Optimization in Relative Scale

Yurii Nesterov, CORE/INMA (UCL)

May 25, 2012 (University of Edinburgh)

Standard approach: Absolute accuracy

Problem: $f^* \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q} f(x)$, where $Q \subseteq R^n$ is a closed convex set.

Def: For $\epsilon > 0$, find $\bar{x} \in Q$ satisfying $\boxed{f(\bar{x}) \leq f^* + \epsilon}$.

Black Box problem classes

- *Bounds on the growth.* (Strong) convexity with $\mu \geq 0$:
  $f(y) \geq f(x) + \langle f'(x), y - x \rangle + \frac{1}{2}\mu \|y - x\|^2, \quad x, y \in Q$
- *Bounds on derivatives.* For example,
  $$\|f'(x)\|_* \leq M, \quad \text{or}, \quad \|f''(x)\| \leq L, \quad \text{etc.}$$

Structural Optimization

- *Functional model of feasible set*: self-concordant barriers.
- *Smoothing technique.*

Note: operation $f \Rightarrow f + \text{const}$ does not change complexity.

Standard approach: Complexity Bounds

Sublinear convergence

- *Nonsmooth functions:* $f(x_k) - f^* \leq \frac{MR}{\sqrt{k}} \;\Rightarrow\; k \leq \frac{M^2R^2}{\epsilon^2}$.
- *Smooth functions:* $f(x_k) - f^* \leq \frac{LR^2}{k^2} \;\Rightarrow\; k \leq \frac{L^{1/2}R}{\epsilon^{1/2}}$.
- *Smoothing technique:* $f(x_k) - f^* \leq \mu + \frac{M^2R^2}{\mu k^2} \;\Rightarrow\; k \leq \frac{MR}{\epsilon}$.

Linear convergence

- *Cutting plane:* $f(x_k) - f^* \leq MR \cdot e^{-k/n} \;\Rightarrow\; k \leq n \ln \frac{MR}{\epsilon}$.
- *Interior point:* $f(x_k) - f^* \leq \nu \cdot e^{-k/\sqrt{\nu}} \;\Rightarrow\; k \leq \sqrt{\nu} \ln \frac{\nu}{\epsilon}$.

In both cases, complexity of each iteration is very high ($n^3 \dots n^5$).

Some criticism

Sublinear convergence

- Constants L, M, and R are unknown. They can be very big.

Linear convergence

Dependence $\ln \frac{1}{\epsilon}$ is very weak. We can reach any accuracy. **But,**

- In many situations, $n \approx \frac{1}{\xi^p}$, where ξ is the *accuracy of the model*.
- If $\epsilon \approx \xi$, then the notion of polynomial solvability looses any sense.

Alternative approach: Relative accuracy of the solution.

Optimal method for smooth functions

Problem: $\min_x \{f(x) : \ x \in Q\}$ with $f \in C^{1,1}(Q)$.

Prox-function: strongly convex $d(x)$, $x \in Q$:

$$d(x_0) = 0, \quad d(x) \geq 0 \ \forall x \in Q.$$

Gradient mapping:

$$T(x) = \arg\min_{y \in Q} \left\{ \langle \nabla f(x), y - x \rangle + \tfrac{1}{2} L(f) \|y - x\|^2 \right\}.$$

Method FGM: **For** $k \geq 0$ **do**

1. Compute $f(x_k), \nabla f(x_k)$. Find $y_k = T(x_k)$.
2. Find $z_k = \arg\min_{x \in Q} \{ L(f) d(x) + \sum_{i=0}^{k} \frac{i+1}{2} \langle \nabla f(x_i), x \rangle \}$.
3. Set $x_{k+1} = \frac{2}{k+3} z_k + \frac{k+1}{k+3} y_k$.

Convergence: $f(y_k) - f(x^*) \leq \frac{4L(f)d(x^*)}{(k+1)^2}$.

Relative accuracy (RA)

Problem: $f^* \stackrel{\mathrm{def}}{=} \min\limits_{x \in Q} f(x) \; > 0,$ where Q is a closed convex set.

Definition:

For $\delta \in (0,1)$, find $\bar{x} \in Q$ satisfying $(1-\delta) f(\bar{x}) \leq f^* \leq f(\bar{x})$.

