Lecture 2: Looking into the Black Box. Structural Optimization.

Yuriii Nesterov, CORE/INMA (UCL)

March 2, 2012
Outline

1. Nonsmooth Optimization
2. Smoothing technique
3. Application examples
4. Interior-point methods: standard problem
5. Newton method
6. Self-concordant functions and barriers
7. Application examples
8. Further directions
Nonsmooth Unconstrained Optimization

**Problem:** \( \min \{ f(x) : x \in R^n \} \Rightarrow x^*, f^* = f(x^*) \), where \( f(x) \) is a nonsmooth convex function.

**Subgradients:** \( g \in \partial f(x) \iff f(y) \geq f(x) + \langle g, y - x \rangle \ \forall y \in R^n \).

**Main difficulties:**
- \( g \in \partial f(x) \) is *not* a descent direction at \( x \).
- \( g \in \partial f(x^*) \) does not imply \( g = 0 \).

**Example**
\[
f(x) = \max_{1 \leq j \leq m} \{ \langle a_j, x \rangle + b_j \},
\]
\[
\partial f(x) = \text{Conv} \ \{ a_j : \langle a_j, x \rangle + b_j = f(x) \}.
\]
Advantages

- Very simple iteration scheme.
- Low memory requirements.
- Optimal rate of convergence (uniformly in the dimension).
- Interpretation of the process.

Objections:

- Low rate of convergence. (Confirmed by theory!)
- No acceleration.
- High sensitivity to the step-size strategy.
Lower complexity bounds

Nemirovsky, Yudin 1976

If $f(x)$ is given by a local black-box, it is impossible to converge faster than $O \left( \frac{1}{\sqrt{k}} \right)$ uniformly in $n$. ($k$ is the # of calls of oracle.)

NB: Convergence is very slow.

Question: We want to find an $\epsilon$-solution of the problem

$$\max_{1 \leq j \leq m} \{ \langle a_j, x \rangle + b_j \} \rightarrow \min_x : x \in \mathbb{R}^n,$$

by a gradient scheme ($n$ and $m$ are big).

What is the worst-case complexity bound?

“Right answer” (Complexity Theory): $O \left( \frac{1}{\epsilon^2} \right)$ calls of oracle.

Our target: A gradient scheme with $O \left( \frac{1}{\epsilon} \right)$ complexity bound.

Reason of speed up: our problem is not in a black box.
**Complexity of Smooth Minimization**

**Problem:** $f(x) \rightarrow \min \; x \in \mathbb{R}^n$, where $f$ is a convex function and $\|\nabla f(x) - \nabla f(y)\|_* \leq L(f)\|x - y\|$ for all $x, y \in \mathbb{R}^n$.

(For measuring gradients we use dual norms: $\|s\|_* = \max_{\|x\| = 1} \langle s, x \rangle$.)

**Rate of convergence:** Optimal method gives $O\left(\frac{L(f)}{k^2}\right)$.

**Complexity:** $O\left(\sqrt{\frac{L(f)}{\epsilon}}\right)$. The difference with $O\left(\frac{1}{\epsilon^2}\right)$ is very big.
Smoothing the convex function

For function $f$ define its Fenchel conjugate:

$$f^*(s) = \max_{x \in \mathbb{R}^n} [\langle s, x \rangle - f(x)].$$

It is a closed convex function with $\text{dom } f^* = \text{Conv}\{f'(x) : x \in \mathbb{R}^n\}$.

Moreover, under very mild conditions $(f^*(s))^* \equiv f(x)$.

Define $f_\mu(x) = \max_{s \in \text{dom } f^*} [\langle s, x \rangle - f^*(s) - \frac{\mu}{2} \|s\|_2^2]$, where $\| \cdot \|_*$ is a Euclidean norm.

