, r)$$

Necessity

- $\operatorname{int} \|r\|\mathbb{B} \cap (P - Q) = \emptyset$
- By optimality of z,

$$\begin{aligned} \inf(P - Q; z') &\leq \inf(P - Q; z) \\ &\leq \langle z, r \rangle \end{aligned}$$



By the S.H.T. $\exists z' \in \mathbb{S}^*$,

$$\begin{aligned} \|r\| &= \sup(\|r\|\mathbb{B}; z') \\ &\leq \inf(P - Q; z') \end{aligned}$$

Duality of convex relaxations

Theorem

Vectors $r \in P - Q$ and $z \in \mathbb{B}^*$ form a primal–dual optimal pair iff

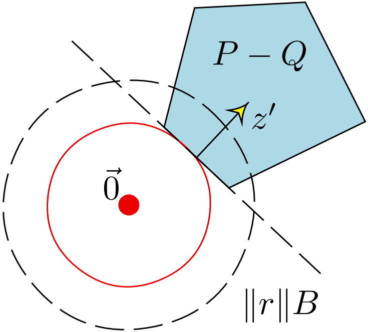
$$\|r\|_{\mathbb{B}} = \langle z, r \rangle \qquad -z \in N(P - Q, r)$$

Necessity

- By optimality of z,

$$\begin{aligned} \inf(P - Q; z') &\leq \inf(P - Q; z) \\ &\leq \langle z, r \rangle \end{aligned}$$

- Now apply Weak Duality.



By the S.H.T. $\exists z' \in \mathbb{S}^*$,

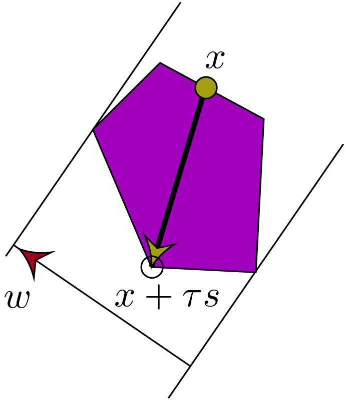
$$\begin{aligned} \|r\| &= \sup(\|r\|\mathbb{B}; z') \\ &\leq \inf(P - Q; z') \end{aligned}$$

Outline

Width of convex bodies

For convex body K, define

$$\text{length}_s(K) := \sup \{\, \tau \mid [x, x + \tau s] \subset K \,\}$$
$$\text{breadth}_w(K) := \sup_{x,y \in K} \langle\, w, x - y \,\rangle = \sup(K - K; w)$$



Width of convex bodies

For convex body K, define

$$\text{length}_s(K) := \sup\{\,\tau \mid [x, x+\tau s] \subset K\,\}$$
$$\text{breadth}_w(K) := \sup_{x,y\in K} \langle\, w, x-y \,\rangle = \sup(K-K; w)$$

Proposition (Gritzmann–Klee 1992)

$$\text{width}(X) := \inf_{w\in\mathbb{S}^*} \text{breadth}_w(X) = \inf_{s\in\mathbb{S}} \text{length}_s(X)$$

Proposition

$$\text{width}(X) = -\text{margin}(X, X)$$

Width of convex bodies

For convex body K, define

$$\text{length}_s(K) := \sup \{ \tau \mid [x, x + \tau s] \subset K \}$$

$$\text{breadth}_w(K) := \sup_{x,y \in K} \langle w, x - y \rangle = \sup(K - K; w)$$

Proposition (Gritzmann–Klee 1992)

$$\text{width}(X) := \inf_{w \in \mathbb{S}^*} \text{breadth}_w(X) = \inf_{s \in \mathbb{S}} \text{length}_s(X)$$

Proposition

$$\text{width}(X) = -\,\text{margin}(X, X)$$

Computing Width is NP-Hard

Theorem (Gritzmann–Klee 1992)

Computing width(K) *is NP-hard, for* K *a simplex, in at least the* L_2 *and* L_∞ *norms.*

Corollary

Computing margin(P, Q) *is NP-hard in at least the* L_2 *and* L_∞ *norms.*

Corollary (GK92, new proof)

Computing width$(\cdot)$ *is solvable in polynomial time for centrally symmetric polytopes given by their facets.*

Computing Width is NP-Hard

Theorem (Gritzmann–Klee 1992)

Computing width(K) *is NP-hard, for* K *a simplex, in at least the* L_2 *and* L_∞ *norms.*

Corollary

Computing margin(P, Q) *is NP-hard in at least the* L_2 *and* L_∞ *norms.*

Corollary (GK92, new proof)

Computing width$(\cdot)$ *is solvable in polynomial time for centrally symmetric polytopes given by their facets.*