Condition $f^* > 0$ must be guaranteed. *How?*

Different approaches:

- Homogeneous model.
- Polyhedral model.
- Barrier subgradient method.
- Minimization of strictly positive functions.

Homogeneous model

Problem: $f(x) = F(A^T x) \;\to\; \min: \; x \in \mathcal{L} = \{x : \; Cx = b\}$,

where $F(y)$ is a convex homogeneous function of degree one:

$$F(y) = \max_{s \in Q_2} \langle s, y \rangle,$$

and $0 \in \mathrm{int}\, Q_2 \subset R^m$. Then $f^* > 0$.

Example: $f(x,\tau) = \max\limits_{1 \le j \le m} |\langle a_j, x\rangle + \tau b_j| \quad \to \quad \min\limits_{x,\tau} : \; \tau = 1.$ $\square$

Let $\| \cdot \|_2$ be a Euclidean norm in R^m, $B(r) = \{y : \; \|y\|_2 \le r\}$,

$$\gamma_0 = \max_r \{r : \; B(r) \subseteq Q_2\}, \; \gamma_1 = \max_r \{r : \; B(r) \supseteq Q_2\}.$$

Then for $\|x\| = \|A^T x\|_2$ we have $\gamma_0 \|x\|_1 \le f(x) \le \gamma_1 \|x\|_1$.

Moreover, for $x_0 = \arg\min\limits_{x \in \mathcal{L}} \|x\|_1$ and any $x \in \mathcal{L}$ we have

$$\|x_0 - x^*\| \le \frac{1}{\gamma_0} f^* \le \frac{1}{\gamma_0} f(x).$$

Optimization strategy

Denote $f_\mu(x) = \max\limits_u \{\langle A^T x, s\rangle - \frac{1}{2}\mu\|s\|_2^2 : \ s \in Q_2\}$,
$$Q(R) = \{x \in \mathcal{L} : \ \|x\| \leq R\}.$$

Let $x_N(R)$ be an output of FGM after N steps as applied to f_μ with
$$\mu = \frac{2R}{\gamma_1 \cdot (N+1)}, \quad Q = Q(R).$$

Denote $\alpha = \frac{\gamma_1}{\gamma_0} \geq 1$, $\tilde{N} = \left\lfloor 2e \cdot \alpha \cdot \left(1 + \frac{1}{\delta}\right) \right\rfloor$. Consider the process:

Set $\hat{x}_0 = x_0$. For $t \geq 1$ iterate

$\hat{x}_t := x_{\tilde{N}}\left(\frac{1}{\gamma_0} f(\hat{x}_{t-1})\right)$. **If** $f(\hat{x}_t) \geq \frac{1}{e} f(\hat{x}_{t-1})$ **then** $T := t$ **and** Stop.

Theorem. $T \leq 1 + \ln \alpha$. *Moreover,* $f(\hat{x}_T) \leq (1+\delta) f^*$, *and the total number of lower-level steps in the process does not exceed*
$$2e \cdot \alpha \cdot \left(1 + \tfrac{1}{\delta}\right) \cdot (1 + \ln \alpha).$$

Example: $f(x) = \max\limits_{1\le j\le m} |\langle a_j, x\rangle|, \; m > n.$

$$F(s) = \max_{1\le j\le m} |s^{(j)}|, \quad \|s\|_2^2 = \sum_{j=1}^m s_j^2,$$
$$\gamma_0 = \tfrac{1}{\sqrt{m}}, \quad \gamma_1 = 1, \quad \alpha = \sqrt{m}.$$

Number of iterations: $2e\sqrt{m}\cdot\left(1+\frac{1}{\delta}\right)\cdot\left(1+\frac{1}{2}\ln m\right)$.

Each iteration takes $O(mn)$ operations. Thus, the total complexity is

$$O\left(mn^2 + \tfrac{m^{1.5}n}{\delta}\ln m\right) \quad \text{a.o.}$$

For IPM the theoretical bound is $O\left((m^{1.5}n + m^{0.5}n^3)\ln\frac{1}{\delta}\right)$ a.o.
The switching rule is $\frac{m}{n^2} \le \delta \ln \frac{1}{\delta}$.

