Minimization of convex compositions

Dmitriy Drusvyatskiy
Mathematics, University of Washington

Joint work with
A.S. Lewis (Cornell), C. Paquette (UW)

FOCM 2017, Continuous Optimization

AFOSR, NSF

Convex composition:

$$\min_{x} \; F(x) = g(x) + h(c(x))$$

Nonsmooth & Nonconvex minimization

Convex composition:

$$\min_{x} \; F(x) = g(x) + h(c(x))$$

where

- $g\colon \mathbf{R}^d \to \overline{\mathbf{R}}$ is closed, convex.
- $h\colon \mathbf{R}^m \to \mathbf{R}$ is convex and 1-Lipschitz.
- $c\colon \mathbf{R}^d \to \mathbf{R}^m$ is C^1-smooth and ∇c is β-Lipschitz.

Nonsmooth & Nonconvex minimization

Convex composition:

$$\min_x \; F(x) = g(x) + h(c(x))$$

where

- $g\colon \mathbf{R}^d \to \overline{\mathbf{R}}$ is closed, convex.
- $h\colon \mathbf{R}^m \to \mathbf{R}$ is convex and 1-Lipschitz.
- $c\colon \mathbf{R}^d \to \mathbf{R}^m$ is C^1-smooth and ∇c is β-Lipschitz.

(Burke '85, Cartis-Gould-Toint '11, Fletcher '82, Lewis-Wright '15, Nesterov '06, Powell '84, Wright '90, Yuan '83, ...)

Examples

$$\boxed{\min_x \; g(x) + h(c(x))}$$

Examples:

- Additive composition:

$$\min_x \; g(x) + c(x)$$

- Least-norm misfit: $\min_{x \in Q} \; \|c(x)\|$

  - Robust Phase Retrieval: $\min_x \sum_{i=1}^{m} |(a_i^T x)^2 - b_i|$
  - Nonneg. Factorization: $\min_{X,Y \geq 0} \; \|XY^T - D\|$

Prox-linear algorithm

$$\boxed{\min_{x} \; F(x) = g(x) + h\big(c(x)\big)}$$

Prox-linear algorithm

$$\boxed{\min_x \; F(x) = g(x) + h(c(x))}$$

Local Model:

$$F_x(y) := g(y) + h\Big(c(x) + \nabla c(x)(y - x)\Big)$$

Prox-linear algorithm

$$\boxed{\min_x \; F(x) = g(x) + h(c(x))}$$

Local Model:

$$F_x(y) := g(y) + h\Big(c(x) + \nabla c(x)(y - x)\Big)$$

Accuracy: $$|F_x(y) - F(y)| \leq \frac{\beta}{2}\|y - x\|^2 \qquad \forall x, y$$

Prox-linear algorithm

$$\boxed{\min_x \; F(x) = g(x) + h(c(x))}$$

Local Model:

$$F_x(y) := g(y) + h\Big(c(x) + \nabla c(x)(y - x)\Big)$$

Accuracy: $$|F_x(y) - F(y)| \leq \frac{\beta}{2}\|y - x\|^2 \qquad \forall x, y$$

Prox-linear method (Burke, Fletcher, Nesterov, Powell, ...):

$$x^+ = \operatorname*{argmin}_y \; \left\{F_x(y) + \frac{\beta}{2}\|y - x\|^2\right\}$$

Prox-linear algorithm

$$\boxed{\min_x \; F(x) = g(x) + h(c(x))}$$

Local Model:

$$F_x(y) := g(y) + h\Big(c(x) + \nabla c(x)(y - x)\Big)$$

Accuracy: $$|F_x(y) - F(y)| \le \tfrac{\beta}{2}\|y - x\|^2 \qquad \forall x, y$$

Prox-linear method (Burke, Fletcher, Nesterov, Powell, ...):

$$x^+ = \operatorname*{argmin}_y \; \left\{F_x(y) + \tfrac{\beta}{2}\|y - x\|^2\right\}$$

Big assumption: x^+ is computable (for now)

