# Worst-case evaluation complexity for nonconvex optimization: adventures in the jungle of high-order nonlinear optimization

Philippe Toint (with Coralia Cartis and Nick Gould)

Namur Center for Complex Systems (naXys), University of Namur, Belgium

( philippe.toint@unamur.be )

ICM 2018, Rio de Janeiro, Brazil, August 2018

# Thanks

- University of Oxford and Balliol College, UK
- University of Firenze, Italy
- University of Namur, Belgium
- Belgian National Fund for Scientific Research

# The problem (again)

We consider the unconstrained nonlinear programming problem:

$$\text{minimize} \quad f(x)$$

for $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ smooth.

Important special case: the nonlinear least-squares problem

$$\text{minimize} \quad f(x) = \tfrac{1}{2}\|F(x)\|^2$$

for $x \in \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ smooth.

# A useful observation

Note the following: if

- $f$ has gradient $g$ and globally Lipschitz continuous Hessian $H$ with constant $2L$

Taylor, Cauchy-Schwarz and Lipschitz imply

$$\begin{aligned} f(x+s) &= f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle \\ &\quad + \int_0^1 (1-\alpha)\langle s, [H(x+\alpha s) - H(x)]s\rangle \, d\alpha \\ &\leq \underbrace{f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle + \tfrac{1}{3}L\|s\|_2^3}_{m(s)} \end{aligned}$$

$\Longrightarrow$ reducing $m$ from $s = 0$ improves $f$ since $m(0) = f(x)$.

Griewank, 1981

# Approximate model minimization

Lipschitz constant $L$ unknown $\Rightarrow$ replace by adaptive parameter $\sigma_k$ in the model :

$$m(s) \stackrel{\text{def}}{=} f(x) + s^T g(x) + \tfrac{1}{2} s^T H(x) s + \tfrac{1}{3}\sigma_k \|s\|_2^3 = T_{f,2}(x,s) + \tfrac{1}{3}\sigma_k \|s\|_2^3$$

Computation of the step:

1. minimize $m(s)$ until an approximate first-order minimizer is obtained:

$$\|\nabla_s m(s)\| \le \kappa_{\text{stop}} \|s\|^2$$

(s-rule)
Note: no global optimization involved.

# Second-order Adaptive Regularization (AR2)

**Algorithm 1.1: The AR2 Algorithm**

Step 0: Initialization: $x_0$ and $\sigma_0 > 0$ given. Set $k = 0$

Step 1: Termination: If $\|g_k\| \leq \epsilon$, terminate.

Step 2: Step computation:

Compute $s_k$ such that $m_k(s_k) \leq m_k(0)$ and $\|\nabla_s m(s_k)\| \leq \kappa_{\rm stop}\|s_k\|^2$.

Step 3: Step acceptance:

Compute $\rho_k = \dfrac{f(x_k) - f(x_k + s_k)}{f(x_k) - T_{f,2}(x_k, s_k)}$

and set $x_{k+1} = \begin{cases} x_k + s_k & \text{if } \rho_k > 0.1 \\ x_k & \text{otherwise} \end{cases}$

Step 4: Update the regularization parameter:

$$\sigma_{k+1} \in \begin{cases} [\sigma_{\min}, \sigma_k] & = \frac{1}{2}\sigma_k \quad \text{if } \rho_k > 0.9 & \textit{very successful} \\ [\sigma_k, \gamma_1\sigma_k] & = \sigma_k \quad \text{if } 0.1 \leq \rho_k \leq 0.9 & \textit{successful} \\ [\gamma_1\sigma_k, \gamma_2\sigma_k] & = 2\sigma_k \quad \text{otherwise} & \textit{unsuccessful} \end{cases}$$

# Second-order regularization highlights

$$f(x+s) \leq m(s) \equiv f(x) + s^T g(x) + \tfrac{1}{2} s^T H(x) s + \tfrac{1}{3} L \|s\|_2^3$$

- Nesterov and Polyak minimize $m$ globally and exactly
  - N.B. $m$ may be non-convex!
  - efficient scheme to do so if $H$ has sparse factors
- global (ultimately rapid) convergence to a 2nd-order critical point of $f$
- better worst-case function-evaluation complexity than previously known

