

A path and some 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@fundp.ac.be`)

Florence, Italy, 24 October 2017

Thanks

- University of Firenze, Italy
- University of Namur, Belgium

The problem (again)

We consider the unconstrained nonlinear programming problem:

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

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

Important special case: the **nonlinear least-squares problem**

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

for $x \in \mathbb{R}^n$ and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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

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

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

Approximate model minimization

Lipschitz constant L **unknown** \Rightarrow replace by **adaptive parameter** σ_k in the model :

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

Computation of the step:

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

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

(s-rule)

Note: **no global optimization involved.**

Adaptive Regularization with Cubics (ARC2 or AR2)

Algorithm 1.1: The ARC2 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 $\|\nabla_s m(s_k)\| \leq \kappa_{\text{stop}} \|s_k\|^2$.

Step 3: Step acceptance:

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

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:

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

Cubic regularization highlights

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

- 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

$$\|g_k\| \leq \epsilon?$$

If H is globally Lipschitz and the s-rule is applied, the ARC2 algorithm requires at most

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

for some κ_S independent of ϵ .

c.f. Nesterov & Polyak

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

Evaluation complexity: proof (1)

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

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

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

$$\forall k \geq 0 \quad f(x_k) - T_{f,2}(x_k, s_k) \geq \frac{1}{6} \sigma_{\min} \|s_k\|^3$$

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

Evaluation complexity: proof (2)

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

Assume that $\sigma_k \geq \frac{L_f(\rho + 1)}{\rho(1 - \eta_2)}$. Then

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

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

Evaluation complexity: proof (3)

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

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

Evaluation complexity: proof (4)

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

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

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

Evaluation complexity: proof (5)

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

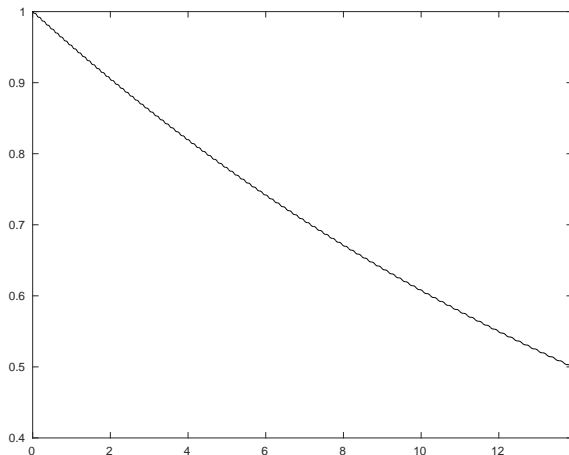
$\sigma_k \in [\sigma_{\min}, \sigma_{\max}]$ + mechanism of the σ_k update.

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

One evaluation per iteration (successful or unsuccessful).

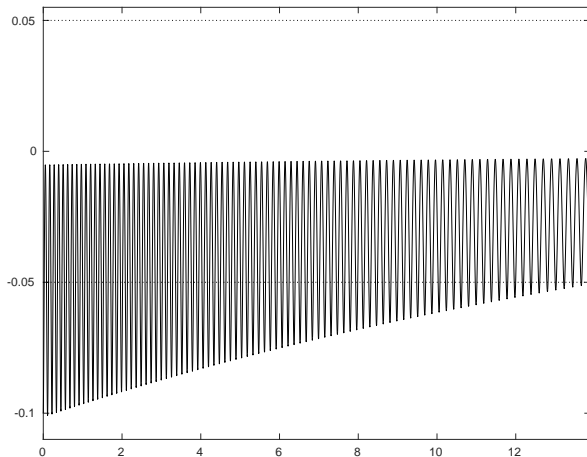
Evaluation complexity: sharpness

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



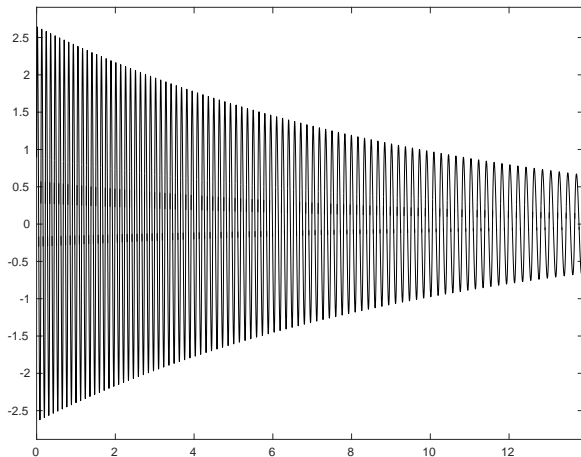
The objective function

An example of slow ARC2 (2)



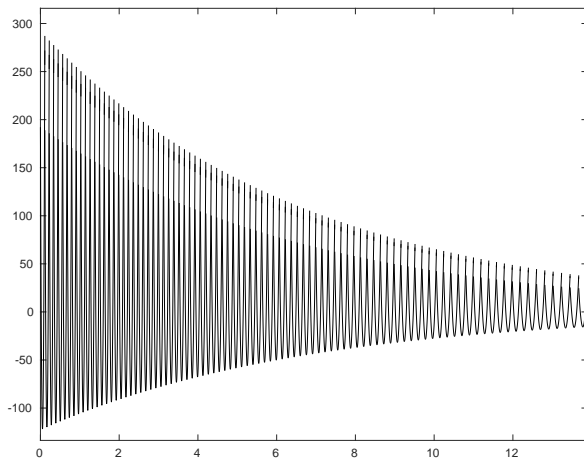
The first derivative

An example of slow ARC2 (3)