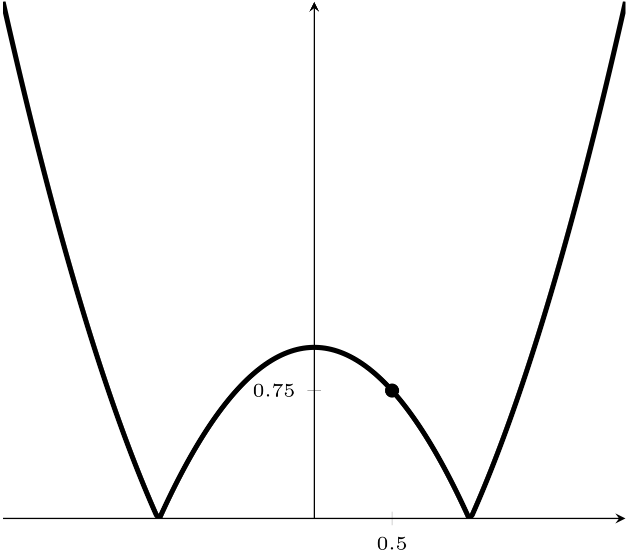


Figure: $f(x) = |x^2 - 1|$

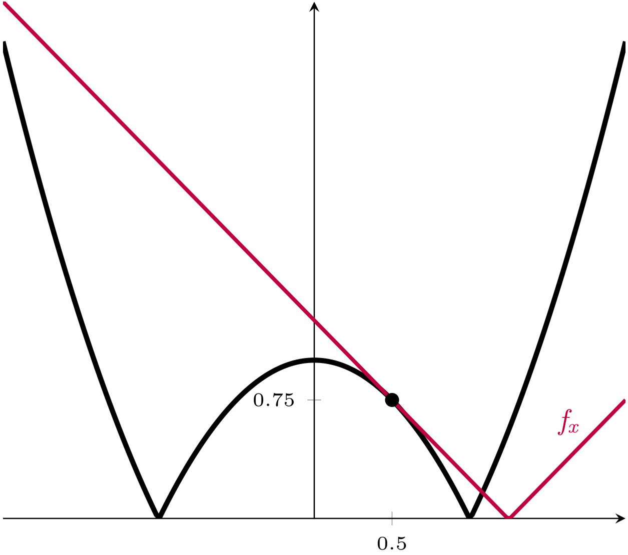


Figure: $f(x) = |x^2 - 1|$

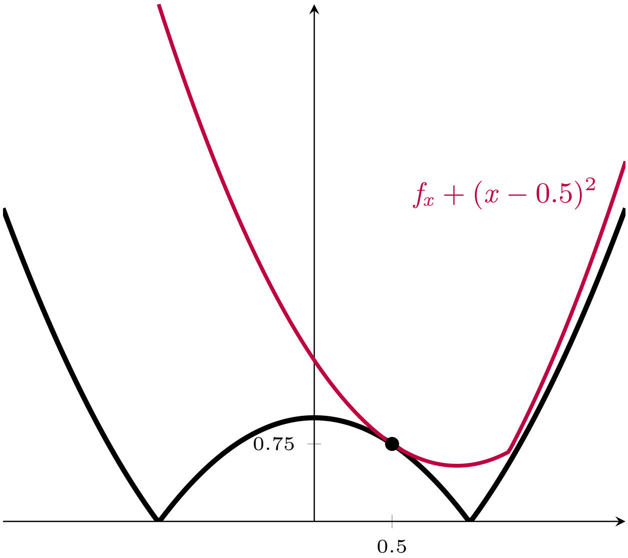


Figure: $f(x) = |x^2 - 1|$

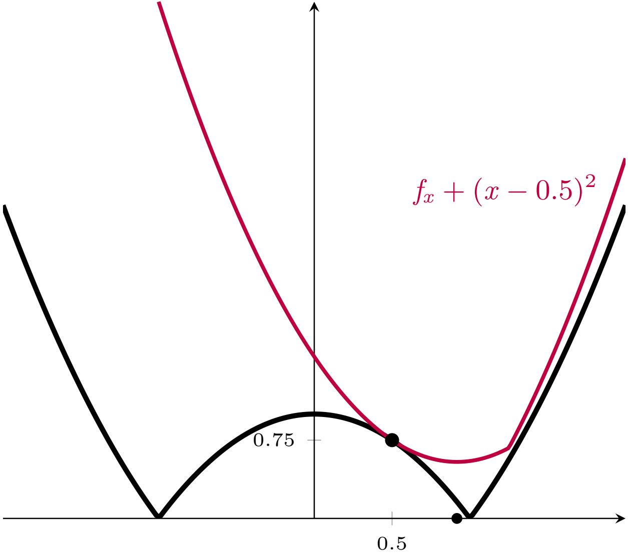


Figure: $f(x) = |x^2 - 1|$

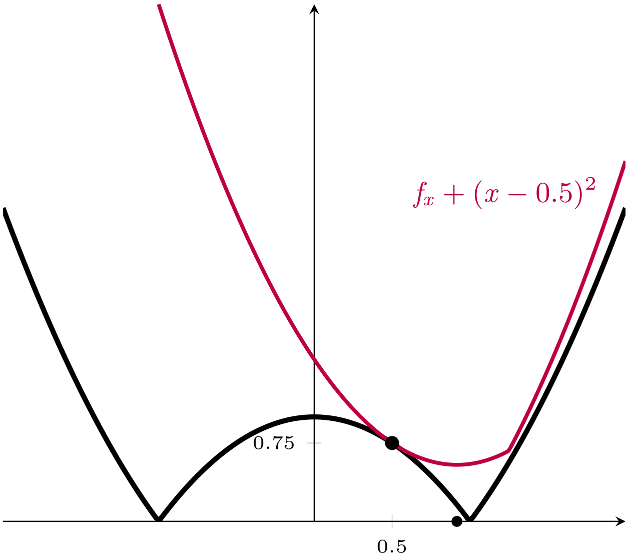


Figure: $f(x) = |x^2 - 1|$

- No finite termination

Sublinear rate

Prox-gradient:

$$\boxed{\mathcal{G}(x) := \beta(x - x^+)}$$

Sublinear rate

Prox-gradient:

$$\boxed{\mathcal{G}(x) := \beta(x - x^+)}$$

Philosophy (Nesterov '13): $x^+ = x - \frac{1}{\beta}\mathcal{G}(x)$ and so

$$\|\mathcal{G}(x)\| \qquad \approx \qquad \text{``}\|\nabla F(x)\|'' \qquad \text{[more later]}$$

Sublinear rate

Prox-gradient:

$$\boxed{\mathcal{G}(x) := \beta(x - x^+)}$$

Philosophy (Nesterov '13): $x^+ = x - \frac{1}{\beta}\mathcal{G}(x)$ and so

$$\|\mathcal{G}(x)\| \qquad \approx \qquad \text{``}\|\nabla F(x)\|'' \qquad \text{[more later]}$$

Sublinear rate:

