IE 521 Convex Optimization

Lecture 22: Constrained Subgradient Methods

Niao He

1st May 2019

Outline

Recap: Simple Constrained Convex Problems

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in X \end{aligned}$$

Subgradient Method

$$x_{t+1} = \Pi_X(x_t - \gamma_t g_t), \quad g_t \in \partial f(x_t)$$

Problem Class	Stepsize	Convergence
Convex Lipschitz	$O(\frac{1}{\sqrt{t}})$	$O(\frac{D_X M_f}{\sqrt{t}})$
Strongly Convex Lipschitz	$O(\frac{1}{\mu t})$	$O(\frac{M_f^2}{\mu t})$

Bundle Methods

- Kelley method
- Level-set method

General Constrained Convex Problems

We will focus on the general convex problem:

$$\begin{aligned} \min_{x \in X} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, i = 1, \ldots m \end{aligned}$$

Assumptions

- X is simple and admits easy-to-compute projections
- First-order oracles for $f_0(x), f_i(x)$ are available

Note $f_0(x), f_i(x)$ are not necessarily differentiable or smooth

Problem Reformulation

The problem can be rewritten as

$$\begin{aligned} \min_{x \in X} \quad & f_0(x) \\ \text{s.t.} \quad & f(x) \leq 0 \end{aligned} \tag{P}$$

where

$$f(x) = \max_{1 \leq i \leq m} f_i(x).$$

Remark. We can easily compute the subgradient for $f(x)$.

Problem Reformulation

The problem can be rewritten as

$$\begin{aligned} \min_{x \in X} \quad & f_0(x) \\ \text{s.t.} \quad & f(x) \leq 0 \end{aligned} \tag{P}$$

where

$$f(x) = \max_{1 \leq i \leq m} f_i(x).$$

Remark. We can easily compute the subgradient for $f(x)$.

So how to solve (P) given access to $f_0'(f), f'(x)$?

Constrained Subgradient Method

---

0. Initialize $x_1 \in X$
1. For $t \geq 1$, compute

$$g_t = \begin{cases} f_0'(x_t), & \text{if } f(x_t) < \gamma_t \|f'(x_t)\|_2 \\ f'(x_t), & \text{if } f(x_t) \geq \gamma_t \|f'(x_t)\|_2 \end{cases}$$

$$x_{t+1} = \Pi_X(x_t - \gamma_t \frac{g_t}{\|g_t\|_2})$$

Convergence of Constrained Subgradient Method

Theorem. Let f_0 and f be Lipschitz continuous with constant $M_{f_0} > 0$ and $M_f > 0$. Let D_X be the diameter of the set X such that $\forall x, y \in X, \|x - y\|_2 \leq D_X$. Set the stepsize $\gamma_t = \frac{D_X}{\sqrt{t+0.5}}$. We have, for any $T \geq 3$,

$$\min_{1 \leq t \leq T} f_0(x_t) - f_0^* \leq \frac{\sqrt{3} D_X M_{f_0}}{\sqrt{T - 1.5}}$$

$$\min_{1 \leq t \leq T} f(x_t) \leq \frac{\sqrt{3} D_X M_f}{\sqrt{T - 1.5}}$$

Proof of Convergence

- First, we have

$$\|x_{t+1} - x^*\|_2^2 \leq \|x_t - x^*\|_2^2 - 2\gamma_t \underbrace{\frac{g_t^T(x_t - x^*)}{\|g_t\|_2^2}}_{\nu_t} + \gamma_t^2$$

- Define

$$I = \{t \in [\frac{T}{3}, T] : \ g_t = f_0'(x_t)\}$$
$$I^c = \{t \in [\frac{T}{3}, T] : \ g_t = f'(x_t)\}$$

- For $t \in I^c, \nu_t \geq \gamma_t$. This is because

$$f'(x_t)^T(x_t - x^*) \geq f(x_t) - f(x^*) \geq f(x_t) \geq \gamma_t \|f'(x_t)\|_2$$

Proof of Convergence (cont'd)

