New Computational Guarantees for Solving Convex Optimization Problems with First Order Methods, via a Function Growth Condition Measure

Robert M. Freund and Haihao Lu

MIT

ICCOPT Tokyo, August 2016

Outline

- Review of Basic First-Order Methods (FOMs)
- Motivation: Renegar's Recent Work
- Function Growth Constant
- New Computational Guarantees for Non-smooth Optimization
- New Computational Guarantees for Smooth Optimization
- Remarks, Extensions, Next Steps

Review of Projected Subgradient Descent

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

Assume easy to compute the (Euclidean) projection $\Pi_Q(x)$ of x onto Q

Projected Subgradient Descent

Given $x^0 \in Q$, $k \leftarrow 0$, $x_b^0 \leftarrow x^0$, $f_b^0 \leftarrow f(x^0)$
At iteration k:

1. Compute a subgradient of $f(\cdot)$ at x^k: $\quad g^k \in \partial f(x_k)$
2. Perform update : $\quad x^{k+1} \leftarrow \Pi_Q(x_k - \alpha_k g^k)$

$$f_b^{k+1} \leftarrow \min\{f_b^k, f(x^{k+1})\}$$
$$x_b^{k+1} \leftarrow \arg \min_{x \in \{x_b^k,\ x^{k+1}\}} \{f(x)\} \ .$$

Computational Guarantee for Subgradient Descent

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

$\text{Opt} := \{x \in Q : f(x) = f^*\}$

M-Lipschitz continuity : $|f(y) - f(x)| \leq M\|y - x\| \quad$ for all $x, y \in Q$

Theorem: Convergence Bound for Subgradient Descent [Polyak, Nesterov]

Given $\varepsilon > 0$, let us use the step-size sequence $\alpha_i = \varepsilon/\|g^i\|^2$ for all i. Define:

$$N := \frac{M^2 \text{Dist}(x^0, \text{Opt})^2}{\varepsilon^2} - 1 \ .$$

Then for all $k \geq N$ it holds that $f_b^k \leq f^* + \varepsilon$.

Review of Accelerated Gradient Method

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

Lipschitz gradient: $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ for all $x, y \in Q$

Accelerated Gradient Method

Given $x^0 \in Q$ and $z^0 := x^0$, and $i \leftarrow 0$. Define step-size parameters $\theta_i \in (0, 1]$ recursively by $\theta_0 := 1$ and θ_{i+1} satisfies $\frac{1}{\theta_{i+1}^2} - \frac{1}{\theta_{i+1}} = \frac{1}{\theta_i^2}$.

At iteration k:

1. Update : $y^i \leftarrow (1 - \theta_i)x^i + \theta_i z^i$

$$z^{k+1} \leftarrow \arg\min_{x \in Q}\{f(y^k) + \nabla f(y^k)^T(x - z^k) + \tfrac{1}{2}\theta_k L\|x - z^k\|^2\}$$

$$x^{k+1} \leftarrow (1 - \theta_k)x^k + \theta_k z^{k+1}$$

Computational Guarantee for Accelerated Gradient Method

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

$$\text{Opt} := \{x \in Q : f(x) = f^*\}$$

Theorem: Convergence Bound for Accelerated Gradient Method [Nesterov, Tseng]

For all $k \geq 0$ it holds that:

$$f(x^k) \quad \leq \quad f^* \; + \; \frac{2L\text{Dist}(x^0, \text{Opt})^2}{(k+1)^2} \; .$$

Quantities in these Analyses

- squared distance to the optimal solution set: $\text{Dist}(x^0, \text{Opt})^2$
- M-Lipschitz function : $|f(y) - f(x)| \leq M\|y - x\|$ for all $x, y \in Q$
- L-Lipschitz gradient : $\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|$ for all $x, y \in Q$
- absolute optimality accuracy ε : $f(x^k) \leq f^* + \varepsilon$

Renegar's recent paper

"A Framework for Applying Subgradient Methods to Conic Optimization Problems" by James Renegar

June, 2015 (earlier versions September 2014, March 2015)

arXiv:1503.02611

- the paper considers SDP in conic format and its extensions
- here we present the results only stated for LP for ease of presentation

Linear Optimization

Given LP data A, b, c

We have the standard linear problem:

$$\begin{array}{rrl} z^* := & \text{minimum}_x & c^T x \\ & \text{s.t.} & Ax = b \\ & & x \geq 0 \end{array}$$

We are also given $\bar{x}$ for which $\bar{x} > 0$ and $A\bar{x} = b$

Herein we (re-)define our linear problem as:

$$\begin{array}{rrrl} LP: & z^* := & \text{minimum}_x & c^T x \\ & & \text{s.t.} & Ax = b \\ & & & c^T x < c^T \bar{x} \\ & & & x \geq 0 \end{array}$$

Transformed Problem [≡ Renegar]

$$
\begin{array}{llll}
LP: & z^* := & \text{minimum}_x & c^T x \\
 & & \text{s.t.} & Ax = b \\
 & & & c^T x < c^T \bar{x} \\
 & & & x \geq 0
\end{array}
$$

Notation: $\bar{X} := \text{diag}(\bar{x}_1, \ldots, \bar{x}_n)$

Given the scalar $\delta > 0$:

$$
\begin{array}{lll}
TP: & \text{minimum}_d & \max_j (\bar{X}^{-1} d)_j \\
 & \text{s.t.} & Ad = 0 \\
 & & c^T d = \delta
\end{array}
$$

