### Minimum norm points on polytope boundaries

David Bremner

UNB

#### December 20, 2013

Joint work with Yan Cui, Zhan Gao

#### Outline



2 Maximum margin classifiers

- 3 Enumerative Algorithms
- 4 Cutting Plane Approaches

#### Support functionals

Linear Programs with varying objectives

$$\langle x, y \rangle = x^T y = y^T x$$

For 
$$K \subset \mathbb{R}^d$$
,  $w \in \mathbb{R}^d$ ,  
 $\sup(K; w) := \sup_{k \in K} \langle w, k \rangle$  (upper)  
 $\inf(K; w) := \inf_{k \in K} \langle w, k \rangle$  (lower)  
 $= -\sup(K; -w)$ 

Let polyhedron  $K = \{Ax \le b\}$   $\sup(K; c) = \max c^T x$ Ax < b

#### Support functionals

Linear Programs with varying objectives

$$\langle x, y \rangle = x^{T}y = y^{T}x$$
  
For  $K \subset \mathbb{R}^{d}, w \in \mathbb{R}^{d},$   

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

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

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

$$\sup(K; w) = \sup_{k \in K} \langle w, k \rangle = \{Ax \le b\}$$

#### Support functionals

Linear Programs with varying objectives

$$\langle x, y \rangle = x^{T} y = y^{T} x$$
For  $K \subset \mathbb{R}^{d}, w \in \mathbb{R}^{d},$ 

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

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

$$= -\sup(K; -w)$$
Let polyhedron  $K = \{Ax \le b\}$ 

$$\sup(K; c) = \max c^{T} x$$

 $Ax \leq b$ 

## Outer Normal Cone

Geometric LP Duality



• N(K; p) is the set of all objective functions optimal at x

## Outer Normal Cone

Geometric LP Duality



• N(K; p) is the set of all objective functions optimal at x

#### Polarity



#### Minkowski Norms

$$\|x\|_{\mathbb{B}} = \inf \{ \gamma \ge 0 \mid x \in \gamma \mathbb{B} \}$$
(1)  
= sup(\mathbb{B}^\*; x) (2)

#### Minkowski Norms

$$\|x\|_{\mathbb{B}} = \inf \{ \gamma \ge 0 \mid x \in \gamma \mathbb{B} \}$$
(1)  
= sup(\mathbb{B}^\*; x) (2)



#### Minkowski Norms

$$\|x\|_{\mathbb{B}} = \inf \{ \gamma \ge 0 \mid x \in \gamma \mathbb{B} \}$$
(1)  
= sup(\mathbb{B}^\*; x) (2)



 $\mathbb{S} = \mathsf{bd}\,\mathbb{B}$   $\mathbb{S}^* = \mathsf{bd}(\mathbb{B}^*)$ 

#### Separable

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



#### Margin

 $\begin{array}{ll} \max\gamma & {\sf such that} \\ \exists z\in \mathbb{S}^* \\ {\sf sup}({\it Q};z)+\gamma\leq {\sf inf}({\it P};z) \end{array}$ 

- For the Euclidean norm
   z ∈ S\* just means z is a unit vector.
- Both that constraint and  $z \neq \vec{0}$  are non-convex

#### Separable

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



# $\begin{array}{l} \mathsf{Margin} \\ \mathsf{max}\,\gamma \quad \mathsf{such \ that} \\ \exists z \in \mathbb{S}^* \\ \mathsf{sup}(\textit{Q}; z) + \gamma \leq \mathsf{inf}(\textit{P}; z) \end{array}$

- For the Euclidean norm
   z ∈ S\* just means z is a unit vector.
- Both that constraint and  $z \neq \vec{0}$  are non-convex

#### Separable

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



# $\begin{array}{l} \mathsf{Margin} \\ \mathsf{max}\,\gamma \quad \mathsf{such \ that} \\ \exists z \in \mathbb{S}^* \\ \mathsf{sup}(\textit{Q}; z) + \gamma \leq \mathsf{inf}(\textit{P}; z) \end{array}$

