Part IV. Structural Programming.

Lecture 6.

Self-Concordant Functions.

– Do we really have a black box?

– What the Newton Method actually does?

– Definition of Self-Concordant Functions.

– Main Properties.

– Newton Method.

Black Box in Convex Programming

Suppose we want to solve the problem:

$$\min_{x \in R^n} f(x).$$

Note:

1. In order to apply the Convex Programming methods we have to be *sure* that f is convex.

2. In order to prove that f is convex, we have to check its *structure*.

3. If f is obtained from the *basic* convex functions by *convex* operations (summation, maximum, etc.), we conclude that it is convex.

Conclusion:

1. The function *cannot* be in a Black Box, when we check its convexity.

2. We *put* it there only for numerical methods.

That is the main contradiction of the Convex Programming Theory.

What we are going to do?

We replace the Black Box concept by the concept of *mediator*.

For a problem $\mathcal{P}$, a mediator is a new problem instance $\mathcal{M}$, which:

1. Properly reflects all properties of problem $\mathcal{P}$ as a minimization problem.

2. Is easier to solve than the initial problem.

Note:

1. The creation of the mediator can be seen as a *part* of the whole process of solving the initial problem.

2. The data exchange between the mediator and the numerical method can be described by an oracle.

However, this oracle is *not local* anymore, since a nontrivial mediator reflects some *global* properties of the initial problem.

End points:

$$\mathcal{M} \equiv \mathcal{P} \longleftarrow \ldots \longrightarrow \mathcal{M} \equiv (f^*, x^*).$$

But we need something in the middle!

Basic Method

Note:

The class of mediators can be oriented on a specific numerical method.

In this case, we apply the following scheme:

1. Choose a basic method.

2. Describe a set of problems, for which the basic method is very efficient.

3. Prove that the diversity of these problems is sufficient to be used as mediators for our initial problem class.

4. Describe the class of problems, for which the mediator can be created in a computable form.

In the theory of Self-Concordant Functions the basic method is:

> The *Newton Method* as applied in the framework of *Sequential Unconstrained Minimization.*

What can be good for Newton Method?

Standard theorem:

Assume that:

- $f''(x^*) \geq l I_n$
- $\| f''(x) - f''(y) \| \leq M \| x - y \|, \forall x, y \in R^n$.
- The starting point x_0 is close enough to x^*:

$$\| x_0 - x^* \| < \bar{r} = \frac{2l}{3M}.$$

Then $\| x_k - x^* \| < \bar{r}$ for all k and the Newton method converges quadratically:

$$\| x_{k+1} - x^* \| \leq \frac{M \| x_k - x^* \|^2}{2(l - M \| x_k - x^* \|)}.$$

Note:

1. The description of the *region* of quadratic convergence is given in the metric $\langle \cdot, \cdot \rangle$.

2. This region is changing when we choose another metric.

Affine invariance of the Newton Method

Let f satisfy our assumptions. Consider the function

$$\phi(y) = f(Ay),$$

where A is nondegenerate $(n \times n)$-matrix.

Lemma 6.1 *Let $\{x_k\}$ be a sequence, generated by the Newton Method for function f:*

$$x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k), \quad k \geq 0.$$

Consider the sequence $\{y_k\}$, generated by the Newton Method for function ϕ:

$$y_{k+1} = y_k - [\phi''(y_k)]^{-1} \phi'(y_k), \quad k \geq 0,$$

with $y_0 = A^{-1} x_0$. Then

$$y_k = A^{-1} x_k$$

for all $k \geq 0$.

Proof:
Let $y_k = A^{-1} x_k$ for some $k \geq 0$. Then

$$\begin{aligned} y_{k+1} &= y_k - [\phi''(y_k)]^{-1} \phi'(y_k) \\ &= y_k - [A^T f''(Ay_k) A]^{-1} A^T f'(Ay_k) \\ &= A^{-1} x_k - A^{-1} [f''(x_k)]^{-1} f'(x_k) = A^{-1} x_{k+1} \end{aligned}$$

□

Conclusion: The real region of quadratic convergence *does not depend on the metric*!