Obvious questions:

- can we avoid the global Lipschitz requirement? YES!
- can we approximately minimize $m$ and retain good worst-case function-evaluation complexity? YES !
- does this work well in practice? yes

# Evaluation complexity: an important result

How many function evaluations (iterations) are needed to ensure that

$$\|g_k\| \le \epsilon?$$

If $H$ is globally Lipschitz and the s-rule is applied, the AR2 algorithm requires at most

$$\left\lceil \frac{\kappa_{\rm S}}{\epsilon^{3/2}} \right\rceil \text{ evaluations}$$

for some $\kappa_{\rm S}$ independent of $\epsilon$.

"Nesterov & Polyak"

Note: an $O(\epsilon^{-3})$ bound holds for convergence to second-order critical points.

# Evaluation complexity: proof (1)

$$f(x_k + s_k) \le T_{f,2}(x_k, s_k) + \frac{L_f}{p}\|s_k\|^3$$

$$\|g(x_k + s_k) - \nabla_s T_{f,2}(x_k, s_k)\| \le L_f\|s_k\|^2$$

Lipschitz continuity of $H(x) = \nabla_x^2 f(x)$

$$\forall k \ge 0 \qquad f(x_k) - T_{f,2}(x_k, s_k) \ge \tfrac{1}{6}\sigma_{\min}\|s_k\|^3$$

$$f(x_k) = m_k(0) \ge m_k(s_k) = T_{f,2}(x_k, s_k) + \tfrac{1}{6}\sigma_k\|s_k\|^3$$

# Evaluation complexity: proof (2)

$$\exists \sigma_{\max} \quad \forall k \geq 0 \qquad \sigma_k \leq \sigma_{\max}$$

Assume that $\sigma_k \geq \dfrac{L_f(p+1)}{p\,(1-\eta_2)}$. Then

$$|\rho_k - 1| \leq \frac{|f(x_k+s_k) - T_{f,2}(x_k,s_k)|}{|T_{f,2}(x_k,0) - T_{f,2}(x_k,s_k)|} \leq \frac{L_f(p+1)}{p\,\sigma_k} \leq 1 - \eta_2$$

and thus $\rho_k \geq \eta_2$ and $\sigma_{k+1} \leq \sigma_k$.

# Evaluation complexity: proof (3)

$$\forall k \text{ successful} \qquad \|s_k\| \geq \left( \frac{\|g(x_{k+1})\|}{L_f + \kappa_{\rm stop} + \sigma_{\max}} \right)^{\frac{1}{2}}$$

$$\begin{aligned}
\|g(x_k+s_k)\| &\leq \|g(x_k+s_k) - \nabla_s T_{f,2}(x_k,s_k)\| \\
&\quad + \Big\| \nabla_s T_{f,2}(x_k,s_k) + \sigma_k \|s_k\| s_k \Big\| + \sigma_k \|s_k\|^2 \\
&\leq L_f \|s_k\|^2 + \|\nabla_s m(s_k)\| + \sigma_k \|s_k\|^2 \\
&\leq [L_f + \kappa_{\rm stop} + \sigma_k] \, \|s_k\|^2
\end{aligned}$$

# Evaluation complexity: proof (4)

$$\|g(x_{k+1})\| \leq \epsilon \text{ after at most } \frac{f(x_0) - f_{\rm low}}{\kappa} \, \epsilon^{-3/2} \text{ successful iterations}$$

Let $\mathcal{S}_k = \{j \leq k \geq 0 \mid \text{iteration } j \text{ is successful}\}$.

$$\begin{aligned}
f(x_0) - f_{\rm low} &\geq f(x_0) - f(x_{k+1}) \geq \sum_{i \in \mathcal{S}_k} \Big[ f(x_i) - f(x_i + s_i) \Big] \\
&\geq \tfrac{1}{10} \sum_{i \in \mathcal{S}_k} \Big[ f(x_i) - T_{f,2}(x_i, s_i) \Big] \geq |\mathcal{S}_k| \frac{\sigma_{\min}}{60} \min_i \|s_i\|^3 \\
&\geq |\mathcal{S}_k| \frac{\sigma_{\min}}{60 \Big( L_f + \kappa_{\rm stop} + \sigma_{\max} \Big)^{3/2}} \min_i \|g(x_{i+1})\|^{3/2} \\
&\geq |\mathcal{S}_k| \frac{\sigma_{\min}}{60 \Big( L_f + \kappa_{\rm stop} + \sigma_{\max} \Big)^{3/2}} \epsilon^{3/2}
\end{aligned}$$

