Structural Optimization for Large-Scale Problems

Lecture 6: Optimization in Relative Scale

Yurii Nesterov

Minicourse: November 15, 16, 22, 23, 2024 (SDS, Shenzhen)

Outline

§1 回顾先前的结论

Complexity of Convex Optimization

Problem: $\min_x \{f(x) : \; x \in Q \subseteq \mathbb{R}^n\} \stackrel{\mathrm{def}}{=} f^*$,

with *convex* f and Q. Assume $\exists\, x^* : \; f(x^*) = f^*$.

Solution: $\bar{x} \in Q : \quad f(\bar{x}) - f^* \leq \epsilon$.

Model of the problem:

1. Black box. $\hat{x} \in Q \;\Rightarrow\; \boxed{\text{Oracle}} \;\Rightarrow\; f(\hat{x}),\, f'(\hat{x})$.

Analytical complexity	
Problem Class	**Calls of Oracle**
$(a)^*$ $\quad n \to \infty, \; f$ is Lipschitz	$\approx O\left(\frac{L^2R^2}{\epsilon^2}\right)$
(b) $\quad n \to \infty, \; f'$ is Lipschitz	$\approx O\left(\sqrt{\frac{LR^2}{\epsilon}}\right)$
$(c)^*$ $\quad n << \infty, \; f$ is Lipschitz	$\approx O\left(n \ln \frac{LR}{\epsilon}\right)$

where L is the Lipschitz constant and $R = \|x_0 - x^*\|$.

Interior-Point Methods

- $f(x) = \langle c, x \rangle$.
- Set Q is described by a computable self-concordant barrier $F(\cdot)$ with parameter ν.

Complexity: $O(\sqrt{\nu} \ln \frac{1}{\epsilon})$ $(*)$ iterations of a Newton-type method.

Note:

- For any Q there exists $F(\cdot)$ with $\nu = O(n)$.
- From the view point of Black-Box Theory, (*) is impossible.
- In order to form $F(\cdot)$, we need to look *inside* Q.

Can we do better?

Do we always need polynomial-time methods?

1. Polynomial-time methods have complexity

$$O\left(p(n)\ln\frac{1}{\epsilon}\right) \qquad \left(\text{instead of } O\left(p\left(\frac{1}{\epsilon}\right)\right)\right)$$

where $p(\cdot)$ is a polynomial.

2. Dependence $\ln\frac{1}{\epsilon}$ is very weak. Hence, *any* accuracy is achievable.

3. The *higher* is performance of a method, the *smaller* is its field of applications.

4. Accepting the solutions with *reasonable* accuracy, we significantly increase the class of *solvable* problems.

5. In many situations $n = p_1\left(\frac{1}{\xi}\right)$, where ξ is the *accuracy of the model*.
Then, we should choose

$$\epsilon = \varphi(\xi) \quad \Leftrightarrow \quad \xi = \varphi^{-1}(\epsilon)$$

and the notion of polynomial-time complexity looses any sense.

Smoothing technique

Main idea: Use the huge difference in complexity of smooth and non-smooth optimization,

$$O\left(\sqrt{\frac{LR^2}{\epsilon}}\right) \quad \Leftrightarrow \quad O\left(\frac{L^2R^2}{\epsilon^2}\right).$$

Primal problem: Find $f^* = \min_x \{f(x) : \; x \in Q_1\}$,

where $Q_1 \subset E_1$ is convex closed and bounded.

Model of objective function:

$$f(x) = \max_u \{\langle Ax + b, u \rangle_2 : \; u \in Q_2\},$$

where $Q_2 \subset E_2$ is a closed convex bounded set.

Adjoint problem: $\max_u \{\phi(u) : \; u \in Q_2\}$,

$$\phi(u) = \min_x \{\langle Ax + b, u \rangle_2 : \; x \in Q_1\}.$$

(Adjoint problem is not uniquely defined.)

Smooth approximations

Prox-function: $d_2(\cdot)$ is continuous and *strongly convex* on Q_2:

$$d_2(v) \geq d_2(u) + \langle \nabla d_2(u), v - u \rangle_2 + \tfrac{1}{2}\sigma_2 \|v - u\|_2^2.$$

Assume: $d_2(u_0) = 0$ and $d_2(u) \geq 0 \; \forall u \in Q_2$.

