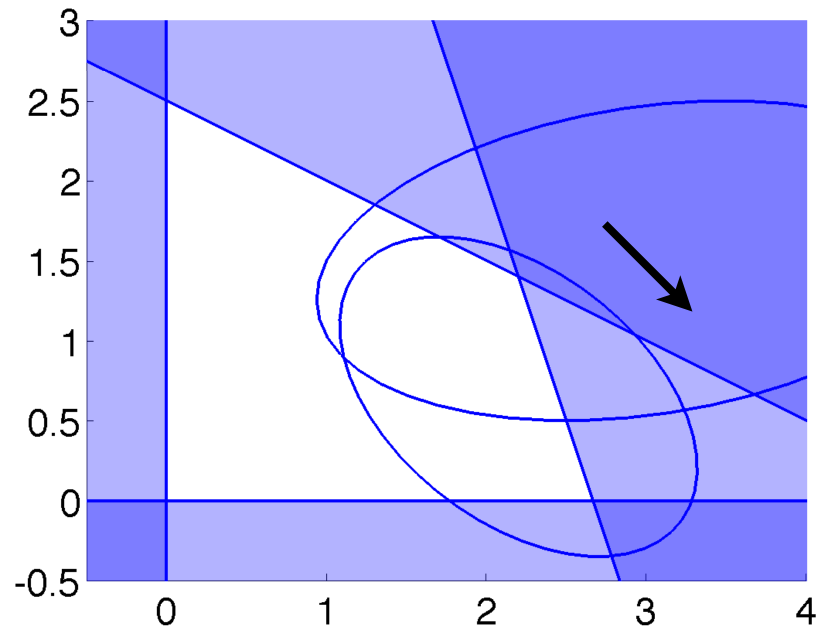
QP & cone program duality
Support vector machines

10-725 Optimization
Geoff Gordon
Ryan Tibshirani

Review

- Quadratic programs
- Cone programs
  - SOCP, SDP
  - QP ⊆ SOCP ⊆ SDP
  - SOC, S_+ are self-dual



- Poly-time algos (but not strongly poly-time, yet)
- Examples: group lasso, Huber regression, matrix completion

Matrix completion

- Observe A_{ij} for $ij \in E$, write $\mathcal{O}_{ij} = \begin{cases} 1 & ij \in E \\ 0 & \text{o/w} \end{cases}$

- $\min_X \|(X-A)\circ \cancel{P}\,\mathcal{O}\|_F^2 + \underbrace{\lambda\|X\|_*}_{\lambda(\mathrm{tr}(P)+\mathrm{tr}(Q))/2}$

$M = \begin{bmatrix} P & X \\ X^T & Q \end{bmatrix} \succeq 0$ s.t.

$X = U\Sigma V^T$

$M \succeq 0 \iff \mathrm{tr}(B^T M) \geq 0 \;\; \forall B \succeq 0$ take $B = \begin{pmatrix} UU^T & -UV^T \\ -VU^T & VV^T \end{pmatrix}$

$0 \leq \mathrm{tr}(B^T M) = \mathrm{tr}(\cancel{UU^T}P) + \mathrm{tr}(\cancel{VV^T}Q) - 2\,\mathrm{tr}(VU^T X)$

$\mathrm{tr}(P) + \mathrm{tr}(Q) \geq 2\,\mathrm{tr}(\underbrace{U^T X V}_{\Sigma}) = 2\|X\|_*$

