

## 5. Subgradient method

- subgradient method
- convergence analysis
- optimal step size when  $f^*$  is known
- alternating projections
- optimality

# Subgradient method

to minimize a nondifferentiable convex function  $f$ : choose  $x^{(0)}$  and repeat

$$x^{(k)} = x^{(k-1)} - t_k g^{(k-1)}, \quad k = 1, 2, \dots$$

$g^{(k-1)}$  is any subgradient of  $f$  at  $x^{(k-1)}$

## Step size rules

- fixed step:  $t_k$  constant
- fixed length:  $t_k \|g^{(k-1)}\|_2 = \|x^{(k)} - x^{(k-1)}\|_2$  is constant
- diminishing:  $t_k \rightarrow 0$ ,  $\sum_{k=1}^{\infty} t_k = \infty$

# Assumptions

- $f$  has finite optimal value  $f^*$ , minimizer  $x^*$
- $f$  is convex,  $\text{dom } f = \mathbf{R}^n$
- $f$  is Lipschitz continuous with constant  $G > 0$ :

$$|f(x) - f(y)| \leq G\|x - y\|_2 \quad \forall x, y$$

this is equivalent to  $\|g\|_2 \leq G$  for all  $x$  and  $g \in \partial f(x)$  (see next page)

*Proof.*

- assume  $\|g\|_2 \leq G$  for all subgradients; choose  $g_y \in \partial f(y)$ ,  $g_x \in \partial f(x)$ :

$$g_x^T(x - y) \geq f(x) - f(y) \geq g_y^T(x - y)$$

by the Cauchy-Schwarz inequality

$$G\|x - y\|_2 \geq f(x) - f(y) \geq -G\|x - y\|_2$$

- assume  $\|g\|_2 > G$  for some  $g \in \partial f(x)$ ; take  $y = x + g/\|g\|_2$ :

$$\begin{aligned} f(y) &\geq f(x) + g^T(y - x) \\ &= f(x) + \|g\|_2 \\ &> f(x) + G \end{aligned}$$

# Analysis

- the subgradient method is not a descent method
- the key quantity in the analysis is the distance to the optimal set

with  $x^+ = x^{(i)}$ ,  $x = x^{(i-1)}$ ,  $g = g^{(i-1)}$ ,  $t = t_i$ :

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

combine inequalities for  $i = 1, \dots, k$ , and define  $f_{\text{best}}^{(k)} = \min_{0 \leq i < k} f(x^{(i)})$ :

$$\begin{aligned}2\left(\sum_{i=1}^k t_i\right)(f_{\text{best}}^{(k)} - f^*) &\leq \|x^{(0)} - x^*\|_2^2 - \|x^{(k)} - x^*\|_2^2 + \sum_{i=1}^k t_i^2 \|g^{(i-1)}\|_2^2 \\ &\leq \|x^{(0)} - x^*\|_2^2 + \sum_{i=1}^k t_i^2 \|g^{(i-1)}\|_2^2\end{aligned}$$

**Fixed step size:**  $t_i = t$

$$f_{\text{best}}^{(k)} - f^* \leq \frac{\|x^{(0)} - x^*\|_2^2}{2kt} + \frac{G^2 t}{2}$$

- does not guarantee convergence of  $f_{\text{best}}^{(k)}$
- for large  $k$ ,  $f_{\text{best}}^{(k)}$  is approximately  $G^2 t/2$ -suboptimal

**Fixed step length:**  $t_i = s/\|g^{(i-1)}\|_2$

$$f_{\text{best}}^{(k)} - f^* \leq \frac{G\|x^{(0)} - x^*\|_2^2}{2ks} + \frac{Gs}{2}$$

- does not guarantee convergence of  $f_{\text{best}}^{(k)}$
- for large  $k$ ,  $f_{\text{best}}^{(k)}$  is approximately  $Gs/2$ -suboptimal

**Diminishing step size:**  $t_i \rightarrow 0$ ,  $\sum_{i=1}^{\infty} t_i = \infty$

$$f_{\text{best}}^{(k)} - f^* \leq \frac{\|x^{(0)} - x^*\|_2^2 + G^2 \sum_{i=1}^k t_i^2}{2 \sum_{i=1}^k t_i}$$

