IPAM OPWS2: Numerical Methods for Continuous Optimization

First-order Methods for Convex Optimization with Inexact Oracle

Olivier Devolder, François Glineur and Yurii Nesterov: UCLouvain, CORE & INMA

Why an inexact oracle for first-order methods ?

In smooth convex optimization:

- First-order methods = Methods of choice for large-scale problems due to their cheap iteration cost
- Sometime impossible/costly to compute *exact* first-order information (function and gradient values) at each iteration

  Possible causes: numerical errors (see E1.) ; need to solve another (simpler) optimization problem, which can only be done approximately (see E2.) ; non smoothness! (see E3., E4.)
- **Our goal**: Study the effect of inexact first-order information on two usual first-order methods:

  Classical Gradient Method (CGM)
  and Fast Gradient Method (FGM)
- **Important issue**: link between desired objective function accuracy and accuracy needed for oracle first-order information ?

A definition of inexact oracle.

Consider the convex optimization problem:

$$f^* = \min_{x\in Q} f(x)$$

where f is convex and Q is a closed convex set $\subset \mathbb{R}^n$.
Definition:
f is equipped with a **first-order** (ϵ, L)**-oracle** $(f_{y,\epsilon}, g_{y,\epsilon}) = \mathcal{O}_{\epsilon,L}[f](y)$
$\Leftrightarrow$ for any $y \in Q$, we can compute $(f_{y,\epsilon}, g_{y,\epsilon})$ satisfying:

$$f_{y,\epsilon} + \langle g_{y,\epsilon}, x-y\rangle \leq f(x) \leq f_{y,\epsilon} + \langle g_{y,\epsilon}, x-y\rangle + \frac{L}{2}\|x-y\|^2 + \epsilon \quad \forall x \in Q.$$

Properties: $f_{y,\epsilon}$ is ϵ-accurate, $g_{y,\epsilon}$ is an ϵ-subgradient + upper bound

First-order methods with a (ϵ, L)-oracle

Let $\bar\epsilon$ = desired accuracy for the solution (SA),
let ϵ = accuracy of the oracle (OA) and define $R = \|x^0 - x^*\|$:

1. **Classical Gradient Method**

$$f(x^k) - f^* \leq \frac{C_1 LR^2}{k} + \epsilon$$

   - No accumulation of errors
     Error asymptotically tends to ϵ (OA)
   - OA=SA : $\epsilon = \Theta(\bar\epsilon)$
   - Complexity: $O\left(\frac{LR^2}{\bar\epsilon}\right)$ (not optimal)

2. **Fast Gradient Method**

$$f(x^k) - f^* \leq \frac{C_2 LR^2}{k^2} + C_3 k\epsilon$$

   - Accumulation of errors
     Error asymptotically tends to ∞
     (decreases at first, then increases linearly)
   - OA must be smaller than SA: $\epsilon = \Theta(\bar\epsilon^{3/2})$
   - Optimal complexity: $O\left(\sqrt{\frac{L}{\bar\epsilon}}R\right)$.

Both methods are optimal (in a different way)

1. CGM is the fastest first-order method without error accumulation
2. Any first-order method with convergence rate $\frac{1}{k^2}$ must suffer from error accumulation, and FGM has the lowest possible error accumulation for such a method: $\Theta(k\epsilon)$.

Examples of (ϵ, L)-oracles

E1. ***Exact computation at shifted points***

Assumptions: f is convex with a Lipschitz-continuous gradient (constant $\bar L$)
Oracle: At each point $\bar x \in Q$, the oracle provides exact value of f and ∇f but computed at a different point $\bar x_\epsilon$.
$\Rightarrow$ (ϵ, L)-oracle with $\epsilon = \bar L \|\bar x - \bar x_\epsilon\|_2^2$ and $L = 2\bar L$.

E2. ***Smooth saddle point problem***

Assumptions:

$$f(x) = \max_{u\in U} \Psi(x,u)$$

where

- U is a closed, convex set
- $\Psi(x,u) = G(u) + \langle Au, x\rangle$
- $G(u)$ is a differentiable, strongly concave function with parameter $\kappa > 0$

Denoting $u_x = \arg\min_{u\in U} \Psi(x,u)$, we have:

$$f(x) = \Psi(x, u_x), \quad \nabla f(x) = Au_x.$$

Oracle: At each point $x \in Q$, the oracle provides

$$f_{x,\epsilon} = \Psi(x, \bar u_x), \quad g_{x,\epsilon} = A\bar u_x$$

where $\bar u_x$ is an approximate solution of $\max_{u\in U} \Psi(x,u)$.
$\Rightarrow$ (ϵ, L)-oracle with $\epsilon = 2(\Psi(x,u_x) - \Psi(x,\bar u_x))$ and $L = \frac{2\|A\|}{\kappa}$.

E3. ***Non-smooth convex function***

Assumptions: f is convex, subdifferentiable with bounded variation of the subgradients:

$$\|g(x) - g(y)\|_* \leq M \quad \forall g(x) \in \partial f(x), g(y) \in \partial f(y), \quad \forall x, y \in Q$$

Oracle: At each point $\bar x$, the oracle provides $f(\bar x)$ and $g(\bar x) \in \partial f(x)$.
$\Rightarrow$ (ϵ, L)-oracle with arbitrary ϵ and $L = \frac{M^2}{2\epsilon}$ (i.e. a whole family of oracles with arbitrary value of ϵ)
Consequence:
Application of CGM or FGM to f with right choice of ϵ solves non-smooth problem with an optimal rate of convergence $\Theta\left(\frac{LR}{\sqrt{k}}\right)$.

E4. ***Smooth convex function with Hôlder continuous gradient***

Assumptions: f is convex, differentiable with Hôlder continuous gradient:

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L_\nu \|x-y\|^\nu \quad \forall x, y \in Q$$

for a given $0 \leq \nu < 1$.
Oracle: At each point $\bar x$, oracle provides $f(\bar x)$ and $\nabla f(\bar x)$
$\Rightarrow$ (ϵ, L)-oracle with arbitrary ϵ and

$$L = L_\nu^{\frac{2}{1+\nu}}\left(\frac{1}{(1+\nu)^{\frac{1-\nu}{1+\nu}}} - \frac{1}{2^{\frac{1-\nu}{1+\nu}}}\right)\frac{1}{\epsilon^{\frac{1-\nu}{1+\nu}}}.$$

Consequence:
Application of FGM to f with right choice of ϵ solves 'weakly' smooth problem with an optimal rate of convergence : $\Theta\left(\frac{L_\nu R^{1+\nu}}{k^{\frac{1+3\nu}{2}}}\right)$.

WE OBTAIN **UNIVERSAL OPTIMAL METHOD** BOTH FOR SMOOTH, WEAKLY SMOOTH AND NON-SMOOTH CONVEX PROBLEMS.

Acknowledgements

First author is an F.R.S-FNRS research fellow. Poster presents research results of the Belgian Network DYSCO. Scientific responsibility rests with its authors.