$P = U\Sigma U^T \quad Q = V\Sigma V^T$

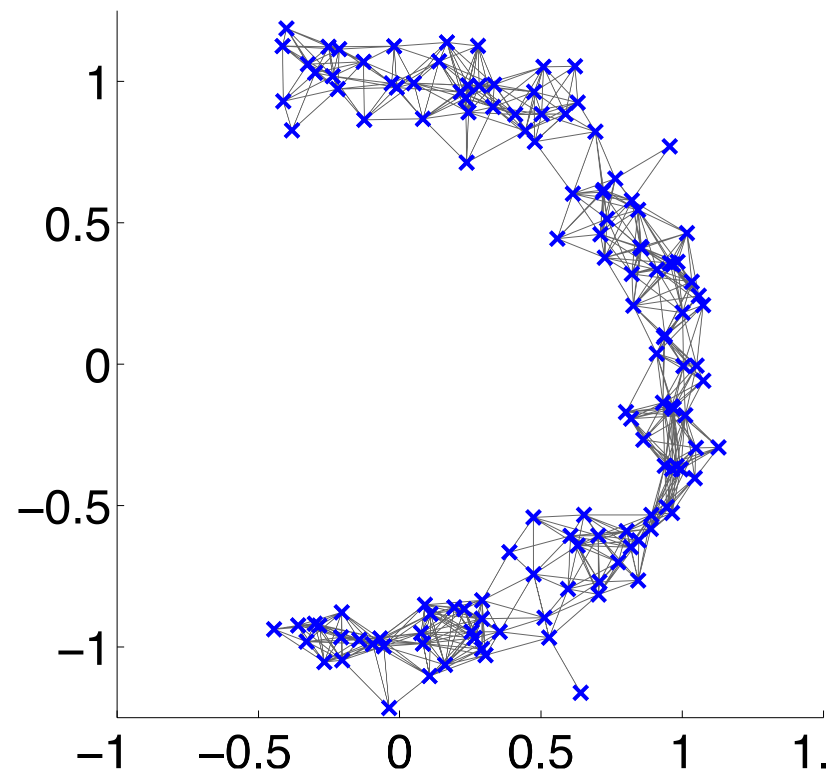
$\mathrm{tr}(P) = \mathrm{tr}(\cancel{U^T U}\Sigma) \quad \mathrm{tr}(P) + \mathrm{tr}(Q) = 2\|X\|_*$

$\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^T \begin{pmatrix} U\Sigma U^T & -U\Sigma V^T \\ -V\Sigma U^T & V\Sigma V^T \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} \Sigma & -\Sigma \\ -\Sigma & \Sigma \end{pmatrix} \succeq 0$

Max-variance unfolding

aka semidefinite embedding

- Goal: given $x_1, \ldots x_T \in R^n$
  - find $y_1, \ldots, y_T \in R^k$ $(k \ll n)$
  - $||y_i - y_j|| \approx ||x_i - x_j||$ $\forall i,j \in E$
- If x_i were near a k-dim subspace of R^n, PCA!
- Instead, two steps:
  - first look for $z_1, \ldots z_T \in R^n$ with
    - $||z_i - z_j|| = ||x_i - x_j||$ $\forall i,j \in E$
    - and var(z) as big as possible
  - then use PCA to get y_i from z_i



MVU/SDE

- $\max_z \operatorname{tr}(\operatorname{cov}(z))$ s.t. $\|z_i - z_j\| = \|x_i - x_j\| \;\; \forall i,j \in E$

$X = (x_1\ x_2\ \dots\ x_T)$ $\quad Z = (z_1 \dots z_T)$ $\quad \underline{P = X^T X}$ $\quad Q = Z^T Z$ $\quad \boxed{Q \succeq 0}$