Transformed Problem, continued

Given the scalar $\delta > 0$:

$$LP: \quad \text{minimum}_x \quad c^T x \qquad \text{s.t.} \quad Ax = b, \quad c^T x < c^T \bar{x}, \quad x \geq 0$$

$$TP: \quad \text{minimum}_d \quad \max_j (\bar{X}^{-1} d)_j \qquad \text{s.t.} \quad Ad = 0, \quad c^T d = \delta$$

$$x \leftarrow \bar{x} - \frac{d}{\max_j (\bar{X}^{-1} d)_j} \qquad d \leftarrow \frac{\delta(\bar{x} - x)}{c^T \bar{x} - c^T x}$$

$$c^T x \leftarrow c^T \bar{x} - \frac{\delta}{\max_j (\bar{X}^{-1} d)_j} \qquad \max_j (\bar{X}^{-1} d)_j \leftarrow \frac{\delta \left(1 - \min_j (\bar{X}^{-1} x)_j\right)}{c^T \bar{x} - c^T x}$$

The Non-smooth Optimization Problem, continued

$$
\begin{aligned}
TP: \quad \text{minimum}_d \quad & f(d) := \max_j (\bar{X}^{-1} d)_j \\
\text{s.t.} \quad & Ad = 0 \\
& c^T d = \delta
\end{aligned}
$$

TP is in an excellent format for solution via a first-order method (FOM):

$$
\begin{aligned}
P: \quad \text{minimum}_x \quad & f(x) \\
\text{s.t.} \quad & x \in Q
\end{aligned}
$$

Here $Q = \{d : Ad = 0, \ c^T d = \delta\}$

Note that $f(\cdot)$ in TP is non-smooth convex with $M = \max_j \{1/\bar{x}_j\}$

Aspiration: Compute an ε'-Relative Solution of LP

Aspiration: Compute x feasible for LP that satisfies:

$$\frac{c^T x - z^*}{c^T \bar{x} - z^*} \ \leq \ \varepsilon'$$

Computing an ε'-Relative Solution of LP via Subgradient Descent

Algorithm for computing an ε'-relative solution of LP

- Given LP for which z^* is finite .
- Given $\bar{x}$ satisfying $A\bar{x} = b$, $\bar{x} > 0$, and $\varepsilon' \in (0, 1)$
- Given x^0 feasible for LP with corresponding value d^0 feasible for TP:
- Run the Subgradient Descent method on the transformed problem TP starting at d^0 with a particular step-size sequence $\{\alpha_i\}$, generating iterates $\{d^i\}$ for TP with corresponding sequence $\{x^i\}$ of re-transformed iterates for LP

Computational Guarantee for the Algorithm

Theorem: A Computational Guarantee [Renegar]

Let the number of Subgradient Descent iterations k satisfy:

$$k \ \geq \ 8L^2 \text{Diam}_{\max}^2 \left(\left(\frac{1}{\varepsilon'} \right) \times 3.5 \times \ln \left(\frac{c^T \bar{x} - z^*}{c^T \bar{x} - c^T x^0} \right) \ + \ \left(\frac{1}{\varepsilon'} \right)^2 + 1 \right) \ .$$

Then using a particular step-size rule, the following holds:

$$\frac{c^T x_b^k - z^*}{c^T \bar{x} - z^*} \ \leq \ \varepsilon' \ .$$

Reneger's step-size rule is a minor variant of a standard step-size rule for Subgradient Descent

Level slices: $\text{Slice}_\alpha := \{x : Ax = b, \ x \geq 0, \ c^T x = \alpha\}$

$\text{Diam}(\text{Slice}_\alpha) := \max\{\|x - y\| : x, y \in \text{Slice}_\alpha\}$

$\text{Diam}_{\max} := \max\{\text{Diam}(\text{Slice}_\alpha) : \alpha \in [z^*, c^T x^0]\}$

Computational Guarantee for the Algorithm

Theorem: A Computational Guarantee [Renegar]

Let the number of Subgradient Descent iterations k satisfy:

$$k \ \geq \ 8L^2 \mathrm{Diam}_{\max}^2 \left(\left(\frac{1}{\varepsilon'}\right) \times 3.5 \times \ln\left(\frac{c^T\bar{x} - z^*}{c^T\bar{x} - c^Tx^0}\right) \ + \ \left(\frac{1}{\varepsilon'}\right)^2 + 1 \right) \ .$$

Then using a particular step-size rule, the following holds:

$$\frac{c^T x_b^k - z^*}{c^T\bar{x} - z^*} \ \leq \ \varepsilon' \ .$$

Reneger's step-size rule is a minor variant of a standard step-size rule for Subgradient Descent

Level slices: $\mathrm{Slice}_\alpha := \{x : Ax = b, \ x \geq 0, \ c^T x = \alpha\}$

$\mathrm{Diam}(\mathrm{Slice}_\alpha) := \max\{\|x - y\| : x, y \in \mathrm{Slice}_\alpha\}$

$\mathrm{Diam}_{\max} := \max\{\mathrm{Diam}(\mathrm{Slice}_\alpha) : \alpha \in [z^*, c^T x^0]\}$

Computational Guarantee for the Algorithm

Theorem: A Computational Guarantee [Renegar]

