IE 521 Convex Optimization

Lecture 23: Dual Methods

Niao He

29th April 2019

Outline

Recap: Subgradient Methods

Simple Constrained Convex Problems

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

Subgradient Method

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

Bundle Methods

- Kelley method
- Level-set method

Recap: Constrained Subgradient Methods

General Constrained Convex Problems

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

Constrained Subgradient Method

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

where

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

Generic Two-stage Schemes

- Constrained Level Method

Now consider some linear structure...

Linearly Constrained Convex Problems

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

Conic Constrained Convex Problems

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax - b \in \mathcal{K} \end{aligned}$$

Examples

- Sparse recovery:

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

- Generalized Lasso:

$$\min_x \ \|Ax - b\|_2^2 + \lambda \|Dx\|_1 \iff \begin{aligned} \min_{x,y} \ & \|Ax - b\|_2^2 + \lambda \|y\|_1 \\ \text{s.t.} \ & Dx = y \end{aligned}$$

- Minimizing over Intersection of Convex Sets:

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t.} \ & x \in C_1 \cap C_2 \end{aligned} \iff \begin{aligned} \min_{x,y} \ & f(x) + I_{C_1}(x) + I_{C_2}(y) \\ \text{s.t.} \ & x = y \end{aligned}$$

Dual Problem

Linearly Constrained Convex Problems

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

Dual Problem

$$\max_\lambda \quad h(\lambda) := -f^*(-A^T\lambda) - b^T\lambda$$

Subgradient of the Dual

- $\partial h(\lambda) = A\partial f^*(-A^T\lambda) - b$
- $u \in \partial f^*(-A^T\lambda) \Leftrightarrow u \in \mathrm{argmin}_x\{f(x) + \lambda^T Ax\}$
- If f is strictly convex, f^* is differentiable.

Dual Subgradient Method

---

0. Initialize λ^1
1. For $t \geq 1$, compute

$$x^{t+1} \in \underset{x}{\operatorname{argmin}}\{f(x) + (\lambda^t)^T Ax\}$$
$$\lambda^{t+1} = \lambda^t + \beta_t(Ax^{t+1} - b)$$

- If f is strictly convex, we get dual gradient ascent.
- If f is μ-strongly convex, we can set constant $\beta_t \leq \mu$.
- Otherwise, we need to set β_t to be diminishing.
- Convergence follows what have for subgradient method.

Dual Decomposition (1960s)

$$\begin{aligned} \min_x \quad & \sum_{i=1}^n f_i(x_i) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

---

0. Initialize λ^1
1. For $t \geq 1$, compute

$$x_i^{t+1} \in \operatorname*{argmin}_{x_i}\{f_i(x_i) + (\lambda^t)^T A_i x_i\}, i = 1, \ldots, n$$

$$\lambda^{t+1} = \lambda^t + \beta_t\left(\sum_{i=1}^n A_i x_i^{t+1} - b\right)$$

- **Broadcast:** send λ to each processor to find x_i in parallel
- **Gather:** collect $A_i x_i$ from each processor to update λ

Augmented Lagrangian

Original Problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

Transformed Problem

$$\begin{aligned} \min_x \quad & f(x) + \frac{\rho}{2}\|Ax - b\|_2^2 \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

Augmented Lagrangian

$$L_\rho(x;\lambda) = f(x) + \lambda^T(Ax - b) + \frac{\rho}{2}\|Ax - b\|_2^2$$

- Adding the term $\frac{\rho}{2}\|Ax - b\|_2^2$ does not change problem
- Note if A is full rank, primal becomes strongly convex

Augmented Lagrangian Method (1960s)

---

0. Initialize λ^1
1. For $t \geq 1$, compute

$$x^{t+1} = \operatorname*{argmin}_x \{f(x) + (\lambda^t)^T Ax + \frac{\rho}{2}\|Ax - b\|_2^2\}$$
$$\lambda^{t+1} = \lambda^t + \rho(Ax^{t+1} - b)$$