$$\|\mathcal{G}(x_k)\| < \epsilon \quad \text{after} \quad \mathcal{O}\left(\frac{\beta}{\varepsilon^2} \cdot (F(x_0) - F^*)\right) \text{ iterations}$$

Sublinear rate

Prox-gradient:

$$\boxed{\mathcal{G}(x) := \beta(x - x^+)}$$

Philosophy (Nesterov '13): $x^+ = x - \frac{1}{\beta}\mathcal{G}(x)$ and so

$$\|\mathcal{G}(x)\| \qquad \approx \qquad \text{``}\|\nabla F(x)\|'' \qquad \text{[more later]}$$

Sublinear rate:

$$\|\mathcal{G}(x_k)\| < \epsilon \quad \text{after} \quad \mathcal{O}\left(\frac{\beta}{\varepsilon^2} \cdot (F(x_0) - F^*)\right) \quad \text{iterations}$$

- d, m are large $\implies$ x^+ inexactly computable

First-order complexity

First-order complexity of a method:

Max # of

- $c(x), \nabla c(x) v$, $\nabla c(x)^T v$,
- $\operatorname{prox}_{th}(x), \operatorname{prox}_{tg}(x)$

a method evaluates to produce $\|\mathcal{G}(x)\| \leq \varepsilon$.

First-order complexity

First-order complexity of a method:

Max # of

- $c(x), \nabla c(x) v$, $\nabla c(x)^T v$,
- $\text{prox}_{th}(x), \text{prox}_{tg}(x)$

a method evaluates to produce $\|\mathcal{G}(x)\| \leq \varepsilon$.