Old assumption:

$$\| f''(x) - f''(y) \| \leq M \| x - y \| .$$

Let $f \in C^3(R^n)$. Denote

$$f'''(x)[u] = \lim_{\alpha \to 0} \frac{1}{\alpha}[f''(x + \alpha u) - f''(x)].$$

That is a matrix!

Then our assumption is equivalent to the following:

$$\| f'''(x)[u] \| \leq M \| u \| .$$

This means that at any point $x \in R^n$ we have

$$| \langle f'''(x)[u]v, v \rangle | \leq M \| u \| \cdot \| v \|^2$$

for all $u, v \in R^n$.

Note:

1. The norm $\| \cdot \|$ has nothing common with our concrete f.

2. However, there exists a norm, which is closely related to f. That is $\| \cdot \|_{f''(x)}$.

3. Let us make a similar assumption, which is written in terms of $\| \cdot \|_{f''(x)}$.

Definition of Self-Concordant Function

Let $f(x) \in C^3(\mathrm{dom}\, f)$ be a *closed convex* function with *open* domain.

Let us fix a point $x \in \mathrm{dom}\, f$ and a direction $u \in R^n$. Consider the function

$$\phi(x;t) = f(x + tu),$$

with the variable $t \in \mathrm{dom}\, \phi(x;\cdot) \subseteq R^1$.

Denote

$$\begin{aligned} Df(x)[u] &= \phi'(x;t) = \langle f'(x), u \rangle, \\ D^2 f(x)[u,u] &= \phi''(x;t) = \langle f''(x)u, u \rangle = \| u \|^2_{f''(x)} \\ D^3 f(x)[u,u,u] &= \phi'''(x;t) = \langle f'''[u]u, u \rangle. \end{aligned}$$

Definition 6.1 *We call a function f* self-concordant *if the inequality*

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

with some $M_f \geq 0$ holds for any $x \in \mathrm{dom}\, f$ and $u \in R^n$.

Note:

- We cannot expect that these functions are very common.
- We need them to construct the mediators.
- Our main goal now is to show that they are suitable for the Newton Method.

Lemma 6.2 *If f is a self-concordant function, then for any u_1, u_2 and $u_3 \in R^n$ we have*

$$\begin{aligned} &| \; D^3 f(x)[u_1, u_2, u_3] \; | \\ &\leq M_f \parallel u_1 \parallel_{f''(x)} \cdot \parallel u_2 \parallel_{f''(x)} \cdot \parallel u_3 \parallel_{f''(x)} . \end{aligned} \tag{6.1}$$

*__Proof:__
Accept without proof. □

Note: Very often

- We use Definition 6.1 to prove that some f is self-concordant.
- We use Lemma 6.2 to establish the properties of self-concordant functions.

Examples

1. Linear function.

$$\begin{aligned} f(x) &= \alpha + \langle a, x \rangle, \quad \operatorname{dom} f = R^n, \\ f'(x) &= a, \quad f''(x) = 0, \quad f'''(x) = 0, \\ M_f &= 0. \end{aligned}$$

2. Convex quadratic function.

$$\begin{aligned} f(x) = {} & \alpha + \langle a, x \rangle + \tfrac{1}{2} \langle Ax, x \rangle, \\ & A = A^T \geq 0, \quad \operatorname{dom} f = R^n, \\ f'(x) = {} & a + Ax, \quad f''(x) = A, \quad f'''(x) = 0, \\ M_f = {} & 0. \end{aligned}$$

3. Logarithmic barrier for a ray.

$$\begin{aligned} f(x) &= -\ln x, \quad \operatorname{dom} f = \{ x \in R^1 \mid \ x > 0 \}, \\ f'(x) &= -\tfrac{1}{x}, \quad f''(x) = \tfrac{1}{x^2}, \quad f'''(x) = -\tfrac{2}{x^3}, \\ M_f &= 2. \end{aligned}$$

4. Logarithmic barrier for a quadratic region.

Let $A = A^T \geq 0$. Consider the *concave* function

$$\phi(x) = \alpha + \langle a, x\rangle - \tfrac{1}{2}\langle Ax, x\rangle.$$

Define

$$f(x) = -\ln \phi(x), \quad \mathrm{dom}\, f = \{x \in R^n \mid \ \phi(x) > 0\}.$$