- For the Euclidean norm
   z ∈ S\* just means z is a unit vector.
- Both that constraint and  $z \neq \vec{0}$  are non-convex

#### Separable

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



#### Margin

- $\begin{array}{l} \max\gamma \quad \text{such that} \\ \exists z\in \mathbb{S}^* \\ \sup(Q;z)+\gamma\leq \inf(P;z) \end{array}$
- For the Euclidean norm
   z ∈ S\* just means z is a unit vector.
- Both that constraint and  $z \neq \vec{0}$  are non-convex

#### Separable

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



#### Margin

- $\begin{array}{l} \max\gamma \quad \text{such that} \\ \exists z\in \mathbb{S}^* \\ \sup(\textit{Q};z)+\gamma\leq \inf(\textit{P};z) \end{array}$
- For the Euclidean norm
   z ∈ S\* just means z is a unit vector.
- Both that constraint and  $z \neq \vec{0}$  are non-convex

$$egin{aligned} \min \|w\|^2 + C\gamma \ s.t. \ \langle \, w, p-q \, 
angle \geq 1 - \gamma \quad p \in P, q \in Q \ \gamma \geq 0 \end{aligned}$$

- Need C suff. large to force  $\gamma < 1$  at opt.
- Given optimal (w,  $\gamma < 1$ ),  $\rho > 1$ , define

$$w' = \rho w$$
  
$$\gamma' = 1 - \rho(1 - \gamma)$$

• If 
$$\gamma < 1$$
  
$$\gamma - \gamma' = (\rho - 1)(1 - \gamma) > 0$$

• Therefore

$$\|w'\| - C\gamma' \le \|w\| + C\gamma$$

$$\begin{split} \min \|w\|^2 + C\gamma \\ s.t. \\ \langle \, w, p-q \, \rangle \geq 1 - \gamma \quad p \in P, q \in Q \\ \gamma \geq 0 \end{split}$$

- Need C suff. large to force  $\gamma < 1$  at opt.
- Given optimal ( $w, \gamma < 1$ ),  $\rho > 1$ , define

$$w' = \rho w$$
  
$$\gamma' = 1 - \rho (1 - \gamma)$$

• If 
$$\gamma < 1$$
  
 $\gamma - \gamma' = (\rho - 1)(1 - \gamma) > 0$ 

• Therefore

$$\|w'\| - C\gamma' \le \|w\| + C\gamma$$

$$egin{aligned} \min \|w\|^2 + C\gamma \ s.t. \ \langle \, w, p-q \, 
angle \geq 1 - \gamma \quad p \in P, q \in Q \ \gamma \geq 0 \end{aligned}$$

- Need C suff. large to force  $\gamma < 1$  at opt.
- Given optimal (w,  $\gamma <$  1),  $\rho >$  1, define

$$egin{aligned} & \mathbf{w}' = 
ho \mathbf{w} \ & \gamma' = 1 - 
ho (1 - \gamma) \end{aligned}$$

• If 
$$\gamma < 1$$
  
$$\gamma - \gamma' = (\rho - 1)(1 - \gamma) > 0$$

• Therefore

$$\|w'\| - C\gamma' \le \|w\| + C\gamma$$

$$egin{aligned} \min \|w\|^2 + C\gamma \ s.t. \ \langle \, w, p-q \, 
angle \geq 1 - \gamma \quad p \in P, q \in Q \ \gamma \geq 0 \end{aligned}$$

- Need C suff. large to force  $\gamma < 1$  at opt.
- Given optimal (w,  $\gamma <$  1),  $\rho >$  1, define

$$egin{aligned} & \mathbf{w}' = 
ho \mathbf{w} \ & \gamma' = 1 - 
ho (1 - \gamma) \end{aligned}$$

