Convex Optimization

(EE227A: UC Berkeley)

Lecture 12
(Subgradient methods)

28 Feb, 2013

Stefanie J. (for Suvrit Sra)

Announcements

- Midterm: final date is 19th March

ALGORITHMS

Unconstrained convex problem

$$\min_x \quad f(x)$$

How to optimize?

Unconstrained convex problem

$$\min_x \quad f(x)$$

How to optimize?

1. Start with some guess x^0; set $k = 0$

Unconstrained convex problem

$$\min_x \quad f(x)$$

How to optimize?

1. Start with some guess x^0; set $k = 0$
2. If $0 \in \partial f(x^k)$, **stop**; output x^k

Unconstrained convex problem

$$\min_x \quad f(x)$$

How to optimize?

1. Start with some guess x^0; set $k = 0$
2. If $0 \in \partial f(x^k)$, **stop**; output x^k
3. Otherwise, generate next guess x^{k+1}

Unconstrained convex problem

$$\min_x \quad f(x)$$

How to optimize?

1. Start with some guess x^0; set $k = 0$
2. If $0 \in \partial f(x^k)$, **stop**; output x^k
3. Otherwise, generate next guess x^{k+1}
4. Repeat above procedure

Unconstrained convex problem

$$\min_x \quad f(x)$$

How to optimize?

1. Start with some guess x^0; set $k = 0$
2. If $0 \in \partial f(x^k)$, **stop**; output x^k
3. Otherwise, generate next guess x^{k+1}
4. Repeat above procedure

- In reality: we stop in finite time
- Only solve problem approximately
- $f(x^k) \leq f(x^*) + \varepsilon$
- **shorthand** $f^k \leq f^* + \varepsilon$

Subgradient method

$$x^{k+1} = x^k - \alpha_k g^k$$

where $g^k \in \partial f(x^k)$ is **any** subgradient

Subgradient method

$$x^{k+1} = x^k - \alpha_k g^k$$

where $g^k \in \partial f(x^k)$ is **any** subgradient

Stepsize $\alpha_k > 0$ must be chosen

Subgradient method

$$x^{k+1} = x^k - \alpha_k g^k$$

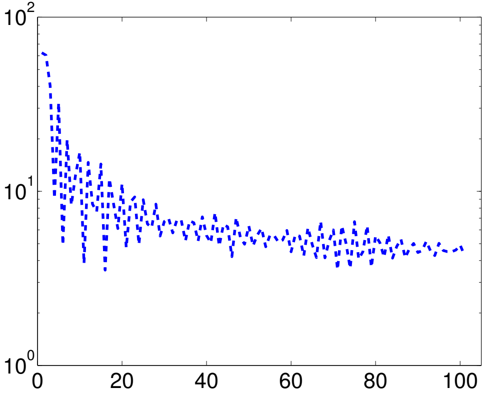
where $g^k \in \partial f(x^k)$ is **any** subgradient

Stepsize $\alpha_k > 0$ must be chosen

- Method generates sequence $\{x^k\}_{k\geq 0}$
- Does this sequence converge to an optimal solution x^*?
- If yes, then how fast?
- What if have constraints: $x \in \mathcal{C}$?

Example

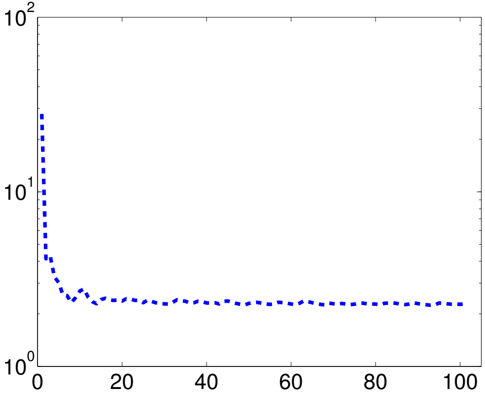
$$\min \quad \frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1$$
$$x^{k+1}=x^k-\alpha_k(A^T(Ax^k-b)+\lambda\,\mathrm{sgn}(x^k))$$