Then

$$\begin{aligned} Df(x)[u] = & -\tfrac{1}{\phi(x)}[\langle a, u\rangle - \langle Ax, u\rangle], \\ D^2 f(x)[u, u] = & \tfrac{1}{\phi^2(x)}[\langle a, u\rangle - \langle Ax, u\rangle]^2 \\ & + \tfrac{1}{\phi(x)}\langle Au, u\rangle, \\ D^3 f(x)[u, u, u] = & -\tfrac{2}{\phi^3(x)}[\langle a, u\rangle - \langle Ax, u\rangle]^3 \\ & - \tfrac{3}{\phi^2(x)}[\langle a, u\rangle - \langle Ax, u\rangle]\langle Au, u\rangle \end{aligned}$$

Denote $\omega_1 = Df(x)[u]$, $\omega_2 = \frac{1}{\phi(x)}\langle Au, u\rangle$. Then

$$\begin{aligned} D^2 f(x)[u, u] \ &= \omega_1^2 + \omega_2 \geq 0, \\ \mid D^3 f(x)[u, u, u] \mid \ &= \mid 2\omega_1^3 + 3\omega_1\omega_2 \mid . \end{aligned}$$

Nontrivial case: $\omega_1 \neq 0$. Denote $\alpha = \frac{\omega_2}{\omega_1^2}$. Then

$$\frac{|D^3 f(x)[u,u,u]|}{(D^2 f(x)[u,u])^{3/2}} \leq \frac{2|\omega_1|^3 + 3|\omega_1|\omega_2}{(\omega_1^2 + \omega_2)^{3/2}} = \frac{2(1+\frac{3}{2}\alpha)}{(1+\alpha)^{3/2}} \leq 2.$$

Thus, $M_f = 2$.

Simple Properties

Theorem 6.1 *Let f_i are self-concordant with constants M_i, $i = 1, 2$ and α, $\beta > 0$.*

Then $f(x) = \alpha f_1(x) + \beta f_2(x)$ is self-concordant with the constant

$$M_f = \max\left\{ \frac{1}{\sqrt{\alpha}} M_1, \frac{1}{\sqrt{\beta}} M_2 \right\}$$

and $\mathrm{dom}\, f = \mathrm{dom}\, f_1 \cap \mathrm{dom}\, f_2$.

Proof:
Note that f is a closed convex function in view of Theorem 7.5.

Let us fix $x \in \mathrm{dom}\, f$ and $u \in R^n$. Then

$$\mid D^3 f_i(x)[u,u,u] \mid \leq M_i \left[D^2 f_i(x)[u,u] \right]^{3/2}, \quad i = 1, 2.$$

Denote $\omega_i = D^2 f_i(x)[u,u] \geq 0$. Then

$$\frac{|D^3 f(x)[u,u,u]|}{[D^2 f(x)[u,u]]^{3/2}} \leq \frac{\alpha |D^3 f_1(x)[u,u,u]| + \beta |D^3 f_2(x)[u,u,u]|}{[\alpha D^2 f_1(x)[u,u] + \beta D^2 f_2(x)[u,u]]^{3/2}}$$

$$\leq \frac{\alpha M_1 \omega_1^{3/2} + \beta M_2 \omega_2^{3/2}}{[\alpha \omega_1 + \beta \omega_2]^{3/2}}.$$

The RHS is not changing when $(\omega_1, \omega_2) \to (t\omega_1, t\omega_2)$ with $t > 0$. Therefore we can assume that

$$\alpha \omega_1 + \beta \omega_2 = 1.$$

Denote $\xi = \alpha \omega_1$. Then the RHS is

$$\frac{M_1}{\sqrt{\alpha}} \xi^{3/2} + \frac{M_2}{\sqrt{\beta}} (1 - \xi)^{3/2}, \quad \xi \in [0, 1]$$

This function is convex in ξ. Therefore its maximum is either $\xi = 0$ or $\xi = 1$ (C.4.1). $\square$

