First-order methods
Convexity

10-725 Optimization
Geoff Gordon
Ryan Tibshirani
Gradient descent
When do we stop?

- Using holdout set, if \( f(x) = E(f_i(x) \mid i \sim P) \)

\[
\frac{1}{N} \sum_{i=1}^{N} (a \cdot x - 5)^2
\]

- Using convergence bounds (later)
  - usual form is:
    - \( K_f (f(x_0) - f(x^*)) \) [some fn of \( 1/\epsilon \)]
    - need estimates of first two terms
- For \( f(x^*) \), duality (later); for \( K_f \), properties of \( f \):
  - convex? strongly convex? Lipschitz?
Convex sets

For all $x, y \in C$, for all $t \in [0, 1]$: $$tx + (1-t)y \in C$$
Examples

\[ \| x \|_p \leq 1 \quad \text{for} \quad p \geq 1 \]

For \( p < 1 \), the unit ball is a star-shaped region.

\[ \| x \|_p = \sqrt[p]{\sum |x_i|^p} \]
Boundaries

- **x on boundary of C** (\(\partial C\)) if:
  \[ B_p(x, \varepsilon) \cap C \neq \emptyset \quad \text{and} \quad B_p(x, \varepsilon) \cap C^c \neq \emptyset \]

- **x in interior of C** if:
  \[ B_p(x, \varepsilon) \subseteq C \quad \text{for small}\ \varepsilon \]

- **x in relative interior** (rel int C) if:
  \[ \text{subspace S \subset C restrict to } \text{int C restrict} \]

- **C closed if**:
  \[ \bar{C} \subseteq C \]

- **C open if**:
  \[ \emptyset \subseteq \bar{C} \]

- **C compact if**:
  \[ \text{closed bounded} \]
Convex hull

Any closed, convex $C$

\[ \text{conv} (X) \]

$|X| = \infty$
Dual representation

\[ C = \bigwedge \{ x : a_i x + b_i \leq 0 \} \]

Any closed convex \( C \) = this
Supporting hyperplane thm

- For any point $x_0$ on boundary of convex $C$:
  - exist $(w, b)$ with
    - $x_0 \cdot w = b$
    - $x \cdot w \leq b \quad \forall x \in C$
Supporting hyperplane exs

$x^2 + y^2 < 1$

$|x| + |y| \leq 1$
Separating hyperplane thm

- For any convex $C$ and $D$ with $C \cap D = \emptyset$
  - exist $(w, b)$ with
    - $w \cdot x \leq b \quad x \in C$
    - $w \cdot x \geq b \quad x \in D$

- If both $C, D$ are closed, and at least one compact:
  - strict $w \cdot x < b$
  - $w \cdot x \geq b$
Separating hyperplane exs

halfspace & complement

$y \geq \frac{1}{x}$

$y < 0$
Proving a set convex

- Use definition directly
- Represent as convex hull or \( \cap \) of halfspaces
- Supporting hyperplane partial converse
  - \( C \) closed, nonempty interior, has supporting hyperplane at all boundary points \( \Rightarrow \) \( C \) convex
- Build \( C \) up from simpler sets using convexity-preserving operations
Convexity-preserving set ops

- Translation \( \{ x + b \mid x \in C \} = C + b \)
- Scaling \( \{ \lambda x \mid x \in C \} = \lambda C \)
- Affine fn \( \{ A x + b \mid x \in C \} = AC + b \)
  - projection (e.g., dropping coords)
- Intersection \( C \cap D \)
- Set sum \( C + D = \{ x + y \mid x \in C, y \in D \} \)
- Perspective \( (x, y, 1) \in C \Rightarrow (x/z, y/z) \)
Ex: symmetric PSD matrices

- Two proofs that \{A \mid A = A^T, A \succeq 0\} is convex
  - \(x^T (tA + (1-t)B) x = t x^T A x + (1-t) x^T B x\)
  - \(x^T A x = \sum_{i \leq j} A_{ij} x_i x_j \geq 0\)
Ex: conditionals

- Given a convex set of dist’ns $P(x_{1:7})$,
  - $P(x_{1:5} \mid x_{6:7}) = \frac{P(x_{1:7})}{P(x_{6:7})}$
  - numerator:
  - denominator:

- Convex? $\checkmark$

$x \in \{0,1\}^7$
Epigraph

\[ \{(x, t) \mid t \geq f(x)\} \]

