IMPROVED ALGORITHMS FOR CONVEX MINIMIZATION IN RELATIVE SCALE

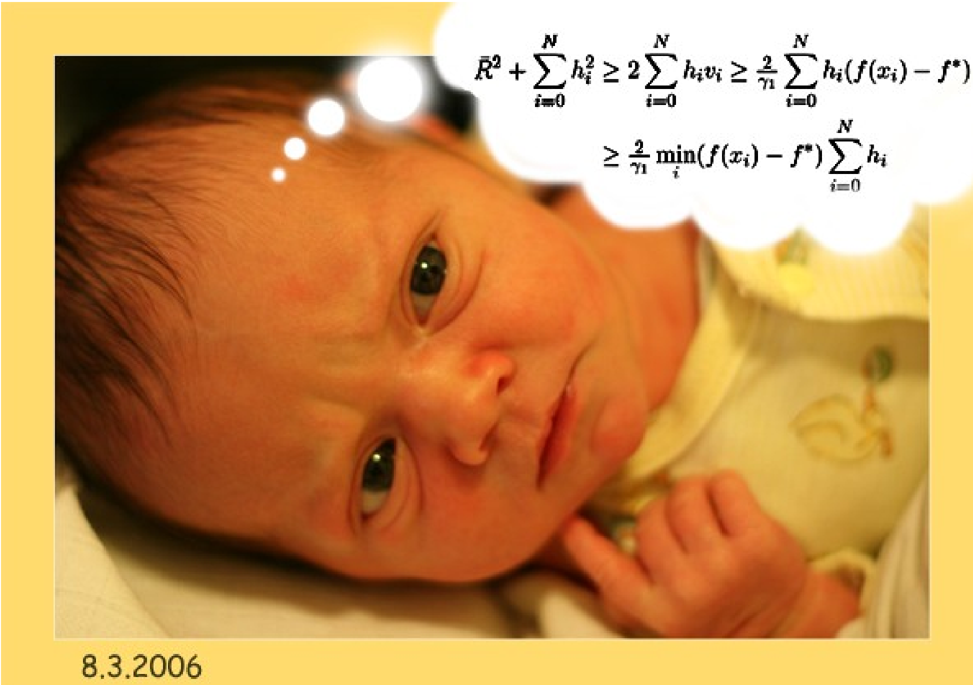
Peter Richtárik
Cornell University

International Symposium on Mathematical Programming
Rio de Janeiro
July 30-August 4, 2006

1. OUTLINE

- The problem; sublinearity
- Ellipsoidal rounding $\Rightarrow$ first aprox. alg.
- Subgradient method $\Rightarrow$ any accuracy
- Preliminary computational experiments
- Smoothing $\Rightarrow$ faster algorithms
- Applications, future work

2. WHERE DO I GET MY IDEAS FROM?



3. THE PROBLEM

Minimize **sublinear** function f over **affine subspace** $\mathcal{L}$

$$f^* \longleftarrow \min\{f(x) \mid x \in \mathcal{L}\}$$

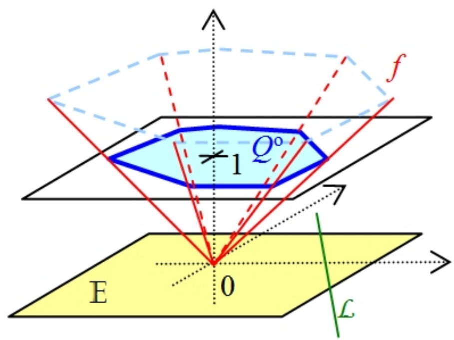
Goal: find solution x with **relative error** δ: $f(x) - f^* \le \delta f^*$

Correspondence: $f(x) = \max\{\langle s, x\rangle \mid s \in Q\}$, where

finite sublinear f $\leftrightarrow$ nonempty convex compact Q

Assumptions:

