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

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

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

**step size rules**

- fixed step: $t_k$ constant
- fixed length: $t_k \| g^{(k-1)} \|_2$ constant (i.e., $\| x^{(k)} - x^{(k-1)} \|_2$ constant)
- diminishing: $t_k \to 0$, $\sum_{k=1}^{\infty} t_k = \infty$
Assumptions

• $f$ has finite optimal value $f^*$, minimizer $x^*$

• $f$ is convex, $\text{dom } f = \mathbb{R}^n$

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

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

this is equivalent to

\[
\|g\|_2 \leq G \quad \forall g \in \partial f(x), \forall x
\]

(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 $y = x + g/\|g\|_2$:

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

$$= f(x) + \|g\|_2$$

$$> f(x) + G$$
Analysis

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

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

\[
\|x^+ - x^*\|_2^2 = \|x - tg - x^*\|_2^2 \\
= \|x - x^*\|_2^2 - 2tg^T(x - x^*) + t^2\|g\|_2^2 \\
\leq \|x - x^*\|_2^2 - 2t(f(x) - f^*) + t^2\|g\|_2^2
\]

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

\[
2\left( \sum_{i=1}^{k} t_i \right) \left( f_{\text{best}}^{(k)} - f^* \right) \leq \|x^{(0)} - x^*\|_2^2 - \|x^{(k)} - x^*\|_2^2 + \sum_{i=1}^{k} t_i^2\|g^{(i-1)}\|_2^2 \\
\leq \|x^{(0)} - x^*\|_2^2 + \sum_{i=1}^{k} t_i^2\|g^{(i-1)}\|_2^2
\]
fixed step size $t_i = t$

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

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

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

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

- does not guarantee convergence of $f^{(k)}_{\text{best}}$
- for large $k$, $f^{(k)}_{\text{best}}$ is approximately $Gs/2$-suboptimal
diminishing step size \( t_i \to 0, \sum_{i=1}^{\infty} t_i = \infty \)

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

can show that \( (\sum_{i=1}^{k} t_i^2)/(\sum_{i=1}^{k} t_i) \to 0 \); hence, \( f^{(k)}_{\text{best}} \) converges to \( f^* \)
Example: 1-norm minimization

\[
\text{minimize } \|Ax - b\|_1 \quad (A \in \mathbb{R}^{500 \times 100}, b \in \mathbb{R}^{500})
\]

subgradient is given by \( A^T \text{sign}(Ax - b) \)

**fixed steplength** \( t_k = s/\|g^{(k-1)}\|_2, \ s = 0.1, 0.01, 0.001 \)
**diminishing step size** $t_k = 0.01/\sqrt{k}$, $t_k = 0.01/k$
Optimal step size for fixed number of iterations

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

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

• for given \( k \), bound is minimized by fixed step length \( s_i = s = R/\sqrt{k} \)

• resulting bound after \( k \) steps is

\[
f^{(k)}_{\text{best}} - f^* \leq \frac{GR}{\sqrt{k}} \]

• guarantees accuracy \( f^{(k)}_{\text{best}} - f^* \leq \epsilon \) 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

$$f^{(k)}_{\text{best}} - f^* \leq \frac{GR}{\sqrt{k}}$$
Exercise: find point in intersection of convex sets

to find a point in the intersection of \( m \) closed convex sets \( C_1, \ldots, C_m \),

\[
\text{minimize } \quad f(x) = \max\{d_1(x), \ldots, d_m(x)\}
\]

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

- \( f^* = 0 \) if the intersection is nonempty
- (from p. 4-15): \( g \in \partial f(\hat{x}) \) if \( g \in \partial d_j(\hat{x}) \) and \( C_j \) is farthest set from \( \hat{x} \)
- (from p. 4-21) subgradient \( g \in \partial d_j(\hat{x}) \) from projection \( P_j(\hat{x}) \) on \( C_j \):

\[
g = 0 \quad (\text{if } \hat{x} \in C_j), \quad g = \frac{1}{d(\hat{x}, C_j)}(\hat{x} - P_j(\hat{x})) \quad (\text{if } \hat{x} \notin C_j)
\]

note that \( \|g\|_2 = 1 \) if \( \hat{x} \notin C_j \)
subgradient method with optimal step size

- optimal step size 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 $f(x^{(k-1)}) = d_j(x^{(k-1)})$); take

$$x^{(k)} = x^{(k-1)} - \frac{f(x^{(k-1)})}{d_j(x^{(k-1)})}(x^{(k-1)} - P_j(x^{(k-1)}))$$

$$= P_j(x^{(k-1)})$$

- a version of the alternating projections algorithm

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

- for $m = 2$, projections alternate onto one set, then the other
Example: Positive semidefinite matrix completion

some entries of $X \in S^n$ fixed; find values for others so $X \succeq 0$

- $C_1 = S^n_+$, $C_2$ is (affine) set in $S^n$ with specified fixed entries
- projection onto $C_1$ by eigenvalue decomposition, truncation

$$P_1(X) = \sum_{i=1}^{n} \max\{0, \lambda_i\} q_i q_i^T \quad \text{if } X = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

- projection of $X$ onto $C_2$ by re-setting specified entries to fixed values

$100 \times 100$ matrix missing 71% entries

Subgradient method

\[ \|X(k+1) - X^*(k)\|_F \]

\[ 0 \quad 10^{-9} \quad 10^{-8} \quad 10^{-7} \quad 10^{-6} \quad 10^{-5} \quad 10^{-4} \quad 10^{-3} \quad 10^{-2} \quad 10^{-1} \quad 10^0 \quad 10^1 \]

\[ 0 \quad 10 \quad 20 \quad 30 \quad 40 \quad 50 \]
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

$$x^{(0)} + \text{span}\{g^{(0)}, g^{(1)}, \ldots, g^{(i-1)}\}$$
test problem and oracle

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

- solution: $x^* = -\frac{1}{k} (1, \ldots, 1, 0, \ldots, 0)$ and $f^* = -\frac{1}{2k}$

- $R = \|x^{(0)} - x^*\|_2 = 1/\sqrt{k}$ and $G = 1 + 1/\sqrt{k}$

- oracle returns subgradient $e_j + x$ where $\hat{j} = \min\{j \mid x_j = \max_{i=1,\ldots,k} x_i\}$

iteration: for $i = 0, \ldots, k - 1$, entries $x_{i+1}^{(i)}, \ldots, x_k^{(i)}$ are zero

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

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

  
  §3.2.1 with the example on page 5-16 of this lecture