Question: Is it possible to improve α?

Asphericity of convex sets

Main inequality: $\gamma_0\|x\| \leq f(x) \leq \gamma_1\|x\|, \ x \in R^n,$ is used for

- bounding of the dual set $\partial f(0)$ (f is homogeneous);
- controlling the distance to the solution by

$$\gamma_0\|x_0 - x^*\| \leq f^* \leq f(x), \quad x \in \mathcal{L}.$$

John Theorem: For any bounded convex *symmetric* set $Q \subset R^n$ there exists a Euclidean norm $\|\cdot\|$ such that

$$B_{\|\cdot\|}(1) \subseteq Q \subseteq B_{\|\cdot\|}(\sqrt{n}).$$

Thus, if $f(x) = f(-x)$, we can expect $\alpha \approx \sqrt{n}$.

In which cases such a norm is computable?

Application example: Rounding

Consider $f(x) = \max\limits_{1 \le j \le m} |\langle a_j, x \rangle|$. Then

$$Q \equiv \partial f(0) = \mathrm{Conv}\,\{\pm a_j,\ j = 1, \dots, m\}.$$

Denote $G_0 = \frac{1}{m} \sum\limits_{j=1}^{m} a_j a_j^T$, $\|a\|_G^* = \langle G^{-1} a, a \rangle^{1/2}$.

Choose a tolerance $\gamma > 1$. **For** $k \ge 0$ **iterate:**

1. Compute $g_k \in Q : \ \|g_k\|_{G_k}^* = r_k \stackrel{\mathrm{def}}{=} \max\limits_g \{\|g\|_{G_k}^* : \ g \in Q\}$.

2. If $r_k \le \gamma n^{1/2}$ **then** Stop **else**

$$\alpha_k = \frac{1}{n} \cdot \frac{r_k^2 - n}{r_k^2 - 1}, \quad G_{k+1} = (1 - \alpha_k) G_k + \alpha_k g_k g_k^*.$$

Theorem. This scheme terminates after at most $N = \frac{n \ln m}{2 \ln \gamma - 1 + \gamma^{-2}}$ iterations with $B_{\|\cdot\|_{G_N}^*}(1) \subset Q \subset B_{\|\cdot\|_{G_N}^*}(\gamma\sqrt{n})$.

Note: Complexity of each iteration is $O(mn)$ a.o.

Application example: Complexity

Problem: $f(x) = \max\limits_{1 \le j \le m} |\langle a_j, x \rangle| \quad \to \quad \min\limits_{x \in R^n} : \langle c, x \rangle = 1.$

Phase 1: find a rounding norm $\| \cdot \|^*$ for the set $Q \equiv \partial f(0) = \mathrm{Conv}\,\{\pm a_j,\ j = 1, \dots, m\}$ of asphericity $\gamma > 1$.

Complexity: $O(mn^2 \ln m)$ a.o.

Phase 2: using this norm, solve our problem up to a relative accuracy δ by a smoothing technique.

Complexity: $O\left(\frac{\sqrt{n}}{\delta} \ln n \sqrt{\ln m} \right)$ iterations of a gradient scheme. In total,

$$O\left(\frac{mn^{1.5}}{\delta} \ln n \sqrt{\ln m} \right) \quad \text{a.o.}$$

Competitors:

- *Ellipsoid method:* $\left(n^2 \ln \frac{1}{\delta} \right) \times mn.$
- *Interior Point:* $\left(\sqrt{m} \ln \frac{m}{\delta} \right) \times mn^2.$

LP-problems with nonnegative components

Problem: $f(x) = \max\limits_{1 \le j \le m} \langle a_j, x \rangle \quad \to \quad \min\limits_{x \ge 0 \in R^n} : \ \langle a_0, x \rangle = 1,$

where $a_j \ge 0$, $j = 1, \ldots, m$, and $a_0 > 0$.

Phase 1: find a diagonal norm $\| \cdot \|_D$ such

$$\|x\|_D \le f(x) \le \gamma \sqrt{n} \cdot \|x\|_D$$

for asphericity $\gamma > 1$. **Complexity**: $O\left(mn^2(\ln n + 2\ln m)\right)$ a.o.