Fix $\mu > 0$, the *smoothness* parameter, and define

$$f_\mu(x) = \max_u \{ \langle Ax + b, u \rangle_2 - \mu d_2(u) : \; u \in Q_2 \}.$$

Denote by $u_\mu(x)$ the solution of this problem.

Theorem: $f_\mu(x)$ is convex and differentiable for $x \in E_1$. For its gradient $\nabla f_\mu(x) = A^* u_\mu(x)$ we have $L_\mu = \frac{1}{\mu \sigma_2} \|A\|_{1,2}^2$, where

$$\|A\|_{1,2} = \max_{x,u} \{ \langle Ax, u \rangle_2 : \; \|x\|_1 = 1, \; \|u\|_2 = 1 \}.$$

Note: 1. for any $\mu \geq 0$ and $x \in E_1$ we have

$$f_0(x) \geq f_\mu(x) \geq f_0(x) - \mu D_2,$$

where $D_2 = \max_u \{ d_2(u) : \; u \in Q_2 \}$.

2. All the norms are very important.

Smoothing strategy

Smoothed problem: $f_\mu(x) \quad \to \quad \min: \; x \in Q_1.$

Lipschitz constant: $L_\mu = \frac{1}{\mu\sigma_2}\|A\|_{1,2}^2.$

Denote $D_1 = \max\limits_x \{d_1(x): \; x \in Q_1\}.$

Theorem: Let us choose $N \geq 1$. Define

$$\mu = \mu(N) = \frac{2\|A\|_{1,2}}{N+1} \cdot \sqrt{\frac{D_1}{\sigma_1\sigma_2 D_2}}.$$

After N iterations of FGM set $\hat{x} = y_N \in Q_1$ and

$$\hat{u} = \sum_{i=0}^{N} \frac{2(i+1)}{(N+1)(N+2)} \, u_\mu(x_i) \in Q_2.$$

Then $\quad 0 \leq f(\hat{x}) - \phi(\hat{u}) \leq \frac{4\|A\|_{1,2}}{N+1} \cdot \sqrt{\frac{D_1 D_2}{\sigma_1\sigma_2}}.$

Corollary. In order to get ϵ-solution we choose

$$\mu = \frac{\epsilon}{2D_2}, \quad L = \frac{D_2}{2\sigma_2} \cdot \frac{\|A\|_{1,2}^2}{\epsilon}, \quad N \geq 4\|A\|_{1,2}\sqrt{\frac{D_1 D_2}{\sigma_1\sigma_2}} \cdot \frac{1}{\epsilon}.$$

Main question

What can we do if D_1 or D_2 are very big?

Example: $$f(x) = \sum_{j=1}^{m} |\langle a_j, x\rangle + b_j| \quad \rightarrow \quad \min_{x \in \mathbb{R}^n}.$$

Suggestion:

If $f^* > 0$, then we can try to find an approximate solution with <u>*relative*</u> *accuracy $\delta > 0$*:

$$f(\bar{x}) \le (1+\delta) f^*.$$

However:

- We need a new model for our problem.
- This model must ensure $f^* > 0$.

Conic unconstrained minimization problem

Problem: Find $f^* = \min_x \{ f(x) : \; x \in \mathcal{L} \}$,

- $\mathcal{L} = \{ x \in \mathbb{R}^n : \; Cx = b \}$, $C \in \mathbb{R}^{p \times n}$ (full rank), and $b \neq 0$.
- f is a convex homogeneous of degree one function.

Main assumptions: $\mathrm{dom}\, f \equiv \mathbb{R}^n, \quad 0 \in \mathrm{int}\, \partial f(0)$.

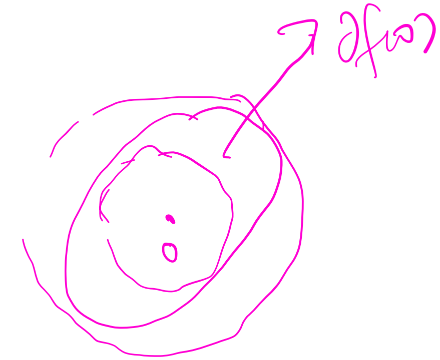
(Hence $f^* > 0$.)