• If 
$$\gamma < 1$$
  
 $\gamma - \gamma' = (\rho - 1)(1 - \gamma) > 0$ 

• Therefore

$$\|w'\| - C\gamma' \le \|w\| + C\gamma$$

$$\begin{split} \min \|w\|^2 + C\gamma \\ s.t. \\ \langle w, p - q \rangle \geq 1 - \gamma \quad p \in P, q \in Q \\ \gamma \geq 0 \end{split}$$

- Need C suff. large to force  $\gamma < 1$  at opt.
- Given optimal (w,  $\gamma <$  1),  $\rho >$  1, define

$$egin{aligned} & \mathbf{w}' = 
ho \mathbf{w} \ & \gamma' = 1 - 
ho (1 - \gamma) \end{aligned}$$

• If 
$$\gamma < 1$$
  
 $\gamma - \gamma' = (
ho - 1)(1 - \gamma) > 0$ 

• Therefore

$$\|w'\| - C\gamma' \le \|w\| + C\gamma$$

$$\begin{split} \min \|w\|^2 + C\gamma \\ s.t. \\ \langle w, p - q \rangle \geq 1 - \gamma \quad p \in P, q \in Q \\ \gamma \geq 0 \end{split}$$

- Need C suff. large to force  $\gamma < 1$  at opt.
- Given optimal (w,  $\gamma <$  1),  $\rho >$  1, define

$$egin{aligned} & \mathbf{w}' = 
ho \mathbf{w} \ & \gamma' = 1 - 
ho (1 - \gamma) \end{aligned}$$

• If 
$$\gamma < 1$$
  
 $\gamma - \gamma' = (
ho - 1)(1 - \gamma) > 0$ 

• Therefore

$$\|w'\| - C\gamma' \le \|w\| + C\gamma$$

#### Translate until separable

Minimize Translation Let P and Q convex bodies. inf ||t|| $\sup(P; z) \leq \inf(Q + t; z)$  $z \neq \vec{0}, \quad q \in Q, \quad q+t \in P$ inf ||r|| $z \in N(P-Q;r)$  $z \neq \vec{0}, r \in P - Q$ 



#### Translate until separable

Minimize Translation Let P and Q convex bodies. inf ||t|| $\sup(P; z) \leq \inf(Q + t; z)$  $z \neq \vec{0}, \quad q \in Q, \quad q+t \in P$ inf ||r|| $z \in N(P-Q;r)$  $z \neq \vec{0}, r \in P - Q$ 



#### Duality

#### Proposition

If  $\vec{0} \in P - Q$ ,

$$shift(P, Q) = -margin(P, Q)$$

Choose  $w \in \mathbb{S}^*$ ,  $z \in N(\mathbb{B}^*; w) \cap \mathsf{bd}(P-Q)$ 

$$\sup(P - Q, w) \ge \langle z, w \rangle = ||z||$$
$$\ge \inf_{k \in bd P - Q} ||k||$$



If  $k \in \text{bd } P - Q$ ,  $w \in N(P - Q; k) \cap \mathbb{S}^*$ 

 $||k|| \ge \langle w, k \rangle = \sup(P - Q; w)$ 

#### Duality

#### Proposition

If  $\vec{0} \in P - Q$ ,

$$shift(P, Q) = -margin(P, Q)$$

Choose 
$$w \in \mathbb{S}^*$$
,  
 $z \in N(\mathbb{B}^*; w) \cap bd(P - Q)$   
 $sup(P - Q, w) \ge \langle z, w \rangle = ||z||$   
 $\ge \inf_{k \in bd P - Q} ||k||$ 



If  $k \in \operatorname{bd} P - Q$ ,  $w \in N(P - Q; k) \cap \mathbb{S}^*$ 

 $\|k\| \geq \langle w, k \rangle = \sup(P - Q; w)$ 

## Duality

#### Proposition

If  $\vec{0} \in P - Q$ ,

$$shift(P, Q) = -margin(P, Q)$$

Choose 
$$w \in \mathbb{S}^*$$
,  
 $z \in N(\mathbb{B}^*; w) \cap bd(P - Q)$   
 $sup(P - Q, w) \ge \langle z, w \rangle = ||z||$   
 $\ge \inf_{k \in bd P - Q} ||k||$ 