Example

$$\min \quad \tfrac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1$$

$$x^{k+1} = x^k - \alpha_k(A^T(Ax^k - b) + \lambda\,\mathrm{sgn}(x^k))$$



(More careful implementation)

Subgradient method – stepsizes

- **Constant** Set $\alpha_k = \alpha > 0$, for $k \geq 0$
- **Scaled constant** $\alpha_k = \alpha/\|g^k\|_2 \quad (\|x^{k+1} - x^k\|_2 = \alpha)$

Subgradient method – stepsizes

- **Constant** Set $\alpha_k = \alpha > 0$, for $k \geq 0$
- **Scaled constant** $\alpha_k = \alpha/\|g^k\|_2 \quad (\|x^{k+1} - x^k\|_2 = \alpha)$
- **Square summable but not summable**

$$\sum_k \alpha_k^2 < \infty, \qquad \sum_k \alpha_k = \infty$$

- **Diminishing scalar**

$$\lim_k \alpha_k = 0, \qquad \sum_k \alpha_k = \infty$$

- **Adaptive stepsizes** (not covered)

Not a descent method!
Work with best f^k so far: $f^k_{\min} := \min_{0 \leq i \leq k} f^i$

Convergence analysis

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$

Convergence analysis

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$
- Bounded subgradients: $\|g\|_2 \leq G$ for all $g \in \partial f$
  ($f(x) - f(y) = \langle g_\xi, x - y \rangle$; use Cauchy-Schwarz or Hölder)

Convergence analysis

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$
- Bounded subgradients: $\|g\|_2 \leq G$ for all $g \in \partial f$
  ($f(x) - f(y) = \langle g_\xi, x - y \rangle$; use Cauchy-Schwarz or Hölder)
- Bounded domain: $\|x^0 - x^*\|_2 \leq R$

Convergence analysis

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$
- Bounded subgradients: $\|g\|_2 \le G$ for all $g \in \partial f$
  $(f(x) - f(y) = \langle g_\xi, x - y \rangle$; use Cauchy-Schwarz or Hölder)
- Bounded domain: $\|x^0 - x^*\|_2 \le R$

Convergence results for: $f^k_{\min} := \min_{0 \le i \le k} f^i$

Subgradient method – convergence

Lyapunov function: **Distance to x^***, not function values

Subgradient method – convergence

Lyapunov function: Distance to x^*, not function values

$$\|x^{k+1} - x^*\|_2^2 \quad = \quad \|x^k - \alpha_k g^k - x^*\|_2^2$$

Subgradient method – convergence

Lyapunov function: **Distance to x^***, not function values

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \alpha_k g^k - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\langle \alpha_k g^k, x^k - x^* \rangle
\end{aligned}
$$

Subgradient method – convergence

Lyapunov function: **Distance to x^***, not function values

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \alpha_k g^k - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\langle \alpha_k g^k, x^k - x^* \rangle \\
&\leq \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\alpha_k (f(x^k) - f^*),
\end{aligned}
$$

since $f^* = f(x^*) \geq f(x^k) + \langle g^k, x^* - x^k \rangle$

Subgradient method – convergence

Lyapunov function: **Distance to x^***, not function values

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \alpha_k g^k - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\langle \alpha_k g^k, x^k - x^* \rangle \\
&\leq \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\alpha_k (f(x^k) - f^*),
\end{aligned}
$$

since $f^* = f(x^*) \geq f(x^k) + \langle g^k, x^* - x^k \rangle$

Apply same argument to $\|x^k - x^*\|_2^2$ recursively

Subgradient method – convergence

Lyapunov function: Distance to x^*, not function values

$$\begin{aligned}\|x^{k+1} - x^*\|_2^2 &= \|x^k - \alpha_k g^k - x^*\|_2^2 \\ &= \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\langle \alpha_k g^k,\, x^k - x^* \rangle \\ &\leq \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\alpha_k (f(x^k) - f^*),\end{aligned}$$

since $f^* = f(x^*) \geq f(x^k) + \langle g^k,\, x^* - x^k \rangle$

Apply same argument to $\|x^k - x^*\|_2^2$ recursively

$$\|x^{k+1} - x^*\|_2^2 \leq \|x^0 - x^*\|_2^2 + \sum_{t=1}^{k} \alpha_t^2 \|g^k\|_2^2 - 2 \sum_{t=1}^{k} \alpha_t (f^t - f^*).$$