Corollary 6.1 *Let f be self-concordant with some constant M_f.*

If $A = A^T \geq 0$ then the function

$$\phi(x) = \alpha + \langle a, x\rangle + \tfrac{1}{2}\langle Ax, x\rangle + f(x)$$

is self-concordant with the constant

$$M_\phi = M_f.$$

Proof:
We have seen that any convex quadratic function is self-concordant with the constant equal to zero. □

Corollary 6.2 *Let f be self-concordant with some constant M_f and $\alpha > 0$.*

Then $\phi(x) = \alpha f(x)$ is self-concordant with the constant

$$M_\phi = \frac{1}{\sqrt{\alpha}} M_f.$$

□

Theorem 6.2 *Let $\mathcal{A}(x) = Ax+b$ be a linear operator:*

$$\mathcal{A}(x): \quad R^n \to R^m.$$

Assume that a function $f(y)$ is self-concordant with the constant M_f.

Then $\phi(x) = f(\mathcal{A}(x))$ is self-concordant and

$$M_\phi = M_f.$$

Proof:
The function $\phi(x)$ is closed and convex in view of Theorem 4.6.

Let us fix $x \in \operatorname{dom}\phi = \{x : \mathcal{A}(x) \in \operatorname{dom} f\}$ and $u \in R^n$.

Denote $y = \mathcal{A}(x)$, $v = Au$. Then

$$\begin{aligned} D\phi(x)[u] &= \langle f'(\mathcal{A}(x)), Au\rangle = \langle f'(y), v\rangle, \\ D^2\phi(x)[u,u] &= \langle f''(\mathcal{A}(x))Au, Au\rangle \\ &= \langle f''(y)v, v\rangle, \\ D^3\phi(x)[u,u,u] &= D^3 f(\mathcal{A}(x))[Au, Au, Au] \\ &= D^3 f(y)[v,v,v]. \end{aligned}$$

Therefore

$$\begin{aligned} &\mid D^3\phi(x)[u,u,u] \mid = \mid D^3 f(y)[v,v,v] \mid \\ &\leq M_f \langle f''(y)v, v\rangle^{3/2} = M_f (D^2\phi(x)[u,u])^{3/2}. \end{aligned}$$

□

Theorem 6.3 *Let f be self-concordant.*

If $\operatorname{dom} f$ *contains no straight line, then $F''(x)$ is non-degenerate at any $x \in \operatorname{dom} f$.*

Proof:
Assume $\langle F''(x)u, u\rangle = 0$ for $x \in \operatorname{dom} f$, $u \in R^n$.

Let $y_\alpha = x + \alpha u \in \operatorname{dom} f$ and $\psi(\alpha) = \langle f''(y_\alpha)u, u\rangle$. Then

$$\psi'(\alpha) = D^3 f(y_\alpha)[u,u,u] \le 2\psi(\alpha)^{3/2}, \quad \psi(0) = 0.$$

Since $\psi(\alpha) \ge 0$, we conclude that $\psi'(0) = 0$. Therefore ψ satisfies the system:

$$\psi'(\alpha) + \xi'(\alpha) = 2\psi(\alpha)^{3/2}, \quad \xi'(\alpha) = 0,$$

$$\psi(0) = \xi(0) = 0.$$

The unique solution is $\psi(\alpha) = \xi(\alpha) = 0$. Therefore $\phi(\alpha) = f(y_\alpha)$ is linear:

$$\phi(\alpha) = f(x) + \langle f'(x), y_\alpha - x\rangle + \int_0^\alpha \int_0^\lambda \langle f''(y_\tau)u, u\rangle d\tau d\lambda = f(x) + \alpha\langle f'(x), u\rangle.$$

Assume $\exists \bar\alpha : y_{\bar\alpha} \in \partial(\operatorname{dom} f)$. Consider $\{\alpha_k\}$: $\alpha_k \uparrow \bar\alpha$. Then $z_k = (y_{\alpha_k}, \phi(\alpha_k)) \to \bar z = (y_{\bar\alpha)}, \phi(\bar\alpha))$.

Note that $z_k \in \operatorname{epi} f$, but $\bar z \notin \operatorname{epi} f$ since $y_{\bar\alpha} \notin \operatorname{dom} f$.

That is a contradiction since f is closed.