If  $k \in bd P - Q$ ,  $w \in N(P - Q; k) \cap \mathbb{S}^*$ 

$$||k|| \geq \langle w, k \rangle = \sup(P - Q; w)$$

$$\inf_{r \in bd K} \|r\| = \min_{F \text{ facet of } K} \min_{r \in F} \|r\|$$

$$= \min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{r \in bd K} \{x \mid a^{T}x = 1\} = \frac{1}{\|a\|}$$

$$\operatorname{Assuming } \vec{0} \in \operatorname{int } K$$

$$\inf_{r \in bd K} \|r\| = \max_{v \text{ vertex of } K^{*}} \|v\|$$

$$= \max_{x \in K^{*}} \|x\|$$

$$\operatorname{Assuming } \vec{0} \in \operatorname{int } K$$

$$\inf_{r \in bd K} \|r\| = \min_{F \text{ facet of } K} \min_{r \in F} \|r\|$$

$$= \min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{F \text{ facet of } K} x = 1 \} = \frac{1}{\|a\|}$$
Theorem (Brieden 2002)
For  $2 \le p < \infty \max_{r \in M} \|x\|_{p}^{p} |Ax \le b$ 
cannot be computed in polynomial
time within a factor of 1.090,
unless  $P = NP$ .

$$\inf_{r \in bd K} \|r\| = \min_{F \text{ facet of } K} \min_{r \in F} \|r\|$$

$$= \min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{F \text{ facet of } K} \min_{r \in aff F} \|r\|$$

$$\min_{F \text{ facet of } K} x = 1 \} = \frac{1}{\|a\|}$$

$$\operatorname{Assuming } \vec{0} \in \operatorname{int } K$$

$$\inf_{r \in bd K} \|r\| = \max_{v \text{ vertex of } K^*} \|v\|$$

$$= \max_{x \in K^*} \|x\|$$

$$\operatorname{Assuming } x = 1 \} = \max_{v \text{ vertex of } K^*} \|v\|$$

$$= \max_{x \in K^*} \|x\|$$

$$\inf_{r \in bd \ K} \|r\| = \min_{F \text{ facet of } K} \min_{r \in F} \|r\|$$

$$= \min_{F \text{ facet of } K} \min_{r \in aff \ F} \|r\|$$

$$\min_{F \text{ facet of } K} \min_{r \in aff \ F} \|r\|$$

$$\min_{F \text{ facet of } K} x = 1 \} = \frac{1}{\|a\|}$$

$$\operatorname{Assuming } \vec{0} \in \operatorname{int } K$$

$$\inf_{\substack{r \in bd \ K}} \|r\| = \max_{\substack{v \text{ vertex of } K^* \\ v \text{ vertex of } K^* \\ = \max_{x \in K^*}} \|v\|$$

$$\lim_{r \in bd \ K} \|x\|$$

$$\lim_{r \in bd \ K} \|x\|$$

#### Hardness

$$\mathsf{breadth}_w(X) := \sup_{\substack{x,y \in X}} \langle w, x - y \rangle = \mathsf{sup}(X - X; w)$$
  
 $\mathsf{width}(X) := \inf_{w \in \mathbb{S}^*} \mathsf{breadth}_w(X)$ 

Proposition

width(
$$X$$
) = - margin( $X, X$ )

Proposition (Gritzmann-Klee 1992)

Computing the width is NP-hard.

#### Hardness

$$\mathsf{breadth}_w(X) := \sup_{\substack{x,y \in X}} \langle w, x - y \rangle = \mathsf{sup}(X - X; w)$$
  
 $\mathsf{width}(X) := \inf_{w \in \mathbb{S}^*} \mathsf{breadth}_w(X)$ 

#### Proposition

width(
$$X$$
) = - margin( $X, X$ )

Proposition (Gritzmann-Klee 1992)

Computing the width is NP-hard.