# Evaluation complexity: proof (5)

$$k \le \kappa_u |\mathcal{S}_k|, \quad \text{where} \quad \kappa_u \stackrel{\text{def}}{=} \left(1 + \frac{|\log \gamma_1|}{\log \gamma_2}\right) + \frac{1}{\log \gamma_2} \log \left(\frac{\sigma_{\max}}{\sigma_0}\right),$$

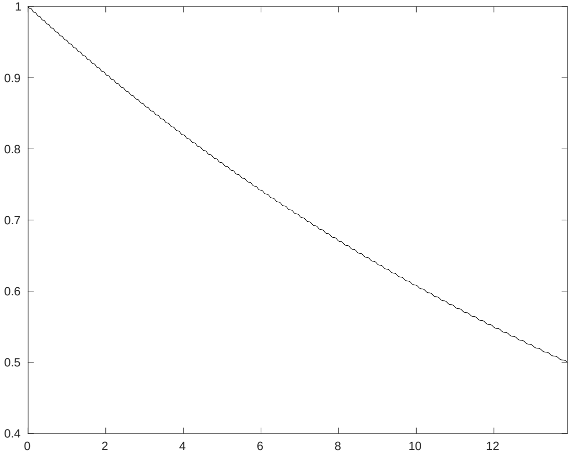
$\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ + mechanism of the $\sigma_k$ update.

$$\|g(x_{k+1})\| \le \epsilon \text{ after at most } \frac{f(x_0) - f_{\rm low}}{\kappa} \epsilon^{-3/2} \text{ successful iterations}$$

One evaluation per iteration (successful or unsuccessful).

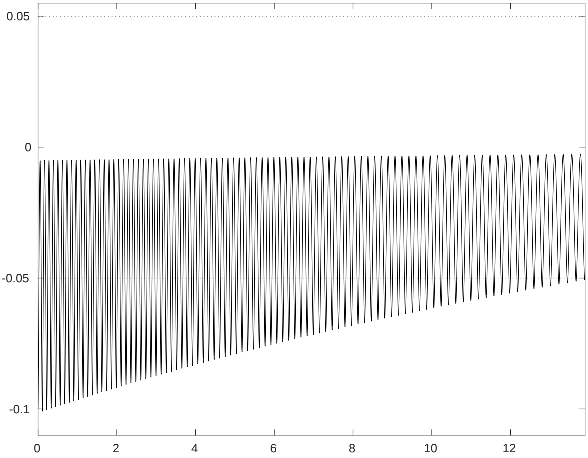
# Evaluation complexity: sharpness

Is the bound in $O(\epsilon^{-3/2})$ sharp? YES!!!