Remark. Any unconstrained minimization problem $\min_{y \in \mathbb{R}^{n-1}} \phi(y)$ can be written in a *homogenized* form:

$$x = (y, \tau) \in \mathbb{R}^{n-1} \times R^1_+, \quad f(x) = \tau \phi(y/\tau), \quad Cx \equiv \tau, \quad b = 1.$$

However, we cannot guarantee $0 \in \mathrm{int}\, \partial f(0)$. □

Asphericity



Let us fix $\| \cdot \|$. Define $\gamma_1 \geq \gamma_0 > 0$ as follows:

$$B_{\|\cdot\|^*}(\gamma_0) \subseteq \partial f(0) \subseteq B_{\|\cdot\|^*}(\gamma_1).$$

Denote $\alpha = \frac{\gamma_0}{\gamma_1} < 1.$

NB: This parameter determines the complexity of finding approximate solutions to our problem with certain relative accuracy.

Ellipsoidal norms:

1. In view of John theorem, we can always ensure $\alpha \geq \frac{1}{n}$.

2. If $\partial f(0)$ is symmetric, then $\alpha \geq \frac{1}{\sqrt{n}}$.

3. Let we know a self-concordant barrier $\psi(v)$ for the convex set $\partial f(0)$ and $\psi'(0) = 0$. Then we can use

$$\|v\|^* = \langle v, \psi''(0)v \rangle^{1/2}, \quad \|x\| = \langle [\psi''(0)]^{-1}x, x \rangle^{1/2}.$$

Hence, $\gamma_0 = 1$, $\gamma_1 = \nu + 2\sqrt{\nu}$, where ν is the parameter of $\psi(\cdot)$.

Polyhedral $\partial f(0)$

Lemma. Let $f(x) = \max\limits_{1 \le i \le m} \langle a_i, x \rangle$, matrix $A = (a_1, \dots, a_m)$ has full row rank, and $\sum\limits_{i=1}^m a_i = 0$. Then the norm $\|x\| = \left[\sum\limits_{i=1}^m \langle a_i, x \rangle^2 \right]^{1/2}$ is well defined, and we can choose $\gamma_1 = 1$, $\gamma_0 = \frac{1}{\sqrt{m(m-1)}}$.

Proof. Since $G = \sum\limits_{i=1}^m a_i a_i^T \succ 0$, then $\|v\|^* = \langle v, G^{-1} v \rangle^{1/2}$ and

$$(\|a_i\|^*)^2 = \langle a_i, G^{-1} a_i \rangle = \max_{x \in \mathbb{R}^n} \{ 2\langle a_i, x \rangle - \langle Gx, x \rangle \}$$

$$= \max_{x \in \mathbb{R}^n} \left\{ 2\langle a_i, x \rangle - \sum_{k=1}^m \langle a_k, x \rangle^2 \right\} \le \max_{x \in \mathbb{R}^n} \{ 2\langle a_i, x \rangle - \langle a_i, x \rangle^2 \} = 1.$$

Since $\partial f(0) = \mathrm{Conv}\,\{a_i,\ i = 1, \dots, m\}$, we can take $\gamma_1 = 1$.

On the other hand, for any $x \in \mathbb{R}^n$ we have $\sum_{i=1}^{m} \langle a_i, x \rangle = 0$. Therefore

$$\langle Gx, x \rangle = \sum_{i=1}^{m} \langle a_i, x \rangle^2$$

$$\leq \max_{s \in \mathbb{R}^m} \left\{ \sum_{i=1}^{m} (s^{(i)})^2 : \ \sum_{i=1}^{m} s^{(i)} = 0, \ s^{(i)} \leq f(x), \ i = 1, \ldots, m \right\}.$$

The extremum in the above maximization problem is attained, for example, at

$$\hat{s} = f(x) \cdot (e - me_1).$$

Hence, $\langle Gx, x \rangle \leq m(m-1) f^2(x)$. That is $f(x) \geq \frac{\|x\|}{\sqrt{m(m-1}}$, and we can take $\gamma_0 = \frac{1}{\sqrt{m(m-1)}}$. $\square$

Projection of the origin

Denote $\|x_0\| = \min_x \{\|x\| : \; Cx = b\} \stackrel{\mathrm{def}}{=} \rho$.