Let the number of Subgradient Descent iterations k satisfy:

$$k \ \geq \ 8L^2 \mathrm{Diam}_{\max}^2 \left(\left(\frac{1}{\varepsilon'}\right) \times 3.5 \times \ln\left(\frac{c^T\bar{x} - z^*}{c^T\bar{x} - c^Tx^0}\right) \ + \ \left(\frac{1}{\varepsilon'}\right)^2 + 1 \right) \ .$$

Then using a particular step-size rule, the following holds:

$$\frac{c^Tx_b^k - z^*}{c^T\bar{x} - z^*} \ \leq \ \varepsilon' \ .$$

Reneger's step-size rule is a minor variant of a standard step-size rule for Subgradient Descent

Level slices: $\mathrm{Slice}_\alpha := \{x : Ax = b, \ x \geq 0, \ c^Tx = \alpha\}$

$\mathrm{Diam}(\mathrm{Slice}_\alpha) := \max\{\|x - y\| : x, y \in \mathrm{Slice}_\alpha\}$

$\mathrm{Diam}_{\max} := \max\{\mathrm{Diam}(\mathrm{Slice}_\alpha) : \alpha \in [z^*, c^Tx^0]\}$

Computational Guarantee for the Algorithm

Theorem: A Computational Guarantee [Renegar]

Let the number of Subgradient Descent iterations k satisfy:

$$k \ \geq \ 8L^2 \text{Diam}^2_{\max} \left(\left(\frac{1}{\varepsilon'}\right) \times 3.5 \times \ln\left(\frac{c^T\bar{x} - z^*}{c^T\bar{x} - c^Tx^0}\right) \ + \ \left(\frac{1}{\varepsilon'}\right)^2 + 1 \right) \ .$$

Then using a particular step-size rule, the following holds:

$$\frac{c^Tx_b^k - z^*}{c^T\bar{x} - z^*} \ \leq \ \varepsilon' \ .$$

Reneger's step-size rule is a minor variant of a standard step-size rule for Subgradient Descent

Level slices: $\text{Slice}_\alpha := \{x : Ax = b, \ x \geq 0, \ c^Tx = \alpha\}$

$\text{Diam}(\text{Slice}_\alpha) := \max\{\|x - y\| : x, y \in \text{Slice}_\alpha\}$

$\text{Diam}_{\max} := \max\{\text{Diam}(\text{Slice}_\alpha) : \alpha \in [z^*, c^Tx^0]\}$

Computational Guarantee for the Algorithm

Theorem: A Computational Guarantee [Renegar]

Let the number of Subgradient Descent iterations k satisfy:

$$k \ \geq \ 8L^2 \mathrm{Diam}_{\max}^2 \left(\left(\frac{1}{\varepsilon'} \right) \times 3.5 \times \ln \left(\frac{c^T\bar{x} - z^*}{c^T\bar{x} - c^Tx^0} \right) \ + \ \left(\frac{1}{\varepsilon'} \right)^2 + 1 \right) \ .$$

Then using a particular step-size rule, the following holds:

$$\frac{c^T x_b^k - z^*}{c^T\bar{x} - z^*} \ \leq \ \varepsilon' \ .$$

Reneger's step-size rule is a minor variant of a standard step-size rule for Subgradient Descent

Level slices: $\mathrm{Slice}_\alpha := \{x : Ax = b, \ x \geq 0, \ c^Tx = \alpha\}$

$\mathrm{Diam}(\mathrm{Slice}_\alpha) := \max\{\|x - y\| : x, y \in \mathrm{Slice}_\alpha\}$

$\mathrm{Diam}_{\max} := \max\{\mathrm{Diam}(\mathrm{Slice}_\alpha) : \alpha \in [z^*, c^Tx^0]\}$

Research Questions

- Is this result only specific to LP and/or TP?
- Or is this result an instance of a more general theory?
- If so, what is the general theory and how does it apply to different optimization problems solved with FOMs?

New Theory for First-Order Methods via a Function Growth Measure

Let us consider the general setting:

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

$f(\cdot)$ is convex on Q

Q is a closed convex set

Strict Lower Bound f_{slb}

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

Let f_{slb} be a known and given strict lower bound on f^*, namely: $f_{slb} < f^*$

f_{slb} arises naturally in optimizing loss functions in statistics and machine learning:

- $f_{slb} = 0$ for exponential loss: $f(x) = \ln\left(\frac{1}{m}\sum_{i=1}^m e^{-A_i x}\right) + \lambda\|x\|_p^r$
- $f_{slb} = 0$ for logistic loss: $f(x) = \frac{1}{m}\sum_{i=1}^m \ln\left(1 + e^{-A_i x}\right) + \lambda\|x\|_p^r$
- $f_{slb} = 0$ for regularized least-squares loss: $f(\beta) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda\|\beta\|_p^r$
- $f_{slb} = 0$ in Renegar's transformed problem TP, when LP primal has an optimum

ε'-Relative Optimal Solution

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

Let $\varepsilon' > 0$ be given.