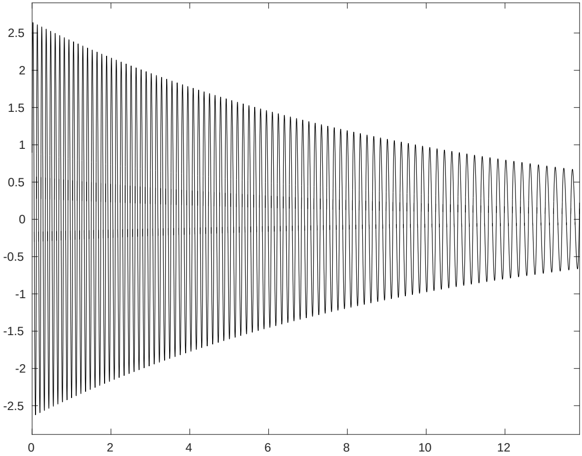

The objective function

# An example of slow AR2 (2)



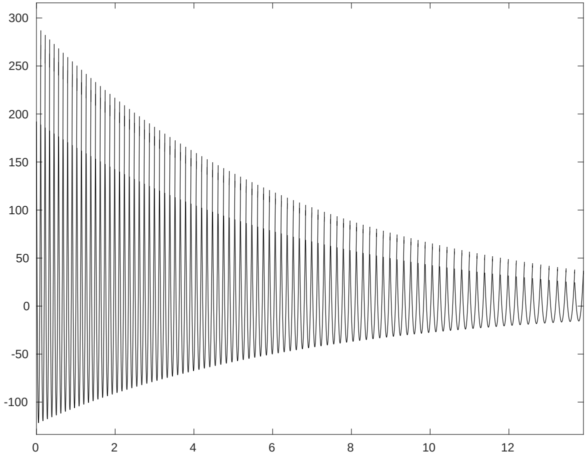

The first derivative

# An example of slow AR2 (3)




The second derivative

# An example of slow AR2 (4)




The third derivative

# Slow steepest descent (1)

The steepest descent method with requires at most

$$\left\lceil \frac{\kappa_C}{\epsilon^2} \right\rceil \text{ evaluations}$$

for obtaining $\|g_k\| \leq \epsilon$.

Nesterov

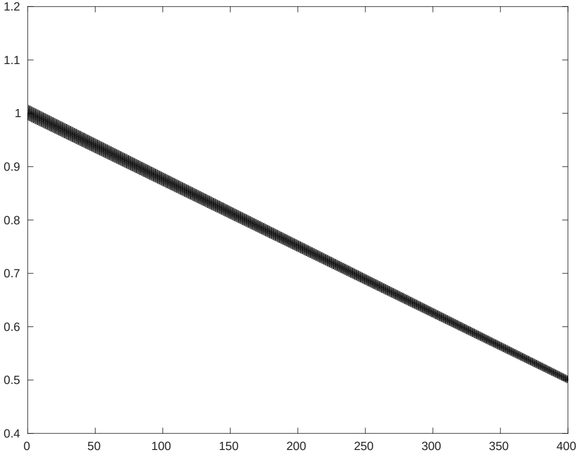
Sharp??? YES

Newton's method (when convergent) requires at most

$$O(\epsilon^{-2}) \text{ evaluations}$$

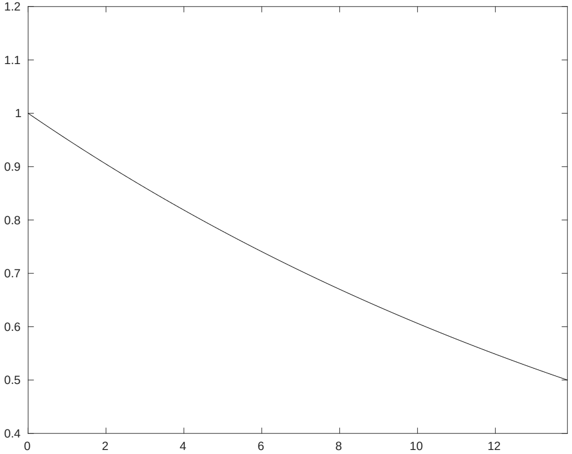
for obtaining $\|g_k\| \leq \epsilon$ !!!!

# Slow Newton




The objective function for slow Newton

# Slow Steepest-Descent




The objective function for slow Steepest descent

# General second-order methods

Assume that, for $\beta \in (0,1]$, the step is computed by

$$(H_k + \lambda_k I)s_k = -g_k \;\text{ and }\; 0 \le \lambda_k \le \kappa_s \|s_k\|^\beta$$

(ex: Newton, AR2, Levenberg-Morrison-Marquardt, (trust-region), Curtis-Robinson-Samadi, Royer-Wright,…)

The corresponding method terminates in at most

$$\left\lceil \frac{\kappa_C}{\epsilon^{(\beta+2)/(\beta+1)}} \right\rceil \text{ evaluations}$$

to obtain $\|g_k\| \le \epsilon$ on functions with bounded and (segment-wise) $\beta$-Hölder continuous Hessians, and the bound is sharp.