Thm: (D-Paquette '16) Define

$$\|\nabla c\| := \max_{x \in \text{dom } g} \|\nabla c(x)\|.$$

There exists a method with complexity

$$\widetilde{\mathcal{O}}\left(\frac{\beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right).$$

First-order complexity

First-order complexity of a method:

Max # of

- $c(x), \nabla c(x)v, \nabla c(x)^T v,$
- $\mathrm{prox}_{th}(x), \mathrm{prox}_{tg}(x)$

a method evaluates to produce $\|\mathcal{G}(x)\| \leq \varepsilon$.

Thm: (D-Paquette '16) Define

$$\|\nabla c\| := \max_{x \in \mathrm{dom}\, g} \|\nabla c(x)\|.$$

There exists a method with complexity

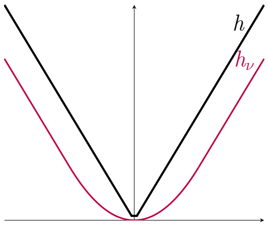
$$\widetilde{\mathcal{O}}\left(\frac{\beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right).$$

Strategy:

- smoothing of h,
- inexact prox-linear method,
- fast-gradient subsolve.

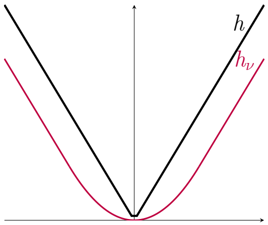
Smoothing

Moreau envelope: $$h_\nu(x) := \min_y \left\{h(y) + \frac{1}{2\nu}\|y - x\|^2\right\}.$$



Smoothing

Moreau envelope: $$h_\nu(x) := \min_y \{h(y) + \frac{1}{2\nu}\|y - x\|^2\}.$$



Lemma: For $\nu > 0$, consider

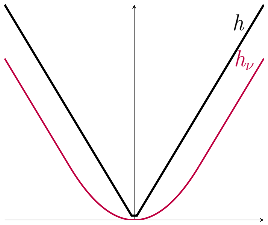
$$\min_x \; g(x) + h_\nu(c(x)).$$

Then

$$\|\mathcal{G}(x)\| \leq \|\mathcal{G}^\nu(x)\| + \sqrt{\frac{\beta\nu}{2}}.$$

Smoothing

Moreau envelope: $$h_\nu(x) := \min_y \{h(y) + \tfrac{1}{2\nu}\|y - x\|^2\}.$$



Lemma: For $\nu > 0$, consider

$$\min_x \; g(x) + h_\nu(c(x)).$$

Then

$$\|\mathcal{G}(x)\| \le \|\mathcal{G}^\nu(x)\| + \sqrt{\frac{\beta\nu}{2}}.$$

Therefore set $\frac{\varepsilon}{2} = \sqrt{\frac{\beta\nu}{2}} \quad \Longrightarrow \quad \nu = \frac{\varepsilon^2}{2\beta}$.

Inexact prox-linear

Input: A point $x_0 \in \operatorname{dom} g$, a method $\mathcal{M}$
Step k: $(k \geq 1)$
Consider: $\min_y \; \varphi(y) := F_{x_k}(y) + \frac{\beta}{2}\|y - x_k\|^2.$

Suppose $\mathcal{M}$ is linearly convergent from $y_0 = x_k$:

$$\varphi(y_i) - \varphi(y^*) \leq \gamma(1-\tau)^i \|y_0 - y^*\|^2 \qquad \forall i = 1, 2 \ldots$$

Set $x_{k+1} = y_T$ where $T := \left\lceil \frac{1}{\tau} \log(4\gamma/\beta) \right\rceil$.

[motivated by Lin-Mairal-Harchauoi '16]

Inexact prox-linear

Input: A point $x_0 \in \operatorname{dom} g$, a method $\mathcal{M}$
Step k: $(k \geq 1)$
Consider: $$\min_y \; \varphi(y) := F_{x_k}(y) + \frac{\beta}{2}\|y - x_k\|^2.$$

Suppose $\mathcal{M}$ is linearly convergent from $y_0 = x_k$:

$$\varphi(y_i) - \varphi(y^*) \leq \gamma(1-\tau)^i \|y_0 - y^*\|^2 \qquad \forall i = 1, 2 \ldots$$

Set $x_{k+1} = y_T$ where $T := \left\lceil \frac{1}{\tau} \log(4\gamma/\beta) \right\rceil$.