Definition: ε'-relative solution of P

An ε'-relative solution of P is a point $x \in Q$ that satisfies:

$$\frac{f(x) - f^*}{f^* - f_{slb}} \leq \varepsilon' .$$

In the often-case when $f_{slb} = 0$, then this becomes:

$$\frac{f(x)}{f^*} \leq 1 + \varepsilon'$$

Function Growth Constant G

$$P: \quad f^* := \quad \text{minimum}_x \quad f(x)$$
$$\text{s.t.} \quad x \in Q$$

Suppose we have a strict lower bound f_{slb} on f^*, namely $f_{slb} < f^*$

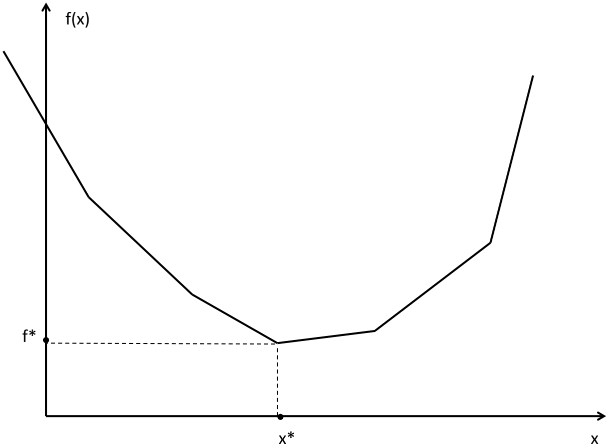
$\text{Opt} := \{x \in Q : f(x) = f^*\}$

$\text{Dist}(x, \text{Opt}) := \min_y \{\|y - x\| : y \in \text{Opt}\}$

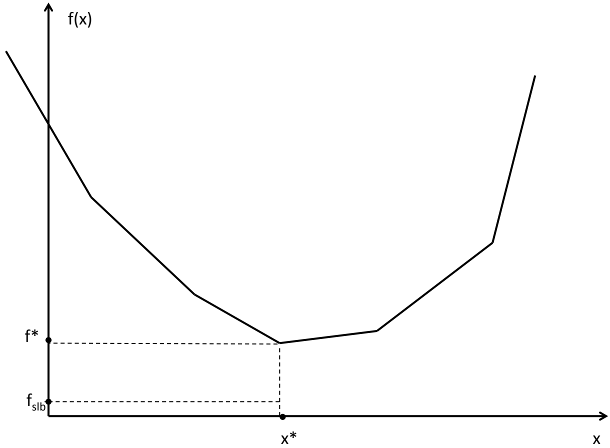
Definition: function growth constant G

$$G := \sup_{x \in Q} \left\{ \frac{\text{Dist}(x, \text{Opt})}{f(x) - f_{slb}} \right\}$$

