# An introduction to complexity analysis for nonconvex optimization

Philippe Toint (with Coralia Cartis and Nick Gould)

Namur Center for Complex Systems

FUNDP – University of Namur, Belgium

Séminaire Résidentiel Interdisciplinaire, Saint Hubert, January 2011

# The problem

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 typical application of nonlinear least-squares

Consider a (physical, chemical, biological, . . . ) process evolving over time:

$$y = P(t)$$

and a parametrized model for this process

$$y = M(t, x)$$

for which observations $\{y_i \approx P(t_i)\}_{i=1}^{nobs}$ are known.
How to choose $x$, the model parameters? Often:

$$x_* = \arg\min_x \tfrac{1}{2} \sum_{i=1}^{nobs} \|y_i - M(t_i, x)\|_2^2$$

Examples in sciences, engineering, economy, medecine, psychlogy, . . .

# Unconstrained optimization algorithms

More generally, how to find

$$x_* = \arg\min_x f(x)$$

(assuming the problem is well-defined) ???
Typically, generate a sequence of iterates $\{x_k\}_{k=0}^{\infty}$ such that

$$\{f(x_k)\}_{k=0}^{\infty} \text{ is decreasing}$$

and "hope" that, for some solution $x_*$, $\{x_k\}_{k=0}^{\infty} \to x_*$ !

# How to generate the iterates?

A (good?) sequence of iterates $\{x_k\}_{k=0}^{\infty}$ is generated by

**unconstrained optimization algorithms**

- (search methods (no derivatives of $f$ used))
- gradient methods (steepest descent)
- Newton methods and its variants ensuring global convergence
  - trust-region methods
  - cubic regularization methods
  - (linesearch methods)

# How is an (approximate) solution recognized?

Stop the algorithm as soon as

- the slope of $f$ is (approximately) zero, i.e.

$$\|\nabla_x f(x_k)\| \leq \epsilon_g \qquad \text{(1rst-order optimality)}$$

- the curvature of $f$ is (approximately) non-negative, i.e.

$$\lambda_{\min}[\nabla_{xx} f(x_k)] \geq -\epsilon_H \qquad \text{(2nd-order optimality)}$$

for some (small) $\epsilon_g > 0$ and $\epsilon_H > 0$.

THE COMPLEXITY QUESTION:
How many iterations are needed *at most*?

# The complexity question

THE COMPLEXITY QUESTION:
How many iterations are needed *at most*?

- needs assumptions on the smoothness of $f$ [unspecified here]
- (convex) vs. NONCONVEX
- strongly depends on the algorithm!
  - the model of $f$ being used (linear/quadratic/cubic)
  - the model minimization (global vs. local)
  - the cost of an iteration
- typically very pessimistic
- (usually quite tricky and technical. . . )

# A first approach to first-order optimality

Consider achieving (approximate) 1rst-order optimality:

SURPRISE nr 1: a bound exists!
(and is independent of problem dimension)

| | | |
|---|---|---|
| Gradient methods | $O(1/\epsilon_g^2)$ | Nesterov |
| 1rst-order trust-region | $O(1/\epsilon_g^2)$ | Gratton, Sartenaer and T. |

# How to prove such results?

1. Prove that, at "successful iterations" ($j \in \mathcal{S}$),

$$f(x_j) - f(x_{j+1}) \geq \kappa_r \|\nabla_x f(x_{j+1})\|^\alpha, \quad \alpha = 2$$

2. Assume $\|\nabla_x f(x_j)\| \geq \epsilon_g$ for all $j = 0, \ldots, k$.
Then, for $n_{\mathcal{S}}(k) = |\mathcal{S} \cap \{1, \ldots, k\}|$,

$$\begin{aligned} n_{\mathcal{S}}(k)\, \epsilon_g^\alpha &\leq \sum_{j=0,\, j\in\mathcal{S}}^{k} \|\nabla_x f(x_{j+1})\|^\alpha \leq \frac{1}{\kappa_r} \sum_{j=0,\, j\in\mathcal{S}}^{k} [f(x_j) - f(x_{j+1})] \\ &\leq \frac{f(x_0) - f(x_{k+1})}{\kappa_r} \leq \frac{f(x_0) - f_*}{\kappa_r} \end{aligned}$$

and thus

$$n_{\mathcal{S}}(k) \leq \frac{f(x_0) - f_*}{\kappa_r \epsilon^\alpha}$$

3. Prove that

$$k \leq \kappa_s n_{\mathcal{S}}(k)$$

# More on first-order optimality (1)

SURPRISE nr 2:
a better bound exists for (cubic) *regularization methods*

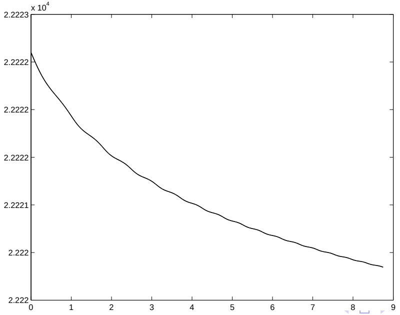
| | | |
|---|---|---|
| With global model min | $O(1/\epsilon_g^{3/2})$ | Nesterov |
| With local model min | $O(1/\epsilon_g^{3/2})$ | Cartis, Gould and T. |

# More on first-order optimality (2)

MOREOVER: the better bound (for cubic regularization) is

- sharp
- optimal for 2nd-order methods

Explicit counter example built by Hermite interpolation (Cartis, Gould and T.)