- Claim: There exists $t^* \in I$, such that $\nu_{t^*} < \gamma_{t^*}$.
  - Otherwise, suppose $\nu_t \geq \gamma_t, \forall t \in I$. Then
$$\|x_{t+1} - x^*\|_2^2 \leq \|x_t - x^*\|_2^2 - \gamma_t^2, \forall t \in [\frac{T}{3}, T].$$
  - This implies that
$$\sum_{t=\lfloor T/3 \rfloor}^{T} \gamma_t^2 \leq D_X^2.$$
  - On the other hand,
$$\sum_{t=\lfloor T/3 \rfloor} \gamma_t^2 \geq D_X^2 \int_{T/3}^{T-1} \frac{1}{t-0.5} dt = D_X^2 \ln 3 > D_X^2.$$

Proof of Convergence (cont'd)

- Since $\nu_{t^*} < \gamma_{t^*} \leq \gamma_{\lfloor T/3 \rfloor}$, we have

$$f_0(x_{t^*}) - f_0^* \leq f_0'(x_{t^*})^T(x_{t^*} - x^*) \leq \nu_{t^*}\|g_{t^*}\|_2 \leq \frac{\sqrt{3}D_X M_{f_0}}{\sqrt{T - 1.5}}.$$

- Since $t^* \in I$, we have

$$f(x_{t^*}) \leq \gamma_{t^*}\|f'(x_{t^*})\|_2 \leq \frac{\sqrt{3}D_X M_f}{\sqrt{T - 1.5}}$$

Constrained Convex Problem

Constrained Convex Problem:

$$\begin{aligned} \min_{x \in X} \quad & f_0(x) \\ \text{s.t.} \quad & f(x) \leq 0 \end{aligned} \qquad \text{(P)}$$

with optimal value being $\alpha^* := \text{Opt}(P)$.

Parametric Function: Define

$$\Phi(x;\alpha) = \max\{f_0(x) - \alpha, f(x)\}$$

$$\phi(\alpha) = \min_{x \in X} \Phi(x;\alpha)$$

Note that $\phi(\alpha)$ is a one-dimensional function.

Properties

Lemma. The following results hold:

1. $\phi(\alpha^*) = 0$
2. $\phi(\alpha) \leq 0, \forall \alpha \geq \alpha^*$.
3. $\phi(\alpha) > 0, \forall \alpha < \alpha^*$
4. $\phi(\alpha) - \beta \leq \phi(\alpha + \beta) \leq \phi(\alpha), \forall \beta \geq 0$.

Remark.

- $\phi(\alpha)$ is non-increasing, 1-Lipschitz continuous, and convex.
- The smallest root of function $\phi(\alpha)$ corresponds to the optimal value of (P).
- Procedures for finding the root: Bisection, Newton's method, etc.

Generic Two-stage Scheme

0. Initialize α_1
1. For $i = 1, \ldots, N$
   - (Approximately) solve the subproblem:
     $$\min_{x \in X} \left\{ \Phi(x; \alpha_i) := \max\{f_0(x) - \alpha_i, f(x)\} \right\}$$
     through some subgradient routines (e.g., subgradient method, bundle methods)
   - Update $\alpha_i \to \alpha_{i+1}$ through some route-finding procedure (e.g., approximate Newton step)

Remark. Total number of complexity is expected to be

$$O\left(\frac{1}{\epsilon^2} \log(\frac{1}{\epsilon})\right).$$

Constrained Level Method

Model:

$$f_{0,t}(x) = \max_{1\leq i\leq t}\left\{f_0(x_i) + f_0'(x_i)^T(x - x_i)\right\}$$

$$f_t(x) = \max_{1\leq i\leq t}\left\{f(x_i) + f'(x_i)^T(x - x_i)\right\}$$

$$\Phi_t(x;\alpha) = \max\{f_{0,t}(x) - \alpha, f_t(x)\}$$

$$\phi_t(\alpha) = \min_{x\in X}\Phi_t(x;\alpha)$$

Model's Smallest Root:

$$\alpha_t^* := \min\{\alpha : \phi_t(\alpha) = 0\} = \min_{x\in X}\{f_{0,t}(x) :\ f_t(x) \leq 0\}$$

Remark.