Note: $f'_\mu(x) = s_\mu(x)$, and $x = f'_*(s_\mu(x)) + \mu s_\mu(x)$. Therefore,

$$\|x^1 - x^2\|^2 = \|f'_*(s^1) - f'_*(s^2)\|^2 + 2\mu \langle f'_*(s^1) - f'_*(s^2), s^1 - s^2 \rangle$$

$$+ \mu^2 \|s^1 - s^2\|^2 \geq \mu^2 \|s^1 - s^2\|^2.$$  

Thus, $f_\mu \in C^{1,1}_{1/\mu}$ and $f(x) \geq f_\mu(x) \geq f(x) - \mu D^2$,

where $D = \text{Diam}(\text{dom } f^*)$. 

Yu. Nesterov () Structural Optimization March 2, 2012 7 / 30
Main questions

1. Given by a non-smooth convex $f(x)$, can we form its computable smooth $\epsilon$-approximation $f_\epsilon(x)$ with

$$L(f_\epsilon) = O \left( \frac{1}{\epsilon} \right)?$$

If yes, we need only $O \left( \sqrt{\frac{L(f_\epsilon)}{\epsilon}} \right) = O \left( \frac{1}{\epsilon} \right)$ iterations.

2. Can we do this in a systematic way?

**Conclusion:** We need a convenient *model* of our problem.
Adjoint problem

**Primal problem:** Find \( f^* = \min_x \{ f(x) : x \in Q_1 \} \), where \( Q_1 \subset E_1 \) is convex closed and bounded.

**Objective:** \( f(x) = \hat{f}(x) + \max_u \left\{ \langle Ax, u \rangle_2 - \hat{\phi}(u) : u \in Q_2 \right\} \), where

- \( \hat{f}(x) \) is differentiable and convex on \( Q_1 \).
- \( Q_2 \subset E_2 \) is a closed convex and bounded.
- \( \hat{\phi}(u) \) is continuous convex function on \( Q_2 \).
- linear operator \( A : E_1 \to E_2^* \).

**Adjoint problem:** \( \max_u \left\{ \phi(u) : u \in Q_2 \right\} \), where

\[
\phi(u) = -\hat{\phi}(u) + \min_x \left\{ \langle Ax, u \rangle_2 + \hat{f}(x) : x \in Q_1 \right\}.
\]

**NB:** Adjoint problem is not unique!
Example

Consider $f(x) = \max_{1 \leq j \leq m} |\langle a_j, x \rangle_1 - b_j|$.

1. $Q_2 = E_1^*$, $A = I$, $\hat{\phi}(u) \equiv f_*(u) = \max_x \{\langle u, x \rangle_1 - f(x) : x \in E_1\}$

   \[
   = \min_{s \in \mathbb{R}^m} \left\{ \sum_{j=1}^m s_j b_j : u = \sum_{j=1}^m s_j a_j, \sum_{j=1}^m |s_j| \leq 1 \right\}.
   \]

2. $E_2 = \mathbb{R}^m$, $\hat{\phi}(u) = \langle b, u \rangle_2$, $f(x) = \max_{1 \leq j \leq m} |\langle a_j, x \rangle_1 - b_j|$

   \[
   = \max_{u \in \mathbb{R}^m} \left\{ \sum_{j=1}^m u_j [\langle a_j, x \rangle_1 - b_j] : \sum_{j=1}^m |u_j| \leq 1 \right\}.
   \]

3. $E_2 = \mathbb{R}^{2m}$, $\hat{\phi}(u)$ is a linear, $Q_2$ is a simplex:

   \[
   f(x) = \max_{u \in \mathbb{R}^{2m}} \left\{ \sum_{j=1}^m (u^1_j - u^2_j) [\langle a_j, x \rangle_1 - b_j] : \sum_{j=1}^m (u^1_j + u^2_j) = 1, u \geq 0 \right\}.
   \]

**NB:** Increase in dim $E_2$ decreases the complexity of representation.
Smooth approximations

**Prox-function:** \( d_2(u) \) is continuous and strongly convex on \( Q_2 \):
\[
d_2(v) \geq d_2(u) + \langle \nabla d_2(u), v - u \rangle_2 + \frac{1}{2} \sigma_2 \| v - u \|_2^2.
\]
Assume: \( d_2(u_0) = 0 \) and \( d_2(u) \geq 0 \ \forall u \in Q_2 \).