[motivated by Lin-Mairal-Harchauoi '16]

Thm:

$$\|\mathcal{G}(x_k)\| < \epsilon \quad \text{after} \quad \mathcal{O}\left(\frac{\beta}{\varepsilon^2} \cdot (F(x_0) - F^*)\right) \text{ outer iterations.}$$

Inexact prox-linear

Input: A point $x_0 \in \operatorname{dom} g$, a method $\mathcal{M}$
Step k: $(k \geq 1)$
Consider: $\min_y \ \varphi(y) := F_{x_k}(y) + \frac{\beta}{2}\|y - x_k\|^2.$

Suppose $\mathcal{M}$ is linearly convergent from $y_0 = x_k$:

$$\varphi(y_i) - \varphi(y^*) \leq \gamma(1-\tau)^i \|y_0 - y^*\|^2 \qquad \forall i = 1, 2 \ldots$$

Set $x_{k+1} = y_T$ where $T := \left\lceil \frac{1}{\tau}\log(4\gamma/\beta) \right\rceil$.

[motivated by Lin-Mairal-Harchauoi '16]

Thm:

$\|\mathcal{G}(x_k)\| < \epsilon$ after $\mathcal{O}\left(\frac{\beta}{\varepsilon^2} \cdot (F(x_0) - F^*)\right)$ outer iterations.

$\Longrightarrow$ smoothing+prox-linear+fast-gradient gives complexity

$$\widetilde{\mathcal{O}}\left(\frac{\beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right).$$

Inexact prox-linear

Input: A point $x_0 \in \operatorname{dom} g$, a method $\mathcal{M}$
Step k: $(k \geq 1)$
Consider: $$\min_{y} \; \varphi(y) := F_{x_k}(y) + \frac{\beta}{2}\|y - x_k\|^2.$$

Suppose $\mathcal{M}$ is linearly convergent from $y_0 = x_k$:

$$\mathbb{E}[\varphi(y_i) - \varphi(y^*)] \leq \gamma(1-\tau)^i \|y_0 - y^*\|^2 \qquad \forall i = 1, 2 \ldots$$

Set $x_{k+1} = y_T$ where $T := \left\lceil \frac{1}{\tau} \log(4\gamma/\beta) \right\rceil$.

[motivated by Lin-Mairal-Harchauoi '16]

Thm:
$$\mathbb{E}[\|\mathcal{G}(x_k)\|] < \epsilon \quad \text{after} \quad \mathcal{O}\left(\frac{\beta}{\varepsilon^2} \cdot (F(x_0) - F^*)\right) \text{ outer iterations.}$$

$\Longrightarrow$ smoothing+prox-linear+fast-gradient gives complexity

$$\widetilde{\mathcal{O}}\left(\frac{\beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right).$$

Finite-sum problem:

Finite-sum problem:

$$\min_x \; F(x) = \frac{1}{m}\sum_{i=1}^{m} h_i(c_i(x)) + g(x)$$

Finite-sum problem:

Finite-sum problem:

$$\min_x \; F(x) = \frac{1}{m}\sum_{i=1}^{m} h_i(c_i(x)) + g(x)$$

where

- g is closed, convex.
- h_i is convex and 1-Lipschitz.
- c_i is C^1-smooth and ∇c_i is β-Lipschitz.

Finite-sum problem:

Finite-sum problem:

$$\min_x \; F(x) = \frac{1}{m}\sum_{i=1}^{m} h_i(c_i(x)) + g(x)$$

where

- g is closed, convex.
- h_i is convex and 1-Lipschitz.
- c_i is C^1-smooth and ∇c_i is β-Lipschitz.

This is the old composite problem with

$$h(z_1, \ldots, z_m) = \frac{1}{m}\sum_{i=1}^{m} h_i(z_i) \qquad c(x) = (c_1(x), c_2(x), \ldots, c_m(x)).$$

Finite-sum problem:

Finite-sum problem:

$$\min_x \; F(x) = \frac{1}{m}\sum_{i=1}^{m} h_i(c_i(x)) + g(x)$$

where

- g is closed, convex.
- h_i is convex and 1-Lipschitz.
- c_i is C^1-smooth and ∇c_i is β-Lipschitz.