Theorem. 1. $\gamma_0 \cdot \|x\| \le f(x) \le \gamma_1 \cdot \|x\|$, $x \in \mathbb{R}^n$.

Hence, $f(\cdot)$ is Lipschitz continuous with constant γ_1.

2. $\alpha f(x_0) \le \gamma_0 \cdot \|x_0\| \le f^* \le f(x_0) \le \gamma_1 \cdot \|x_0\|$.

3. For any x^*, we have $\|x_0 - x^*\| \le \frac{2}{\gamma_0} f^* \le \frac{2}{\gamma_0} f(x_0)$.

If $\|\cdot\|$ is Euclidean, then $\|x_0 - x^*\| \le \frac{1}{\gamma_0} f^* \le \frac{1}{\gamma_0} f(x_0)$.

Proof. For any $x \in \mathbb{R}^n$ we have

$$
\begin{aligned}
f(x) &= \max_v \{\langle v, x\rangle : \; v \in \partial f(0)\} \\
&\ge \max_u \{\langle v, x\rangle : \; v \in B_{\|\cdot\|^*}(\gamma_0)\} = \gamma_0 \cdot \|x\|,
\end{aligned}
$$

$$f(x) = \max_v \{\langle v, x\rangle : \ v \in \partial f(0)\}$$
$$\leq \max_u \{\langle v, x\rangle : \ v \in B_{\|\cdot\|^*}(\gamma_1)\} = \gamma_1 \cdot \|x\|.$$

Therefore for any x and $h \in \mathbb{R}^n$ we have

$$f(x+h) \leq f(x) + f(h) \leq f(x) + \gamma_1 \cdot \|h\|.$$

Moreover,

$$f^* = \min_x \{f(x) : \ Cx = b\} \geq \min_x \{\gamma_0 \|x\| : \ Cx = b\} = \gamma_0 \cdot \rho.$$

Hence, $f^* \geq \gamma_0 \cdot \|x_0\| \geq \alpha f(x_0), \ \ f^* \leq f(x_0) \leq \gamma_1 \cdot \|x_0\|$.

3. Note that $\|x_0 - x^*\| \leq \|x_0\| + \|x^*\| \leq \frac{2}{\gamma_0} \cdot f^*$.

If the norm is Euclidean, then

$$\|x_0 - x^*\|^2 = \|x^*\|^2 - \|x_0\|^2 < \|x^*\|^2.$$

$\square$

Subgradient approximation scheme $G_N(R)$

for $k := 0$ **to** N **do** Compute $f(x_k)$ and $g(x_k)$. Define

$$x_{k+1} := \pi_{\mathcal{L}} \left(x_k - \frac{R}{\sqrt{N+1}} \cdot \frac{g(x_k)}{\|g(x_k)\|^*} \right).$$

$R \sim \|x^0 - x^*\|$

Output: $G_N(R) = \arg\min\{ f(x) : \ x = x_0, \dots, x_N \}$.

Rate of convergence: $\quad f(G_N(R)) - f^* \le \frac{\gamma_1}{\sqrt{N+1}} \cdot \frac{\|x_0 - x^*\|^2 + R^2}{2R}$. $\le \delta f^*$

We need to choose R properly! What about $\hat{\rho} \stackrel{\text{def}}{=} \frac{1}{\gamma_0} f(x_0)$?

Theorem. For $\delta \in (0,1)$, let us choose $N = \left\lfloor \frac{1}{\alpha^4 \delta^2} \right\rfloor$.
Then $f(G_N(\hat{\rho})) \le (1+\delta) \cdot f^*$.

$\frac{\|x_0 - x^*\|^2}{R} + R$

Proof. $\quad f(G_N(\hat{\rho})) - f^* \le \alpha^2 \delta \gamma_1 \cdot \frac{\|x_0 - x^*\|^2 + \hat{\rho}^2}{2\hat{\rho}}$

$$\le \alpha^2 \delta \gamma_1 \hat{\rho} = \alpha \delta f(x_0) \le \delta \cdot f^*. \square$$

NB: Bad dependence in α.