Note: ranges form $\epsilon^{-2}$ to $\epsilon^{-3/2}$

**AR2 is optimal within this class**

# High-order models (1)

What happens if one considers the model

$$m_k(s) = T_{f,p}(x_k, s) + \frac{\sigma_k}{p!}\|s\|_2^{p+1}$$

where

$$T_{f,p}(x,s) = f(x) + \sum_{j=1}^{p} \frac{1}{j!}\nabla_x^j f(x)[s]^j$$

terminating the step computation when

$$\|\nabla_s m(s_k)\| \le \kappa_{\text{stop}}\|s_k\|^p$$

???

now the AR$p$ method!

# High-order models (2)

$\epsilon$-approx 1rst-order critical point after at most

$$\frac{f(x_0) - f_{\rm low}}{\kappa} \, \epsilon^{-\frac{p+1}{p}}$$

successful iterations

also for convexly constrained problems!

Moreover (using the correct subproblem termination rule)

$\epsilon$-approx "2nd order critical point" after at most

$$\frac{f(x_0) - f_{\rm low}}{\kappa} \, \epsilon^{-\frac{p+1}{p-1}}$$

successful iterations

for unconstrained problems only!

Much better than the standard $\mathcal{O}(\epsilon^{-3})$ result!!!

# Derivative tensors for partially separable problems

$f$ is partially separable if

$$f(x) = \sum_{i=1}^{m} f_i(U_i x) = \sum_{i=1}^{m} f_i(x_i) \quad \text{where} \quad \mathrm{rank}(U_i) \ll n$$

Then

$$\nabla_x^p f(x)[s]^p = \sum_{i=1}^{m} \nabla_{x_i}^p f_i(x)[U_i x]^p$$

Note:

$$\text{size}(\ \nabla_{x_i}^p f_i(x)\ ) \ll \text{size}(\ \nabla_x^p f(x)\ )!!!$$

# A (not so) obvious question

If one uses a model of degree $p$ ($T_{f,p}(x,s)$), why be satisfied with first- or second-order critical points???

What do we mean by critical points of order larger than 2 ???

What are necessary optimality conditions for order larger than 2 ???

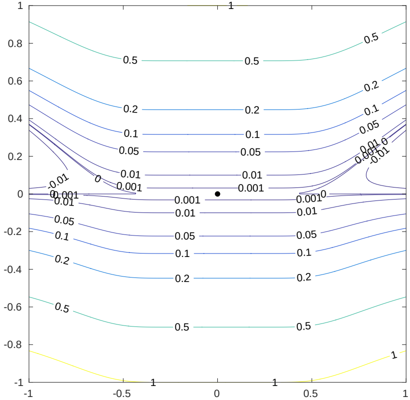
Not an obvious question!

# A sobering example (1)

Consider the unconstrained minimization of

$$f(x_1, x_2) = \begin{cases} x_2 \left( x_2 - e^{-1/x_1^2} \right) & \text{if } x_1 \neq 0, \\ x_2^2 & \text{if } x_1 = 0, \end{cases}$$

Peano (1884), Hancock (1917)



# A sobering example (2)

Conclusions:

- looking at optimality along straight lines is not enough
- depending on Taylor's expansion for necessary conditions is not always possible

Even worse:

$$f(x_1, x_2) = \begin{cases} x_2\left(x_2 - \sin(1/x_1)e^{-1/x_1^2}\right) & \text{if } x_1 \neq 0, \\ x_2^2 & \text{if } x_1 = 0, \end{cases}$$

(no continuous descent path from 0, although not a local minimizer!!!)

Hopeless?

# Limiting one's ambitions...

Note: the non-existence of continuous descent paths remains a necessary condition! Focus on polynomial paths

$$x(\alpha) = x_* + \sum_{i=1}^{q} \alpha^i s_i + o(\alpha^q)$$

Suppose that $x_*$ is a local minimizer. Then, for $j \in \{1, \ldots, q\}$,

$$\sum_{k=1}^{j} \frac{1}{k!} \Bigg( \sum_{(\ell_1,\ldots,\ell_k)\in\mathcal{P}(j,k)} \nabla_x^k f(x_*)[s_{\ell_1}, \ldots, s_{\ell_k}] \Bigg) \geq 0$$

holds for all $(s_1, \ldots, s_j)$ such that, for $i \in \{1, \ldots, j-1\}$,

$$\sum_{k=1}^{i} \frac{1}{k!} \Bigg( \sum_{(\ell_1,\ldots,\ell_k)\in\mathcal{P}(i,k)} \nabla_x^k f(x_*)[s_{\ell_1}, \ldots, s_{\ell_k}] \Bigg) = 0.$$

# And then?

In short:

- reduces to (in)equalities on $\nabla_x^j f(x)[s]^j$ in the kernel of $\nabla_x^{j-1}$ for $j = 1,2,3$
- inherently more complicated for orders 4 and above
(conditions involving a mix of $\nabla_x^j f(x)[s]^j$ of different orders)

Desperate?