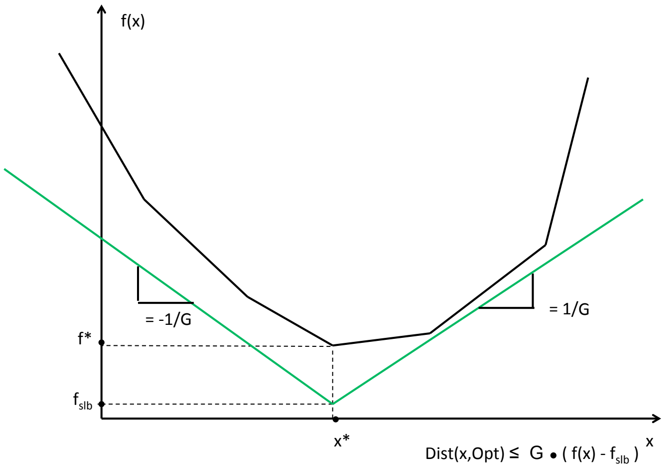
Geometric Picture of G



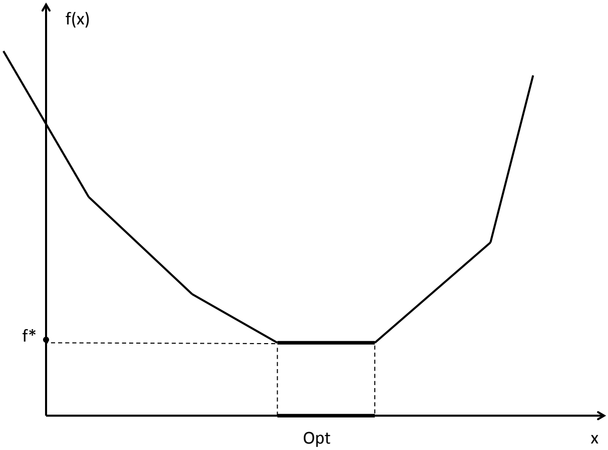
Geometric Picture of G



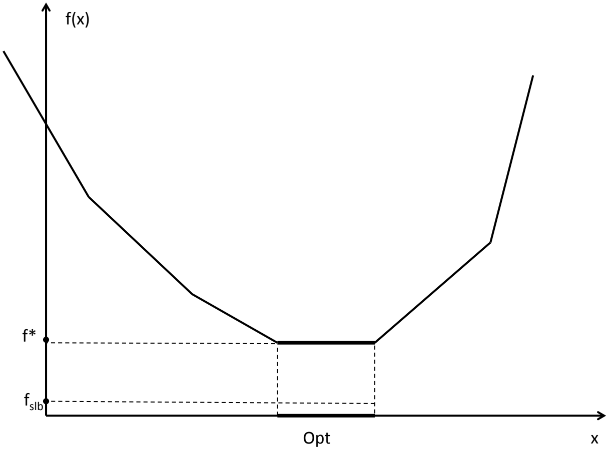
Geometric Picture of G



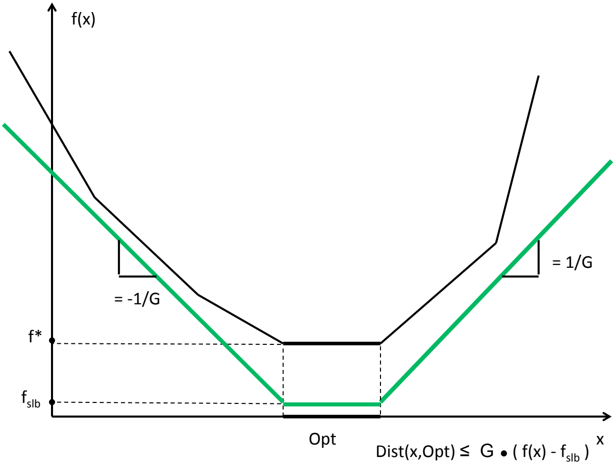
Geometric Picture of G



Geometric Picture of G



Geometric Picture of G



Function Growth Constant G, continued

$$G := \sup_{x \in Q} \left\{ \frac{\text{Dist}(x, \text{Opt})}{f(x) - f_{slb}} \right\}$$

Then G is the smallest value of $\bar{G}$ satisfying:

$$\text{Dist}(x, \text{Opt}) \leq \bar{G} \cdot (f(x) - f_{slb}) \text{ for all } x \in Q$$

G measures how quickly the distances from the optimal solutions grow with increasing function values.

More Interpretation of G

$$\text{Dist}(x, \text{Opt}) \le G \cdot (f(x) - f_{slb}) \quad \text{for all } x \in Q$$

This rearranges to:

$$f(x) \;\ge\; \bar{f}(x) := f_{slb} + G^{-1}\text{Dist}(x, \text{Opt}) \quad \text{for all } x \in Q$$

The convex function $\bar{f}(\cdot) := f_{slb} + G^{-1}\text{Dist}(\cdot, \text{Opt})$ lies below $f(\cdot)$

Q: When is G finite? A: "Almost always."

ε-optimal level set: $\text{Opt}_\varepsilon := \{x \in Q : \ f(x) \le f^* + \varepsilon\}$

Theorem: Sufficient Conditions for $G < +\infty$

Suppose that for some $\varepsilon > 0$ there exists a bounded set E_ε for which $\text{Opt}_\varepsilon \subset E_\varepsilon + S$, where S is the recession cone of Opt_ε. Then for any given strict lower bound $f_{slb} < f^*$, the growth constant G is finite.

Implication:

- If Opt is bounded, then G is finite.
- If $\text{Opt} = E + T$ where E is bounded and T is a subspace, then G is finite.

An instance where $G = +\infty$:

$$Q := \{(x_1, x_2) : x_1 \ge 1\}$$
$$f(x_1, x_2) := \frac{x_2^2}{x_1}$$

Non-Smooth Optimization

New Computational Guarantees for Non-smooth Optimization

New Computational Guarantees for Subgradient Descent

Theorem: Computational Guarantee for Subgradient Descent

Let $\varepsilon' > 0$ be given, and let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \left(\frac{f_b^i - f_{slb}}{\sqrt[3]{e}\|g^i\|^2}\right)\left(\frac{\varepsilon'}{1+\varepsilon'}\right) ,$$

and suppose that

$$k \ \geq \ M^2 G^2 \left[16\left(\frac{1+\varepsilon'}{\varepsilon'}\right)\ln\left(\frac{f(x^0)-f^*}{f^*-f_{slb}}\right) \ + \ 11\left(\frac{1+\varepsilon'}{\varepsilon'}\right)^2\right] .$$

Then:

$$\frac{f(x_b^k)-f^*}{f^*-f_{slb}} \ \leq \varepsilon' .$$

Here $e = 2.718\ldots$

New Computational Guarantees for Subgradient Descent

Theorem: Computational Guarantee for Subgradient Descent

Let $\varepsilon' > 0$ be given, and let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \left(\frac{f_b^i - f_{slb}}{\sqrt[3]{e}\|g^i\|^2}\right)\left(\frac{\varepsilon'}{1+\varepsilon'}\right) ,$$

and suppose that

$$k \ \geq \ M^2 G^2 \left[16\left(\frac{1+\varepsilon'}{\varepsilon'}\right)\ln\left(\frac{f(x^0)-f^*}{f^*-f_{slb}}\right) \ + \ 11\left(\frac{1+\varepsilon'}{\varepsilon'}\right)^2\right] .$$

Then:

$$\frac{f(x_b^k)-f^*}{f^*-f_{slb}} \ \leq \varepsilon' .$$

Here $e = 2.718...$

New Computational Guarantees for Subgradient Descent

Theorem: Computational Guarantee for Subgradient Descent

Let $\varepsilon' > 0$ be given, and let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \left(\frac{f_b^i - f_{slb}}{\sqrt[3]{e}\|g^i\|^2}\right)\left(\frac{\varepsilon'}{1+\varepsilon'}\right) ,$$

and suppose that

$$k \ \geq \ M^2 G^2 \left[16\left(\frac{1+\varepsilon'}{\varepsilon'}\right)\ln\left(\frac{f(x^0) - f^*}{f^* - f_{slb}}\right) \ + \ 11\left(\frac{1+\varepsilon'}{\varepsilon'}\right)^2\right] .$$