Subgradient method – convergence

Lyapunov function: **Distance to x^***, not function values

$$
\begin{aligned}
\|x^{k+1} - x^*\|_2^2 &= \|x^k - \alpha_k g^k - x^*\|_2^2 \\
&= \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\langle \alpha_k g^k, x^k - x^* \rangle \\
&\le \|x^k - x^*\|_2^2 + \alpha_k^2 \|g^k\|_2^2 - 2\alpha_k (f(x^k) - f^*),
\end{aligned}
$$

since $f^* = f(x^*) \ge f(x^k) + \langle g^k, x^* - x^k \rangle$

Apply same argument to $\|x^k - x^*\|_2^2$ recursively

$$
\|x^{k+1} - x^*\|_2^2 \le \|x^0 - x^*\|_2^2 + \sum_{t=1}^{k} \alpha_t^2 \|g^k\|_2^2 - 2 \sum_{t=1}^{k} \alpha_t (f^t - f^*).
$$

Now use our convenient assumptions!

Subgradient method – convergence

$$\|x^{k+1} - x^*\|_2^2 \leq R^2 + G^2 \sum\nolimits_{t=1}^{k} \alpha_t^2 - 2 \sum\nolimits_{t=1}^{k} \alpha_t (f^t - f^*).$$

- To get a bound on the last term, simply notice (for $t \leq k$)

$$f^t \geq f_{\min}^t \geq f_{\min}^k \qquad \text{since} \quad f_{\min}^t := \min_{0 \leq i \leq t} f(x^i)$$

Subgradient method – convergence

$$\|x^{k+1} - x^*\|_2^2 \leq R^2 + G^2 \sum_{t=1}^{k} \alpha_t^2 - 2 \sum_{t=1}^{k} \alpha_t (f^t - f^*).$$

- To get a bound on the last term, simply notice (for $t \leq k$)

$$f^t \geq f^t_{\min} \geq f^k_{\min} \qquad \text{since} \quad f^t_{\min} := \min_{0 \leq i \leq t} f(x^i)$$

- Plugging this in yields the bound

$$2 \sum_{t=1}^{k} \alpha_t (f^t - f^*) \geq 2(f^k_{\min} - f^*) \sum_{t=1}^{k} \alpha_t.$$

Subgradient method – convergence

$$\|x^{k+1} - x^*\|_2^2 \leq R^2 + G^2 \sum_{t=1}^{k} \alpha_t^2 - 2 \sum_{t=1}^{k} \alpha_t (f^t - f^*).$$

- To get a bound on the last term, simply notice (for $t \leq k$)

$$f^t \geq f_{\min}^t \geq f_{\min}^k \qquad \text{since} \quad f_{\min}^t := \min_{0 \leq i \leq t} f(x^i)$$

- Plugging this in yields the bound

$$2 \sum_{t=1}^{k} \alpha_t (f^t - f^*) \geq 2(f_{\min}^k - f^*) \sum_{t=1}^{k} \alpha_t.$$

- So that we finally have

$$0 \leq \|x^{k+1} - x^*\|_2 \leq R^2 + G^2 \sum_{t=1}^{k} \alpha_t^2 - 2(f_{\min}^k - f^*) \sum_{t=1}^{k} \alpha_t$$

Subgradient method – convergence

$$\|x^{k+1} - x^*\|_2^2 \leq R^2 + G^2 \sum_{t=1}^{k} \alpha_t^2 - 2 \sum_{t=1}^{k} \alpha_t (f^t - f^*).$$

- To get a bound on the last term, simply notice (for $t \leq k$)

$$f^t \geq f^t_{\min} \geq f^k_{\min} \qquad \text{since} \quad f^t_{\min} := \min_{0 \leq i \leq t} f(x^i)$$