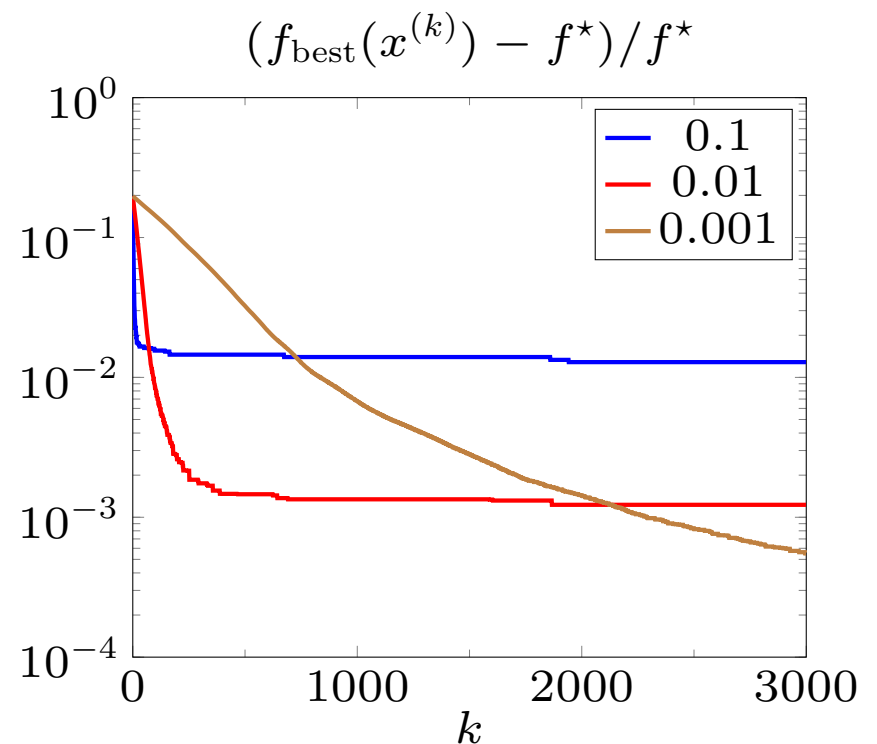
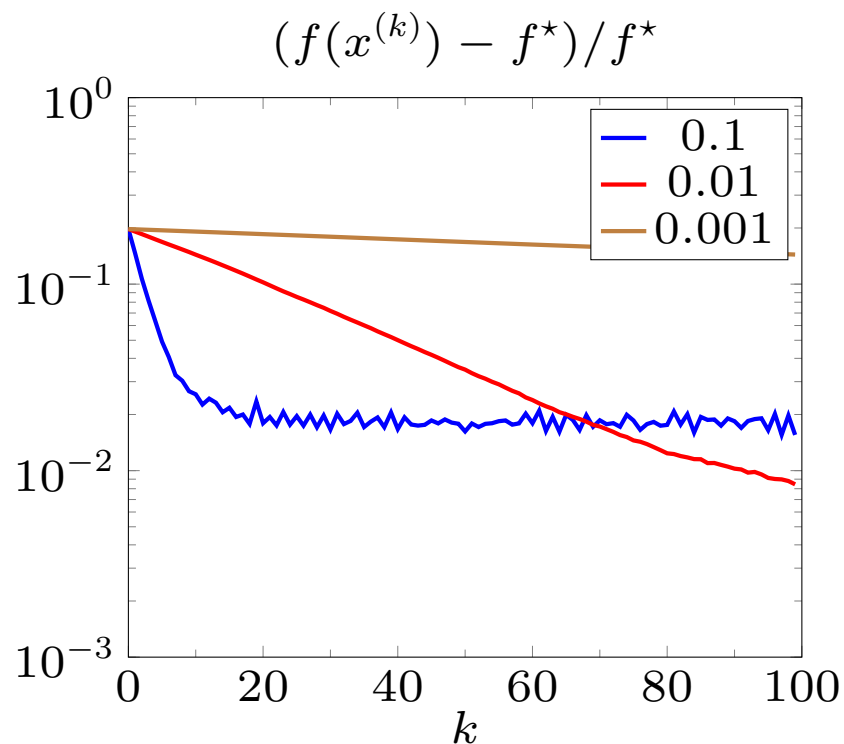
can show that  $(\sum_{i=1}^k t_i^2) / (\sum_{i=1}^k t_i) \rightarrow 0$ ; hence,  $f_{\text{best}}^{(k)}$  converges to  $f^*$

