

# Subgradient Descent

David Rosenberg

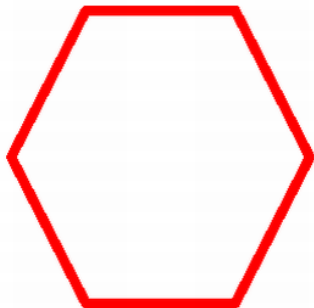
New York University

October 29, 2016

# Convex Sets

## Definition

A set  $C$  is **convex** if the line segment between any two points in  $C$  lies in  $C$ .

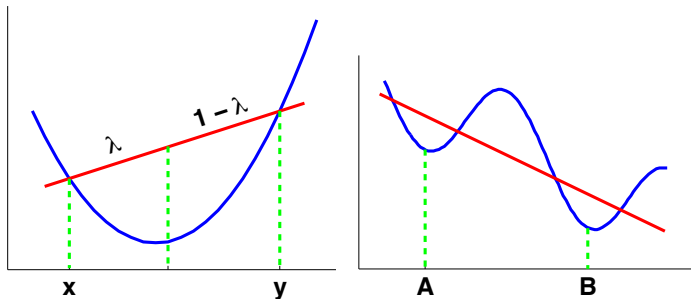


KPM Fig. 7.4

# Convex and Concave Functions

## Definition

A function  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is **convex** if the line segment connecting any two points on the graph of  $f$  lies above the graph.  $f$  is **concave** if  $-f$  is convex.

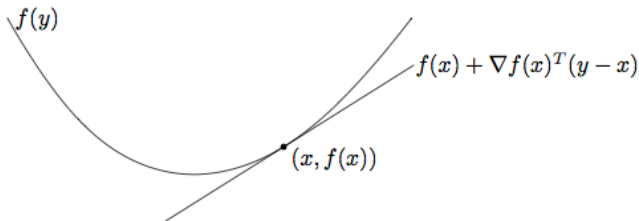


KPM Fig. 7.5

# First-Order Approximation

- Suppose  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is **differentiable**.
- Predict  $f(y)$  given  $f(x)$  and  $\nabla f(x)$ ?
- Linear (i.e. “**first order**”) approximation:

$$f(y) \approx f(x) + \nabla f(x)^T (y - x)$$



Boyd & Vandenberghe Fig. 3.2

# First-Order Condition for Convex, Differentiable Function

- Suppose  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is **convex** and **differentiable**.
- Then for any  $x, y \in \mathbf{R}^n$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

- The linear approximation to  $f$  at  $x$  is a **global underestimator** of  $f$ :

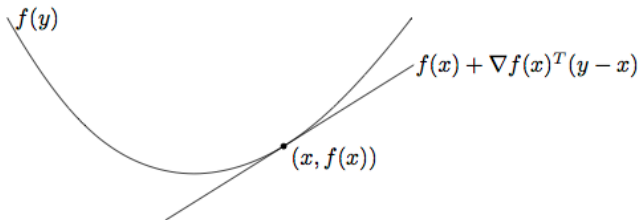


Figure from Boyd & Vandenberghe Fig. 3.2; Proof in Section 3.1.3

# First-Order Condition for Convex, Differentiable Function

- Suppose  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is **convex** and **differentiable**
- Then for any  $x, y \in \mathbf{R}^n$

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

## Corollary

*If  $\nabla f(x) = 0$  then  $x$  is a global minimizer of  $f$ .*

For convex functions, **local information gives global information.**

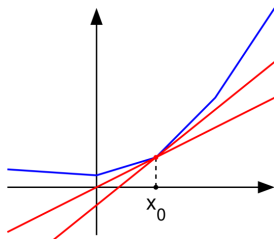
# Subgradients

## Definition

A vector  $g \in \mathbf{R}^n$  is a **subgradient** of  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  at  $x$  if for all  $z$ ,

$$f(z) \geq f(x) + g^T(z - x).$$

$g$  is a subgradient iff  $f(x) + g^T(z - x)$  is a global underestimator of  $f$



Blue is a graph of  $f(x)$ .

Each red line is a lower bound:  $x \mapsto f(x_0) + g^T(x - x_0)$

# Subdifferential

## Definitions

- $f$  is **subdifferentiable** at  $x$  if  $\exists$  at least one subgradient at  $x$ .
- The set of all subgradients at  $x$  is called the **subdifferential**:  $\partial f(x)$

## Basic Facts

- $f$  is convex and differentiable  $\implies \partial f(x) = \{\nabla f(x)\}$ .
- Any point  $x$ , there can be 0, 1, or infinitely many subgradients.
- $\partial f(x) = \emptyset \implies f$  is not convex.