- Plugging this in yields the bound

$$2 \sum_{t=1}^{k} \alpha_t (f^t - f^*) \geq 2(f^k_{\min} - f^*) \sum_{t=1}^{k} \alpha_t.$$

- So that we finally have

$$0 \leq \|x^{k+1} - x^*\|_2 \leq R^2 + G^2 \sum_{t=1}^{k} \alpha_t^2 - 2(f^k_{\min} - f^*) \sum_{t=1}^{k} \alpha_t$$

$$\boxed{f^k_{\min} - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^{k} \alpha_t^2}{2 \sum_{t=1}^{k} \alpha_t}}$$

Subgradient method – convergence

$$f^k_{\min} - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t}$$

Exercise: Analyze $\lim_{k \to \infty} f^k_{\min} - f^*$ for the different choices of stepsize that we mentioned.

Subgradient method – convergence

$$f_{\min}^k - f^* \leq \frac{R^2+G^2\sum_{t=1}^k \alpha_t^2}{2\sum_{t=1}^k \alpha_t}$$

Exercise: Analyze $\lim_{k\to\infty} f_{\min}^k - f^*$ for the different choices of stepsize that we mentioned.

Constant step: $\alpha_k = \alpha$; We obtain

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 k\alpha^2}{2k\alpha}$$

Subgradient method – convergence

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t}$$

Exercise: Analyze $\lim_{k\to\infty} f_{\min}^k - f^*$ for the different choices of stepsize that we mentioned.

Constant step: $\alpha_k = \alpha$; We obtain

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 k \alpha^2}{2k\alpha} \to \frac{G^2\alpha}{2} \quad \text{as } k \to \infty.$$

Subgradient method – convergence

$$\boxed{f_{\min}^k - f^* \le \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2\sum_{t=1}^k \alpha_t}}$$

Exercise: Analyze $\lim_{k\to\infty} f_{\min}^k - f^*$ for the different choices of stepsize that we mentioned.

Constant step: $\alpha_k = \alpha$; We obtain

$$f_{\min}^k - f^* \le \frac{R^2 + G^2 k\alpha^2}{2k\alpha} \to \frac{G^2\alpha}{2} \quad \text{as } k \to \infty.$$

Square summable, not summable: $\sum_k \alpha_k^2 < \infty$, $\sum_k \alpha_k = \infty$

Subgradient method – convergence

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t}$$

Exercise: Analyze $\lim_{k\to\infty} f_{\min}^k - f^*$ for the different choices of stepsize that we mentioned.

Constant step: $\alpha_k = \alpha$; We obtain

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 k \alpha^2}{2k\alpha} \to \frac{G^2 \alpha}{2} \quad \text{as } k \to \infty.$$

Square summable, not summable: $\sum_k \alpha_k^2 < \infty$, $\sum_k \alpha_k = \infty$
As $k \to \infty$, numerator $< \infty$ but denominator $\to \infty$; so $f_{\min}^k \to f^*$

Subgradient method – convergence

$$f_{\min}^k - f^* \le \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t}$$

Exercise: Analyze $\lim_{k\to\infty} f_{\min}^k - f^*$ for the different choices of stepsize that we mentioned.

Constant step: $\alpha_k = \alpha$; We obtain

$$f_{\min}^k - f^* \le \frac{R^2 + G^2 k \alpha^2}{2k\alpha} \to \frac{G^2\alpha}{2} \quad \text{as } k \to \infty.$$

Square summable, not summable: $\sum_k \alpha_k^2 < \infty$, $\sum_k \alpha_k = \infty$

As $k \to \infty$, numerator $< \infty$ but denominator $\to \infty$; so $f_{\min}^k \to f^*$

In practice, fair bit of stepsize tuning needed, e.g. $\alpha_t = a/(b+t)$

Subgradient method – convergence

- Suppose we want $f_{\min}^{k} - f^{*} \leq \varepsilon$, how big should k be?