The second derivative

An example of slow ARC2 (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

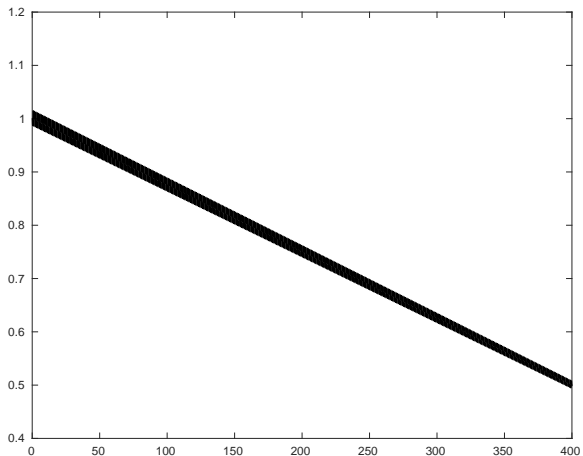
Sharp??? YES

Newton's method (when convergent) requires at most

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

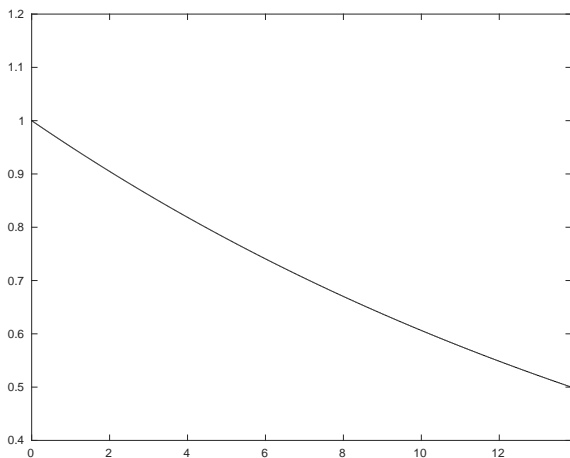
for obtaining $\|g_k\| \leq \epsilon$!!!!

Slow Newton



The objective function for slow Newton

Slow Steepest-Descent



The objective function for slow Steepest descent

More general second-order methods

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

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

(ex: Newton, ARC2, 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 $\|g_k\| \leq \epsilon$ on functions with bounded and (segment-wise) β -Hölder continuous Hessians.

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

ARC2 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

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

???

now the AR_p method!

High-order models (2)

ϵ -approx 1st-order critical point after at most

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

successful iterations

also for convexly constrained problems!

Moreover (using the correct subproblem termination rule)

ϵ -approx “2nd order critical point” after at most

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

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

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

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 ???

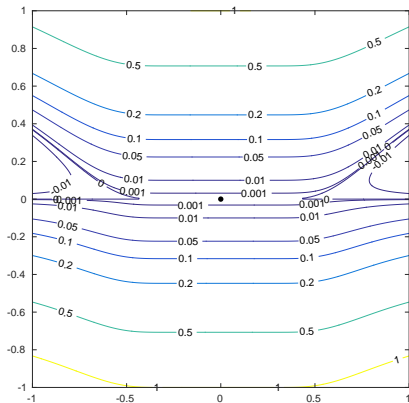
Not an obvious question!

A sobering example (1)

Consider the unconstrained minimization of

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

(cf. 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 $j \in \{1, \dots, q\}$,

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

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

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

And then?

In short:

- reduces to (in)equalities on $\nabla_x^j f(x)[s]^j$ in the kernel of ∇_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 $\Delta > 0$,

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

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

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

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

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 $j = 1, \dots, q$,

(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 ϵ -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?

\vdots	—	—	—	—		?
q	—	—	—	$O(\epsilon^{-(q+1)})$?
\vdots	—	—		?		?
2		$O(\epsilon^{-3})$	$[O(\epsilon^{-(p+1)/(p-1)})]$...
1	$O(\epsilon^{-2})$	$O(\epsilon^{-3/2})$	$O(\epsilon^{-(p+1)/p})$...
$\uparrow q/p \rightarrow$	1	2	p	...

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

Trust-region algo

Regularization algo

[] for unconstrained problems only!

The equality-constrained case

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

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

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

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

and

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

→ even more complicated to handle

Necessary conditions for EC-NLO

Verification essentially (even more) hopeless because of

- dependence of $c_{k,j}(x)$ on $s_{\ell_1}, \dots, s_{\ell_k}$
- 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”

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

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

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

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

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

- 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

$$"O"(\epsilon_P^{1-\pi} \epsilon_D^{-\pi}) \text{ evaluations}$$

to find an iterate x_k with either

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

for some Lagrange multiplier y , or

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

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

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

More difficult but maybe not impossible 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

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

Higher order complexity for EC-NLO? (2)

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

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

A **possibly important** consequence:

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

⇒ Need for a completely fresh point of view!

Conclusions

- Complexity analysis for general q -th order critical points

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

- Complexity analysis for first-order critical points

$$O(\epsilon_P^{1-\pi} \epsilon_D^{-\pi}) \text{ (equality and general constraints)}$$

- 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>