# Global Optimality Condition

## Definition

A vector  $g \in \mathbf{R}^n$  is a **subgradient** of  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  at  $x$  if for all  $z$ ,

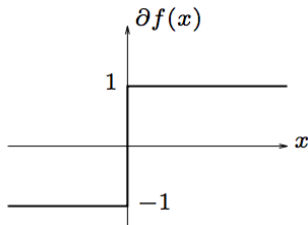
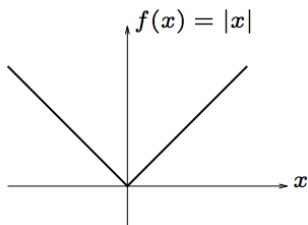
$$f(z) \geq f(x) + g^T(z - x).$$

## Corollary

If  $0 \in \partial f(x)$ , then  $x$  is a **global minimizer** of  $f$ .

# Subdifferential of Absolute Value

- Consider  $f(x) = |x|$



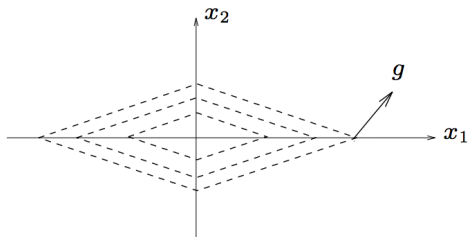
- Plot on right shows  $\{(x, g) \mid x \in \mathbf{R}, g \in \partial f(x)\}$

# Descent Directions

- For differentiable  $f$ ,  $-\nabla f(x)$  is a descent direction.
- What can we do for non-differentiable  $f$ ?
- Can we use  $-g$  as a step, for some  $g \in \partial f(x)$ ?
- Is  $-g$  a descent direction?

# Subgradient Not a Descent Direction

$$f(x) = |x_1| + 2|x_2|$$



- Diamonds are level sets of  $f(x)$ . ( $f$  minimized at origin)
- $g$  is a subgradient at the point it's drawn.
- Moving in  $-g$  direction increases the function.

---

Figure from Boyd EE364b: Subgradients Slides,  
[http://web.stanford.edu/class/ee364b/lectures/subgradients\\_slides.pdf](http://web.stanford.edu/class/ee364b/lectures/subgradients_slides.pdf), slide 28.

# Subgradient Descent

- Suppose  $f$  is convex, and we start optimizing at  $x_0$ .
- Repeat
  - Step in a negative subgradient direction:

$$x = x_0 - tg,$$

where  $t > 0$  is the step size and  $g \in \partial f(x_0)$ .

$-g$  not a descent direction – can this work?

# Subgradient Gets Us Closer To Minimizer

## Theorem

Suppose  $f$  is convex.

- Let  $x = x_0 - tg$ , for  $g \in \partial f(x_0)$ .
- Let  $z$  be any point for which  $f(z) < f(x_0)$ .
- Then for small enough  $t > 0$ ,

$$\|x - z\|_2 < \|x_0 - z\|_2.$$

- Apply this with  $z = x^* \in \arg \min_x f(x)$ .

$\implies$  **Negative subgradient step gets us closer to minimizer.**

# Subgradient Gets Us Closer To Minimizer (Proof)

- Let  $x = x_0 - tg$ , for  $g \in \partial f(x_0)$  and  $t > 0$ .
- Let  $z$  be any point for which  $f(z) < f(x_0)$ .
- Then

$$\begin{aligned} \|x - z\|_2^2 &= \|x_0 - tg - z\|_2^2 \\ &= \|x_0 - z\|_2^2 - 2tg^T(x_0 - z) + t^2\|g\|_2^2 \\ &\leq \|x_0 - z\|_2^2 - 2t[f(x_0) - f(z)] + t^2\|g\|_2^2 \end{aligned}$$

- Consider  $-2t[f(x_0) - f(z)] + t^2\|g\|_2^2$ .
  - It's a convex quadratic (facing upwards).
  - Has zeros at  $t = 0$  and  $t = 2(f(x_0) - f(z))/\|g\|_2^2 > 0$ .
  - Therefore, it's negative for any

$$t \in \left(0, \frac{2(f(x_0) - f(z))}{\|g\|_2^2}\right).$$

# Convergence Theorem for Fixed Step Size

Assume  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is convex and

- $f$  is Lipschitz continuous with constant  $G > 0$ :

$$|f(x) - f(y)| \leq G\|x - y\| \text{ for all } x, y$$

## Theorem

*For fixed step size  $t$ , subgradient method satisfies:*

$$\lim_{k \rightarrow \infty} f(x_{best}^{(k)}) \leq f(x^*) + G^2 t / 2$$



# Convergence Theorems for Decreasing Step Sizes

Assume  $f : \mathbf{R}^n \rightarrow \mathbf{R}$  is convex and

- $f$  is Lipschitz continuous with constant  $G > 0$ :

$$|f(x) - f(y)| \leq G\|x - y\| \text{ for all } x, y$$

## Theorem

*For step size respecting Robbins-Monro conditions,*

$$\lim_{k \rightarrow \infty} f(x_{best}^{(k)}) \leq f(x^*)$$