Then:

$$\frac{f(x_b^k) - f^*}{f^* - f_{slb}} \ \leq \varepsilon' .$$

Here $e = 2.718...$

New Computational Guarantees for Subgradient Descent

Theorem: Computational Guarantee for Subgradient Descent

Let $\varepsilon' > 0$ be given, and let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \left(\frac{f_b^i - f_{slb}}{\sqrt[3]{e}\|g^i\|^2}\right)\left(\frac{\varepsilon'}{1+\varepsilon'}\right) ,$$

and suppose that

$$k \ \geq \ M^2 G^2 \left[16\left(\frac{1+\varepsilon'}{\varepsilon'}\right) \ln\left(\frac{f(x^0) - f^*}{f^* - f_{slb}}\right) \ + \ 11\left(\frac{1+\varepsilon'}{\varepsilon'}\right)^2\right] .$$

Then:

$$\frac{f(x_b^k) - f^*}{f^* - f_{slb}} \ \leq \varepsilon' .$$

Here $e = 2.718...$

Comparison with the Standard Computational Guarantee for Subgradient Descent

Define: $\bar{C} := \dfrac{\text{Dist}(x^0, \text{Opt})}{G(f^* - f_{slb})}$

$$\frac{\text{New Guarantee}}{\text{Standard Guarantee for Sub.Descent}} \leq \frac{16\varepsilon' \ln\left(1 + MG\bar{C}\right)}{\bar{C}^2} + \frac{11}{\bar{C}^2}$$

This ratio $\rightarrow 0$ when $\text{Dist}(x^0, \text{Opt})$ is sufficiently large

And this is true for any problem instance

New Computational Guarantees for Subgradient Descent when f^* is known

Theorem: Computational Guarantee for Subgradient Descent when f^* is known

Let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \frac{f(x^i) - f^*}{\|g^i\|^2} \ ,$$

and suppose that

$$k \ \geq \ 2M^2G^2 \left[1 + 2.9 \ln\left(\frac{f(x^0) - f^*}{f^* - f_{slb}}\right) + 2.9 \ln\left(\frac{1}{\varepsilon'}\right) + 6.8 \left(\frac{1}{\varepsilon'}\right) + 2\left(\frac{1}{\varepsilon'}\right)^2 \right] .$$

Then:

$$\frac{f(x_b^k) - f^*}{f^* - f_{slb}} \ \leq \varepsilon' \ .$$

New Computational Guarantees for Subgradient Descent when f^* is known

Theorem: Computational Guarantee for Subgradient Descent when f^* is known

Let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \frac{f(x^i) - f^*}{\|g^i\|^2} \ ,$$

and suppose that

$$k \ \geq \ 2M^2 G^2 \left[1 + 2.9 \ln\left(\frac{f(x^0) - f^*}{f^* - f_{slb}}\right) + 2.9 \ln\left(\frac{1}{\varepsilon'}\right) + 6.8\left(\frac{1}{\varepsilon'}\right) + 2\left(\frac{1}{\varepsilon'}\right)^2\right] .$$

Then:

$$\frac{f(x_b^k) - f^*}{f^* - f_{slb}} \ \leq \varepsilon' \ .$$

New Computational Guarantees for Subgradient Descent when f^* is known

Theorem: Computational Guarantee for Subgradient Descent when f^* is known

Let the step-sizes for Subgradient Descent Method applied to solve P be chosen as:

$$\alpha_i := \frac{f(x^i) - f^*}{\|g^i\|^2} \ ,$$

and suppose that

$$k \ \geq \ 2M^2G^2 \left[1 + 2.9\ln\left(\frac{f(x^0) - f^*}{f^* - f_{slb}}\right) + 2.9\ln\left(\frac{1}{\varepsilon'}\right) + 6.8\left(\frac{1}{\varepsilon'}\right) + 2\left(\frac{1}{\varepsilon'}\right)^2\right] \ .$$

Then:

$$\frac{f(x_b^k) - f^*}{f^* - f_{slb}} \ \leq \varepsilon' \ .$$

Improving the Guarantee for Non-Smooth Optimization using (Nesterov-style) Smooth Approximations

Suppose that there is a smoothing technique with the following two properties:

1. there is a known constant $\bar{D} > 0$ such that for any given $\mu > 0$ we can construct a smooth convex function $f_\mu(\cdot) : Q \to \mathbb{R}$ which satisfies:

$$f(x) \le f_\mu(x) \le f(x) + \bar{D}\mu \text{ for all } x \in Q \ , \text{ and}$$

2. $f_\mu(\cdot)$ has Lipschitz continuous gradient on Q with Lipschitz constant $L_\mu \le A/\mu$ for some known constant A

Nesterov [2005] showed how to optimize $f(\cdot)$ by instead working with the smooth function $f_\mu(\cdot)$ for a well-chosen value of μ

Smooth Approximations Method

Smooth Approximations Method

Initialize with $x^0 \in Q$ and $\varepsilon' > 0$.
Set $x_{1,0} \leftarrow x^0$, $i \leftarrow 1$.

At outer iteration i:

1. **Set smoothing parameter.** $\mu_i \leftarrow \dfrac{\varepsilon' \cdot (f(x_{i,0}) - f_{slb})}{5\bar{D}}$.
2. **Initialize inner iteration.** $j \leftarrow 0$
3. **Run inner iterations.** At inner iteration j:

   If $\dfrac{f(x_{i,j}) - f_{slb}}{f(x_{i,0}) - f_{slb}} \geq 0.8$, then:

   $$x_{i,j+1} \leftarrow \text{AGM}(f_{\mu_i}(\cdot),\ x_{i,0},\ j+1) ,$$

   $j \leftarrow j+1$, and Goto step 3.