This is the old composite problem with

$$h(z_1, \ldots, z_m) = \frac{1}{m}\sum_{i=1}^{m} h_i(z_i) \qquad c(x) = (c_1(x), c_2(x), \ldots, c_m(x)).$$

First-order complexity of a method:

Max # of

- $\boxed{c_i(x), \; \nabla c_i(x)^T v}$,
- $\mathrm{prox}_{th}(x), \mathrm{prox}_{tg}(x)$

a method evaluates to produce $\|\mathcal{G}(x)\| \leq \varepsilon$.

Finite-sum problem:

Define:

$$\|\nabla c\| := \max_{i=1,\ldots,m} \max_{x \in \operatorname{dom} g} \|\nabla c_i(x)\|.$$

Finite-sum problem:

Define:

$$\|\nabla c\| := \max_{i=1,\ldots,m} \max_{x \in \mathrm{dom}\, g} \|\nabla c_i(x)\|.$$

Baseline complexity (from before):

$$\widetilde{\mathcal{O}}\left(\frac{m \cdot \beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right)$$

Finite-sum problem:

Define:

$$\|\nabla c\| := \max_{i=1,\ldots,m} \max_{x \in \mathrm{dom}\, g} \|\nabla c_i(x)\|.$$

Baseline complexity (from before):

$$\widetilde{\mathcal{O}}\left(\frac{m \cdot \beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right)$$

Thm: (D-Paquette '16) Better complexity in expectation:

$$\widetilde{\mathcal{O}}\left(\left(\frac{m \cdot \beta}{\varepsilon^2} + \frac{\sqrt{m} \cdot \beta \cdot \|\nabla c\|}{\varepsilon^3}\right) \cdot (F(x_0) - F^*)\right)$$

Finite-sum problem:

Define:

$$\|\nabla c\| := \max_{i=1,\ldots,m} \max_{x \in \mathrm{dom}\, g} \|\nabla c_i(x)\|.$$

Baseline complexity (from before):

$$\widetilde{\mathcal{O}}\left(\frac{m \cdot \beta \cdot \|\nabla c\|}{\varepsilon^3} \cdot (F(x_0) - F^*)\right)$$

Thm: (D-Paquette '16) Better complexity in expectation:

$$\widetilde{\mathcal{O}}\left(\left(\frac{m \cdot \beta}{\varepsilon^2} + \frac{\sqrt{m} \cdot \beta \cdot \|\nabla c\|}{\varepsilon^3}\right) \cdot (F(x_0) - F^*)\right)$$

Strategy:

- smoothing of h_i,
- inexact prox-linear method,
- accelerated incremental subsolves:
  Lin-Mairal-Harchaui '15, Shalev-Shwartz '15,
  Frostig-Ge-Kakade-Sidford '15, Lan '15, Allen-Zhu '16

What does $\|\mathcal{G}(x)\| < \epsilon$ mean?

Additive composite $F = c + g$ setting:

$$\operatorname{dist}(0; \partial F(x_{k+1})) \leq 2\|\mathcal{G}(x_k)\|.$$

What does $\|\mathcal{G}(x)\| < \epsilon$ mean?

Additive composite $F = c + g$ setting:

$$\mathrm{dist}(0; \partial F(x_{k+1})) \leq 2\|\mathcal{G}(x_k)\|.$$

False in general!

What does $\|\mathcal{G}(x)\| < \epsilon$ mean?

Additive composite $F = c + g$ setting:

$$\operatorname{dist}(0; \partial F(x_{k+1})) \leq 2\|\mathcal{G}(x_k)\|.$$

False in general!

Different motivation: $F + \frac{\beta}{2}\|\cdot\|^2$ is a convex function $\Longrightarrow$

$$\boxed{\nabla F_t(x) = t^{-1}(x - \operatorname{prox}_{tF}(x))} \qquad \forall t < \beta^{-1}.$$

What does $\|\mathcal{G}(x)\| < \epsilon$ mean?

Additive composite $F = c + g$ setting:

$$\operatorname{dist}(0; \partial F(x_{k+1})) \le 2\|\mathcal{G}(x_k)\|.$$

False in general!

Different motivation: $F + \frac{\beta}{2}\|\cdot\|^2$ is a convex function $\Longrightarrow$

$$\boxed{\nabla F_t(x) = t^{-1}(x - \operatorname{prox}_{tF}(x))} \qquad \forall t < \beta^{-1}.$$

Thm: (D-Paquette '16)

$$\frac{1}{4}\|\nabla F_{\frac{1}{2\beta}}(x)\| \le \|\mathcal{G}(x)\| \le \frac{3}{2}\left(1 + \frac{1}{\sqrt{2}}\right)\|\nabla F_{\frac{1}{2\beta}}(x)\|$$

What does $\|\mathcal{G}(x)\| < \epsilon$ mean?