Subgradient method – convergence

- Suppose we want $f_{\min}^k - f^* \leq \varepsilon$, how big should k be?
- Optimize the bound for α_t: want

$$f_{\min}^k - f^* \leq \qquad\qquad\qquad\qquad \varepsilon$$

Subgradient method – convergence

- Suppose we want $f_{\min}^k - f^* \leq \varepsilon$, how big should k be?
- Optimize the bound for α_t: want

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t} \quad \leq \quad \varepsilon$$

Subgradient method – convergence

- Suppose we want $f_{\min}^k - f^* \leq \varepsilon$, how big should k be?
- Optimize the bound for α_t: want

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t} \quad \leq \quad \varepsilon$$

- For fixed k: best possible stepsize is constant α

$$\frac{R^2 + G^2 k \alpha^2}{2k\alpha} \leq \epsilon \quad \Rightarrow \quad \alpha = \frac{R}{G\sqrt{k}}$$

Subgradient method – convergence

- Suppose we want $f_{\min}^k - f^* \le \varepsilon$, how big should k be?
- Optimize the bound for α_t: want

$$f_{\min}^k - f^* \le \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2\sum_{t=1}^k \alpha_t} \quad \le \quad \varepsilon$$

- For fixed k: best possible stepsize is constant α

$$\frac{R^2 + G^2 k\alpha^2}{2k\alpha} \le \epsilon \quad \Rightarrow \quad \alpha = \frac{R}{G\sqrt{k}}$$

- Then, after k steps $f_{\min}^k - f^* \le RG/\sqrt{k}$.
- For accuracy ϵ, we need at least $(RG/\epsilon)^2 = O(1/\epsilon^2)$ steps

Subgradient method – convergence

- Suppose we want $f_{\min}^k - f^* \leq \varepsilon$, how big should k be?
- Optimize the bound for α_t: want

$$f_{\min}^k - f^* \leq \frac{R^2 + G^2 \sum_{t=1}^k \alpha_t^2}{2 \sum_{t=1}^k \alpha_t} \quad \leq \quad \varepsilon$$

- For fixed k: best possible stepsize is constant α

$$\frac{R^2 + G^2 k \alpha^2}{2k\alpha} \leq \epsilon \quad \Rightarrow \quad \alpha = \frac{R}{G\sqrt{k}}$$

- Then, after k steps $f_{\min}^k - f^* \leq RG/\sqrt{k}$.
- For accuracy ϵ, we need at least $(RG/\epsilon)^2 = O(1/\epsilon^2)$ steps
- (quite slow)

Exercise

Support vector machines

- Let $\mathcal{D} := \{(x_i, y_i) \mid x_i \in \mathbb{R}^n, y_i \in \{\pm 1\}\}$
- We wish to find $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that

$$\min_{w,b} \quad \tfrac{1}{2}\|w\|_2^2 + C \sum_{i=1}^{m} \max[0, 1 - y_i(w^T x_i + b)]$$

- Derive and implement a subgradient method
- Plot evolution of objective function
- Experiment with different values of $C > 0$
- Plot and keep track of $f_{\text{min}}^k := \min_{0 \le t \le k} f(x^t)$

Polyak's stepsize

- Assume f^* is known (or can be estimated). Then use

$$\alpha_t = \frac{f^t - f^*}{\|g^t\|_2^2}$$

Polyak's stepsize

- Assume f^* is known (or can be estimated). Then use

$$\alpha_t = \frac{f^t - f^*}{\|g^t\|_2^2}$$

- Motivation: recall bound

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - 2\alpha_t(f^t - f^*) + \alpha_t^2\|g^t\|^2$$

and minimize RHS.

Polyak's stepsize

- Assume f^* is known (or can be estimated). Then use

$$\alpha_t = \frac{f^t - f^*}{\|g^t\|_2^2}$$

- Motivation: recall bound

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - 2\alpha_t(f^t - f^*) + \alpha_t^2 \|g^t\|^2$$

  and minimize RHS.
- Let's plug in α_t:

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

Polyak's stepsize

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

Polyak's stepsize

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

- **Observation 1** $\|x^t - x^*\|$ decreases

Polyak's stepsize

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

- **Observation 1** $\|x^t - x^*\|$ decreases
- Recursion:

$$\sum_{t=1}^{k} \frac{(f^t - f^*)^2}{\|g_t\|^2} \quad \leq \quad \|x^1 - x^*\|^2 \quad \leq \quad R^2$$