optimize over Q

$\|z_i - z_j\|^2 = z_i^T z_i - 2 z_i^T z_j + z_j^T z_j$

$$\boxed{Q_{ii} - 2Q_{ij} + Q_{jj} = P_{ii} - 2P_{ij} + P_{jj} \quad ij \in E}$$

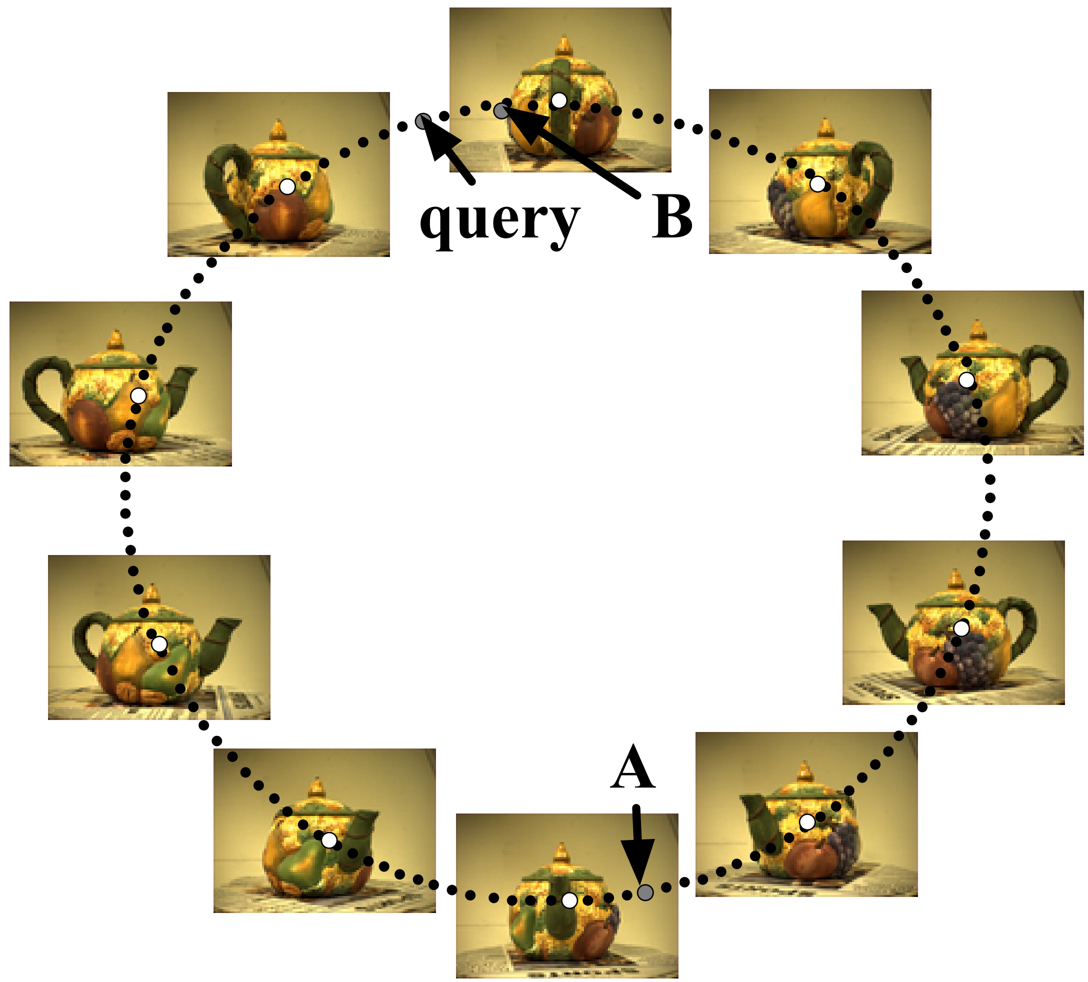
$\operatorname{cov}(z) = \frac{1}{n}\sum_i (z_i - \bar z)(z_i - \bar z)^T = \frac{1}{n}\sum z_i z_i^T - \bar z \bar z^T = \frac{1}{n} ZZ^T - Z11^T Z^T / n^2$

$$\operatorname{tr}(\operatorname{cov}(z)) = \boxed{\frac{1}{n}\operatorname{tr}(Q) - \operatorname{tr}(Q11^T)/n^2 \leftarrow \text{max}}$$

Result

- Embed 400 images of a teapot into 2d

Euclidean distance from query to A is smaller; after MVU, distance to B is smaller



[Weinberger & Saul, AAAI, 2006]

Duality for QPs and Cone Ps

- Combined QP/CP:
  - min $c^T x + x^T H x/2$ s.t. $Ax + b \in K$ $x \in L$
  - cones K, L implement any/all of equality, inequality, generalized inequality
  - assume K, L proper (closed, convex, ~~solid~~, ~~pointed~~)

$y \in K^*$ $s \in L^*$

$y^T(Ax+b) \geq 0$ $x^T s \geq 0$

$$c^T x + x^T H x/2 \geq c^T x + x^T H x/2 - y^T(Ax+b) - s^T x$$
$$\geq \min_z c^T z + z^T H z/2 - y^T(Az+b) - s^T z$$
$$= -z^T H z/2 - y^T b$$

$0 = c + Hz - A^T y - s$

$Hz = s + A^T y - c$

$$\max_{y \in K^*} -z^T H z/2 - y^T b \quad \text{st} \begin{cases} 0 = c + Hz - A^T y - s \\ s \in L^* \end{cases}$$

$\hookrightarrow c + Hz - A^T y \in L^*$

Primal-dual pair

- Primal:
  - min $c^T x + x^T H x/2$ s.t. $Ax + b \in K$ $x \in L$
- Dual:
  - max $-z^T H z/2 - b^T y$ s.t. $Hz + c - A^T y \in L^*$ $y \in K^*$

KKT conditions

primal-dual pair

- min $c^Tx + x^THx/2$ s.t. $Ax + b \in K$ $\quad x \in L$
- max $-b^Ty - z^THz/2$ s.t. $Hz + c - A^Ty \in L^*$ $\quad y \in K^*$