Considering direction $-u$, we conclude: $y_\alpha \in \operatorname{dom} f$ for all α. That contradicts to assumptions of the theorem. □

Theorem 6.4 *Let f be a self-concordant function. Then for any point $\bar{x} \in \partial(\mathrm{dom}\, f)$ and any sequence*

$$\{x_k\} \subset \mathrm{dom}\, f : \quad x_k \to \bar{x}$$

we have $f(x_k) \to +\infty$.

Proof:
Note that the sequence $\{f(x_k)\}$ is bounded below:

$$f(x_k) \geq f(x_0) + \langle f'(x_0), x_k - x_0 \rangle.$$

Assume that it is bounded from above.

Then it has a limit point $\bar{f}$. We can think that this is a unique limit point.

Therefore

$$z_k = (x_k, f(x_k)) \to \bar{z} = (\bar{x}, \bar{f}).$$

Note that $z_k \in \mathrm{epi}\, f$, but $\bar{z} \notin \mathrm{epi}\, f$ since $\bar{x} \notin \mathrm{dom}\, f$.

That is a contradiction since f is closed. $\square$

Thus, $f(x)$ is a *barrier function* for $\mathrm{cl}\,(\mathrm{dom}\, f)$.

Main Inequalities

From now on:

- We fix some self-concordant function $f(x)$.
- We assume that $M_f = 2$ (otherwise we can scale it, see Corollary 6.2).
- We assume that dom f contains no straight line (this implies that all $f''(x)$ are nondegenerate, see Theorem 6.3).
- Denote:

$$\begin{aligned} \| u \|_x &= \langle f''(x)u, u\rangle^{1/2}, \\ \| v \|_x^* &= \langle [f''(x)]^{-1}v, v\rangle^{1/2}, \\ \lambda_f(x) &= \langle [f''(x)]^{-1}f'(x), f'(x)\rangle^{1/2}. \end{aligned}$$

Clearly, $| \langle v, u\rangle | \leq \| v \|_x^* \cdot \| u \|_x$.

We call:

- $\| u \|_x$ the local norm of a point u with respect to x.
- $\lambda_f(x) = \| f'(x) \|_x^*$ the local norm of

  the gradient $f'(x)$.

Let us fix $x \in \operatorname{dom} f$ and $u \in R^n$, $u \neq 0$. Consider the function

$$\phi(t) = \frac{1}{\langle f''(x+tu)u, u\rangle^{1/2}}, \quad t: \ x + tu \in \operatorname{dom} f.$$

Lemma 6.3 *For all feasible t we have:*

$$\mid \phi'(t) \mid \leq 1.$$

Proof:
Indeed,

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

Therefore $\mid \phi'(t) \mid \leq 1$ in view of Definition 6.1. □

Corollary 6.3 *The domain of function $\phi(t)$ contains the interval $(-\phi(0), \phi(0))$.*

Proof:
Since $f(x+tu) \to \infty$ as $x+tu$ approaches the boundary of $\operatorname{dom} f$ (see Theorem 6.4), the same is true for $\langle f''(x+tu)u, u\rangle$.

Therefore

$$\operatorname{dom} \phi(t) \equiv \{t \mid \ \phi(t) > 0\}.$$

It remains to note that $\phi(t) \geq \phi(0) - \mid t \mid$ in view of Lemma 6.3. □

Denote

$$W^0(x;r) = \{y \in R^n \mid \| y - x \|_x < r\},$$
$$W(x;r) = \mathrm{cl}\left(W^0(x;r)\right).$$

Theorem 6.5 *1. For any $x \in \mathrm{dom}\, f$ we have:*

$$W^0(x;1) \subseteq \mathrm{dom}\, f.$$

2. For all x, $y \in \mathrm{dom}\, f$ the following inequality holds:

$$\| y - x \|_y \geq \frac{\|y-x\|_x}{1+\|y-x\|_x}. \tag{6.2}$$

3. If $\| y - x \|_x < 1$ then

$$\| y - x \|_y \leq \frac{\|y-x\|_x}{1-\|y-x\|_x}. \tag{6.3}$$

Proof:
1. In view of Corollary 6.3, $\mathrm{dom}\, f$ contains the set

$$\{y = x + tu \mid \; t^2 \| u \|_x^2 < 1\}$$

(since $\phi(0) = 1/ \| u \|_x$). That is exactly $W^0(x;1)$.
2. Let us choose $u = y - x$. Then