Polyak's stepsize

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

- **Observation 1** $\|x^t - x^*\|$ decreases
- Recursion:

$$\sum_{t=1}^{k} \frac{(f^t - f^*)^2}{\|g_t\|^2} \quad \leq \quad \|x^1 - x^*\|^2 \quad \leq \quad R^2$$

- Now use $\|g^t\| \leq G$

$$\sum_{t=1}^{k} (f^t - f^*)^2 \leq R^2 G^2$$

Polyak's stepsize

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

- **Observation 1** $\|x^t - x^*\|$ decreases
- Recursion:

$$\sum_{t=1}^{k} \frac{(f^t - f^*)^2}{\|g_t\|^2} \quad \leq \quad \|x^1 - x^*\|^2 \quad \leq \quad R^2$$

- Now use $\|g^t\| \leq G$

$$\sum_{t=1}^{k} (f^t - f^*)^2 \leq R^2 G^2$$

- **Observation 2** $f^t \to f^*$

Polyak's stepsize

$$\|x^{t+1} - x^*\|^2 \leq \|x^t - x^*\|^2 - \frac{(f^t - f^*)^2}{\|g_t\|^2}$$

- **Observation 1** $\|x^t - x^*\|$ decreases
- Recursion:

$$\sum_{t=1}^{k} \frac{(f^t - f^*)^2}{\|g_t\|^2} \quad \leq \quad \|x^1 - x^*\|^2 \quad \leq \quad R^2$$

- Now use $\|g^t\| \leq G$

$$\sum_{t=1}^{k} (f^t - f^*)^2 \leq R^2 G^2$$

- **Observation 2** $f^t \to f^*$
- for accuracy ε, need $k = (RG/\varepsilon)^2$

Constrained optimization

$$\min \quad f(x) \qquad \text{s.t.} \quad x \in \mathcal{C}$$

Constrained optimization

$$\min \quad f(x) \qquad \text{s.t.} \quad x \in \mathcal{C}$$

- Previously:

$$x^{t+1} = x^t - \alpha_t g^t$$

- This could be infeasible!
  Solution: projection

Projected subgradient method

$$x^{k+1} = P_{\mathcal{C}}(x^k - \alpha_k g^k)$$

where $g^k \in \partial f(x^k)$ is any subgradient

Projected subgradient method

$$x^{k+1} = P_{\mathcal{C}}(x^k - \alpha_k g^k)$$

where $g^k \in \partial f(x^k)$ is any subgradient

- **Projection** closest feasible point

$$P_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|^2$$

(Assume $\mathcal{C}$ is closed and convex, then projection is unique)

Projected subgradient method

$$x^{k+1} = P_{\mathcal{C}}(x^k - \alpha_k g^k)$$

where $g^k \in \partial f(x^k)$ is any subgradient

- **Projection** closest feasible point

$$P_{\mathcal{C}}(x) = \arg\min_{y \in \mathcal{C}} \|x - y\|^2$$

(Assume $\mathcal{C}$ is closed and convex, then projection is unique)

- Great as long as projection is "easy"
- Same questions as before:
  - Does it converge?
  - For which stepsizes?
  - How fast?

Convergence

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$
- Bounded subgradients: $\|g\|_2 \leq G$ for all $g \in \partial f$
- Bounded domain: $\|x^0 - x^*\|_2 \leq R$

Convergence

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$
- Bounded subgradients: $\|g\|_2 \leq G$ for all $g \in \partial f$
- Bounded domain: $\|x^0 - x^*\|_2 \leq R$

Analysis

- Let $z^{t+1} = x^t - \alpha_t g^t$.
- Then $x^{t+1} = P_{\mathcal{C}}(z^{t+1})$.