# Using Taylor's models, nevertheless

Define, for some small $\Delta > 0$,

$$\phi_{f,j}^{\Delta}(x) \stackrel{\rm def}{=} f(x) - \text{globmin}_{\substack{x+d\in\mathcal{F}\\ \|d\|\leq\Delta}} T_{f,j}(x,d),$$

$$\left[\lim_{\Delta\to 0} \frac{\phi_{f,j}^{\Delta}(x)}{\Delta^j} = 0\right] \Rightarrow \quad \text{path-based necessary conditions at } x \;.$$

$\nabla_x^q f$ Lipschitz continous near $x_\epsilon \in \mathcal{F}$. Suppose that

$$\phi_{f,j}^{\Delta}(x_\epsilon) \leq \epsilon\Delta^j \quad \text{for} \quad j = 1, \ldots, q$$

Then

$$f(x_\epsilon + d) \geq f(x_\epsilon) - 2\epsilon\Delta^q \quad \forall x_\epsilon + d \;\; \text{with} \;\; \|d\| \leq \left(\frac{q!\,\epsilon\Delta^q}{L_{f,q}}\right)^{\frac{1}{q+1}}$$

# A (theoretical) trust-region algorithm

**Algorithm 3.1: Trust-region with adaptive order models (TR$q$)**

Step 0: Initialization: $q$, $\epsilon \in (0,1]$, $x_0$ and $\Delta_1 \in [\epsilon, 1]$, $\Delta_{\max} \in [\Delta_1, 1]$.

Step 1: Step computation: For $j = 1, \ldots, q$,

(i) evaluate $\nabla^j f(x_k)$ and $\phi_{f,j}^{\Delta_k}(x_k)$

(ii) if $\phi_{f,j}^{\Delta_k}(x_k) > \epsilon \Delta_k^j$, go to Step 2 with $s_k = d$

Terminate with $x_\epsilon = x_k$ and $\Delta_\epsilon = \Delta_k$.

Step 2: Accept the new iterate: Compute $f(x_k + s_k)$ and

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{T_{f,j}(x_k, 0) - T_{f,j}(x_k, s_k)}.$$

If $\rho_k \geq \eta_1$, set $x_{k+1} = x_k + s_k$. Otherwise set $x_{k+1} = x_k$.

Step 4: Update the trust-region radius. Set

$$\Delta_{k+1} \in \begin{cases} [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{if } \rho_k < \eta_1, \\ [\gamma_2 \Delta_k, \Delta_k] & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\Delta_k, \min(\Delta_{\max}, \gamma_3 \Delta_k)] & \text{if } \rho_k \geq \eta_2, \end{cases}$$

# An evaluation complexity bound

TR$q$ computes an $\epsilon$-approx "$q$-th order critical point" after at most

$$\kappa_S \; \epsilon^{-(q+1)}$$

(successful) iterations.

Same results for problems involving convex constraints!

# Complexity of convexly constrained problems

**Where do we stand?**

| $\uparrow q/p \rightarrow$ | 1 | 2 | $\cdots$ | $\cdots$ | $p$ | $\cdots$ |
|---|---|---|---|---|---|---|
| $\vdots$ | – | – | – | – | | **?** |
| $q$ | – | – | – | $O(\epsilon^{-(q+1)})$ | **?** | **?** |
| $\vdots$ | – | – | | **?** | **?** | **?** |
| 2 | | $O(\epsilon^{-3})$ | $\cdots$ | $\cdots$ | $[O(\epsilon^{-(p+1)/(p-1)})]$ | $\cdots$ |
| 1 | $O(\epsilon^{-2})$ | $O(\epsilon^{-3/2})$ | $\cdots$ | $\cdots$ | $O(\epsilon^{-(p+1)/p})$ | $\cdots$ |
| | $\leftarrow$ *sharp* $\rightarrow$ | | | | | |

Complexity of optimality order $q$ as a function of model degree $p$

Trust-region algo Regularization algo

[ ] for unconstrained problems only!