- $\phi_1(\alpha) \leq \phi_2(\alpha) \leq \cdots \leq \phi(\alpha)$
- $\alpha_1^* \leq \alpha_2^* \leq \cdots \leq \alpha^*$

Constrained Level Method

0. Initialize $\alpha_1 < \alpha^*, \kappa \in (0, 1/2)$
1. For $i = 1, \ldots, N$
   - Run Level-set Method to solve $\min_{x\in X} \ \Phi(x;\alpha_i)$ and generate $\{x_\tau\}_{\tau=1}^t \in X$ until

$$\underbrace{\min_{x\in X} \Phi_t(x;\alpha_i)}_{\text{model's best value}} \geq (1-\kappa) \underbrace{\min_{1\leq\tau\leq t} \Phi_t(x_\tau;\alpha_i)}_{\text{model's record value}}$$

   - Set $t(i) = t$
   - Terminate if $\min_{1\leq\tau\leq t} \Phi_t(x_\tau;\alpha_i) \leq \epsilon$
   - Otherwise, set $\alpha_{i+1} = \alpha^*_{t(i)}$

Complexity of Constrained Level Method

- Number of master steps is at most

$$N(\epsilon) \leq \frac{1}{\ln(2-2\kappa)} \ln \frac{t^* - t_0}{(1-\kappa)\epsilon}$$

- Number of Level Method iterations at each step is at most

$$t(i) \leq \frac{D_X^2 M_f^2}{\kappa^2 \epsilon^2 (1-\alpha)^2 \alpha (2-\alpha)}$$

Lower Bound Complexity

Theorem. (Nemirovski & Yudin, 1979) For any $1 \leq t \leq n$, $x_1 \in \mathbb{R}^n$, there exists a M-Lipschitz continuous convex function f and a convex set X with diameter D_X, such that for any first-order algorithm that generates:

$$x_{t+1} \in x_1 + span(g_1, ..., g_t)$$

where $g_i \in \partial f(x_i)$, we always have

$$\min_{1 \leq s \leq t} f(x_s) - f_* \geq \frac{D_X M_f}{(1 + \sqrt{t})}$$

Lower Bound Complexity

Theorem. (Nemirovski & Yudin, 1979) For any $1 \leq t \leq n$, $x_1 \in \mathbb{R}^n$, there exists a μ-strongly convex, M-Lipschitz continuous function f and a convex set X, for any first-order method that generates

$$x_{t+1} \in x_1 + span(g_1, ..., g_t)$$

where $g_i \in \partial f(x_i)$, we always have

$$\min_{1 \leq s \leq t} f(x_s) - f_* \geq \frac{M^2}{8\mu t}$$

Worse-case Example

Consider the problem $\min_{x \in X} \ f(x)$ where

$$f(x) = C \cdot \max_{1 \le i \le t} x_i + \frac{\mu}{2}\|x\|_2^2,$$

$$X = \{x \in \mathbb{R}^n : \|x\|_2 \le R := D_X/2\}$$

- The subdifferential set of function f:

$$\partial f(x) = \mu x + C \cdot \mathrm{Conv}\{e_i : i \text{ that } x_i = \max_{1 \le j \le t} x_j\}$$

  - Oracle example: Given an input x, it returns $g = C \cdot e_i + \mu x$, with $i = \min\{i : x_i = \max_{1 \le j \le t} x_j\}$.

- The optimal solution and optimal value:

$$x_{*i} = \begin{cases} -\frac{C}{\mu t} & 1 \le i \le t \\ 0 & t < i \le n \end{cases} \quad \text{and} \quad f_* = -\frac{C^2}{2\mu t}.$$

Worst-case Example

- W.l.o.g., set $x_1 = 0$.
- By induction, we can show that $x_t \in span(e_1, ..., e_{t-1})$.
- This implies for $1 \leq s \leq t$, $f(x_s) \geq 0$.

$$\min_{1\leq s\leq t} f(x_s) - f_* \geq \frac{C^2}{2\mu t}.$$

  - If $C = \frac{M\sqrt{t}}{1+\sqrt{t}}$, $\mu = \frac{M}{R(1+\sqrt{t})}$, then $f(x)$ is M-Lipschitz continuous. Moreover,

$$\min_{1\leq s\leq t} f(x_s) - f_* \geq \frac{C^2}{2\mu t} = \frac{MD_X}{4(1+\sqrt{t})}$$

  - If $C = \frac{M}{2}$, $\mu = \frac{M}{2R}$, then $f(x)$ is M-Lipschitz continuous and μ-strongly convex. Moreover,

$$\min_{1\leq s\leq t} f(x_s) - f_* \geq \frac{C^2}{2\mu t} = \frac{M^2}{8\mu t}$$

References

- Nesterov (2004), Chapter 3.2-3.3