$$\phi(1) = \frac{1}{\|y-x\|_y}, \quad \phi(0) = \frac{1}{\|y-x\|_x},$$

and $\phi(1) \leq \phi(0) + 1$ in view of Lemma 6.3. That is (6.2).
3. If $\| y - x \|_x < 1$, then $\phi(0) > 1$, and in view of Lemma 6.3 $\phi(1) \geq \phi(0) - 1$. That is (6.3). □

Theorem 6.6 *Let $x \in \operatorname{dom} f$. Then for any*

$$y \in W^0(x;1)$$

we have:

$$(1- \| y - x \|_x)^2 f''(x) \le f''(y) \le \frac{1}{(1-\|y-x\|_x)^2} f''(x). \tag{6.4}$$

Proof:
Let us fix some $u \in R^n$, $u \neq 0$. Consider the function

$$\psi(t) = \langle f''(x + t(y-x))u, u\rangle, \quad t \in [0,1].$$

Denote $x_t = x + t(y-x)$. Then, in view of Lemma 6.2 and (6.3), we have:

$$\begin{aligned} | \psi'(t) | &= | D^3 f(x_t)[y-x, u, u] | \\ &\le 2 \| y - x \|_{x_t} \| u \|_{x_t}^2 = \tfrac{2}{t} \| x_t - x \|_{x_t} \psi(t) \\ &\le \tfrac{2}{t} \cdot \tfrac{\|x_t - x\|_x \psi(t)}{1 - \|x_t - x\|_x} = \tfrac{2\|y-x\|_x \psi(t)}{1 - t\|y-x\|_x}. \end{aligned}$$

Therefore

$$\begin{aligned} 2(\ln(1 - t \| y - x \|_x))' &\le (\ln \psi(t))' \\ &\le -2(\ln(1 - t \| y - x \|_x))'. \end{aligned}$$

Let us integrate this inequality in $t \in [0,1]$. We get:

$$(1- \| y - x \|_x)^2 \le \frac{\psi(1)}{\psi(0)} \le \frac{1}{(1-\|y-x\|_x)^2}.$$

That is (6.4). □

Corollary 6.4 *Let* $x \in \operatorname{dom} f$ *and* $r = \| y - x \|_x < 1$. *Then we can estimate the matrix*

$$G = \int_0^1 f''(x + \tau(y - x))d\tau$$

as follows:

$$(1 - r + \frac{r^2}{3})f''(x) \leq G \leq \frac{1}{1-r} f''(x).$$

Proof:
Indeed, in view of Theorem 6.6 we have:

$$\begin{aligned} G &= \int_0^1 f''(x + \tau(y - x))d\tau \\ &\geq f''(x) \int_0^1 (1 - \tau r)^2 d\tau \\ &= (1 - r + \tfrac{1}{3}r^2) f''(x), \\ G &\leq f''(x) \int_0^1 \frac{d\tau}{(1-\tau r)^2} = \frac{1}{1-r} f''(x). \end{aligned}$$

□

Remarks:

- At any point $x \in \mathrm{dom}\, f$ we can point out an *ellipsoid*

  $W^0(x;1) = \{x \in R^n \mid \ \langle f''(x)(y-x), y-x\rangle < 1\}$,

  belonging to $\mathrm{dom}\, f$.

- Inside the ellipsoid $W(x;r)$ with $r \in [0,1)$ the function f is almost quadratic since

  $$(1-r)^2 f''(x) \le f''(y) \le \frac{1}{(1-r)^2} f''(x)$$

  for all $y \in W(x;r)$.

  Choosing r small enough, we can make the quality of quadratic approximation acceptable for our goals.

Note:

1. The above facts form the basis almost for all consequent results.

2. In Convex Optimization we have never seen such favorable situation.

Theorem 6.7 *For any x, $y \in \operatorname{dom} f$ we have:*

$$\langle f'(y) - f'(x), y - x\rangle \geq \frac{\|y-x\|_x^2}{1+\|y-x\|_x}, \qquad (6.5)$$

$$f(y) \geq f(x) + \langle f'(x), y - x\rangle + \omega(\| y - x \|_x), \quad (6.6)$$

where $\omega(t) = t - \ln(1+t)$.