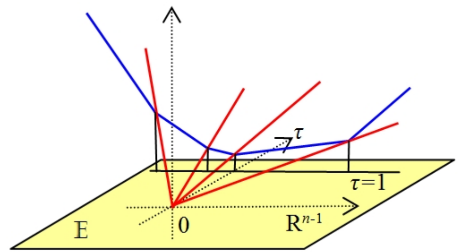
- $f : \mathbb{E} \to \mathbb{R}$
- $0 \in \operatorname{int} Q$
- $0 \notin \mathcal{L} \subset \mathbb{E}$



4. WHY SUBLINEAR FUNCTIONS?

Example: Minimizing the max of abs values of affine functions:

$$\min_{y\in\mathbb{R}^{n-1}} \max_{1\le i\le m} \{|\langle \bar{a}_i, y\rangle - c_i|\}$$



Homogenization:
$a_i = [\bar{a}_i^T; -c_i]$, $x = [y^T, \tau] \in \mathbb{R}^n$ gives

$$\min_{x\in\mathbb{R}^n}\{\max_{1\le i\le m} |\langle a_i, x\rangle| \;:\; x_n = 1\}$$

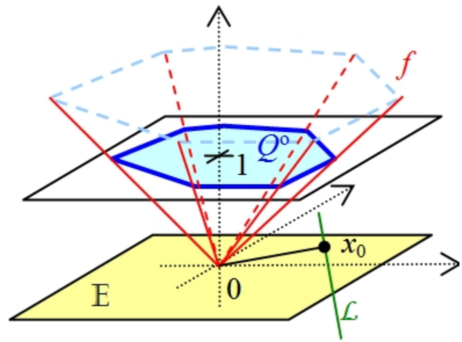
So have $\min f(x)$ subject to $x \in \mathcal{L}$ where

- $f(x) = \max\{\langle s, x\rangle \mid s \in Q\}$
- $Q = \{\pm a_i; i = 1, \ldots, m\}$ (or convex hull of this)
- $\mathcal{L} = \{x \in \mathbb{R}^n \mid x_n = 1\}$

5. FIRST IDEA

Notice:

- f "looks like" a norm
- it is easy to minimize a norm over an affine subspace



Idea:

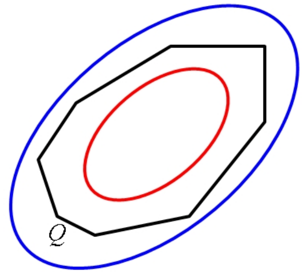
- Approximate f by a Euclidean norm $\|\cdot\|_G$ and compute the projection x_0
- How good is $f(x_0)$ compared to f^*?

6. ELLIPSOIDAL ROUNDING

Assume we have found G and values

$$0 < \gamma_0 \leq \gamma_1$$

such that

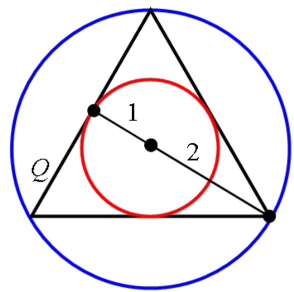


$$E(G, \gamma_0) \subseteq Q \subseteq E(G, \gamma_1),$$

where $E(G, r) = \{s : \sqrt{\langle s, G^{-1}s \rangle} \leq r\}$.

Then $\gamma_0 \|x\|_G \leq f(x) \leq \gamma_1 \|x\|_G$ for all $x \in \mathbb{E}$

Key parameter: $\alpha = \gamma_0 / \gamma_1 \in (0, 1]$
$\Rightarrow$ α-**rounding**



Theorem [John]: Every convex body admits a $1/n$-rounding. Centrally symmetric bodies admit $1/\sqrt{n}$-rounding.

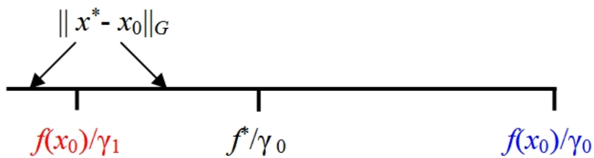
7. KEY CONSEQUENCES

It can be shown that

(1) $\frac{f(x_0)}{\gamma_1} \le \|x_0\|_G \le \frac{f^*}{\gamma_0} \le \frac{f(x_0)}{\gamma_0}$