Geoff Gordon—10-725 Optimization—Fall 2012
Domain

- $\text{dom } f = \{ x \mid f(x) \text{ exists, } f(x) < \infty \}$
- $\text{dom } 1/x = \mathbb{R} - \{0\}$
- $\text{dom } \ln(x) = x > 0$
Convex functions

\[ f \text{ convex } \iff \text{epi } f \text{ convex} \]

\[ \text{dom } f \text{ convex} \]
\[ f(x) + (1-t)f(y) \leq f(tx + (1-t)y) \quad \forall x, y \in \text{dom } f \]
Relating convex sets and fns

- $f(x)$ convex $\Rightarrow \{ x | f(x) \leq c \}$ convex

- Converse?

$$f(x) = \sqrt{|x|}$$

quasi convex
Proving a function convex

- Use definition directly
- Prove that epigraph is convex via set methods
  - e.g., supporting hyperplanes: for all $x, y,$
  - this is first-order convexity condition for fns
- 2nd order: $\frac{d^2}{dx^2} f(x) \geq 0$
- Construct $f$ from simpler convex fns using convexity-preserving ops
Convexity-preserving fn ops

- Nonnegative weighted sum
  \[ \langle f + bg \rangle(x) = a f(x) + b g(x) \]
- Pointwise max/sup
  \[ g(x) = \sup_{i \in I} f_i(x) \]
- Composition w/ affine
  \[ f(Ax + b) \]
- Composition w/ monotone convex
  \[ g(f(x)) \]
- Perspective
  \[ g(x; t) = f(x/t) \quad t > 0 \]

- \( f(x, y) \) convex in \((x, y)\), set \( C \) convex:
  - \( g(y) = \min_{x \in C} f(x, y) \) is convex if \( g(y) > -\infty \)
Example: \( f(x) = |x| \)

\[
|x| \geq x \\
|x| \geq -x
\]

\[
\text{if } x + (1-t)y \geq 0 \\
|x + (1-t)y| = x + (1-t)y \leq |x| + (1-t)|y|
\]

\[
\text{or} \\
|x + (1-t)y| = -x - (1-t)y \\
\leq |x| + (1-t)|y|
\]
In 2 or more dimensions

- All the above, but for 2nd order:
  \[ H = \frac{d^2}{dx^2} f(x) \]
  \[ H \geq 0 \]
  \[ e^{i\pi/H} \]

- Or: reduce to 1D
  \[ f(x_0 + td) \]
Ex: structured classifier

Convex
Ex: structured classifier

\[
\begin{align*}
  \mathbf{x}_i & \quad \text{pixels of char } i \\
  \varphi_j(\mathbf{x}_i) & \quad \text{feature of a char} \\
  \psi_{ijk}(\mathbf{x}_i, \mathbf{y}_i) & \quad \psi_j(\mathbf{x}_i) \mathbb{1}(\mathbf{y}_i = \mathbf{a}) \quad \mathbb{1} (\mathbf{y}_{i-1} = \mathbf{e}) \\
  \chi_{ikl}(\mathbf{y}_i, \mathbf{y}_{i+1}) & \quad \mathbb{1} (\mathbf{y}_i = \mathbf{a}) \mathbb{1} (\mathbf{y}_{i+1} = \mathbf{e}) \\
  L(\mathbf{x}, \mathbf{y}; \mathbf{v}, \mathbf{w}) & = \sum_{i=1}^n \psi_{ijk} \mathbb{1}(\mathbf{y}_i = \mathbf{a}) \mathbb{1} (\mathbf{y}_{i-1} = \mathbf{e}) + \sum_{i=1}^{n-1} \chi_{ikl} (\mathbf{y}_i, \mathbf{y}_{i+1}) \mathbb{1} (\mathbf{y}_i = \mathbf{a}) \mathbb{1} (\mathbf{y}_{i+1} = \mathbf{e}) \\
  \text{Classifier:} & \quad \hat{y} = \underset{y}{\arg\max} L(\mathbf{x}, y; \mathbf{v}, \mathbf{w})
\end{align*}
\]
Learning structured classifier

- Get it right if: \( L(x, y; u, w) > L(x, y', u, w) \)
- So, want: \( L(x, y; u, w) \geq \max_y \left( L(x, y; u, w) + \pi(y, y') \right) \)
- Where \( \pi(y, y') = \{
  0 & y = y', \\
  > 0 & y \neq y'.
\}\)
- RHS: convex in \( y, w \)
- RHS - LHS: convex
- Train: lots of pairs \((x^t, y^t)\)