Phase 2: using this norm, solve the problem up to a relative accuracy δ by the smoothing technique.

Complexity: $O\left(\frac{\sqrt{n}}{\delta} \ln n \sqrt{\ln m}\right)$ iterations. In total,

$$O\left(\frac{mn^{1.5}}{\delta} \ln n \sqrt{\ln m}\right) \quad \text{a.o.}$$

Competitors:

- *Ellipsoid method:* $\quad O\left(n^2 \ln \frac{1}{\delta}\right) \times mn.$
- *Interior point:* $\quad O\left(\sqrt{m} \ln \frac{m}{\delta}\right) \times mn^2.$

Concave Maximization

Primal problem: $\max\limits_{x \in Q_p} f(x)$,

where $f(x) = \min\limits_{w \in Q_d} \langle Ax + b, w \rangle$. (Concave objective!)

Assume: At $x \in Q_p$ we can compute $f(x)$ and $f'(x) = A^T w(x)$, where $w(x) \in \operatorname{Arg} \min\limits_{w \in Q_d} \langle Ax + b, w \rangle$.

Dual problem: $\min\limits_{w \in Q_d} \eta(w)$, where $\eta(w) = \max\limits_{x \in Q_p} \langle Ax + b, w \rangle$.

Lemma. Let $\{\lambda_i \geq 0\}$ and $\{x_i \in Q_p\}$. Define

$$l_k(y) = \sum_{i=0}^{k} \lambda_i \langle f'(x_i), y - x_i \rangle, \quad l_k^* = \max_{y \in Q_p} l_k(y).$$

Let $S_k = \sum\limits_{i=0}^{k} \lambda_i$, $\bar{x}_k = \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i x_i$, $\bar{w}_k = \frac{1}{S_k} \sum\limits_{i=0}^{k} \lambda_i w(x_i)$. Then

$$\eta(\bar{w}_k) - f(\bar{x}_k) \leq \frac{1}{S_k} l_k^*.$$

Barrier subgradient method

Denote by x_0 the *analytic center* of $Q_p \subset E$: $x_0 = \arg\min\limits_{x \in Q_p} F(x)$, where $F(x)$ is a ν-self-concordant barrier for Q_p. Denote

$$u^*_\beta(s) = \arg\max_{x \in Q_p} \{\langle s, x - x_0 \rangle - \beta[F(x) - F(x_0)]\}, \; s \in E^*,$$

where $\beta > 0$ is a smoothing parameter. Consider the method:

Initialization: Set $s_0 = 0 \in E^*$. **Iteration $(k \geq 0)$:**

1. Choose $\beta_k > 0$ and compute $x_k = u^*_{\beta_k}(s_k)$.

2. Choose $\lambda_k > 0$ and set $s_{k+1} = s_k + \lambda_k f'(x_k)$.

Assumption: for all $x \in Q_p$ we have

$$\|f'(x)\|^*_x \stackrel{\mathrm{def}}{=} \langle [F''(x)]^{-1} f'(x), f'(x) \rangle^{1/2} \leq M.$$

Choose: $\lambda_k = 1$, $\beta_0 = \beta_1$, $\beta_k = M \cdot \left(1 + \sqrt{\frac{k}{\nu}}\right)$, $k \geq 1$.

Th: $\frac{1}{S_k} l^*_k \leq M \cdot \left(\sqrt{\frac{\nu}{k+1}} + \frac{\nu}{k+1} \right) \cdot O\left(\ln\left(\nu \cdot (k+1)\right)\right) \to 0.$

Main application

Consider a concave optimization problem

$$\psi_* \stackrel{\mathrm{def}}{=} \max_x \{\psi(x) : \; x \in Q_p\},$$

We assume that ψ is concave and *non-negative* on Q_p:

$$\psi(x) \; > \; 0, \quad \forall x \in \mathrm{int}\, Q_p.$$

Lemma: For any $x \in \mathrm{int}\, Q_p$ we have $\|\psi'(x)\|_x^* \; \leq \; \psi(x)$.