(2) $\|x^* - x_0\|_G \le \frac{f^*}{\gamma_0}$

(3) f is γ_1-Lipschitz



Notice that

- $(1) \quad \Rightarrow \quad f(x_0) \le (1+\delta) f^*$ with $\delta = \frac{1}{\alpha} - 1$
  $\Rightarrow \quad O(1/\alpha)$-**approximation algorithm**
- $(2) + (3)$ suggest further use of **subgradient method** started from x_0

8. A SUBGRADIENT METHOD

Constant step-size subgradient algorithm:

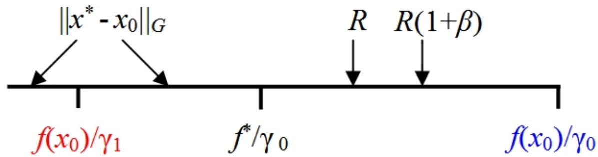
1. Choose R such that $\|x^* - x_0\|_G \leq R$
2. For $k = 0 \ldots N-1$ repeat $x_{k+1} = x_k - \frac{R}{\sqrt{N+1}} g$
   (g is subgradient of f at x_k projected onto $\mathcal{L}$ and normalized)
3. Output best point seen x

Theorem: $f(x) - f^* \leq \frac{\gamma_1 R}{\sqrt{N+1}}$

Aiming for relative error:

- **Available upper bound:** $R = f(x_0)/\gamma_0 \quad \Rightarrow \quad N = \lfloor \frac{1}{\alpha^4 \delta^2} \rfloor$ iterations needed to get within $1+\delta$ of f^*
- **Ideal upper bound:** $R = f^*/\gamma_0 \quad \Rightarrow \quad N = \lfloor \frac{1}{\alpha^2 \delta^2} \rfloor$
- **Nesterov's approach:** Start with the bad bound and iteratively improve it $\quad \Rightarrow \quad N = O(\frac{1}{\alpha^2 \delta^2} \ln \frac{1}{\alpha})$

9. BISECTION IDEA



Key lemma: If $f^*/\gamma_0 \leq R$ then subgradient method after

$$N = \lfloor \frac{1}{\beta^2\alpha^2} \rfloor = O(\frac{1}{\alpha^2})$$

steps outputs x with

$$\frac{f(x)}{\gamma_0} \leq R(1+\beta)$$

This leads to speedup of Nesterov's algorithm:

Approach	Complexity
"Ideal upper bound"	$O(\frac{1}{\alpha^2\delta^2})$
Nesterov's algorithm	$O(\frac{1}{\alpha^2\delta^2}\ln\frac{1}{\alpha})$
Bisection algorithm	$O(\frac{1}{\alpha^2}\ln\ln\frac{1}{\alpha} + \frac{1}{\alpha^2\delta^2})$

10. NON-RESTARTING ALGORITHM

- Subgradient subroutine is always started from x_0
- Can we use collected information to start next routine from a different point?

Key lemma: If $f^*/\gamma_0 \le R$ then subgradient method started from x_-, run for $N = \lfloor \frac{1}{\beta^2\alpha^2} \rfloor$ steps with step lengths $(\|x_-\|_G + R)/\sqrt{N+1}$ outputs x with

$$\frac{f(x)}{\gamma_0} \le R(1+\beta) + \frac{f(x_-)}{\gamma_0}\beta$$

Approach	Complexity
Nesterov's algorithm	$O(\frac{1}{\alpha^2\delta^2} \ln \frac{1}{\alpha})$
Nonrestarting Nesterov's algorithm	$O(\frac{1}{\alpha^2\delta^2} \ln \frac{1}{\alpha})$
Bisection algorithm	$O(\frac{1}{\alpha^2} \ln \ln \frac{1}{\alpha} + \frac{1}{\alpha^2\delta^2})$
Nonrestarting bisection algorithm	$O(\frac{1}{\alpha^2} \ln \frac{1}{\alpha} + \frac{1}{\alpha^2\delta^2})$

11. SOME COMPUTATIONAL EXPERIMENTS

Problem:

$$\min f(x) \equiv \max_{i=1:m} |\langle a_i, x\rangle| \quad \text{subject to} \quad \langle d, x\rangle = 1$$

- We first construct a good and a bad ellipsoidal rounding of the centrally symmetric set

$$Q = \partial f(0) = \mathrm{Conv}\{\pm a_i, i = 1, \dots, m\}$$

- A good rounding has $\alpha \approx 1/\sqrt{n}$ and a bad $\alpha = 1/\sqrt{m}$.
- Random instances with $n = 100, m = 500, \delta = 0.05$.

