15. Conclusions

- structural convex optimization: some rules
- review of three examples
- wrap up

Complexities of convex optimization

$$\underset{x \in Q}{\text{minimize}} \quad f(x),$$

where $Q \subseteq \mathbf{R}^n$ is bounded, closed and convex ($\|x\|_2 \leq R,\ \forall x \in Q$)

problem class	# oracle calls	lower bound
$\|\nabla f(\cdot)\|_2 \leq G$	$\leq O(n)$	$O\left(\frac{G^2 R^2}{\epsilon^2}\right)$
$\|\nabla^2 f(\cdot)\|_2 \leq L$	$\leq O(n)$	$O\left(\frac{\sqrt{L}R}{\sqrt{\epsilon}}\right)$
$\mu \leq \|\nabla^2 f(\cdot)\|_2 \leq L$	$\leq O(n)$	$O\left(\sqrt{\frac{L}{\mu}} \log \frac{LR^2}{\epsilon}\right)$
$\|\nabla f(\cdot)\|_2 \leq G$	$\geq O(n)$	$O\left(n \log \frac{GR}{\epsilon}\right)$

- based on **local black-box** first-order oracle
- optimal methods exist: efficiency estimate proportional to lower bound

A conceptual contradiction

convexity is a global structure

- usually checked by inspection: e.g., composition of basic convex functions
- numerical verification of convexity is extremely difficult

but numerical methods use local black-box

beyond block box: structural convex optimization

- exploiting structure to improve performance of numerical methods
- recent developments:
  - interior-point methods
  - smoothing
  - minimization of composite objective

A classical example

solving system of linear equations

$$Ax = b$$

standard procedure:

- check if A is symmetric and positive definite
- compute Cholesky factorization $A = LL^T$ and form two auxiliary systems
$$Ly = b, \qquad L^T x = y$$
- solve these systems by forward and backward substitution

Useful rules (Nesterov)

1. find a class of problems which can be solved very efficiently
2. describe transformation rules for converting problem into desired form
3. describe class of problems for which transformation rules are applicable

for solving linear systems:

1. **class of easy problems:** linear systems with triangular matrices
2. **transformation rules:** Cholesky factorization (cost $n^3/3$ flops)
3. **applicable problems:** symmetric and positive definite matrices

Three examples in convex optimization

- interior-point methods
- smoothing
- minimization of composite objective

Interior-point methods

- **class of easy problems:** unconstrained minimization of self-concordant functions (using Newton's method)
- **transformation rules:** path following with self-concordant barriers
- **applicable problems:** convex optimization problems with structural constraints (e.g., LP, SOCP, SDP, and asymmetric conic problems)

complexity bound: $O(\sqrt{\theta}\log\frac{\theta}{\epsilon})$

- self-concordant parameter θ can be much smaller than dimension of space (n for LP, 2 for SOCP, n for SDP with $n(n+1)/2$ variables)
- cf. $O(n\log\frac{1}{\epsilon})$ lower bound for problems with separation oracle

Smoothing

- **class of easy problems:** minimizing smooth functions (with Lipschitz continuous gradient) by accelerated gradient method
- **applicable problems:** minimizing nonsmooth functions with structure

$$f(x) = \sup_{y \in \mathbf{dom}\, h} \left((Ax+b)^T y - h(y) \right) = h^*(Ax+b)$$

- **transformation rules:** strongly convex regularization of conjugate

$$f_\mu(x) = \sup_{y \in \mathbf{dom}\, h} \left((Ax+b)^T y - h(y) - \mu d(y) \right) = (h + \mu d)^*(Ax+b)$$

complexity: $O(1/\epsilon)$ with $\mu \propto \epsilon$

(in between optimal bounds of smooth and nonsmooth optimization)

Minimization of composite objective

problem class:

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \left\{ \phi(x) \triangleq f(x) + \Psi(x) \right\}$$

- f is convex and smooth (having Lipschitz-continuous gradient)
- Ψ is convex, but may be nondifferentiable
- using black-box first-order oracle, complexity is $O(1/\epsilon^2)$

structural convex optimization

- assume Ψ is *simple*, e.g., can solve explicitly the auxiliary problem

$$\underset{x \in \mathbf{dom}\, \Psi}{\text{minimize}} \left\{ f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 + \Psi(x) \right\}$$

- accelerated gradient methods achieve reduced complexity $O(1/\sqrt{\epsilon})$

Example: sparse least-squares

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1 \qquad \text{(where } A \in \mathbf{R}^{m \times n}\text{)}$$

- important applications in signal processing, statistics, machine learning
- focus on problem class: $m < n$ and $x^\star$ sparse (compressed sensing)

complexities of structural convex optimization

numerical method	analytical complexity	oracle complexity
subgradient method	$O(1/\epsilon^2)$	$O(mn)$
proximal gradient method	$O(1/\epsilon)$	$O(mn)$
accelerated gradient method	$O(1/\sqrt{\epsilon})$	$O(mn)$
interior-point method	$O(\sqrt{n}\log(1/\epsilon))$	$O(m^2 n)$
proximal gradient homotopy (assume A has RIP)	$O(\log(1/\epsilon))$	$O(mn)$

On the role of complexity theory

complexity analysis plays an important role in convex optimization

- many ideas appeared early, but did not result in significant impact due to lack of convincing complexity analysis

modern algorithms	early prototypes
accelerated gradient methods	heavy ball method
polynomial-time IPMs	classical barrier methods
smoothing	smoothing

- quotes from Yurii Nesterov (in his 2004 book)
  - it became more and more common that the new methods were provided with a complexity analysis, which is considered a better justification of their efficiency than computational experiments

Acknowledgements

Thanks to:

- TA (Reza), guest lecturer (James), project mentors
- Lin Xiao (MSR), and Lieven Vandenberghe (UCLA)
- ee546 students

please fill in online **course feedback form** for instructor and TA. we appreciate your feedback!

References

- Yu. Nesterov, *Introductory Lectures on Convex Optimization. A Basic Course* (2004), Section 1.1.

- Yu. Nesterov, *How to advance in Structural Convex Optimization* (November 2008), OPTIMA 78, Mathematical Programming Society Newsletter, pages 2-5.