   Else $x_{i+1,0} \leftarrow x_{i,j}$, $i \leftarrow i+1$, and Goto step 1.

"$x_{i,j} \leftarrow \text{AGM}(f_{\mu_i}(\cdot),\ x_{i,0},\ j)$" denotes assigning to $x_{i,j}$ the j^{th} iterate of AGM applied with objective function $f_{\mu_i}(\cdot)$ using the initial point $x_{i,0} \in Q$

Computational Guarantee for Smooth Approximations Method

Complexity Bound for Smooth Approximations Method

Let $x^0 \in Q$ be the initial point and let the relative accuracy $\varepsilon' \in (0, 1]$ be given, and let x^k denote the iterate value of the Smooth Approximations Method after a total of k inner iterations. If

$$k \;\geq\; G\sqrt{A}\sqrt{\bar{D}}\left(32\left\lceil \frac{\ln\left(1+\frac{f(x^0)-f^*}{f^*-f_{slb}}\right)}{\sqrt{\varepsilon'}}\right\rceil + 44\left\lceil \frac{1}{\varepsilon'}\right\rceil\right) ,$$

then

$$\frac{f(x^N)-f^*}{f^*-f_{slb}} \;\leq\; \varepsilon' \; .$$

Computational Guarantee for Smooth Approximations Method

Complexity Bound for Smooth Approximations Method

Let $x^0 \in Q$ be the initial point and let the relative accuracy $\varepsilon' \in (0,1]$ be given, and let x^k denote the iterate value of the Smooth Approximations Method after a total of k inner iterations. If

$$k \ \geq \ G\sqrt{A}\sqrt{\bar{D}}\left(32\left[\frac{\ln\left(1+\frac{f(x^0)-f^*}{f^*-f_{slb}}\right)}{\sqrt{\varepsilon'}}\right]+44\left[\frac{1}{\varepsilon'}\right]\right) \ ,$$

then

$$\frac{f(x^N)-f^*}{f^*-f_{slb}} \ \leq \ \varepsilon' \ .$$

Comparison with the Standard Computational Guarantee for Smoothing Method

Define: $\bar{C} := \frac{\text{Dist}(x^0, \text{Opt})}{G(f^* - f_{slb})}$

$$\frac{\text{Guarantee of New Method}}{\text{Standard Guarantee for Smoothing}} \leq \frac{8\sqrt{2}\sqrt{\varepsilon'}\ln\left(1 + MG\bar{C}\right)}{\bar{C}} + \frac{11\sqrt{2}}{\bar{C}} .$$

This ratio $\rightarrow 0$ when $\text{Dist}(x^0, \text{Opt})$ is sufficiently large

And this is true for any problem instance

Smooth Optimization

New Computational Guarantees for Smooth Optimization

Accelerated Gradient Method with Simple Restarting

Accelerated Gradient Method with Simple Restarting

Initialize with $x^0 \in Q$ and $\varepsilon' > 0$.
Set $x_{1,0} \leftarrow x^0$, $i \leftarrow 1$.

At outer iteration i:

1. **Initialize inner iteration.** $j \leftarrow 0$
2. **Run inner iterations.** At inner iteration j:

   If $\dfrac{f(x_{i,j}) - f_{slb}}{f(x_{i,0}) - f_{slb}} \geq 0.8$, then:

   $$x_{i,j+1} \leftarrow \mathrm{AGM}(f(\cdot),\ x_{i,0},\ j+1) ,$$

   $j \leftarrow j+1$, and Goto step 2.

   Else $x_{i+1,0} \leftarrow x_{i,j}$, $i \leftarrow i+1$, and Goto step 1.

"$x_{i,j} \leftarrow \mathrm{AGM}(f_{\mu_i}(\cdot),\ x_{i,0},\ j)$" denotes assigning to $x_{i,j}$ the j^{th} iterate of AGM applied with objective function $f_{\mu_i}(\cdot)$ using the initial point $x_{i,0} \in Q$

Computational Guarantee for AGM with Simple Restarting

Complexity Bound for Accelerated Gradient Method with Simple Restarting

Let $x^0 \in Q$ be the initial point and let x^k denote the iterate value of the Accelerated Gradient Method with Simple Restarting after a total of k inner iterations. If

$$k \ \geq \ G\sqrt{L}\left(17\left\lceil\frac{\sqrt{f^* - f_{slb}}}{\sqrt{\varepsilon'}}\right\rceil \ + \ 22\sqrt{(f(x^0) - f_{slb})}\right) \ ,$$

then

$$\frac{f(x^k) - f^*}{f^* - f_{slb}} \ \leq \ \varepsilon' \ .$$

Computational Guarantee for AGM with Simple Restarting

Complexity Bound for Accelerated Gradient Method with Simple Restarting

Let $x^0 \in Q$ be the initial point and let x^k denote the iterate value of the Accelerated Gradient Method with Simple Restarting after a total of k inner iterations. If

$$k \;\geq\; G\sqrt{L}\left(17\left\lceil\frac{\sqrt{f^* - f_{slb}}}{\sqrt{\varepsilon'}}\right\rceil + 22\sqrt{\left(f(x^0) - f_{slb}\right)}\right) ,$$

then

$$\frac{f(x^k) - f^*}{f^* - f_{slb}} \;\leq\; \varepsilon' .$$

Comparison with the Standard Accelerated Gradient Method

Define: $\bar{C} := \dfrac{\text{Dist}(x^0, \text{Opt})}{G(f^* - f_{slb})}$

$$\frac{\text{New Method Guarantee}}{\text{Std. AGM Guarantee}} \leq \frac{8.5\sqrt{2}}{\bar{C}} + 11\sqrt{\varepsilon'}\sqrt{L}G\sqrt{f^* - f_{slb}} + \frac{11\sqrt{\varepsilon'}}{\bar{C}^2 G\sqrt{L}\sqrt{f^* - f_{slb}}}$$