$\frac{\gamma_1}{\gamma_0} = \alpha$

Accelerated subgradient method

Denote $\hat{N} = \left\lfloor \frac{e}{\alpha^2} \cdot \left(1 + \frac{1}{\delta}\right)^2 \right\rfloor$. Consider the process:

Set $\hat{x}_0 = x_0$, and for $t \geq 1$ iterate

$$\hat{x}_t := G_{\hat{N}}\left(\frac{1}{\gamma_0} f(\hat{x}_{t-1})\right); \text{ **if** } f(\hat{x}_t) \geq \frac{1}{\sqrt{e}} f(\hat{x}_{t-1}) \text{ **then** } \{T := t, \text{ **Stop.**}\}$$

Theorem. $T \leq 1 + 2\ln\frac{1}{\alpha}$ and $f(\hat{x}_T) \leq (1+\delta) f^*$.

The total number of gradient steps does not exceed

$$\frac{e}{\alpha^2} \cdot \left(1 + \frac{1}{\delta}\right)^2 \cdot \left(1 + 2\ln\frac{1}{\alpha}\right).$$

Proof. At the beginning of stage t, $\left(\frac{1}{\sqrt{e}}\right)^{t-1} f(x_0) \geq f(\hat{x}_{t-1})$.

Thus, $\left(\frac{1}{\sqrt{e}}\right)^{T-1} f(x_0) \geq f(\hat{x}_{T-1}) \geq f^* \geq \alpha f(x_0)$.

Since $\|x_0 - x^*\| \leq \frac{1}{\gamma_0} f^* \leq \frac{1}{\gamma_0} f(\hat{x}_{T-1})$, we get

$$f(\hat{x}_T) - f^* \leq \frac{\gamma_1}{\sqrt{\hat{N}+1}} \cdot \frac{1}{\gamma_0} \cdot f(\hat{x}_{T-1}) \leq \frac{\sqrt{e}}{\alpha\sqrt{\hat{N}+1}} \cdot f(\hat{x}_T) \leq \frac{\delta}{1+\delta} \cdot f(\hat{x}_T).$$

$\square$

Smoothing for relative scale

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

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

$$F(y) = \max_{s \in Q_2} \langle s, y \rangle, \quad 0 \in \operatorname{int} Q_2 \subset \mathbb{R}^m.$$

Thus, $f^* > 0$.

Let $\| \cdot \|_2$ be a Euclidean norm in $\mathbb{R}^m$. Define

$$B(r) = \{y : \; \|y\|_2 \leq r\},$$

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

Then for the norm $\|x\|_1 = \|A^T x\|_2$ we have

$$\gamma_0 \|x\|_1 \leq f(x) \leq \gamma_1 \|x\|_1.$$

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

$$\|x_0 - x^*\|_1 \leq \frac{1}{\gamma_0} f^* \leq \frac{1}{\gamma_0} f(x).$$

Denote $Q_1(R) = \{x \in \mathcal{L} : \ \|x\|_1 \leq R\}$ and

$$f_\mu(x) = \max_s \{\langle A^T x, s\rangle - \tfrac{1}{2}\mu\|s\|_2^2 : \ s \in Q_2\}.$$

Let $x_N(R)$ be an output of the method FGM after N steps as applied to function f_μ with

$$\mu = \tfrac{2R}{\gamma_1 \cdot (N+1)}, \quad Q_1 = Q_1(R).$$

Denote $\alpha = \frac{\gamma_0}{\gamma_1} \leq 1$, and $\tilde{N} = \left\lfloor 2\frac{e}{\alpha} \cdot \left(1 + \frac{1}{\delta}\right)\right\rfloor$.

Consider the following 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** $T := t$, **Stop.**

Theorem. $T \leq 1 + \ln\frac{1}{\alpha}$ and $f(\hat{x}_T) \leq (1+\delta) f^*$.

The total number of gradient steps $\leq 2\frac{e}{\alpha} \cdot \left(1 + \frac{1}{\delta}\right) \cdot \left(1 + \ln\frac{1}{\alpha}\right)$.

Example

$f(x) = \max_{1 \le j \le m} |\langle a_j, x \rangle|$, $m > n$.