Additive composite $F = c + g$ setting:

$$\mathrm{dist}(0; \partial F(x_{k+1})) \leq 2\|\mathcal{G}(x_k)\|.$$

False in general!

Different motivation: $F + \frac{\beta}{2}\|\cdot\|^2$ is a convex function $\Longrightarrow$

$$\boxed{\nabla F_t(x) = t^{-1}(x - \mathrm{prox}_{tF}(x))} \qquad \forall t < \beta^{-1}.$$

Thm: (D-Paquette '16)

$$\frac{1}{4}\|\nabla F_{\frac{1}{2\beta}}(x)\| \leq \|\mathcal{G}(x)\| \leq \frac{3}{2}\left(1 + \frac{1}{\sqrt{2}}\right)\|\nabla F_{\frac{1}{2\beta}}(x)\|$$

$$\Longrightarrow \mathrm{dist}\Big((x, 0); \mathrm{gph}\, \partial F\Big) \leq \|\mathcal{G}(x)\|.$$

Local regularity conditions

Setting: $S := \{x : 0 \in \partial F(x)\}$, $\bar{x} \in S$, $\bar{x} \in \operatorname{int} \mathcal{X}$

Local regularity conditions

Setting: $S := \{x : 0 \in \partial F(x)\}, \qquad \bar{x} \in S, \qquad \bar{x} \in \operatorname{int} \mathcal{X}$

1) Error Bound property:

$$\alpha \cdot \operatorname{dist}(x; S) \leq \operatorname{dist}(0; \partial F(x)) \qquad \text{for } x \in \mathcal{X}$$

2) Tilt-stability: (Poliquin-Rockafellar '98)

$$v \mapsto \operatorname*{argmin}_{x \in \mathcal{X}} \{F(x) - \langle v, x \rangle\} \qquad \text{is } \tfrac{1}{\alpha}\text{-Lipschitz near } \bar{v} = 0$$

3) Sharpness: (Burke-Ferris '93)

$$\alpha \leq \operatorname{dist}(0; \partial F(x)) \qquad \text{for noncritical } x \in \mathcal{X}$$

Local regularity conditions

Setting: $S := \{x : 0 \in \partial F(x)\}, \qquad \bar{x} \in S, \qquad \bar{x} \in \operatorname{int} \mathcal{X}$

1) Error Bound property:

$$\alpha \cdot \operatorname{dist}(x; S) \leq \operatorname{dist}(0; \partial F(x)) \qquad \text{for } x \in \mathcal{X}$$

$$\Longrightarrow \qquad F(x) \geq F^* + \frac{\alpha}{2} \cdot \operatorname{dist}^2(x, S) \qquad \text{for } x \in \mathcal{X}$$

2) Tilt-stability: (Poliquin-Rockafellar '98)

$$v \mapsto \operatorname*{argmin}_{x \in \mathcal{X}} \{F(x) - \langle v, x \rangle\} \qquad \text{is } \frac{1}{\alpha}\text{-Lipschitz near } \bar{v} = 0$$

3) Sharpness: (Burke-Ferris '93)

$$\alpha \leq \operatorname{dist}(0; \partial F(x)) \qquad \text{for noncritical } x \in \mathcal{X}$$

Local regularity conditions

Setting: $S := \{x : 0 \in \partial F(x)\}, \qquad \bar{x} \in S, \qquad \bar{x} \in \operatorname{int} \mathcal{X}$

1) Error Bound property:

$$\alpha \cdot \operatorname{dist}(x; S) \leq \operatorname{dist}(0; \partial F(x)) \qquad \text{for } x \in \mathcal{X}$$

$$\Longrightarrow \qquad F(x) \geq F^* + \frac{\alpha}{2} \cdot \operatorname{dist}^2(x, S) \qquad \text{for } x \in \mathcal{X}$$

2) Tilt-stability: (Poliquin-Rockafellar '98)

$$v \mapsto \operatorname*{argmin}_{x \in \mathcal{X}} \{F(x) - \langle v, x \rangle\} \qquad \text{is } \frac{1}{\alpha}\text{-Lipschitz near } \bar{v} = 0$$