Proof: For arbitrary $x \in \mathrm{int}\, Q$ and $r \in [0,1)$ define $y = x - \frac{r}{\|\psi'(x)\|_x^*}[F''(x)]^{-1}\psi'(x)$. Then $y \in \mathrm{int}\, Q$, and

$$0 \leq \psi(y) \; \leq \; \psi(x) + \langle \nabla\psi(x), y - x \rangle = \psi(x) - r\|\nabla\psi(x)\|_x^*. \; \Box$$

Corollary: Define $f(x) \; = \; \ln \psi(x)$. Then

$$\|f'(x)\|_x \; \leq \; 1, \quad x \in Q_p.$$

Hence, we can maximize ψ in *relative scale*!

This leads to *Fully Polynomial-Time Approximation Schemes*.

Problems with nonnegative components

Consider the problem: $\psi_* = \min\limits_{w \in Q_d} \max\limits_{1 \le i \le m} f_i(w)$, where

- Q_d is closed and convex.
- $f_i(w)$ are convex and non-negative on Q_d.

Assume that for any $x \geq 0 \in R^m$ the function

$$\psi(x) = \min_{w \in Q_d} \sum_{i=1}^{m} x^{(i)} f_i(w)$$

is well defined and easily computable.

We can rewrite the problem as

$$\psi_* = \max_x \{\psi(x) : \langle e, x \rangle = 1, \; x \geq 0 \in R^m\},$$

where $e \in R^m$ is the vector of all ones.

Its δ-approximation in relative scale can be found in

$$O^*\left(\frac{m}{\delta^2}\right) \quad \text{iterations.}$$

Application: Semidefinite Relaxation

Let $A \succeq 0$. Consider the problem

$$f_* \stackrel{\mathrm{def}}{=} \max_x \{ \langle Ax, x \rangle : \; x^{(i)} = \pm 1, \; i = 1, \dots, n \},$$

Define SDP-relaxation $\psi_* \;=\; \min_y \{ \langle e, y \rangle : \; D(y) \succeq A \}$,

where $D(y)$ is a diagonal matrix with y on the diagonal.

It is known that $\frac{2}{\pi} \psi_* \leq f_* \leq \psi_*$.

It can be proved that

$$\psi_* = \max_X \{ \; \psi(X) \stackrel{\mathrm{def}}{=} \left[\sum_{i=1}^n \langle X q_i, q_i \rangle^{1/2} \right]^2 : \; \langle I, X \rangle = 1, \; X \succeq 0 \},$$

where q_i are the columns of the matrix L, and $A = L^T L$.

Note that function ψ is concave and positive for $X \succ 0$. We take

$$F(X) \;=\; -\ln \det X, \quad \nu = n.$$

Hence, ψ_* can be approximated in $O^*(\frac{n}{\delta^2})$ iterations.

Each iteration requires a tri-diagonalization of $(n \times n)$-matrix.

Strictly positive functions

Definition

Convex function f is called strictly positive on Q if

$$f(y) + f(x) + \langle f'(x), y - x \rangle \geq 0, \quad x, y \in Q.$$

Corollary: $\quad f(y) \geq |f(x) + \langle f'(x), y - x \rangle|, \quad x, y \in Q.$

Simple properties

- $f(x) \equiv \text{const} > 0$ is strictly positive.
- Strict positivity is an *affine-invariant* property.
- Class of strictly positive functions is a convex cone.

Simple examples

Lemma 1. Let B be bounded, closed, and centrally symmetric.

Then $f(x) = \max\limits_{x \in B} \langle s, x \rangle$ is strictly positive on R^n.

Proof: Since $f(x) = \langle f'(x), x \rangle$ and $-f'(x) \in B$, we have

$$f(y) \;\; \geq \;\; \langle -f'(x), y \rangle \;=\; -f(x) - \langle f'(x), y - x \rangle. \qquad \square$$

The simplest examples of strictly positive functions are *norms*.

Lemma 2. Let $f_1(x)$ and $f_2(x)$ be strictly positive on Q.

Then $f(x) = \max\{f_1(x), f_2(x)\}$ is also strictly positive.