Fix \( \mu > 0 \), the smoothing parameter, and define
\[
f_\mu(x) = \max_u \{ \langle Ax, u \rangle_2 - \hat{\phi}(u) - \mu d_2(u) : u \in Q_2 \}.
\]
Denote by \( u(x) \) the solution of this problem.

**Theorem:** \( f_\mu(x) \) is convex and differentiable for \( x \in E_1 \). Its gradient
\( \nabla f_\mu(x) = A^* u(x) \) is Lipschitz continuous with
\[
L(f_\mu) = \frac{1}{\mu \sigma_2} \| A \|_{1,2}^2,
\]
where \( \| A \|_{1,2} = \max_{x,u} \{ \langle Ax, u \rangle_2 : \| x \|_1 = 1, \| u \|_2 = 1 \} \).

**NB:** 1. For any \( x \in E_1 \) 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 norms are very important.
Optimal method

Problem: \( \min_{x} \{ f(x) : x \in Q_1 \} \) with \( f \in C^{1,1}(Q_1) \).

Prox-function: strongly convex \( d_1(x), d_1(x^0) = 0, d_1(x) \geq 0, x \in Q_1 \).

Gradient mapping: 
\[
T_L(x) = \arg \min_{y \in Q_1} \left\{ \langle \nabla f(x), y - x \rangle_1 + \frac{1}{2} L \| y - x \|_1^2 \right\}.
\]

Method. For \( k \geq 0 \) do:
1. Compute \( f(x^k), \nabla f(x^k) \).
2. Find \( y^k = T_L(f)(x^k) \).
3. Find \( z^k = \arg \min_{x \in Q_1} \left\{ \frac{L(f)}{\sigma} d_1(x) + \sum_{i=0}^{k} \frac{i+1}{2} \langle \nabla f(x^i), x \rangle_1 \right\} \).
4. Set \( x^{k+1} = \frac{2}{k+3} z^k + \frac{k+1}{k+3} y^k \).

Convergence: \( f(y^k) - f(x^*) \leq \frac{4L(f)d_1(x^*)}{\sigma_1(k+1)^2} \), where \( x^* \) is the optimal solution.
Applications

Smooth problem: \( \bar{f}_\mu(x) = \hat{f}(x) + f_\mu(x) \to \min : x \in Q_1. \)

Lipschitz constant: \( L_\mu = L(\hat{f}) + \frac{1}{\mu \sigma_2} \| A \|_{1,2}^2. \) Denote \( D_1 = \max_x \{ d_1(x) : x \in Q_1 \}. \)

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 set \( \hat{x} = y^N \in Q_1 \) and
\[
\hat{u} = \sum_{i=0}^{N} \frac{2(i+1)}{(N+1)(N+2)} u(x^i) \in Q_2.
\]
Then \( 0 \leq f(\hat{x}) - \phi(\hat{u}) \leq \frac{4 \| A \|_{1,2}}{N+1} \cdot \sqrt{\frac{D_1 D_2}{\sigma_1 \sigma_2}} + \frac{4L(\hat{f})D_1}{\sigma_1(N+1)^2}. \)

Corollary. Let \( L(\hat{f}) = 0. \) For getting an \( \epsilon \)-solution, we choose
\[
\mu = \frac{\epsilon}{2D_2}, \quad L = \frac{D_2}{2\sigma_2} \cdot \frac{\| A \|_{1,2}^2}{\epsilon}, \quad N \geq 4 \| A \|_{1,2} \sqrt{\frac{D_1 D_2}{\sigma_1 \sigma_2}} \cdot \frac{1}{\epsilon}.
\]
Example: Equilibrium in matrix games (1)