$+x^THz - x^THz + x^TA^Ty - x^TA^Ty$

$c^Tx + x^THx/2 + b^Ty + z^THz/2 = 0$

$(x-z)^TH(x-z)/2 + (Ax+b)^Ty + x^T(c+Hz-A^Ty) = 0$

$(Ax+b)^Ty = 0$

$x^T(c+Hz-A^Ty) = 0$

$Hx = Hz$

KKT cond's

KKT conditions

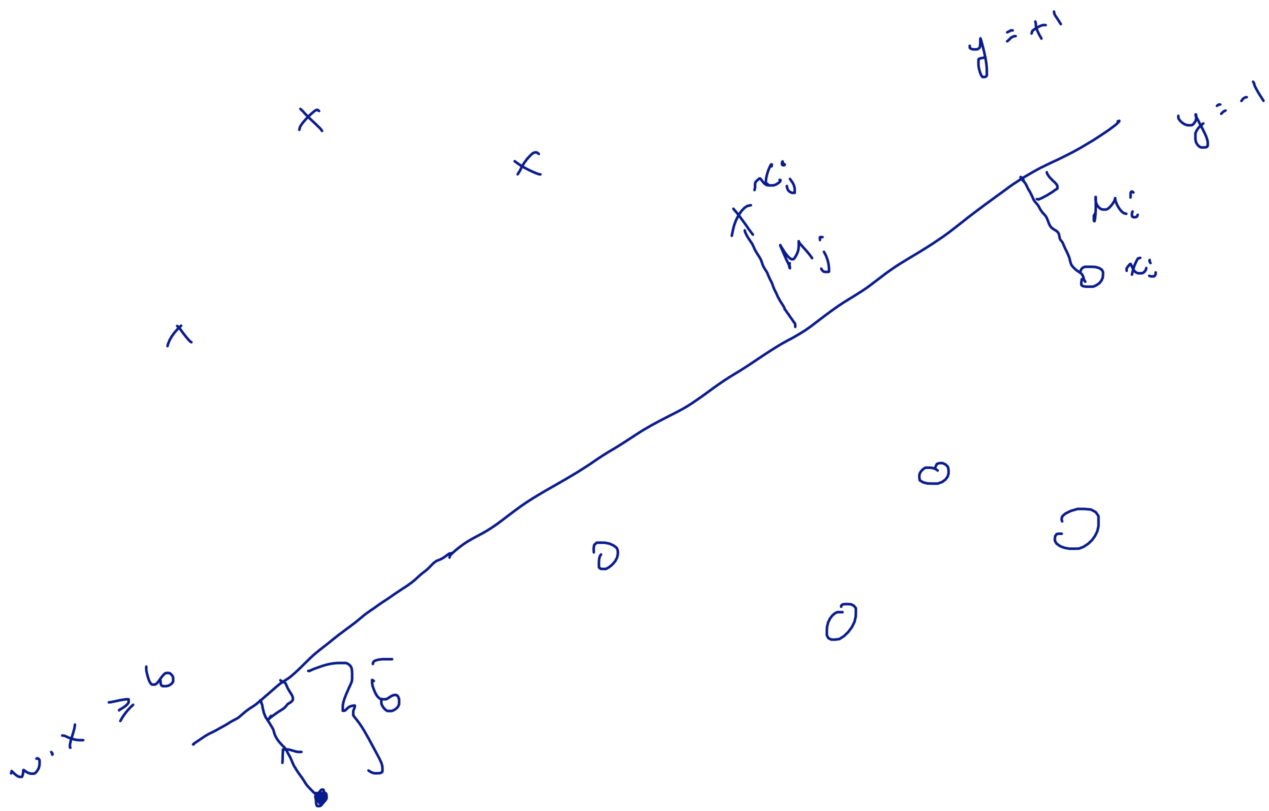
$c - A^T y \in L^*$

- primal: $Ax+b \in K$ $\quad x \in L$
- dual: $Hz + c - A^T y \in L^*$ $\quad y \in K^*$
- quadratic: $Hx = Hz$ $\quad \rightarrow \emptyset$
- comp. slack: $y^T(Ax+b) = 0$ $\quad x^T(Hz+c-A^T y) = 0$

$x^T(c - A^T y) = 0$

Support vector machines

(separable case)



$x_i \in \mathbb{R}^n \qquad y_i \in \{-1, 1\}$

$\bar{w} = w/\|w\| \qquad \bar{b} = b/\|w\|$

$M_i = \bar{b} - \bar{w} \cdot x_i = y_i(\bar{w} \cdot x_i - \bar{b})$

$M_j = \bar{w} \cdot x_j - \bar{b} = y_j(\bar{w} \cdot x_j - \bar{b})$

max M st

$M \leq y_i(x_i \cdot \bar{w} - \bar{b}) \quad \forall i$

Maximizing margin

- margin $M = y_i (x_i \cdot \bar{w} - \bar{b})$
- max M s.t. $M \leq y_i (x_i \cdot \bar{w} - \bar{b})$