#### **Brute Force Solution**

#### function

MINNORMFACET $(P, Q \subset \mathbb{R}^d)$   $\lambda = +\infty$  R = ExtremePoints(P - Q)for all facets F of conv R do  $\lambda = \min(\text{MinNormAff}(F), \lambda)$ end for end function

## Brute Force Solution



#### Enumerating Facets with Reverse Search

- Facets of *P Q* are computed without extra memory
- the number of facets and bases grow quickly.



#### Enumerating Facets with Reverse Search

- Facets of *P Q* are computed without extra memory
- the number of facets and bases grow quickly.





## **Classification Accuracy**



### Maximum Error



David Bremner (UNB)

Minimum norm facet

#### Eccentricity



• Most of the volume is far away from the interesting part

• Hopefully most of the combinatorial complexity is also far away.

David Bremner (UNB)

#### Eccentricity



- Most of the volume is far away from the interesting part
- Hopefully most of the combinatorial complexity is also far away.

David Bremner (UNB)

#### **Cutting Planes**



#### Double description style update (Motzkin)

Let  $E_H$  be the edges of conv V that properly intersect H. The vertex set of  $\operatorname{conv}(V) \cap H^-$  is  $V \cap H^- \cup \bigcup_{e \in E_H} (e \cap H)$ 

#### **Cutting Planes**



#### Double description style update (Motzkin)

Let  $E_H$  be the edges of conv V that properly intersect H. The vertex set of conv $(V) \cap H^-$  is  $V \cap H^- \cup \bigcup_{e \in E_H} (e \cap H)$ 

#### Heuristics

- Adding a cutting plane is (usually) only effective if it cuts off many more vertices than it creates.
- Experimentally, the farthest vertex seems to be a good choice of cutting plane normal, at least for "random-ish" data.



#### Heuristics

- Adding a cutting plane is (usually) only effective if it cuts off many more vertices than it creates.
- Experimentally, the farthest vertex seems to be a good choice of cutting plane normal, at least for "random-ish" data.



- In some sense what we want is polytopal approximation of the sphere.
- Find polytope with small *H* and *V* complexity containing sphere.
- Using (approximately) regular simplex
- or round carefully
- doesn't seem to iterate that well.



- In some sense what we want is polytopal approximation of the sphere.
- Find polytope with small *H* and *V* complexity containing sphere.
- Using (approximately) regular simplex
- or round carefully
- doesn't seem to iterate that well.



- In some sense what we want is polytopal approximation of the sphere.
- Find polytope with small *H* and *V* complexity containing sphere.
- Using (approximately) regular simplex
- or round carefully
- doesn't seem to iterate that well.



- In some sense what we want is polytopal approximation of the sphere.
- Find polytope with small *H* and *V* complexity containing sphere.
- Using (approximately) regular simplex
- round carefully
- doesn't seem to iterate that well.



- In some sense what we want is polytopal approximation of the sphere.
- Find polytope with small *H* and *V* complexity containing sphere.
- Using (approximately) regular simplex
- round carefully
- doesn't seem to iterate that well.



## Performance in $\mathbb{R}^9$



David Bremner (UNB)

#### Average runtime versus dimension



time (seconds)

### Cutting polytope method



#### Conclusions

- Minimum norm boundary point is dual to maximum margin classifying hyperplane
- NP-Hard
- Solvable via (partial) enumerative approaches.
- Cutting planes/polytopes yield some speedup, but plenty of room for improvement.

#### Acknowledgements

- Geometric background: with Peter Gritzmann
- Software: Irslib (Avis), cddlib (Fukuda), gmp, Minkowski sum (Weibel)

Cutting Plane Approaches

#### Filtering redundant points of P - Q



#### Proposition (Fukuda)

Let *m* be the max degree of a vertex of conv( $X_i$ ). The extreme points of  $X_1 - X_2$  can be computed with  $O(m \cdot f_0(X_1 - X_2) + f_0^2(X_1) + f_0^2(X_2))$ LPs with *m* constraints.

#### Vertices of the unit ball