Define $F(s) = \max_{1 \le j \le m} |s^{(j)}|$, $\|s\|_2^2 = \sum_{j=1}^{m} s_j^2$,

$$\gamma_0 = \frac{1}{\sqrt{m}}, \quad \gamma_1 = 1, \quad \alpha = 1/\sqrt{m}.$$

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

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

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*?*

Remarks

Main inequality

$$\gamma_0 \|x\|_1 \leq f(x) \leq \gamma_1 \|x\|_1, \quad x \in \mathbb{R}^n,$$

is used for

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

$$\gamma_0 \|x_0 - x^*\|_1 \leq f^* \leq f(x), \quad x \in \mathcal{L}.$$

John Theorem: For any bounded convex *symmetric* set $Q \subset \mathbb{R}^n$ 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 $\alpha \approx 1/\sqrt{n}$.

In which cases such a norm is computable?

Finding the rounding ellipsoid

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

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

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 tolerance $\gamma > 1$. Consider the process

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**

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

Theorem. The 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.

Idea of the proof

1. Let $\xi \in \Delta_m$. Define $G(\xi) = \sum\limits_{j=1}^{m} \xi^{(i)} a_i a_i^T$. Then

$$\langle G(\xi)x, x\rangle^{1/2} \leq f(x), \; x \in \mathbb{R}^n.$$

This means that $B_{\|\cdot\|^*_{G(\xi)}}(1) \subseteq \partial f(0)$.

2. Consider the function $\psi(\alpha) = \ln\det((1-\alpha)G + \alpha a a^T)$.

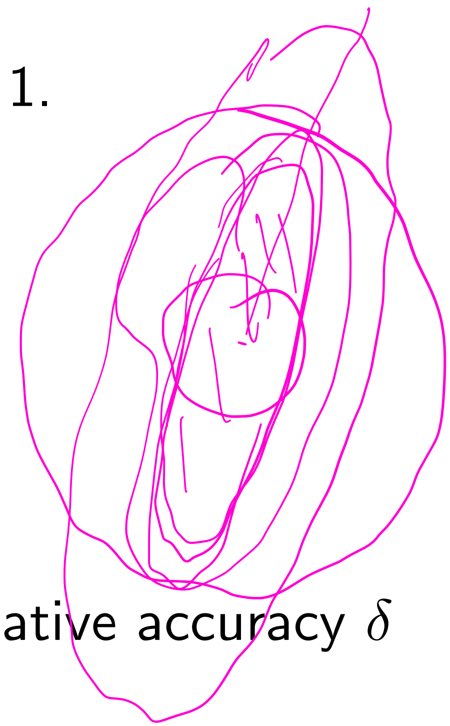
Its derivative is $\psi'(\alpha) = \langle [(1-\alpha)G + \alpha a a^T]^{-1}, a a^T - G\rangle$.

Thus, $\psi'(0) = (\|a\|^*_G)^2 - n$.

3. Denote $\sigma \stackrel{\mathrm{def}}{=} \frac{1}{n}(\|a\|^*_G)^2 - 1 > 0$. Then

$$\max_{\alpha \in [0,1]} \psi(\alpha) - \psi(0) \geq \frac{\sigma^2}{2(1+\sigma)^2}.$$

Application example



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

Phase 1: find a rounding norm $\| \cdot \|^*$ for the set

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

with tolerance parameter $\gamma > 1$.

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

Phase 2: using this norm, solve the 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, $O\left(\frac{mn^{1.5}}{\delta} \ln n \sqrt{\ln m} \right)$ a.o.

Competitors: Ellipsoid method: $\left(n^2 \ln \frac{1}{\delta} \right) \times mn$.

Interior point: $\left(\sqrt{m} \ln \frac{m}{\delta} \right) \times mn^2$.

Conclusion

1. We discussed a new direction in *Structural Optimization*, Optimization with *relative* accuracy.

It is very much "*problem oriented*".

2. In many situations, the complexity of our algorithms is proportional to the *square root* of the number of iterations of the Black Box Schemes.

3. Our bounds *do not* depend on the data.

4. Very often we do not need very high *relative* accuracy.

5. Low complexity of each iteration.

6. Low memory requirements. (Sometimes very low.)