Convergence

Assumptions

- Min is attained: $f^* := \inf_x f(x) > -\infty$, with $f(x^*) = f^*$
- Bounded subgradients: $\|g\|_2 \leq G$ for all $g \in \partial f$
- Bounded domain: $\|x^0 - x^*\|_2 \leq R$

Analysis

- Let $z^{t+1} = x^t - \alpha_t g^t$.
- Then $x^{t+1} = P_{\mathcal{C}}(z^{t+1})$.
- Recall analysis of unconstrained method:

$$\begin{aligned} \|z^{t+1} - x^*\|_2^2 &= \|x^t - \alpha_t g^t - x^*\|_2^2 \\ &\leq \|x^t - x^*\|_2^2 + \alpha_t^2 \|g^t\|_2^2 - 2\alpha_t(f(x^t) - f^*) \\ &\cdots \end{aligned}$$

- Need to relate to $\|x^{t+1} - x^*\|_2^2$, the rest of the proof is the same as above.

Projection Theorem

Let $\mathcal{C}$ be nonempty, closed and convex.

- Optimality conditions: $y^* = P_{\mathcal{C}}(z)$ iff

$$\langle z - y^*, y - y^* \rangle \leq 0 \text{ for all } y \in \mathcal{C}$$

- The projection is nonexpansive:

$$\|P_{\mathcal{C}}(x) - P_{\mathcal{C}}(z)\| \;\leq\; \|x - z\|^2 \quad \text{for all } x, z \in \mathbb{R}^n.$$

Convergence

- Use nonexpansiveness of projection:

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

Convergence

- Use nonexpansiveness of projection:

$$\begin{aligned}\|x^{t+1} - x^*\|_2^2 &= \|P_{\mathcal{C}}(x^t - \alpha_t g^t) - x^*\|_2^2 \\ &\leq \|x^t - \alpha_t g^t - x^*\|_2^2 \\ &\leq \|x^t - x^*\|_2^2 + \alpha_t^2 \|g^t\|_2^2 - 2\alpha_t (f(x^t) - f^*) \\ &\ldots\end{aligned}$$

Same convergence results as in unconstrained case:

- within neighborhood of optimal for constant step size
- converges for diminishing non-summable

Examples

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \\ \text{s.t.} \quad & x \in \mathcal{C} \end{aligned}$$

- **Nonnegativity** $x \geq 0$
  $P_{\mathcal{C}}(z) = [z]_+$
  Update step: $x^{k+1} = [x^k - \alpha_k(A^T(Ax^k - b) + \lambda\,\mathrm{sgn}(x^k))]_+$

Examples

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1 \\ \text{s.t. } & x \in \mathcal{C} \end{aligned}$$

- **Nonnegativity** $x \geq 0$
  $P_{\mathcal{C}}(z) = [z]_+$
  Update step: $x^{k+1} = [x^k - \alpha_k(A^T(Ax^k - b) + \lambda\,\mathrm{sgn}(x^k))]_+$
- ℓ_∞**-ball** $\|x\|_\infty \leq 1$
  Projection: $\min \|x - z\|^2$ s.t. $x \leq 1$ and $x \geq -1$

Examples

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_1 \\ \text{s.t. } & x\in\mathcal{C} \end{aligned}$$

- **Nonnegativity** $x \geq 0$
  $P_{\mathcal{C}}(z) = [z]_+$
  Update step: $x^{k+1} = [x^k - \alpha_k(A^T(Ax^k - b) + \lambda\,\mathrm{sgn}(x^k))]_+$
- ℓ_∞**-ball** $\|x\|_\infty \leq 1$
  Projection: $\min \|x-z\|^2$ s.t. $x \leq 1$ and $x \geq -1$
  this is separable, so do it coordinate-wise:
  $P_{\mathcal{C}}(z) = y$ where $y_i = \mathrm{sgn}(z_i)\min\{|z_i|, 1\}$

Examples

$$\begin{aligned} \min \quad & \tfrac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \\ \text{s.t.} \ \ & x \in \mathcal{C} \end{aligned}$$