$v = \bar{w}/M \qquad d = \bar{b}/M$

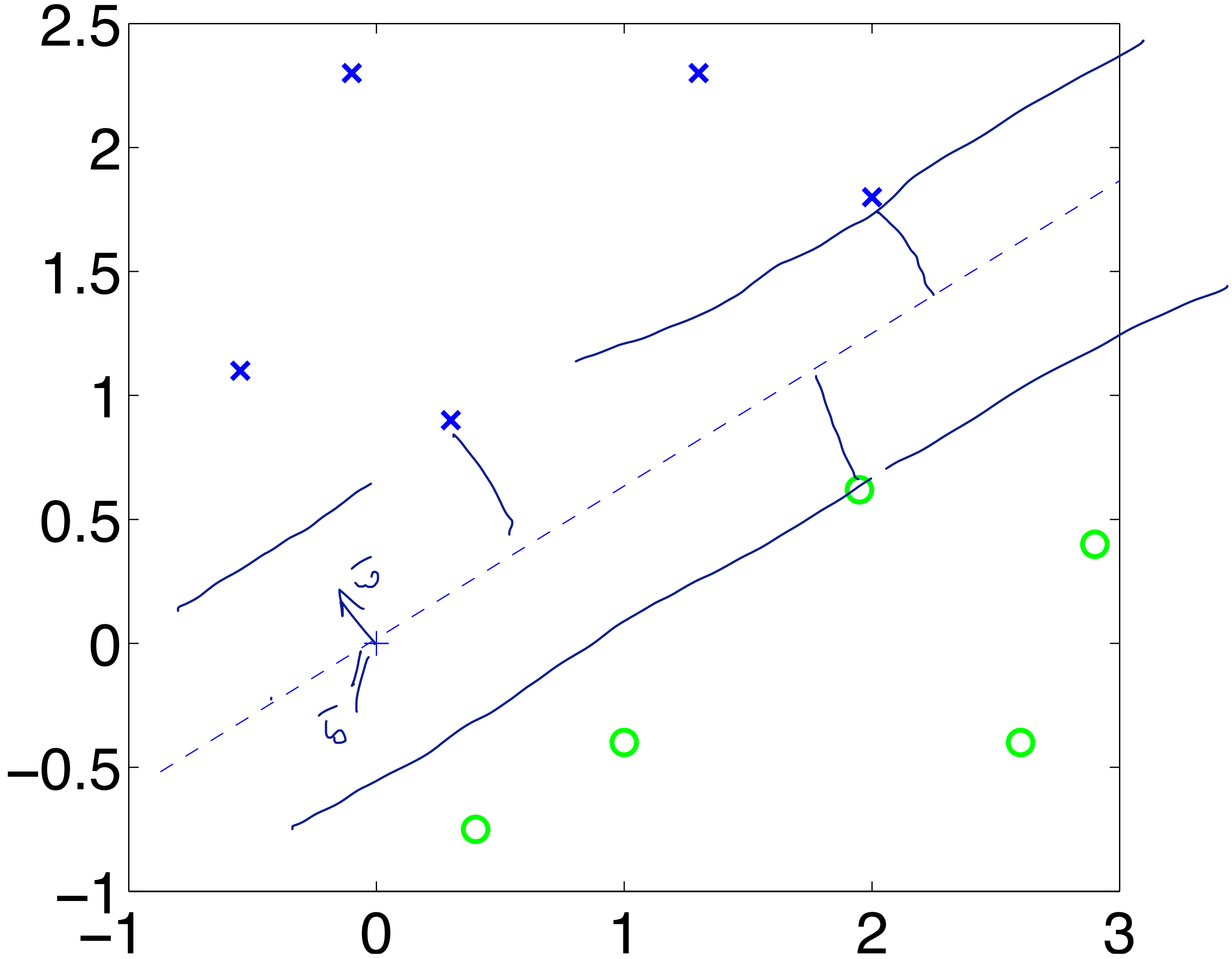
$\bar{w} = Mv \qquad \bar{b} = Md$

$\|v\| = 1/M$

$\max \ 1/\|v\| \quad \text{st} \quad 1 \leq y_i (x_i \cdot v - d)$

$\min \ \|v\|^2 \quad \text{st} \quad$ "

For example



Slacks

- $\min \|v\|^2/2 + C \sum_i s_i$ s.t. $y_i (x_i^T v - d) \geq 1 - s_i \quad \forall i$

$s_i \geq 0$