- Also know as the method of multipliers.
- Improve convergence properties
- Require solving harder subproblems
- Lose decomposability (non-separable subproblems)

Q. why choosing stepsize $\beta_t = \rho$?

Alternating Direction Method of Multipliers (ADMM)

Linear Constrained Convex Problem

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + By = c \end{aligned}$$

Augmented Lagrangian

$$\begin{aligned} L_\rho(x, y; \lambda) = f(x) + g(y) + \lambda^T(Ax + By - c) \\ + \frac{\rho}{2}\|Ax + By - c\|_2^2 \end{aligned}$$

Method of Multiplier performs the primal update

$$(x^{t+1}, y^{t+1}) = \operatorname*{argmin}_{x,y} \; L_\rho(x, y; \lambda^t)$$

ADMM

0. Initialize λ^1, x^1, y^1
1. For $t \geq 1$, compute

$$x^{t+1} = \underset{x}{\operatorname{argmin}}\ L_\rho(x, y^t; \lambda^t)$$
$$y^{t+1} = \underset{y}{\operatorname{argmin}}\ L_\rho(x^{t+1}, y; \lambda^t)$$
$$\lambda^{t+1} = \lambda^t + \rho(Ax^{t+1} + By^{t+1} - c)$$

- Alternative min over x, y rather than joint minimization
- Preserve good convergence as the method of multiplier
- Achieve the decomposability.
- Convergence results are well established now.
- Equivalent to Douglas-Rachford splitting method.

Illustration: LASSO

$$\min \ \|Ax - b\|_2^2 + \lambda\|x\|_1$$

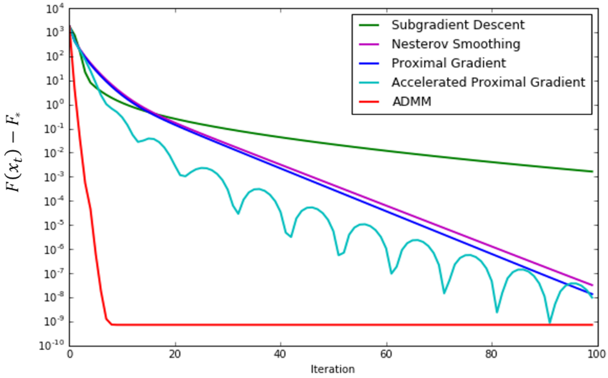


Figure: Algorithms for solving the LASSO problem

Extensions and Variations of ADMM

- Varying penalty parameter
  - increasing ρ when infeasibility error is large
  - decreasing ρ when objective error is large
- General augmented terms
  - $\frac{\rho}{2}\|r\|_2^2 \to \frac{\rho}{2}\|r\|_P^2$ with $P \succeq 0$
  - sometimes lead to easy-to-compute subproblems
- Jacobi style update
  - $x^{t+1} = \mathrm{argmin}_x\, L_\rho(x, y^t; \lambda^t)$
  - $y^{t+1} = \mathrm{argmin}_y\, L_\rho(x^t, y; \lambda^t)$
  - (x, y) can be updated in parallel
- Inexact minimization / Linearization
  - Subproblems are approximately solved through some iterative routines.
- Multi-block extension
  - Active research topic

Relate back to Barrier Method

General Conic Programs

$$\begin{aligned} &\min_x \; f(x) \\ &\text{s.t. } \; Ax - b \in \mathcal{K} \end{aligned}$$

Barrier Method: given a penalty parameter $t > 0$ and barrier function $F(\cdot)$ on $\text{int}(\mathcal{K})$, need to solve subproblems

$$\min \{t \cdot f(x) + F(Ax - b)\}$$

$$\min \{t \cdot f(x) + F(y) : Ax - y = b\}$$

Remark. ADMM could also be used as a subroutine to solve barrier problems (active research topic).