Denote $\Delta_n = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x^{(i)} = 1\}$. Consider the problem

$$\min_{x \in \Delta_n} \max_{u \in \Delta_m} \{\langle Ax, u \rangle_2 + \langle c, x \rangle_1 + \langle b, u \rangle_2\}.$$ 

**Minimization form:**

$$\min_{x \in \Delta_n} f(x), \quad f(x) = \langle c, x \rangle_1 + \max_{1 \leq j \leq m} [\langle a_j, x \rangle_1 + b_j],$$

$$\max_{u \in \Delta_m} \phi(u), \quad \phi(u) = \langle b, u \rangle_2 + \min_{1 \leq i \leq n} [\langle \hat{a}_i, u \rangle_2 + c_i],$$

where $a_j$ are the rows and $\hat{a}_i$ are the columns of $A$.

1. **Euclidean distance:** Let us take

$$\|x\|_1^2 = \sum_{i=1}^n x_i^2, \quad \|u\|_2^2 = \sum_{j=1}^m u_j^2,$$

$$d_1(x) = \frac{1}{2} \|x - \frac{1}{n} e_n\|_1^2, \quad d_2(u) = \frac{1}{2} \|u - \frac{1}{m} e_m\|_2^2.$$

Then $\|A\|_{1,2} = \lambda_\text{max}^{1/2}(A^T A)$ and $f(\hat{x}) - \phi(\hat{u}) \leq \frac{4 \lambda_\text{max}^{1/2}(A^T A)}{N+1}$. 
2. Entropy distance. Let us choose
\[
\|x\|_1 = \sum_{i=1}^{n} |x_i|, \quad d_1(x) = \ln n + \sum_{i=1}^{n} x_i \ln x_i,
\]
\[
\|u\|_2 = \sum_{j=1}^{m} |u_j|, \quad d_2(u) = \ln m + \sum_{j=1}^{m} u_j \ln u_j.
\]

**LM:** $\sigma_1 = \sigma_2 = 1$. (Hint: $\langle d_1''(x)h, h \rangle = \sum_{i=1}^{n} \frac{h_i^2}{x_i} \rightarrow \min_{x \in \Delta_n} = \|h\|_2^2$.)

Moreover, since $D_1 = \ln n$, $D_2 = \ln m$, and
\[
\|A\|_{1,2} = \max_x \{ \max_{1 \leq j \leq m} |\langle a_j, x \rangle| : \|x\|_1 = 1 \} = \max_{i,j} |A_{i,j}|,
\]
we have $f(\hat{x}) - \phi(\hat{u}) \leq \frac{4\sqrt{\ln n \ln m}}{N+1} \cdot \max_{i,j} |A_{i,j}|$.

**NB:** 1. Usually $\max_{i,j} |A_{i,j}| \ll \lambda_{\text{max}}^{1/2}(A^TA)$.

2. We have $\tilde{f}_\mu(x) = \langle c, x \rangle_1 + \mu \ln \left(\frac{1}{m} \sum_{j=1}^{m} e^{[\langle a_j, x \rangle + b_j]/\mu}\right)$. 
**Part II: Interior Point Methods**

**Black-Box Methods:** Main assumptions represent the bounds for the size of certain derivatives.

**Example**

Consider the function $f(x_1, x_2) = \begin{cases} \frac{x_2^2}{x_1}, & x_1 > 0, \\ 0, & x_1 = x_2 = 0. \end{cases}$

It is closed, convex, but discontinuous at the origin.

However, its epigraph $\{ x \in \mathbb{R}^3 : x_1 x_3 \geq x_2^2 \}$ is a simple convex set:

$x_1 = u_1 + u_3, \ x_2 = u_2, \ x_3 = u_1 - u_3 \Rightarrow u_1 \geq \sqrt{u_2^2 + u_3^2}.$

(Lorentz cone)

**Question:** Can we always replace the functional components by convex sets?
Standard formulation

**Problem:** \( f^* = \min_{x \in Q} \langle c, x \rangle \),
where \( Q \subset E \) is a closed convex set with nonempty interior.