# Sharpness revisited for unconstrained problems

- Because the counter-examples discussed above are one-dimensional:

$$\forall \epsilon \; \forall n \quad \exists f : \mathbb{R}^n \rightarrow \mathbb{R}$$

for which most model-based methods with $p = 1, 2$ require at least $O(\epsilon^{-\frac{p+1}{p}})$ evaluations to obtain an $\epsilon$-first-order critical point.

- New result:

$$\forall \epsilon \; \exists n \quad \exists f : \mathbb{R}^n \rightarrow \mathbb{R}$$

for which a larger class of model-based methods with $p \leq n$ require at least $O(\epsilon^{-\frac{p+1}{p}})$ evaluations to obtain an $\epsilon$-first-order critical point.

Carmon, Duchi, Hinder, Sidford (2018)

Note: $n \geq O(\epsilon^{-\frac{p+1}{p}})$!

# The equality-constrained case

Consider now the EC-NLO (general with slack variables formulation):

$$\begin{array}{ll} \text{minimize}_x & f(x) \\ \text{such that} & c(x) = 0 \quad \text{and} \quad x \in \mathcal{F} \end{array}$$

Suppose $x$ is a local minimum of the EC-NLO problem and y the associated multiplier. Then, for every $q > 0$ and $\Lambda(x, y) = f(x) + y^T c(x)$,

$$T_{\Lambda,q}(x, s(\alpha), y) \geq 0$$

for all locally feasible $s(\alpha)$ such that

$$T_{\Lambda,j}(x, s(\alpha), y) = 0 \quad j \in \{1, \ldots, q-1\}$$

and

$$T_{c,j}(x, s(\alpha), y) = 0 \quad j \in \{1, \ldots, q\}$$

$\longrightarrow$ even more complicated to handle!

# Necessary conditions for EC-NLO

Verification essentially (even more) hopeless because of

- dependence of $c_{k,j}(x)$ on $s_{\ell_1}, \ldots, s_{\ell_k}$
- growing number of coefficients
- involves more than $\nabla_x^q f$ for $q \geq 3$!

Ideas for a first-order algorithm:

1. get $\|c(x)\| \leq \epsilon$ (if possible) by minimizing $\|c(x)\|^2$ such that $x \in \mathcal{F}$ (getting $\|J(x)^T c(x)\|$ small unsuitable!)
2. track the "trajectory"

$$\mathcal{T}(t) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n \mid c(x) = 0 \quad \text{and} \quad f(x) = t\}$$

for values of $t$ decreasing from $f$(first feasible iterate) while preserving $x \in \mathcal{F}$

# First-order complexity for EC-NLO

Sketch of a two-phases algorithm:

feasibility: apply a $O(\epsilon^{-\pi})$ method for convex constraints (with specific termination test) to

$$\min_x \nu(x) \stackrel{\rm def}{=} \|c(x)\|^2 \quad \text{such that} \quad x \in \mathcal{F}$$

at most $O(\max[\epsilon_P^{-1}, \epsilon_P^{1-\pi}\epsilon_D^{-\pi}])$ evaluations

tracking: successively

- apply a $O(\epsilon^{-\pi})$ method for convex constraints (with specific termination test) to

$$\min_x \mu(x,t) \stackrel{\rm def}{=} \|c(x)\|^2 + (f(x)-t)^2 \quad \text{such that} \quad x \in \mathcal{F}$$

- decrease $t$ (proportionally to the decrease in $\phi(x)$)

at most $O(\max[\epsilon_P^{-1}, \epsilon_P^{1-\pi}\epsilon_D^{-\pi}])$ evaluations

