Directional Derivatives and Optimality

David Rosenberg

New York University

February 4, 2016

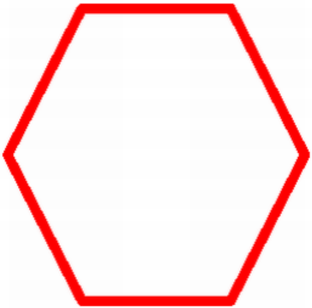
Convex Sets and Functions

Convex Sets

Definition

A set C is **convex** if for any $x_1, x_2 \in C$ and any θ with $0 \leqslant \theta \leqslant 1$ we have

$$\theta x_1 + (1-\theta)x_2 \in C.$$



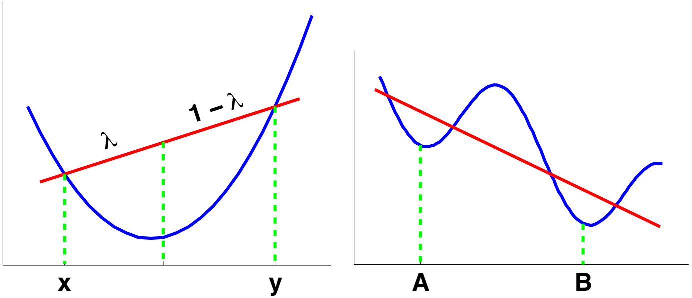
KPM Fig. 7.4

Convex and Concave Functions

Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is **convex** if $\mathbf{dom}\, f$ is a convex set and if for all $x, y \in \mathbf{dom}\, f$, and $0 \leqslant \theta \leqslant 1$, we have

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y).$$



KPM Fig. 7.5

Directional Derivatives and Minima

Directional Derivatives

Definition

A **[one-sided] directional derivative** of f at x in the direction v is

$$f'(x;v) = \lim_{h \downarrow 0} \frac{f(x+hv) - f(x)}{h},$$

and it can be $\pm\infty$ (e.g. for discontinuous functions).

Boyd EE364b: Subgradients Slides

Directional Derivatives

Definition

A **[one-sided] directional derivative** of f at x in the direction v is

$$f'(x;v) = \lim_{h \downarrow 0} \frac{f(x+hv) - f(x)}{h},$$

and it can be $\pm\infty$ (e.g. for discontinuous functions).

- If f is convex and finite near x, then $f'(x;v)$ exists.

Boyd EE364b: Subgradients Slides

Directional Derivatives

Definition

A **[one-sided] directional derivative** of f at x in the direction v is

$$f'(x;v) = \lim_{h \downarrow 0} \frac{f(x+hv) - f(x)}{h},$$

and it can be $\pm\infty$ (e.g. for discontinuous functions).

- If f is convex and finite near x, then $f'(x;v)$ exists.
- f is **differentiable** at x iff for some $g(= \nabla f(x))$ and all v,

$$f'(x;v) = g^T v.$$

Boyd EE364b: Subgradients Slides

Descent Directions and Optimality

Definition

v is a **descent direction** for f at x if $f'(x;v) < 0$.

Boyd EE364b: Subgradients Slides

Descent Directions and Optimality

Definition

v is a **descent direction** for f at x if $f'(x;v) < 0$.

- For differentiable f, if $\nabla f(x) \neq 0$, then $\delta x = -\nabla f(x)$ is a descent direction.

Boyd EE364b: Subgradients Slides

Descent Directions and Optimality

Definition

v is a **descent direction** for f at x if $f'(x;v) < 0$.

- For differentiable f, if $\nabla f(x) \neq 0$, then $\delta x = -\nabla f(x)$ is a descent direction.
- We have a nice characterization for a minimum in terms of directional derivative:

Boyd EE364b: Subgradients Slides

Descent Directions and Optimality

Definition

v is a **descent direction** for f at x if $f'(x;v) < 0$.

- For differentiable f, if $\nabla f(x) \neq 0$, then $\delta x = -\nabla f(x)$ is a descent direction.
- We have a nice characterization for a minimum in terms of directional derivative:

Theorem

If f is convex and finite near x, then either

- *x minimizes f, or*
- *there is a descent direction for f at x.*

Boyd EE364b: Subgradients Slides

λ_{max} for Lasso

- Lasso objective

$$J_\lambda(w) = \sum_{i=1}^{n} \left(w^T x_i - y_i\right)^2 + \lambda \|w\|_1$$

λ_{max} for Lasso

- Lasso objective

$$J_{\lambda}(w) = \sum_{i=1}^{n} \left(w^T x_i - y_i\right)^2 + \lambda \|w\|_1$$

- Is there a λ_{max} such that $\lambda \geqslant \lambda_{max}$ implies $\arg\min_w J_{\lambda}(w) = 0$?

λ_{max} for Lasso

- Lasso objective

$$J_\lambda(w) = \sum_{i=1}^{n} \left(w^T x_i - y_i\right)^2 + \lambda \|w\|_1$$

- Is there a λ_{max} such that $\lambda \geqslant \lambda_{max}$ implies $\arg\min_w J_\lambda(w) = 0$?
- Suppose yes.

$\lambda_{\max}$ for Lasso

- Lasso objective

$$J_\lambda(w) = \sum_{i=1}^{n} \left(w^T x_i - y_i\right)^2 + \lambda \|w\|_1$$

- Is there a $\lambda_{\max}$ such that $\lambda \geqslant \lambda_{\max}$ implies $\arg\min_w J_\lambda(w) = 0$?
- Suppose yes.
- Then $w = 0$ is a minimum of $J_\lambda(w)$.

$\lambda_{\max}$ for Lasso

- Lasso objective

$$J_\lambda(w) = \sum_{i=1}^{n} \left(w^T x_i - y_i\right)^2 + \lambda \|w\|_1$$

- Is there a $\lambda_{\max}$ such that $\lambda \geqslant \lambda_{\max}$ implies $\arg\min_w J_\lambda(w) = 0$?
- Suppose yes.
- Then $w = 0$ is a minimum of $J_\lambda(w)$.
- Let's see what that means in terms of our directional derivative characterization.

Directional Derivative for Lasso

- Consider a step direction v. For convenience, take v s.t. $|v| = 1$.

Directional Derivative for Lasso

- Consider a step direction v. For convenience, take v s.t. $|v| = 1$.
- Then directional derivative at $w = 0$ in direction v is

$$J'_\lambda(0; v) = \lim_{h \downarrow 0} \frac{J(hv) - J(0)}{h}.$$

Directional Derivative for Lasso

- Consider a step direction v. For convenience, take v s.t. $|v| = 1$.
- Then directional derivative at $w = 0$ in direction v is

$$J'_\lambda(0; v) = \lim_{h \downarrow 0} \frac{J(hv) - J(0)}{h}.$$

- For $w = 0$ to be a minimizer, need to have $J'_\lambda(0; v) \geqslant 0$ for every direction v.

Directional Derivative for Lasso

- Consider a step direction v. For convenience, take v s.t. $|v| = 1$.
- Then directional derivative at $w = 0$ in direction v is
$$J_\lambda'(0; v) = \lim_{h \downarrow 0} \frac{J(hv) - J(0)}{h}.$$
- For $w = 0$ to be a minimizer, need to have $J_\lambda'(0; v) \geqslant 0$ for every direction v.
- Can find $\lambda_{\max}$ by finding conditions on λ for this to be the case.