Proof: For arbitrary $x \in Q$, assume $f_1(x) \geq f_2(x)$. Then,

$$\begin{aligned} f(y) \;\; &\geq \;\; f_1(y) \;\geq\; -f_1(x) - \langle f_1'(x), y - x \rangle \\ &= \;\; -f(x) - \langle f'(x), y - x \rangle. \qquad \square \end{aligned}$$

Particular examples

All functions below are strictly positive:

$$
\begin{aligned}
f(x) &= \max_{1 \le i \le m} \|A_i x - b_i\|, \\
f(x) &= \sum_{i=1}^{m} \|A_i x - b_i\|, \\
f(x) &= \sigma_{\max}\left(\sum_{i=1}^{n} A_i x^{(i)}\right), \\
f(x) &= \sum_{j=1}^{m} \sigma_j \left(\sum_{i=1}^{n} A_i x^{(i)}\right),
\end{aligned}
$$

where $A_i \in R^{m \times n}$, and $b_i \in R^m$, $i = 1 \ldots n$.

General convex functions

Theorem 1. Let ϕ be convex function on Q with uniformly bounded subgradients: $\|\phi'(x)\|^* \leq L, \quad x \in Q$.

Then $f(x) = \max\{\phi(x), L\|x\|\}$ is strictly positive on Q.

Proof: Clearly, $\|f'(x)\|^* \leq L$. Therefore,

$$\begin{aligned} f(y) + f(x) + \langle f'(x), y - x \rangle &\geq L\|y\| + L\|x\| + \langle f'(x), y - x \rangle \\ &\geq L\|y\| + L\|x\| - L\|y - x\| \geq 0. \end{aligned}$$

□

Shifted general optimization problem

Consider the problem: $\min_{x \in Q} \phi(x)$, where ϕ has bounded subgradients. Let $x^* \in Q$ be its optimal solution.

Lemma 3. For $x_0 \in Q$ define

$$f(x) = \max\{\phi(x) - \phi(x_0) + 2LR, L\|x - x_0\|\}.$$

It is strictly positive. If $\|x - x_0\| \leq R$ then $f(x) \equiv \phi(x) + \text{const}$.

If $\|x_0 - x^*\| \leq R$, then the optimal value f^* of the equivalent problem $\min_{x \in Q} f(x)$ satisfies $LR \leq f^* \leq 2LR$.

Proof: If $\|x - x_0\| \leq R$, then

$$\phi(x) - \phi(x_0) + 2LR \geq 2LR - L\|x - x_0\| \geq L\|x - x_0\|.$$

Further, $f^* \leq f(x_0) = 2LR$, and

$$f(x) \geq \max\{2LR - L\|x - x_0\|, L\|x - x_0\|\} \geq LR. \qquad \square$$

Optimization problem with squared objective

Problem: $\min\limits_{x \in Q} f(x)$, where f is strictly positive on Q.

New objective: $\hat{f}(x) = \frac{1}{2} f^2(x), \quad \hat{f}'(x) \; = \; f(x) \cdot f'(x).$

Equivalent problem: $\min\limits_{x \in Q} \hat{f}(x).$

Lemma 4. Let f be strictly positive on Q. Then for $x, y \in Q$

$$\hat{f}(y) \;\; \geq \;\; \hat{f}(x) + \langle \hat{f}'(x), y - x \rangle + \tfrac{1}{2} \langle f'(x), y - x \rangle^2.$$

Proof: Indeed,

$$\begin{aligned} \hat{f}(y) \;\; &= \;\; \tfrac{1}{2} f^2(y) \; \geq \; \tfrac{1}{2} [f(x) + \langle f'(x), y - x \rangle]^2 \\ &= \;\; \hat{f}(x) + \langle \hat{f}'(x), y - x \rangle + \tfrac{1}{2} \langle f'(x), y - x \rangle^2. \qquad \square \end{aligned}$$

Important: We have *nonlinear support function*!