- **Nonnegativity** $x \geq 0$
  $P_{\mathcal{C}}(z) = [z]_+$
  Update step: $x^{k+1} = [x^k - \alpha_k(A^T(Ax^k - b) + \lambda \operatorname{sgn}(x^k))]_+$
- **ℓ_∞-ball** $\|x\|_\infty \leq 1$
  Projection: $\min \|x - z\|^2$ s.t. $x \leq 1$ and $x \geq -1$
  this is separable, so do it coordinate-wise:
  $P_{\mathcal{C}}(z) = y$ where $y_i = \operatorname{sgn}(z_i) \min\{|z_i|, 1\}$
  Update step:

$$\begin{aligned} z^{k+1} &= x^k - \alpha_k(A^T(Ax^k - b) + \lambda \operatorname{sgn}(x^k)) \\ x_i^{k+1} &= \operatorname{sgn}(z_i^{k+1}) \min\{|z_i^{k+1}|, 1\} \end{aligned}$$

Examples

- **Linear equality constraints** $Ax = b$ ($A \in \mathbb{R}^{n\times m}$ has rank n)

$$\begin{aligned} P_{\mathcal{C}}(x) &= z - A^\top (AA^\top)^{-1}(Az - b) \\ &= (I - A^\top (A^\top A)^{-1} A) z + A^\top (AA^\top)^{-1} b \end{aligned}$$

Examples

- **Linear equality constraints** $Ax = b$ ($A \in \mathbb{R}^{n\times m}$ has rank n)

$$\begin{aligned} P_{\mathcal{C}}(x) &= z - A^\top(AA^\top)^{-1}(Az - b) \\ &= (I - A^\top(A^\top A)^{-1}A)z + A^\top(AA^\top)^{-1}b \end{aligned}$$

Update step, using $Ax^t = b$:

$$\begin{aligned} x^{t+1} &= P_{\mathcal{C}}(x^t - \alpha_t g^t) \\ &= x^t - \alpha_t(I - A^\top(AA^\top)^{-1}A)g^t \end{aligned}$$

Examples

- **Linear equality constraints** $Ax = b$ ($A \in \mathbb{R}^{n \times m}$ has rank n)

$$\begin{aligned} P_{\mathcal{C}}(x) &= z - A^\top (AA^\top)^{-1}(Az - b) \\ &= (I - A^\top (A^\top A)^{-1} A) z + A^\top (AA^\top)^{-1} b \end{aligned}$$

  Update step, using $Ax^t = b$:

$$\begin{aligned} x^{t+1} &= P_{\mathcal{C}}(x^t - \alpha_t g^t) \\ &= x^t - \alpha_t (I - A^\top (AA^\top)^{-1} A) g^t \end{aligned}$$

- **Simplex** $x^\top 1 = 1$ and $x \geq 0$
  more complex but doable, similarly ℓ_1-norm ball

Some remarks

- Why care?
  - simple
  - low-memory
  - stochastic version possible

Some remarks

- Why care?
  - simple
  - low-memory
  - stochastic version possible
- Another perspective

$$x^{k+1} = \min_{x \in \mathcal{C}} \langle x, g^k \rangle + \frac{1}{2\alpha_k} \|x - x_k\|^2$$

Mirror Descent

Some remarks

- Why care?
  - simple
  - low-memory
  - stochastic version possible
- Another perspective

$$x^{k+1} = \min_{x \in \mathcal{C}} \langle x, g^k \rangle + \frac{1}{2\alpha_k} \|x - x_k\|^2$$

  Mirror Descent
- Improvements using more information (heavy-ball, filtered subgradient, ...)

Some remarks

- Why care?
  - simple
  - low-memory
  - stochastic version possible
- Another perspective

$$x^{k+1} = \min_{x \in \mathcal{C}} \langle x, g^k \rangle + \frac{1}{2\alpha_k} \|x - x_k\|^2$$

  Mirror Descent
- Improvements using more information (heavy-ball, filtered subgradient, ...)
- Don't forget the dual!
  - may be more amenable to optimization
  - duality gap

What we did not cover

- ♠ Adaptive stepsize tricks
- ♠ Space dilation methods, quasi-Newton style subgrads
- ♠ Barrier subgradient method
- ♠ Sparse subgradient method
- ♠ Ellipsoid method, center of gravity, etc. as subgradient methods
- ♠ And many more

References

- ♠ S. Boyd, EE364b Slides
- ♠ Bertsekas, Nonlinear Programming