How we can measure the quality of \( x \in Q \)?

1. The residual \( \langle c, x \rangle - f^* \) is not very informative since it does not depend on *position* of \( x \) inside \( Q \).
2. The boundary of a convex set can be very complicated.
3. It is easy to travel inside provided that we keep a sufficient distance to the boundary.

**Conclusion:** we need a barrier function \( f(x) \):

- \( \text{dom} \ f = \text{int} \ Q \),
- \( f(x) \to \infty \) as \( t \to \partial Q \).
Path-following method

Central path: for $t > 0$ define $x^*(t)$, \[ tc + f'(x^*(t)) = 0 \]
(hence $x^*(t) = \arg \min_{x} [\Psi_t(x) \overset{\text{def}}{=} t\langle c, x \rangle + f(x)]$.)

Lemma. Suppose $\langle f'(x), y - x \rangle \leq A$ for all $x, y \in \text{dom} \ Q$. Then
\[ \langle c, x^*(t) - x^* \rangle = \langle f'(x^*(t)), x^* - x^*(t) \rangle \leq \frac{1}{t}A. \]

Method: $t_k > 0$, $x^k \approx x^*(t_k) \Rightarrow t_{k+1} > t_k$, $x^{k+1} \approx x^*(t_{k+1})$.
For approximating $x^*(t^{k+1})$, we need a powerful minimization scheme.

Main candidate: Newton Method.
(very good local convergence.)
Classical results on the Newton Method

Method: \( x^{k+1} = x^k - [f''(x^k)]^{-1}f'(x^k) \).

Assume that:

- \( f''(x^*) \geq \ell \cdot I_n \)
- \( \|f''(x) - f''(y)\| \leq M\|x - y\|, \forall x, y \in \mathbb{R}^n. \)
- The starting point \( x^0 \) is close to \( x^* \): \( \|x^0 - x^*\| < \bar{r} = \frac{2\ell}{3M} \).

Then \( \|x^k - x^*\| < \bar{r} \) for all \( k \), and the Newton method converges quadratically: \( \|x^{k+1} - x^*\| \leq \frac{M\|x^k - x^*\|^2}{2(\ell - M\|x^k - x^*\|)} \).

Note:

- The description of the region of quadratic convergence is given in terms of the metric \( \langle \cdot, \cdot \rangle \).
- This region is changing when we choose another metric.
Simple observation

Let \( f(x) \) satisfy our assumptions. Consider \( \phi(y) = f(Ay) \), where \( A \) is a non-degenerate \((n \times n)\)-matrix.

**Lemma:** Let \( \{x^k\} \) be a sequence, generated by Newton Method for function \( f \).
Consider the sequence \( \{y^k\} \), generated by the Newton Method for function \( \phi \) with \( y^0 = A^{-1}x^0 \).
Then \( y^k = A^{-1}x^k \) for all \( k \geq 0 \).

**Proof:** Assume \( y_k = A^{-1}x_k \) for some \( k \geq 0 \). Then

\[
y^{k+1} = y^k - \left[ \phi''(y^k) \right]^{-1} \phi'(y^k)
= y^k - \left[ A^T f''(Ay^k)A \right]^{-1} A^T f'(Ay^k)
= A^{-1}x^k - A^{-1}[f''(x^k)]^{-1} f'(x^k) = A^{-1}x^{k+1}.
\]

\(\square\)

**Conclusion:** The method is **affine invariant**. Its region of quadratic convergence **does not depend on the metric**!
What was wrong?

**Old assumption:** $\| f''(x) - f''(y) \| \leq M \| x - y \|.$

Let $f \in C^3(R^n)$. Denote $f'''(x)[u] = \lim_{\alpha \to 0} \frac{1}{\alpha} [f''(x + \alpha u) - f''(x)]$. This is a matrix!