# First-order complexity for EC-NLO

Under the "conditions stated above", the above algorithm takes at most

$$''O''(\epsilon_{\rm P}^{1-\pi}\epsilon_{\rm D}^{-\pi}) \text{ evaluations}$$

to find an iterate $x_k$ with either

$$\|c(x_k)\| \leq \delta\epsilon_P \quad \text{and} \quad \phi^{\Delta}_{\Lambda,1} \leq \|(y,1)\|\epsilon_D\Delta$$

for some Lagrange multiplier $y$, or

$$\|c(x_k)\| > \delta\epsilon \quad \text{and} \quad \phi^{\Delta}_{||c||,1} \leq \epsilon\Delta.$$

# Higher order complexity for EC-NLO? (1)

The above approach for $q = 1$ hinges on

$$\nabla_x^1 \Lambda(x,y) = \frac{1}{f(x) - t} \nabla_x^1 \mu(x,t)$$

Hopeful for $q = 2$ since

$$\nabla_x^2 \Lambda(x,y)[d]^2 = \frac{1}{f(x) - t} \nabla_x^2 \mu(x,t)[d]^2$$

for all

$$d \in \text{span}\left\{\nabla_x^1 f(x)\right\}^\perp \cap \text{span}\left\{\nabla_x^1 c(x)\right\}^\perp \stackrel{\text{def}}{=} \mathcal{M}(x)$$

More difficult but maybe not imposible for $q = 3$ as

$$\nabla_x^3 \Lambda(x,y)[d]^3 = \frac{1}{f(x) - t} \nabla_x^3 \mu(x,t)[d]^3$$

for all

$d \in \mathcal{M}(x) \cap$ [a complicated set depending $\{\nabla_x^1 f\}$, $\{\nabla_x^2 f\}$, $\{\nabla_x^1 c\}$,$\{\nabla_x^2 c_i\}$]

# Higher order complexity for EC-NLO? (2)

But impossible for $q = 4$ (and above) because

$$\begin{aligned}\nabla_x^4 \Lambda(x,y) \;&=\; \frac{1}{f(x)-t}\nabla_x^4 \mu(x,t) \\ &-4\left[\nabla_x^3 f(x)\otimes \nabla_x^1 f(x) + \textstyle\sum_{i=1}^m \nabla_x^3 c_i(x)\otimes \nabla_x^1 c_i(x)\right] \\ &-3\left[\nabla_x^2 f(x)\otimes \nabla_x^2 f(x) + \textstyle\sum_{i=1}^m \nabla_x^2 c_i(x)\otimes \nabla_x^2 c_i(x)\right]\end{aligned}$$

A possibly important consequence:

Every approach based on quadratic (or more general strictly increasing) penalization is probably doomed for $q \geq 4$!

$\Rightarrow$ Need for a completely fresh point of view!

# Conclusions

- Complexity analysis for general $q$-th order critical points

$$O(\epsilon^{-(q+1)}) \quad \text{(unconstrained, convex constraints)}$$

- Complexity analysis for first-order critical points

$$O(\epsilon_{\mathrm{P}}^{1-\pi}\epsilon_{\mathrm{D}}^{-\pi}) \quad \text{(equality and general constraints)}$$

- Jarre's example $\Rightarrow$ global optimization much harder
- Many questions remaining:
  - high-order optimality with high-degree model?
  - beyond first-order for EC-NLO?

# Conclusions

Evaluation complexity improves with the model's degree

Critical points of order higher than 2 are (in general) evasive

Approximate critical points high order can be defined

An evaluation complexity bound for those is available
(more work for higher orders)

The above holds for unconstrained and convexly constrained problems

# Further questions

Can one improve the complexity bound for general $p > q$???

What about high-order criticality for equality constrained problems?

Can this be (more) practical?

Many thanks for your attention!

# Some references

C. Cartis, N. Gould and Ph. L. Toint,
"Second-order optimality and beyond: characterization and evaluation complexity in convexly-constrained nonlinear optimization",
FoCM, to appear.

C. Cartis, N. Gould and Ph. L. Toint,
"Worst-case evaluation complexity and optimality of second-order methods for nonconvex smooth optimization",
Proc ICM 2018, to appear.

C. Cartis, N. Gould and Ph. L. Toint,
"Improved second-order evaluation complexity for unconstrained nonlinear optimization using high-order regularized models",
arXiv:1708.04044, 2017.

Also see http://perso.fundp.ac.be/~phtoint/toint.html