Computing Width is NP-Hard

Theorem (Gritzmann–Klee 1992)

Computing width(K) *is NP-hard, for K a simplex, in at least the L_2 and L_∞ norms.*

Corollary

Computing margin(P, Q) *is NP-hard in at least the L_2 and L_∞ norms.*

Corollary (GK92, new proof)

Computing width($\cdot$) *is solvable in polynomial time for centrally symmetric polytopes given by their facets.*

Duality in the inseparable case

Proposition

For convex body K with $\vec{0} \in K$,

$$\inf_{k \in \operatorname{bd} K} \|k\| = \inf_{w \in \mathbb{S}^*} \sup(K; w) \tag{5}$$

Duality in the inseparable case

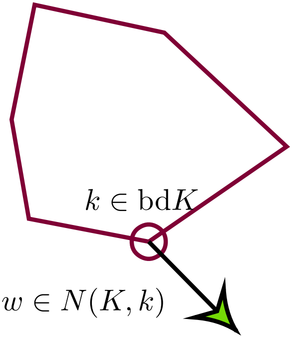
Proposition

For convex body K with $\vec{0} \in K$,

$$\inf_{k \in \mathrm{bd}\, K} \|k\| = \inf_{w \in \mathbb{S}^*} \sup(K; w) \tag{5}$$

- $\forall k \in \mathrm{bd}\, K\ \exists w \in \mathbb{S}^* \cap N(K, k)$

$$\|k\| \geq \langle\, w, k \,\rangle = \sup(K; w)$$



Duality in the inseparable case

Proposition

For convex body K with $\vec{0} \in K$,

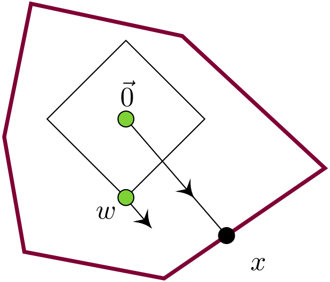
$$\inf_{k \in \operatorname{bd} K} \|k\| = \inf_{w \in \mathbb{S}^*} \sup(K; w) \tag{5}$$

- $\forall k \in \operatorname{bd} K \; \exists w \in \mathbb{S}^* \cap N(K, k)$

  $$\|k\| \geq \langle\, w, k \,\rangle = \sup(K; w)$$

- $\forall w \in \mathbb{S}^* \; \exists x \in N(\mathbb{B}^*, w) \cap \operatorname{bd} K$

  $$\sup(K; w) \geq \langle\, x, w \,\rangle = \|x\|$$



Duality in the inseparable case

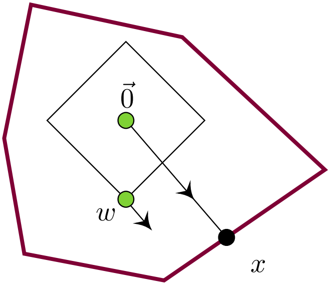
Proposition

For convex body K with $\vec{0} \in K$,

$$\inf_{k \in \operatorname{bd} K} \|k\| = \inf_{w \in \mathbb{S}^*} \sup(K; w) \tag{5}$$

Corollary

If $\vec{0} \in (P - Q)$ then
$\operatorname{shift}(P, Q) = -\operatorname{margin}(P, Q)$



Outline

Margin: separation quantified

Margin as a metric problem

A dual formulation via norm minimization

The separable (convex) case

The inseparable case

Conclusions

Conclusions

- There is a dual relationship between minimum distance, and maximimum margin summarized as

$$\mathrm{shift}(P, Q) = \mathrm{sep}(P, Q) \cdot \mathrm{margin}(P, Q)$$

- In the separable case, both functions are computable by convex minimization in polynomial time, (essentially) arbitrary Minkowski metrics.
- The general (possibly) problem is NP-hard, and only special cases are polynomially solvable (unless P=NP).

Conclusions

- There is a dual relationship between minimum distance, and maximimum margin summarized as

$$\mathrm{shift}(P, Q) = \mathrm{sep}(P, Q) \cdot \mathrm{margin}(P, Q)$$

- In the separable case, both functions are computable by convex minimization in polynomial time, (essentially) arbitrary Minkowski metrics.
- The general (possibly) problem is NP-hard, and only special cases are polynomially solvable (unless P=NP).

Conclusions

- There is a dual relationship between minimum distance, and maximimum margin summarized as

$$\text{shift}(P, Q) = \text{sep}(P, Q) \cdot \text{margin}(P, Q)$$

- In the separable case, both functions are computable by convex minimization in polynomial time, (essentially) arbitrary Minkowski metrics.
- The general (possibly) problem is NP-hard, and only special cases are polynomially solvable (unless P=NP).