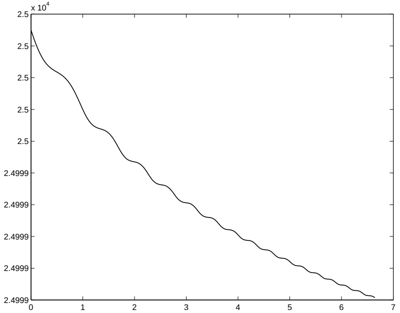


# More on first-order optimality (3)

IN ADDITION:

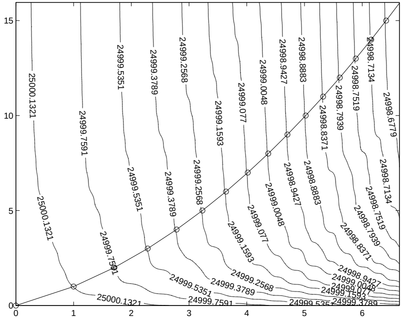
the not-so-good bound for steepest descent is also *sharp*

Another explicit counter example built by Hermite interpolation (Cartis, Gould and T.)



# And then...

SURPRISE nr 3:
Newton's method may need as many iterations
as steepest descent (in its *worst* case)!!!




⇒ Second-order information useless in the worst case!

# Other results

We can also prove that

- the (better) bound for cubic regularization extend to
  - methods using finite-difference gradients
  - derivative-free methods

  (but now depends also on dimension)
- the boundedness of level sets has no impact on the complexity bound
- the not-so-good bound for steepest descent extends to composite non-smooth functions
- much better results hold for the convex case
- also on special function classes (gradient dominated,. . . )

# Finding weak unconstrained minimizers

We are now interested in finding $x_k$ such that

$$\|\nabla_x f(x_k)\| \leq \epsilon_g \quad \textit{and} \quad \lambda_{\min}[\nabla_{xx} f(x_k)] \geq -\epsilon_H$$

(needs second-order information)
For the cubic regularization:

| | | |
|---|---|---|
| With global model min | $O(1/\epsilon_g^3)$ | Nesterov |
| With line model min | $O(1/\epsilon_g^3)$ | Cartis, Gould and T. |

# Finding weak unconstrained minimizers (2)

But also

| | Cubic reg. | Trust-region |
|---|---|---|
| $\epsilon_H \leq \epsilon_g$ | $O(\epsilon_H^{-3})$ sharp | $O(\epsilon_H^{-3})$ sharp |
| $\epsilon_g < \epsilon_H < \sqrt{\epsilon_g}$ | $O(\epsilon_H^{-3})$ sharp | $O(\epsilon_H^{-\{3,5\}})$ |
| $\sqrt{\epsilon_g} \leq \epsilon^H$ | $O(\epsilon_g^{-3/2})$ sharp | $O(\epsilon_g^{-[2,5/2]})$ "sharp" |

Practically sensible: $\epsilon_H \approx \sqrt{\epsilon_g}$

# Constrained optimization

Consider the constrained nonlinear programming problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ & x \in \mathcal{F} \end{array}$$

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

$$\mathcal{F} \text{ is convex.}$$

Typical: bounds on the variables

Main ideas:

- exploit (cheap) projections on convex sets
- prove global convergence + function-evaluation complexity

# A cubic regularization algorithm for the constrained case

For projection-variants to achieve (approximate) 1rst-order optimality

SURPRISE nr 4:
The same bounds hold as for the uncontrained case!!!

| | |
|---|---|
| 1rst-order cubic regularization | $O(1/\epsilon_g^2)$ |
| 2nd-order cubic regularization | $O(1/\epsilon_g^{3/2})$ |

Cartis, Gould and T.

$\Rightarrow$ Convex bounds irrelevant for 1rst-order complexity!

# Conclusions and perspectives

strong position of the cubic regularization approach

worst-case analysis not irrelevant for algorithm design

challenging emerging area with many open questions

**Many thanks for your attention!**

# Further reading


- C. Cartis, N. I. M. Gould and Ph. L. Toint,
  **An adaptive cubic regularisation algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity**,
  IMA Journal of Numerical Analysis, submitted, 2011.
- C. Cartis, N. I. M. Gould and Ph. L. Toint,
  **On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization**,
  SIAM Journal on Optimization, vol. 20(6), pp. 2833–2852, 2010.
- C. Cartis, N. I. M. Gould and Ph. L. Toint,
  **Adaptive cubic overestimation methods for unconstrained optimization. Part II: worst-case function-evaluation complexity**,
  Mathematical Programming A, to appear, 2011.
- C. Cartis, N. I. M. Gould and Ph. L. Toint,
  **On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization**,
  Rapport NAXYS-03-2010, 2010.
- C. Cartis, N. I. M. Gould and Ph. L. Toint,
  **Complexity bounds for second-order optimality in unconstrained optimization**,
  Rapport NAXYS-11-2010, 2010.
- C. Cartis, N. I. M. Gould and Ph. L. Toint,
  **Evaluation complexity of adaptive cubic regularization methods for convex unconstrained optimization**,
  Rapport NAXYS-05-2010, 2010.
- S. Gratton, A. Sartenaer and Ph. L. Toint,
  **Recursive Trust-Region Methods for Multiscale Nonlinear Optimization**,
  SIAM Journal on Optimization, vol. 19(1), pp. 414-444, 2008.
- Yu. Nesterov and B. T. Polyak,
  **Cubic regularization of Newton method and its global performance**,
  Mathematical Programming A, vol. 108(1), pp. 177-205, 2006.