# Example: 1-norm minimization

$$\text{minimize } \|Ax - b\|_1$$

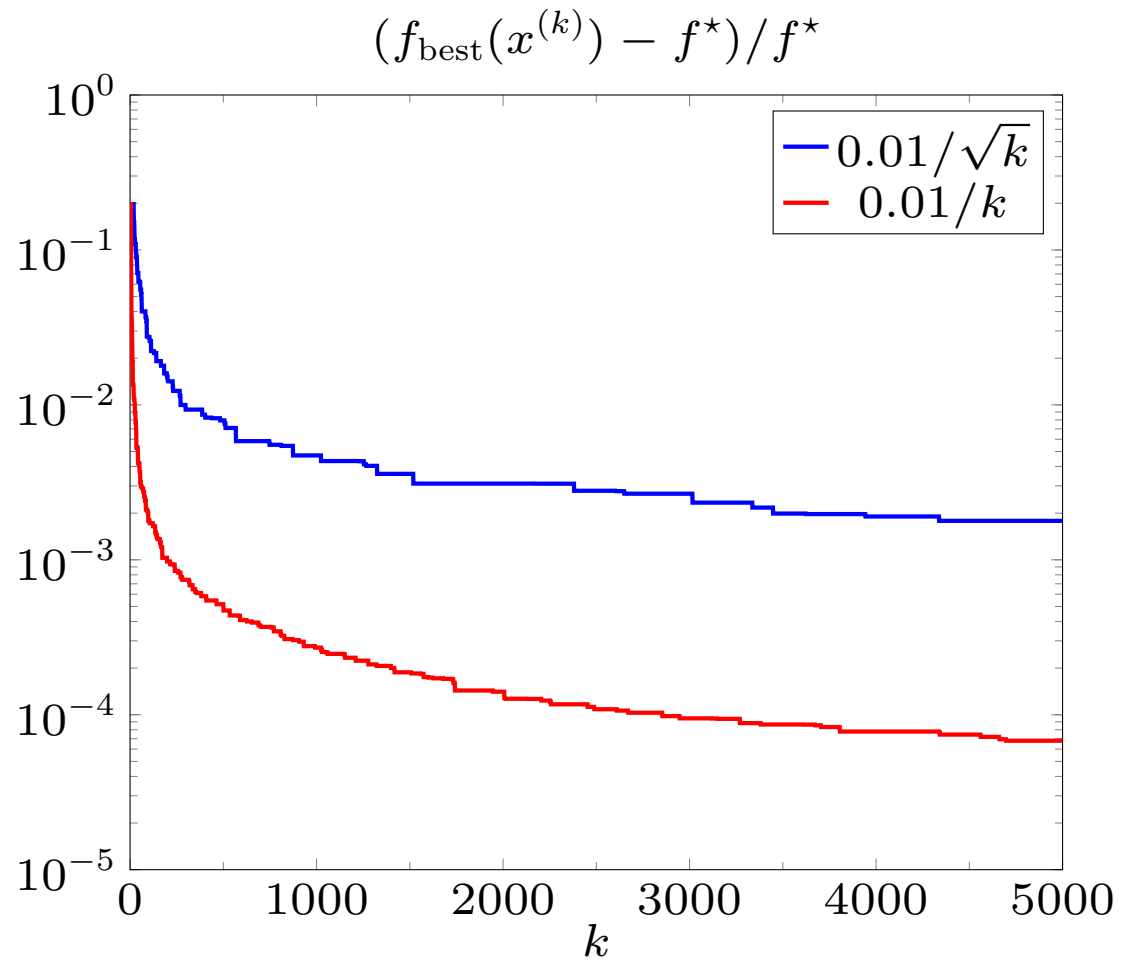
- subgradient is given by  $A^T \text{sign}(Ax - b)$
- example with  $A \in \mathbf{R}^{500 \times 100}$ ,  $b \in \mathbf{R}^{500}$

**Fixed steplength**  $t_k = s/\|g^{(k-1)}\|_2$  for  $s = 0.1, 0.01, 0.001$





**Diminishing step size:**  $t_k = 0.01/\sqrt{k}$  and  $t_k = 0.01/k$



# Optimal step size for fixed number of iterations

from page 5-5: if  $s_i = t_i \|g^{(i-1)}\|_2$  and  $\|x^{(0)} - x^*\|_2 \leq R$ :

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^k s_i^2}{2 \sum_{i=1}^k s_i / G}$$

- for given  $k$ , bound is minimized by fixed step length  $s_i = s = R/\sqrt{k}$
- resulting bound after  $k$  steps is

$$f_{\text{best}}^{(k)} - f^* \leq \frac{GR}{\sqrt{k}}$$

- guarantees accuracy  $f_{\text{best}}^{(k)} - f^* \leq \epsilon$  in  $k = O(1/\epsilon^2)$  iterations

## Optimal step size when $f^*$ is known

- right-hand side in first inequality of page 5-5 is minimized by

$$t_i = \frac{f(x^{(i-1)}) - f^*}{\|g^{(i-1)}\|_2^2}$$

- optimized bound is

$$\frac{(f(x^{(i-1)}) - f^*)^2}{\|g^{(i-1)}\|_2^2} \leq \|x^{(i-1)} - x^*\|_2^2 - \|x^{(i)} - x^*\|_2^2$$

