Algorithms

• Algorithms for LPs
  ‣ Naive: enumerate all bases
  ‣ Simplex
  ‣ Constraint generation (sort of)

• Algorithms for QPs
  ‣ Constraint generation

• Upcoming:
  ‣ Subgradient descent
  ‣ Ellipsoid
  ‣ Interior point (later)
What’s a subgradient?
Residents of Delphi

subgradient

separation oracle for epigraph
Working w/ subgradients

• Subgradients are just like gradients (sort of):
  ‣ when \( f'(x) \) exists: \( \partial f(x) = \)
  ‣ \( \partial (f+g)(x) = \)
  ‣ \( \partial f(Ax+b) = \)
  ‣ \( \partial f(g(x)) = \)
  ‣ \( 0 \in \partial f(x) \iff \)

f, g convex
Proof of chain rule

• $\partial f(g(x)) = (\partial f)(g(x)) \partial g(x)$
  - $f, g$ convex; $f$ monotone
Ex: dual norm

• Given a norm, \( \|x\| \)

• The **dual norm** is
  
  \[ \|y\|_* = \max_{x \in C} x^T y \]

  \[ C = \]

  • ex: dual of \( \|x\|_1 \) is:
  
  • ex: dual of \( \|x\|_2 \) is:

• Suppose we want to solve
  
  \[ \min_w \|w\|_*^2 + \lambda \sum_i (x_i^T w - y_i)^2 \]
Subgradient of dual norm

• $\|w\|_* = \max_{x \in C} x^T w$
  
  ‣ $C = \{ x \mid \|x\| \leq 1 \}$

• $\partial \|w\|^2_* =$

• Let $x_w \in \arg \max_{x \in C} x^T w$
  
  ‣ $\|w\|_* =$
  
  ‣ $\|v\|_* \geq$

• $\partial \|w\| =$
Ex: SVM subgradient

• $\min_w L(w) = ||w||^2/2 + (C/m) \sum_i h(y_i(x_i^Tw-b))$

  ‣ where $h(z) =$

• $\partial h(z) =$

• $\partial L(w) =$
SVM loss: hinge portion
SVM loss: hinge portion
SVM loss: $C = 10$
Subgradient descent

- Greedily try to decrease objective
  - fastest local decrease: negative subgradient
Subgradient descent

- Initialize $x_1$
- for $t = 1$ to “I’m tired”
  - $g_t = $ any element of $\partial f(x_t)$
  - $x_{t+1} = x_t - \eta_t g_t$
- Questions:
  - how to initialize?
  - when are we tired?
  - what to use for $\eta_t$?
Numerical example

- SVM, training examples:
  - (3, +), (0.1, -)

- \( L(w,b) = \frac{w^2}{2} + \left( \frac{C}{2} \right) (h(z_1) + h(z_2)) \)
  - \( z_1 = \)
  - \( z_2 = \)

- \( \partial L(w,b) = \)
Subgradient path
Subgradient projection

• Initialize $x_1$

• for $t = 1$ to “I’m tired”:
  ‣ $g_t = \text{any element of } \partial f(x_t)$
  ‣ $x_{t+0.5} = x_t - \eta_t g_t$
  ‣ $x_{t+1} = \arg \min_{x \in F} ||x - x_{t+0.5}||$

Problem:
$\min_x f(x)$
$s.t. \ x \in F$
Stochastic subgradient

- Recall SVM subgradient:
  \[ \partial L = w + \frac{C}{m} \sum_{i=1}^{m} y_i x_i \partial h(y_i(x_i^T w - b)) \]

- If many examples: 

- Simple trick: to get an unbiased estimate of \( \partial L \), select \( m_0 \ll m \) examples at random, compute
  \[ \partial L \approx w + \frac{C}{m_0} \sum y_i x_i \partial h(y_i(x_i^T w - b)) \]

  \[ L \approx \|w\|^2/2 + \frac{C}{m_0} \sum y_i x_i h(y_i(x_i^T w - b)) \]
Example: $m_0 = 10$ of 100
Stochastic subgradient (or stoch. subg. projection)

- Initialize $x_1$

- for $t = 1$ to “I’m tired”:
  - $f_t = \text{estimate of } f$
  - $g_t = \text{any element of } \partial f_t(x_t)$
  - $x_{t+0.5} = x_t - \eta_t g_t$
  - $x_{t+1} = \text{arg min}_{x \in F} \| x - x_{t+0.5} \|$ (or just $x_{t+0.5}$)

Problem: $\min_x f(x)$ (optionally, s.t. $x \in F$)
Strict convexity

• Def’n: f is $\lambda$-\textit{strictly convex} if, for $g \in \partial f(x)$, 

\begin{itemize}
  \item note: $(\lambda = 0) \iff$
\end{itemize}
A useful inequality
Convergence preview

• For strictly convex $f(x)$ (i.e., $\lambda>0$):
  ‣ set $\eta_t =$
  ‣ $f(x_t) - f(x^*) =$

• For non-strictly convex $f(x)$ (i.e., $\lambda=0$):
  ‣ set $\eta_t =$
  ‣ $f(x_t) - f(x^*) =$

• To get accuracy $\epsilon$:
  ‣ $\lambda>0$: $T =$
  ‣ $\lambda=0$: $T =$
How to prove convergence?

• Might hope \( f(x_{t+1}) < f(x_t) \)
  ‣ or \( E(f(x_{t+1})) < E(f(x_t)) \) for stochastic version

• Not true!
  ‣ corners:

  ‣ randomness:
How to prove?

• Corners: we will show that we decrease *Euclidean distance to minimizer* instead

• Randomness:
Potential function

- $Q(x) = \ldots$
- $Q(x_{t+1}) = \ldots$
Projected subgradient

- $Q(x_{t+0.5}) - Q(x_t) = \eta_t (x^* - x_t)^T g_t + \eta_t^2 \|g_t\|^2/2$
- $Q(x_{t+1}) - Q(x_{t+0.5})$

- So:
Main bound

• Have \( Q(x_{t+1}) - Q(x_t) \leq \eta_t(x^* - x_t)^Tg_t + \eta_t^2||g_t||^2/2 \)
  \( \triangleright \) \( \eta_t(x^* - x_t)^Tg_t \geq Q(x_{t+1}) - Q(x_t) - \eta_t^2||g_t||^2/2 \)

• At step \( t \), compare \( f_t(x^*) \) to \( f_t(x_t) \)

• Sum over \( t = 1..T: \)
Deterministic, \( \lambda > 0 \) case

- \( \sum_{t=1}^{T} f_t(x^*) \geq \sum_{t=1}^{T} f_t(x_t) - Q(x_1)/\eta_1 \)
  
  \( + \sum_{t=2}^{T} Q(x_t)[1/\eta_t - 1/\eta_{t+1} + \lambda] \)
  
  \( - \sum_{t=1}^{T} \eta_t ||g_t||^2/2 \)

- Suppose \( ||g_t||^2/2 \leq G \)

- Let \( \eta_t = \)
Deterministic, $\lambda=0$ case

- $\sum_{t=1}^{T} f_t(x^*) \geq$
  - $\sum_{t=1}^{T} f_t(x_t) - Q(x_1)/\eta_1$
  - $\sum_{t=2}^{T} Q(x_t)[1/\eta_t - 1/\eta_{t+1}]$
  - $\sum_{t=1}^{T} \eta_t G$

- Let $\eta_t =$