Then the old assumption is equivalent to: $\| f'''(x)[u] \| \leq M \| u \|.$

Hence, at any point $x \in R^n$ we have

$$ (*) : \quad |\langle f'''(x)[u]v, \nu \rangle| \leq M \| u \| \cdot \| \nu \|^2 \text{ for all } u, \nu \in R^n. $$

**Note:**

- The LHS of $(*)$ is an affine invariant directional derivative.
- The norm $\| \cdot \|$ has nothing common with our particular $f$.
- However, there exists a local norm, which is closely related to $f$. This is $\| u \|_f''(x) = \langle f''(x) u, u \rangle^{1/2}$.
- Let us make a similar assumption in terms of $\| \cdot \|_{f''(x)}$. 
Definition of Self-Concordant Function

Let $f(x) \in C^3(\text{dom } f)$ be a closed and convex, with open domain. Let us fix a point $x \in \text{dom } f$ and a direction $u \in \mathbb{R}^n$.

Consider the function $\phi(x; t) = f(x + tu)$. Denote

$$Df(x)[u] = \phi_t'(x; 0) = \langle f'(x), u \rangle,$$

$$D^2 f(x)[u, u] = \phi_{tt}(x; 0) = \langle f''(x)u, u \rangle = \|u\|_{f''(x)}^2,$$

$$D^3 f(x)[u, u, u] = \phi_{ttt}(x; 0) = \langle f''''[u]u, u \rangle.$$

**Def.** We call function $f$ self-concordant if the inequality

$$|D^3 f(x)[u, u, u]| \leq 2 \|u\|_{f''(x)}^3$$

holds for any $x \in \text{dom } f$, $u \in \mathbb{R}^n$.

**Note:**

- We cannot expect that these functions are very common.
- We hope that they are good for the Newton Method.
Examples

1. Linear function is s.c. since $f''(x) \equiv 0$, $f'''(x) \equiv 0$.

2. Convex quadratic function is s.c. ($f'''(x) \equiv 0$).

3. Logarithmic barrier for a ray $\{x > 0\}$:
   
   $$f(x) = -\ln x, \quad f'(x) = -\frac{1}{x}, \quad f''(x) = \frac{1}{x^2}, \quad f'''(x) = -\frac{2}{x^3}. \quad \text{(1)}$$

4. Logarithmic barrier for a quadratic region. Consider a concave function $\phi(x) = \alpha + \langle a, x \rangle - \frac{1}{2} \langle Ax, x \rangle$. Define $f(x) = -\ln \phi(x)$.

   $$Df(x)[u] = -\frac{1}{\phi(x)} [\langle a, u \rangle - \langle Ax, u \rangle] \overset{\text{def}}{=} \omega_1, \quad \text{(2)}$$

   $$D^2 f(x)[u]^2 = \frac{1}{\phi^2(x)} [\langle a, u \rangle - \langle Ax, u \rangle]^2 + \frac{1}{\phi(x)} \langle Au, u \rangle, \quad \text{(3)}$$

   $$D^3 f(x)[u]^3 = -\frac{2}{\phi^3(x)} [\langle a, u \rangle - \langle Ax, u \rangle]^3 - \frac{3 \langle Au, u \rangle}{\phi^2(x)} [\langle a, u \rangle - \langle Ax, u \rangle]. \quad \text{(4)}$$

   $$D_2 = \omega_1^2 + \omega_2, \quad D_3 = 2\omega_1^3 - 3\omega_1\omega_2. \quad \text{Hence, } |D_3| \leq 2 |D_2|^{3/2}. \quad \text{(5)}$$
Simple properties

1. If \( f_1, f_2 \) are s.c.f., then \( f_1 + f_2 \) is s.c. function.

2. If \( f(y) \) is s.c.f., then \( \phi(x) = f(Ax + b) \) is also a s.c. function.

Proof: Denote \( y = y(x) = Ax + b, \ v = Au. \) Then

\[
D\phi(x)[u] = \langle f'(y(x)), Au \rangle = \langle f'(y), v \rangle,
\]

\[
D^2\phi(x)[u]^2 = \langle f''(y(x))Au, Au \rangle = \langle f''(y)v, v \rangle,
\]

\[
D^3\phi(x)[u]^3 = D^3f(y(x))[Au]^3 = D^3f(y)[v]^3. \]