$\Longleftrightarrow$ For all small v, there is $x_v \in \mathcal{X}$ satisfying

$$F(x) \geq F(x_v) + \langle v, x - x_v \rangle + \frac{\alpha}{2} \cdot \|x - x_v\|^2 \qquad \text{for } x \in \mathcal{X}$$

3) Sharpness: (Burke-Ferris '93)

$$\alpha \leq \operatorname{dist}(0; \partial F(x)) \qquad \text{for noncritical } x \in \mathcal{X}$$

Local regularity conditions

Setting: $S := \{x : 0 \in \partial F(x)\}, \qquad \bar{x} \in S, \qquad \bar{x} \in \operatorname{int} \mathcal{X}$

1) Error Bound property:

$$\alpha \cdot \operatorname{dist}(x; S) \le \operatorname{dist}(0; \partial F(x)) \qquad \text{for } x \in \mathcal{X}$$

$$\Longrightarrow \qquad F(x) \ge F^* + \frac{\alpha}{2} \cdot \operatorname{dist}^2(x, S) \qquad \text{for } x \in \mathcal{X}$$

2) Tilt-stability: (Poliquin-Rockafellar '98)

$$v \mapsto \operatorname*{argmin}_{x \in \mathcal{X}} \{F(x) - \langle v, x \rangle\} \qquad \text{is } \frac{1}{\alpha}\text{-Lipschitz near } \bar{v} = 0$$

$\Longleftrightarrow$ For all small v, there is $x_v \in \mathcal{X}$ satisfying

$$F(x) \ge F(x_v) + \langle v, x - x_v \rangle + \frac{\alpha}{2} \cdot \|x - x_v\|^2 \qquad \text{for } x \in \mathcal{X}$$

3) Sharpness: (Burke-Ferris '93)

$$\alpha \le \operatorname{dist}(0; \partial F(x)) \qquad \text{for noncritical } x \in \mathcal{X}$$

$$\Longleftrightarrow \qquad F(x) \ge F^* + \alpha \cdot \operatorname{dist}(x, S_0) \qquad \text{for } x \in \mathcal{X}.$$

Local rates

Regularity	**Convergence Guarantee**
Error-bound	$\frac{F(x_{k+1})-F^*}{F(x_k)-F^*} \leq 1 - \left(\frac{\alpha}{\beta}\right)^2$
Tilt-stability	$\frac{F(x_{k+1})-F^*}{F(x_k)-F^*} \leq 1 - \frac{\alpha}{\beta}$
Sharpness	$\|x_{k+1} - x^*\| \leq \mathcal{O}(\|x_k - x^*\|^2)$

(Nesterov '06, D-Lewis '15)

Recent related work

"Accelerated" prox-linear method (D.-Paquette '16):

$$\min_{j=1,\ldots,k} \|\mathcal{G}(x_i)\|^2 \leq (\beta M)^2 \cdot \mathcal{O}\left(k^{-3} + \rho \cdot k^{-1}\right),$$

assuming $M = \operatorname{diam}(\operatorname{dom} g)$ finite and $\rho \in (0,1)$ satisfies

$$F_x(y) \leq F(y) + \rho \cdot \tfrac{\beta}{2}\|y - x\|^2 \qquad \forall x, y$$

Sampling methods (Duchi-Ruan '16): Almost sure conv. on

$$\min_x \; g(x) + \int_{\mathcal{S}} h(c(x,s), s)\, dP(s).$$

Inner-outer subgradient methods w/ rates (Davis-Grimmer '17)

Robust Phase Retrieval (Duchi-Ruan '17):
Quadratic convergence w.h.p on $\quad \min_x \; \frac{1}{m}\sum_{i=1}^{m} |(a_i^T x)^2 - b_i|$
[uses Eldar-Mendelson '12]

Summary

"Efficiency of minimizing compositions of convex functions and smooth maps", D-Paquette, 2016, arXiv:1605.00125.

"Error bounds, quadratic growth, and linear convergence of proximal methods", D-Lewis, Math. Oper. Res. 2017.

- prox-linear method
- first-order complexity
- finite-sum problems
- termination conditions
- local linear/quadratic rates

Thank you!