This ratio $\rightarrow 0$ when $\text{Dist}(x^0, \text{Opt})$ is sufficiently large and $\varepsilon' \rightarrow 0$

And this is true for any problem instance

Improving the Guarantee using Parametric Increased Smoothing

Suppose that $f(\cdot)$ has the representation:

$$f(x) = \max_{\lambda \in P} \{\lambda^T A x - d(\lambda)\}$$

where P is a convex set

$d(\cdot)$ is a σ-strongly convex function on P

$\min_{\lambda \in P} d(\lambda) \geq 0$

Then $f(\cdot)$ is a smooth convex function with $L \leq \|A\|^2/\sigma$ [Nesterov2005]

Parametric Increased Smoothing, continued

$$f(x) = \max_{\lambda \in P}\{\lambda^T Ax - d(\lambda)\}$$

Define:

$$f_\mu(x) = \max_{\lambda \in P}\{\lambda^T Ax - (1+\mu)d(\lambda)\}$$

Then $f(\cdot) = f_0(\cdot)$

$f_\mu(\cdot)$ has a Lipschitz gradient with constant at most $L_\mu := L/(1+\mu)$

If P is bounded, then $\bar{D} := \max_{\lambda \in P}\{d(\lambda)\}$ is finite, and:

$$f(x) - \mu\bar{D} \le f_\mu(x) \le f(x) \text{ for all } x$$

AGM with Parametric Increased Smoothing

AGM with Parametric Increased Smoothing

Initialize with $x^0 \in Q$ and $\varepsilon' > 0$.
Set $x_{1,0} \leftarrow x^0$, $i \leftarrow 1$.

At outer iteration i:

1. **Set smoothing parameter.** $\mu_i \leftarrow \dfrac{\varepsilon' \cdot (f(x_{i,0}) - f_{slb})}{5\bar{D}}$.
2. **Initialize inner iteration.** $j \leftarrow 0$
3. **Run inner iterations.** At inner iteration j:

   If $\dfrac{f(x_{i,j}) - f_{slb}}{f(x_{i,0}) - f_{slb}} \geq 0.8$, then:

   $$x_{i,j+1} \leftarrow \mathrm{AGM}(f_{\mu_i}(\cdot),\ x_{i,0},\ j+1) \ ,$$

   $j \leftarrow j+1$, and Goto step 3.

   Else $x_{i+1,0} \leftarrow x_{i,j}$, $i \leftarrow i+1$, and Goto step 1.

"$x_{i,j} \leftarrow \mathrm{AGM}(f_{\mu_i}(\cdot),\ x_{i,0},\ j)$" denotes assigning to $x_{i,j}$ the j^{th} iterate of AGM applied with objective function $f_{\mu_i}(\cdot)$ using the initial point $x_{i,0} \in Q$

Computational Guarantee for AGM with Parametric Increased Smoothing

Complexity Bound for Accelerated Gradient Method with Parametric Increased Smoothing

Let $x^0 \in Q$ be the initial point and let the relative accuracy $\varepsilon' \in (0,1]$ be given, and let x^k denote the iterate value of the Accelerated Gradient Method with Parametric Increased Smoothing after a total of k inner iterations. If

$$k \ \geq \ G\sqrt{L}\left(24\left[\frac{\sqrt{f^* - f_{slb}}}{\sqrt{\varepsilon'}}\right] + 32\left[\frac{\sqrt{\bar{D}}\ln\left(1 + \frac{f(x^0)-f^*}{f^*-f_{slb}}\right)}{\sqrt{\varepsilon'}}\right]\right) \ ,$$

then

$$\frac{f(x^N) - f^*}{f^* - f_{slb}} \ \leq \ \varepsilon' \ .$$

Computational Guarantee for AGM with Parametric Increased Smoothing

Complexity Bound for Accelerated Gradient Method with Parametric Increased Smoothing

Let $x^0 \in Q$ be the initial point and let the relative accuracy $\varepsilon' \in (0,1]$ be given, and let x^k denote the iterate value of the Accelerated Gradient Method with Parametric Increased Smoothing after a total of k inner iterations. If

$$k \ \geq \ G\sqrt{L}\left(24\left[\frac{\sqrt{f^* - f_{slb}}}{\sqrt{\varepsilon'}}\right] + 32\left[\frac{\sqrt{\bar{D}}\ln\left(1 + \frac{f(x^0)-f^*}{f^*-f_{slb}}\right)}{\sqrt{\varepsilon'}}\right]\right) \ ,$$

then

$$\frac{f(x^N) - f^*}{f^* - f_{slb}} \ \leq \ \varepsilon' \ .$$

Remarks, Extensions, Next Steps

Computational Testing:

- Non-smooth Optimization: LASSO, Support Vector Machines (dual problem)
- Smooth Optimization: logistic regression, binary classificaition
- Conic Optimization (which engendered Renegar's research)
  - Homogeneous self-dual embedding
  - SDP problems in particular (discussions with Franz Rendl)