Summary and Outlook

Topics covered so far

- Classical Convex Optimization (– 1970s)
  - Convex sets, convex functions, convex programs
  - Convex geometry
  - Separation theorems and theorems on alternatives
  - Subgradient, conjugacy, optimality
  - Duality and minimax theorems
  - Polynomial solvability, Ellipsoid method
- Modern Convex Optimization (1980s –)
  - Conic programs (LP, SOCP, SDP)
  - Conic duality
  - SDP relaxations
  - Barriers, self-concordance
  - Interior Point Method
- Large-scale Convex Optimization
  - Subgradient method and bundle methods
  - Constrained subgradient methods
  - Dual subgradient methods
  - Augmented Lagrangian and ADMM

Algorithms covered so far

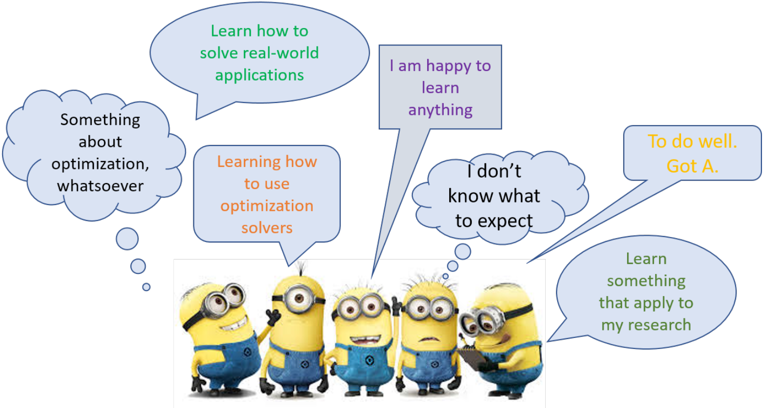
- **Polynomial algorithms**
  - Ellipsoid method
  - Interior point method
- **(Super) Linear-convergent algorithms**
  - Center of gravity method
  - Ellipsoid method
  - (Damped) Newton method
  - GD/ALM/ADMM under smooth and strong convexity
- **First-order algorithms**
  - Subgradient methods, Mirror Descent
  - Constrained subgradient methods
  - Localization methods
- **Second-order algorithms**
  - (Damped) Newton method

What's beyond

- First-order algorithms
  - Accelerated Gradient Descent and its cousins
  - Conditional gradient (Frank-Wolfe) and its variants
  - Coordinate descent method and its variants
  - Proximal gradient method and its acceleration
  - Stochastic gradient methods (SGD, AdaGrad, etc)
  - Variance-reduced methods (SVRG, SAGA, etc.)
  - Distributed and asynchronous gradient methods
  - Online gradient methods
  - ... ...
- Other algorithms
  - Proximal point algorithms
  - Splitting algorithms
  - Sample average approximation
  - ... ...

Check my lecture notes for IE 598: Big Data Optimization.

What were your expectations for the course?



Something you should have learned

- ☑ Understand basic concepts
- ☑ Detect convexity of sets, functions, and programs
- ☑ Reformulate problems into convex forms
- ☑ Characterize optimality and duality of convex programs
- ☑ Know to appropriately choose algorithms in practice
  - Should I use Ellipsoid Method or Interior Point Method?
  - When to use first-order or second-order method?
  - Should I expect linear or sublinear convergence?
  - How should I choose hyperparameters?
  - What can I possibly do to improve the performance?

Where to go from here?

Theory/Modeling/Algorithm/Application

Book:

- https://web.stanford.edu/ boyd/cvxbook/

Blogs:

- https://blogs.princeton.edu/imabandit/
- http://www.offconvex.org/
- https://sunju.org/research/nonconvex/

Conferences:

- INFORMS, IOS, SIAM-OPT, ICCOPT, ICSP, MOPTA
- ICML/NIPS workshops, Simons Institute workshops

Read lots and lots of papers!

Last Thing

Final Exam
7:00-10:00 p.m., Monday, May 6

Project Final Report
11:59 p.m., Friday, May 10

Fin

Thank you!

Please fill in the ICES form for me.