Example: \( f(x) = -\sum_{i=1}^{m} \ln(a_i - \|A_ix - b_i\|^2) \) is a s.c.-function.
Main properties

Let \( x \in \text{dom} \, f \) and \( u \in \mathbb{R}^n, u \neq 0 \). For \( x + tu \in \text{dom} \, f \), consider
\[
\phi(t) = \frac{1}{\langle f''(x+tu)u, u \rangle^{1/2}}.
\]

Lemma. For all feasible \( t \) we have: \(| \phi'(t) | \leq 1 \).

Proof: Indeed, \( \phi'(t) = -\frac{f'''(x+tu)[u]^3}{2\langle f''(x+tu)u,u \rangle^{3/2}} \).

Corollary 1: \( \text{dom} \, \phi \) contains the interval \((-\phi(0), \phi(0))\).

Proof: Since \( f(x + tu) \to \infty \) as \( x + tu \to \partial \text{dom} \, f \), the same is true for \( \langle f''(x+tu)u,u \rangle \). Hence \( \text{dom} \, \phi(t) \equiv \{ t \mid \phi(t) > 0 \} \).

Denote \( W^0(x; r) = \{ y \in \mathbb{R}^n \mid \| y - x \|_f < r \} \). Then
\[
W^0(x; r) \subseteq \text{dom} \, f \text{ for } r < 1.
\]

Main Theorem: for any \( y \in W(x; r), r \in [0, 1) \), we have
\[
(1 - r)^2 F''(x) \leq F''(y) \leq \frac{1}{(1-r)^2} F''(x).
\]
Local convergence

For $x$ close to $x^*$, $f'(x^*) = 0$, function $f(x)$ is almost quadratic:

$$f(x) \approx f^* + \frac{1}{2} \langle f''(x^*)(x - x^*), x - x^* \rangle.$$

Therefore, $f(x) - f^* \approx \frac{1}{2} \| x - x^* \|_2^2 f''(x^*) \approx \frac{1}{2} \| x - x^* \|^2_{f''(x)}$

$$\approx \frac{1}{2} \langle f'(x), [f''(x)]^{-1} f'(x) \rangle \overset{\text{def}}{=} \frac{1}{2} (\| f'(x) \|_x^*)^2 \overset{\text{def}}{=} \lambda_f^2(x).$$

The last value is the local norm of the gradient. It is computable!

**Theorem:** Let $x \in \text{dom} \ f$ and $\lambda_f(x) < 1$.

Then the point $x_+ = x - [f''(x)]^{-1} f'(x)$ belongs to $\text{dom} \ f$ and

$$\lambda_f(x_+) \leq \left( \frac{\lambda_f(x)}{1 - \lambda_f(x)} \right)^2.$$

**NB:** Region of quadratic convergence is $\lambda_f(x) < \bar{\lambda}$, $\frac{\bar{\lambda}}{(1 - \bar{\lambda})^2} = 1$.

It is affine-invariant!
Following the central path

Consider $\Psi_t(x) = t\langle c, x \rangle + f(x)$ with s.c. function $f$.

- For $\Psi_t$, Newton Method has local quadratic convergence.
- The region of quadratic convergence (RQC) is given by $\lambda_{\Psi_t}(x) \leq \beta < \bar{\lambda}$.

Assume we know $x = x^*(t)$. We want to update $t$, $t_+ = t + \Delta$, keeping $x$ in RQC of function $\Psi_{t+\Delta}$: $\lambda_{\Psi_{t+\Delta}}(x) \leq \beta$.

**Question:** How large can be $\Delta$? Since $tc + f'(x) = 0$, we have:

$$\lambda_{\Psi_{t+\Delta}}(x) = \|t_+c + f'(x)\|_x^* = |\Delta| \|c\|_x^* = \frac{|\Delta|}{t} \|f'(x)\|_x^* \leq \beta.$$ 

**Conclusion:** for the linear rate, we need to assume that

$$\langle [f''(x)]^{-1}f'(x), f'(x) \rangle$$

is **uniformly bounded** on $\text{dom } f$.