α	Nest	Nest NR	Bis	decrease in f
$1/11$	$290100, 28, 2^{\dagger}$	$725250, 70, 2$	$146654, 14, 5$	$6.26 \downarrow 3.46$
$1/11$	$145050, 15, 1$	$145050, 15, 1$	$147055, 14, 6$	$4.97 \downarrow 3.05$
$1/22$	$1160400, 117, 2$	$2901003, 291, 2$	$588235, 60, 6$	$6.53 \downarrow 3.15$

† number of lower level iterations; time in seconds and number of calls of the subgradient method.

12. SMOOTHING - GENERAL IDEA

Some methods for minimizing convex functions:

f	Method	Complexity
non-smooth	Black-box subgradient method	$O(\frac{1}{\epsilon^2})$
smooth, ∇f Lipschitz	Efficient smooth method	$O(\sqrt{\frac{L}{\epsilon}})$
non-smooth	Nesterov's smoothing method	$O(\frac{1}{\epsilon})$

Yu. Nesterov. *Smooth Minimization of Nonsmooth Functions, 2003*

Basic Idea: Find smooth ϵ-approximation of f with $O(1/\epsilon)$-Lipschitz gradient and then apply efficient smooth method

$$"O(\sqrt{O(1/\epsilon)/\epsilon}) = O(1/\epsilon)"$$

13. SMOOTHING

Assumptions:

- $Q_1 \subset \mathbb{E}_1, Q_2 \subset \mathbb{E}_2$; closed compact
- $A : \mathbb{E}_1 \to \mathbb{E}_2^*$, linear
- $f : \mathbb{E}_1 \to \mathbb{R}, \quad f(x) = \max\{\langle Ax, u \rangle_2 \mid u \in Q_2\}$

The problem: minimize $f(x)$ subject to $x \in Q_1$

Smoothing: Let d_2 be nonnegative continuous and strongly convex on Q_2 with convexity parameter σ_2. For $\mu > 0$ define

$$f_\mu(x) = \max\{\langle Ax, u \rangle_2 - \mu d_2(u) \mid u \in Q_2\}, \quad \text{then}$$

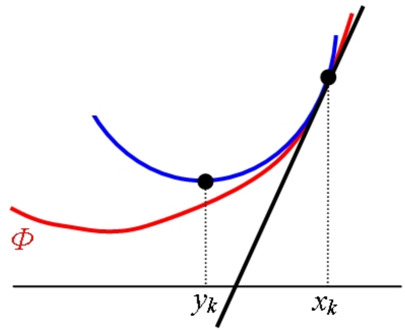
$f_\mu(x) \leq f(x) \leq f_\mu(x) + \mu D_2$, where $D_2 = \max\{d_2(u) \mid u \in Q_2\}$

Theorem [Nesterov, 2003]: f_μ is smooth with Lipschitz continuous gradient with constant $L_\mu = \frac{\|A\|^2}{\mu \sigma_2}$

14. EFFICIENT SMOOTH METHOD

Problem: $\min_x\{\phi(x) : x \in Q\}$

- Q - convex compact set
- $\phi(x)$ - convex & smooth
- $\nabla\phi(x)$ - L-Lipschitz in $\|\cdot\|_G$



Method

For $k = 0, 1, \ldots, N$ repeat

- $y_k := \arg\min_{y\in Q}\{\langle\nabla\phi(x_k), y - x_k\rangle + \frac{L}{2}\|y - x_k\|_G^2\}$
- $z_k := \arg\min_{z\in Q}\{\langle\sum_{i=0}^{k}\frac{i+1}{2}\nabla\phi(x_i), z - x_i\rangle + \frac{L}{2}\|z - x_0\|_G^2\}$
- $x_{k+1} := \frac{2}{k+3}z_k + \frac{k+1}{k+3}y_k$