- applying recursively (with  $\|x^{(0)} - x^*\|_2 \leq R$  and  $\|g^{(i)}\|_2 \leq G$ ) gives

$$f_{\text{best}}^{(k)} - f^* \leq \frac{GR}{\sqrt{k}}$$

## Exercise: find point in intersection of convex sets

find a point in the intersection of  $m$  closed convex sets  $C_1, \dots, C_m$ :

$$\text{minimize } f(x) = \max \{f_1(x), \dots, f_m(x)\}$$

where  $f_j(x) = \inf_{y \in C_j} \|x - y\|_2$  is Euclidean distance of  $x$  to  $C_j$

- $f^* = 0$  if the intersection is nonempty
- (from p. 4-14):  $g \in \partial f(\hat{x})$  if  $g \in \partial f_j(\hat{x})$  and  $C_j$  is farthest set from  $\hat{x}$
- (from p. 4-20) subgradient  $g \in \partial f_j(\hat{x})$  follows from projection  $P_j(\hat{x})$  on  $C_j$ :

$$g = 0 \quad (\text{if } \hat{x} \in C_j), \quad g = \frac{1}{\|\hat{x} - P_j(\hat{x})\|_2} (\hat{x} - P_j(\hat{x})) \quad (\text{if } \hat{x} \notin C_j)$$

note that  $\|g\|_2 = 1$  if  $\hat{x} \notin C_j$

## Subgradient method

- optimal step size (page 5-11) for  $f^* = 0$  and  $\|g^{(i-1)}\|_2 = 1$  is  $t_i = f(x^{(i-1)})$
- at iteration  $k$ , find farthest set  $C_j$  (with  $f(x^{(k-1)}) = f_j(x^{(k-1)})$ ), and take

$$\begin{aligned}x^{(k)} &= x^{(k-1)} - \frac{f(x^{(k-1)})}{f_j(x^{(k-1)})} (x^{(k-1)} - P_j(x^{(k-1)})) \\ &= P_j(x^{(k-1)})\end{aligned}$$

at each step, we project the current point onto the farthest set

- a version of the *alternating projections* algorithm
- for  $m = 2$ , projections alternate onto one set, then the other
- later, we will see faster versions of this that are almost as simple

# Optimality of the subgradient method

can the  $f_{\text{best}}^{(k)} - f^* \leq GR/\sqrt{k}$  bound on page 5-10 be improved?

## Problem class

- $f$  is convex, with a minimizer  $x^*$
- we know a starting point  $x^{(0)}$  with  $\|x^{(0)} - x^*\|_2 \leq R$
- we know the Lipschitz constant  $G$  of  $f$  on  $\{x \mid \|x - x^{(0)}\|_2 \leq R\}$
- $f$  is defined by an oracle: given  $x$ , oracle returns  $f(x)$  and a subgradient

**Algorithm class:**  $k$  iterations of any method that chooses  $x^{(i)}$  in

$$x^{(0)} + \text{span}\{g^{(0)}, g^{(1)}, \dots, g^{(i-1)}\}$$

## Test problem and oracle

$$f(x) = \max_{i=1,\dots,k} x_i + \frac{1}{2}\|x\|_2^2, \quad x^{(0)} = 0$$

- solution:  $x^* = -\frac{1}{k}(\underbrace{1, \dots, 1}_k, \underbrace{0, \dots, 0}_{n-k})$  and  $f^* = -\frac{1}{2k}$
- $R = \|x^{(0)} - x^*\|_2 = 1/\sqrt{k}$  and  $G = 1 + 1/\sqrt{k}$
- oracle returns subgradient  $e_{\hat{j}} + x$  where  $\hat{j} = \min\{j \mid x_j = \max_{i=1,\dots,k} x_i\}$

**Iteration:** for  $i = 0, \dots, k - 1$ , entries  $x_{i+1}^{(i)}, \dots, x_k^{(i)}$  are zero; therefore

$$f_{\text{best}}^{(k)} - f^* = \min_{i < k} f(x^{(i)}) - f^* \geq -f^* = \frac{GR}{2(1 + \sqrt{k})}$$

**Conclusion:**  $O(1/\sqrt{k})$  bound cannot be improved

# Summary: subgradient method

- handles general nondifferentiable convex problem
- often leads to very simple algorithms
- convergence can be very slow
- no good stopping criterion
- theoretical complexity:  $O(1/\epsilon^2)$  iterations to find  $\epsilon$ -suboptimal point
- an ‘optimal’ 1st-order method:  $O(1/\epsilon^2)$  bound cannot be improved



# References

- S. Boyd, lecture notes and slides for EE364b, Convex Optimization II
- Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (2004)  
§3.2.1 with the example on page 5-15 of this lecture