Thus, we come to the definition of **self-concordant barrier**.
Definition of Self-Concordant Barrier

Let $F(x)$ be a s.c.-function. It is a $\nu$-self-concordant barrier, if

$$\max_{u \in \mathbb{R}^n} [2\langle F'(x), u \rangle - \langle F''(x)u, u \rangle] \leq \nu$$

for all $x \in \text{dom } F$.

The value $\nu$ is called the parameter of the barrier.

If $F''(x)$ is non-degenerate, then $\langle [F''(x)]^{-1}F'(x), F'(x) \rangle \leq \nu$.

Another form: $\langle F'(x), u \rangle^2 \leq \nu \langle F''(x)u, u \rangle$.

Main property: $\langle F'(x), y - x \rangle \leq \nu, x, y \in \text{int } Q$.

NB: $\nu$ is responsible for the rate of p.-f. method: $t_+ = t \pm \frac{\alpha \cdot t}{\nu^{1/2}}$.

Complexity: $O\left(\sqrt{\nu} \ln \frac{\nu}{\epsilon}\right)$ iterations of the Newton method.

Calculus: 1. Affine transformations do not change $\nu$.
2. Restriction on a subspace can only decrease $\nu$.
3. $F = F_1 + F_2 \quad \Rightarrow \quad \nu = \nu_1 + \nu_2$. 
Examples

1. Barrier for a ray: \( F(t) = -\ln t, \quad F'(t) = -\frac{1}{t}, \quad F''(t) = \frac{1}{t^2}, \quad \nu = 1. \)

2. Polytop \( \{x : \langle a_i, x \rangle \leq b_i\} \), \( F(x) = -\sum_{i=1}^{m} \ln(b_i - \langle a_i, x \rangle) \), \( \nu = m. \)

3. \( l_2\)-ball: \( F(x) = -\ln(1 - \|x\|^2) \), \( D_1 = \omega_1, \quad D_2 = \omega_1^2 + \omega_2, \quad \nu = 1. \)

4. Intersection of ellipsoids: \( F(x) = -\sum_{i=1}^{m} \ln(r_i^2 - \|A_i x - b_i\|^2) \), \( \nu = m. \)

5. Lorentz cone \( \{t \geq \|x\|\} \), \( F(x, t) = -\ln(t^2 - \|x\|^2) \), \( \nu = 2. \)

6. LMI-cone \( \{X = X^T \succeq 0\} \), \( F(X) = -\ln \det X \), \( \nu = n. \)

7. Epigraph \( \{t \geq e^x\} \), \( F(x, t) = -\ln(t - e^x) - \ln(\ln t - x) \), \( \nu = 4. \)

8. Universal barrier. Define the polar set
\[
P(x) = \{s : \langle s, y - x \rangle \leq 1, \quad y \in Q\}.
\]
Then \( F(x) = -\ln \text{vol}_n P(x) \) is an \( O(n) \)-s.c. barrier for \( Q. \)
Further directions: specification of the model description

Path-following methods

- Conic problems. Gain: primal-dual IPM.
- Self-scaled cones: $F^*(F''(x)u) \equiv F(u) - 2F(x) - \nu$. Gain: long-step methods, very good search directions.
- Positive polynomials: $p(t) \geq 0, \ t \in R$ iff $p_k = \sum_{i+j=k} Y^i j, \ Y \succeq 0$. Gain: very cheap computation of determinants.

Black-box methods

- Composite functions: $f(x) + h(x)$, where $f$ is smooth but complex, and $h$ is nonsmooth and simple. Gain: rate $O\left(\frac{1}{k^2}\right)$.
- Huge-scale problem: very sparse linear operators. Gain: extremely cheap iterations. (Next Lecture.)