Output $x \leftarrow y_N$

Theorem [Nesterov]: $\phi(x) - \phi(x^*) \leq \frac{2L\|x_0 - x^*\|_G^2}{(N+1)^2}$

15. PUTTING IT ALL TOGETHER

Problem: min $f(x) = F(Ax)$ subject to $x \in \mathcal{L}$ where

$$F(v) = \max\{\langle v, u\rangle_2 \mid u \in Q_2\}$$

where $A : R^n \to R^m$ full column rank, $0 \in \operatorname{int} \partial F(0) = \operatorname{int} Q_2$

Step 1: rounding

- Note: $\partial F(0) = Q_2 \quad \Rightarrow \quad \partial f(0) = A^T Q_2$
- find ball α-rounding: $B_{\|\cdot\|_2}(1) \subseteq \partial F(0) \subseteq B_{\|\cdot\|_2}(1/\alpha)$ so that $B_{\|\cdot\|_G^*}(1) \subseteq \partial f(0) \subseteq B_{\|\cdot\|_G^*}(1/\alpha)$ if $G = A^T A$

Step 2: smoothing $\Rightarrow L_\mu = 1/\mu$

Step 3: apply smooth method

$$f^* \le R \quad \Rightarrow \quad x^* \in Q(R) = \{x \mid \|x - x_0\|_G \le R,\ x \in \mathcal{L}\}$$

Use bisection to find good R as before!

16. ALGORITHM COMPARISON

Theorem [R.05]: There is an algorithm for finding point within $(1+\delta)$ of f^* in $O(\frac{1}{\alpha}\ln\ln\frac{1}{\alpha}+\frac{1}{\alpha\delta})$ iterations of the efficient smooth method.

Approach	Complexity
Nesterov's algorithm	$O(\frac{1}{\alpha^2\delta^2}\ln\frac{1}{\alpha})$
Nonrestarting Nesterov's algorithm	$O(\frac{1}{\alpha^2\delta^2}\ln\frac{1}{\alpha})$
Bisection algorithm	$O(\frac{1}{\alpha^2}\ln\ln\frac{1}{\alpha}+\frac{1}{\alpha^2\delta^2})$
Nonrestarting bisection algorithm	$O(\frac{1}{\alpha^2}\ln\frac{1}{\alpha}+\frac{1}{\alpha^2\delta^2})$
Nesterov's smoothing algorithm	$O(\frac{1}{\alpha\delta}\ln\frac{1}{\alpha})$
Smoothing bisection algorithm	$O(\frac{1}{\alpha}\ln\ln\frac{1}{\alpha}+\frac{1}{\alpha\delta})$

Note: The bisection improvement of the smoothing method has been earlier independently obtained by Fabián Chudak and Vânia Eleutério [2005] in the context of combinatorial problems (facility location, packing, scheduling unrelated parallel machines, . . .).

17. APPLICATION EXAMPLES

- **minimizing the max of abs values of affine functions:**

$$\min_{y\in\mathbb{R}^{n-1}} \max_{1\le i\le m} \{|\langle \bar{a}_i, y\rangle - c_i|\}$$

Rounding: $O(n^2(m+n)\ln m)$

Optimization: $O(\sqrt{n\ln m}(\ln\ln n + \frac{1}{\delta}))$ iters of order $O(mn)$

- minimization of largest eigenvalue
- minimization of the sum of largest eigenvalues
- minimization of spectral radius
- bilinear matrix games with nonnegative coefficients, and more

18. CURRENT AND FUTURE WORK

- **Merging rounding and optimization phases**
- Making the subgradient algorithms more practical: variable step lengths/line search.
- Non-ellipsoidal rounding. Sparse rounding.

19. ACKNOWLEDGEMENT

Big thanks to

- **Yurii Nesterov** for his papers!
  - *Smooth minimization of nonsmooth functions, 2003*
  - *Unconstrained convex minimization in relative scale, 2003*
  - *Rounding of convex sets and efficient gradient methods for LP problems, 2004*
- **Mike Todd** for enlightening discussions !

20. One more picture...