Linear feasibility problem

\[
\begin{align*}
\text{min} & \quad c^T x & \text{dual} & \quad \max & \quad -b^T y \\
\text{s.t.} & \quad Ax + b \geq 0 & \quad \text{optimal} & \quad A^T y = c, \ y \geq 0 \\
\end{align*}
\]

find \( x \) s.t.
\( Ax + b \geq 0 \)
linear feasibility

Separation oracle
Difficulties

• How do we get bounding sphere?

• How do we know when to stop?
  \[ \forall i: a_i x + b_i^T y + c_i \geq 0 \quad \text{and} \quad \eta > \text{small constant} \]

• Bound region gets complicated—how do we find its center?
Bounding a half-ellipsoid

- General ellipsoid w/ center $x_C$, shape $A$:
  $$(x - x_c)^T A (x - x_c) \leq 1$$
  $A = U^T U$

- Halfspace: $p^T x \leq p^T x_c$ 
  $p^T (U^T y + x_c) \leq p^T x_c$

- Translate to origin, scale to be spherical
  $$y = U (x - x_c)$$
  $$x = U^{-1} y + x_c$$
  $$(y - x_c)^T U^T U (y - x_c) \leq 1$$
  $q^T = U^T p / \|U^T p\|$
  $q^T y \leq 1$
  $p^T U y \leq 0$
  $q^T y < 0$

Bounding a half-sphere

- Rotate so hyperplane is axis-normal
  $$(0 \ldots 0)$$
  $e_i = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$
  $R = q + e_i$
  $R = \frac{2rr^T}{r^T r} - I$
  $p^T R = I$
  $z = R y$ 
  $z^T \leq 1$
  $\varepsilon_i z \leq 0$
  $(z - z_c)^T B (z - z_c) \leq 1$
  $z^T z \leq 1 \land \varepsilon_i z \leq 0$

- New center $z_c$:
  $$\frac{1}{n-1} \sum_{i=1}^{n} e_i$$

- New shape $B$:
  $$\frac{z^2}{n-1} \begin{pmatrix} \varepsilon_1^2 \\ \vdots \\ \varepsilon_n^2 \end{pmatrix}$$
For example

Ellipsoid algorithm

• Want to find $x$ s.t. $Ax + b + \eta \geq 0$
• Pick $E_0$ s.t. $x^* \in E_0$
• for $t := 1, 2, \ldots$
  – $x_t :=$ center of $E_t$
  – ask whether $Ax_t + b + \eta \geq 0$
    • yes: declare feasible!
    • no: get new constraint w/ normal $p_t$
  – $E_{t+1} := \text{bound}(E_t \cap \{ x | p_t^T x \leq p_t^T x_t \})$
  – if $\text{vol}(E_{t+1}) \leq \varepsilon \text{vol}(E_0)$: declare infeasible!
Getting bounds

- How big does $E_0$ need to be?
- What should $\eta$ be?
- How small does $\varepsilon$ need to be?

- Dotting i’s, crossing t’s

- What if LF was unbounded? So what?

- What about numerical precision?

$$A \cdot U^T U \Rightarrow seven \ roots \ \Rightarrow \ infinite \ precision ??$$

$$O(n^3 M)$$
Comparison to constraint generation

- Ellipsoid is polynomial, but slow
- Constraint generation has no non-trivial runtime bound, but often much faster

Other algorithms

- Interior point: polynomial, can be very fast
- Simplex: exponential in worst case, but often fast in practice
- Randomized simplex: polynomial [Kelner & Spielman, 2006]
- Subgradient descent: weakly polynomial, but really simple, and fast for some purposes
What’s a subgradient?

Subgradients for SVMs

- \( \min_{s,w,b} \|w\|^2 + C \sum_i s_i \) s.t.
  \[ y_i (x_i^T w - b) \geq 1 - s_i \]
  \( s_i \geq 0 \)

- Equivalently, \( h(z) = \max \{ 0, 1 - z \} \)
  \[ \min_{w,b} \|w\|^2 + C \sum_i h(-z_i) = \|w\|^2 + C \sum_i h(-y_i (x_i^T w - b)) \]
Subgradients for SVMs

- $\min_w L(w) = ||w||^2 + \frac{C}{m} \sum_i h(y_i x_i^T w)$

- Subgradient of $h(z)$:
  
  $\nabla h(z) = \begin{cases} 
  0 & z < -1 \\
  1 & z > -1 \\
  \left[0,1\right] & z = -1 \\
  0 & z > 1 
  \end{cases}$

- Subgradient of $L(w)$ wrt $w$:
  
  $\nabla L(w) = 2w + \frac{C}{m} \sum_i h(-y_i x_i^T w) \cdot (-y_i x_i)$
Subgradient descent

- While not tired:
  
  \[ g_t = \text{(estimate of) } \frac{\partial f(x_t)}{\partial x_t} \]
  
  \[ x_{t+1} = x_t - \eta_t g_t \]
  
  \[ x_{t+1}^* = \prod_F x_{t+1} \]
  
  \( \eta_t \) = learning rate
Subgradient questions

• How to choose learning rate?

• How to decide when we’re tired?

• How to estimate $\partial f(x_i)$?