Quasi-Newton Method

Let us fix $G_0 \succ 0$, starting point $x_0 \in Q$, and accuracy $\delta \in (0,1)$. Define $\psi_0(x) = \frac{1}{2}\|x - x_0\|^2_{G_0}$. For $k \geq 0$, consider the process:

$$x_k = \arg\min_{x \in Q} \psi_k(x),$$
$$\psi_{k+1}(x) = \psi_k(x) + a_k \left[\hat{f}(x_k) + \langle \hat{f}'(x_k), x - x_k \rangle + \tfrac{1}{2}\langle f'(x_k), x - x_k \rangle^2 \right],$$

where

$$a_k = \frac{\delta}{1-\delta} \cdot \frac{1}{(\|f'(x_k)\|^*_{G_k})^2}, \quad G_k = \psi''_k(x), \quad k \geq 0,$$

and $\|h\|_G = \langle Gh, h \rangle^{1/2}$, $\|g\|^*_G = \langle g, G^{-1} g \rangle^{1/2}$.

Denote $A_k = \sum_{i=0}^{k-1} a_i$. Clearly, $\boxed{\psi_k(x) \leq A_k \hat{f}(x) + \psi_0(x), \; x \in Q.}$

We can use the technique of estimate sequences!

Main Results

1. For any $k \geq 0$, $$\boxed{\psi_k^* \stackrel{\text{def}}{=} \min_{x \in Q} \psi_k(x) \;\geq\; (1-\delta) \sum_{i=0}^{k-1} a_i \hat{f}(x_i).}$$

2. Since $\psi_k(x)$ are quadratic, their Hessians $G_k \succ 0$ are updated as $$G_{k+1} = G_k + a_k \cdot f'(x_k) f'(x_k)^T = G_k + \frac{\delta}{1-\delta} \cdot \frac{f'(x_k) f'(x_k)^T}{(\|f'(x_k)\|_{G_k}^*)^2}, \quad k \geq 0.$$

Important: $\det G_{k+1} = \frac{1}{1-\delta} \det G_k = \frac{1}{(1-\delta)^{k+1}} \det G_0.$

3. Rate of convergence.

Denote $\tilde{x}_k = \frac{1}{A_k} \sum_{i=0}^{k-1} a_i x_i$. Recall: $G_{k+1} = G_k + a_k \cdot f'(x_k) f'(x_k)^T$.

Theorem: Assume that for SP-function f, $\|f'(\cdot)\|_{G_0}^* \leq L$.

Then, $$(1-\delta)\hat{f}(\tilde{x}_k) \;\leq\; \hat{f}(x^*) + \frac{L^2 \|x_0 - x^*\|_{G_0}^2}{2n\left[e^{\delta(k+1)/n} - 1\right]}.$$

Mixed accuracy

Definition: point $\bar{x} \in Q$ is a solution with *mixed* (ϵ, δ)-accuracy if

$$(1-\delta)\hat{f}(\bar{x}) \quad \leq \quad \hat{f}(x^*) + \epsilon.$$

- $\epsilon > 0$ serves as an absolute accuracy.
- $\delta \in (0,1)$ represents the relative accuracy.

Complexity: $N_n(\epsilon, \delta) \stackrel{\mathrm{def}}{=} \frac{n}{\delta} \ln \left(1 + \frac{L^2 R^2}{2n \cdot \epsilon}\right)$ iterations of Q-N scheme.

Note:

- High absolute accuracy is *easy* to achieve.
- High relative accuracy is *difficult*. (No need?)
- $\#$ of iterations is proportional to $\frac{n}{\delta}$. (Compare with BSM.)
- We have a uniform bound: $N_n(\epsilon, \delta) < N_\infty(\epsilon, \delta) \stackrel{\mathrm{def}}{=} \frac{L^2 R^2}{2\epsilon\delta}$.

Conclusion

- Depending on the *model* of our problem, the relative accuracy can be addressed in different ways.
- This is a flexible notion, which allows finer complexity analysis.
- Corresponding methods have a small number tractable parameters.
- This is a new research direction with interesting perspectives.

References

1. Yu. Nesterov. Rounding of convex sets and efficient gradient methods for linear programming problems. *OMS*, **23**(1), 109-135 (2008).
2. Yu. Nesterov. Unconstrained Convex Minimization in Relative Scale. *MOR*, **34**(1), 180-193 (2009)
3. Yu. Nesterov. Barrier subgradient method. *Mathematical Programming*, **127**(1): 31-56 (2011).
4. Yu. Nesterov. Efficiency of quasi-Newton methods on strictly positive functions. CORE Discussion Paper (2012).