Proof:
Denote $y_\tau = x + \tau(y - x)$, $\tau \in [0, 1]$, and $r = \| y - x \|_x$. Then, in view of (6.2) we have:

$$\langle f'(y) - f'(x), y - x\rangle$$

$$= \int_0^1 \langle f''(y_\tau)(y - x), y - x\rangle d\tau = \int_0^1 \frac{1}{\tau^2} \| y_\tau - x \|_{y_\tau}^2 \, d\tau$$

$$\geq \int_0^1 \frac{r^2}{(1+\tau r)^2} d\tau = r \int_0^r \frac{1}{(1+t)^2} dt = \frac{r^2}{1+r}$$

Further, using (6.5), we obtain:

$$f(y) - f(x) - \langle f'(x), y - x\rangle$$

$$= \int_0^1 \langle f'(y_\tau) - f'(x), y - x\rangle d\tau$$

$$= \int_0^1 \frac{1}{\tau} \langle f'(y_\tau) - f'(x), y_\tau - x\rangle d\tau$$

$$\geq \int_0^1 \frac{\|y_\tau - x\|_x^2}{\tau(1+\|y_\tau - x\|_x)} d\tau = \int_0^1 \frac{\tau r^2}{1+\tau r} d\tau = \int_0^r \frac{t dt}{1+t} = \omega(r).$$

□

Theorem 6.8 *Let* $x \in \operatorname{dom} f$ *and* $\| y - x \|_x < 1$. *Then*

$$\langle f'(y) - f'(x), y - x \rangle \leq \frac{\|y-x\|_x^2}{1-\|y-x\|_x}, \qquad (6.7)$$

$$f(y) \leq f(x) + \langle f'(x), y - x \rangle + \omega_*(\| y - x \|_x), \quad (6.8)$$

where $\omega_*(t) = -t - \ln(1-t)$.

Proof:
Denote $y_\tau = x + \tau(y - x)$, $\tau \in [0, 1]$, and $r = \| y - x \|_x$. Since $\| y_\tau - x \| < 1$, in view of (6.3) we have:

$$\langle f'(y) - f'(x), y - x \rangle$$

$$= \int_0^1 \langle f''(y_\tau)(y - x), y - x \rangle d\tau = \int_0^1 \frac{1}{\tau^2} \| y_\tau - x \|_{y_\tau}^2 \, d\tau$$

$$\leq \int_0^1 \frac{r^2}{(1-\tau r)^2} d\tau = r \int_0^r \frac{1}{(1-t)^2} dt = \frac{r^2}{1-r}$$

Further, using (6.7), we obtain:

$$f(y) - f(x) - \langle f'(x), y - x \rangle$$

$$= \int_0^1 \langle f'(y_\tau) - f'(x), y - x \rangle d\tau$$

$$= \int_0^1 \frac{1}{\tau} \langle f'(y_\tau) - f'(x), y_\tau - x \rangle d\tau$$

$$\leq \int_0^1 \frac{\|y_\tau - x\|_x^2}{\tau(1-\|y_\tau - x\|_x)} d\tau = \int_0^1 \frac{\tau r^2}{1-\tau r} d\tau = \int_0^r \frac{t dt}{1-t} = \omega_*(r).$$

□

Functions $\omega(t)$ and $\omega_*(\tau)$

Note that

$$\omega(t) = t - \ln(1+t), \quad \omega_*(\tau) = -\tau - \ln(1-\tau).$$

Therefore

$$\begin{aligned} \omega'(t) &= \tfrac{t}{1+t} \geq 0, & \omega''(t) &= \tfrac{1}{(1+t)^2} > 0, \\ \omega_*'(\tau) &= \tfrac{\tau}{1-\tau} \geq 0, & \omega_*''(\tau) &= \tfrac{1}{(1-\tau)^2} > 0. \end{aligned}$$

Thus, $\omega(t)$ and $\omega_*(\tau)$ are convex functions.

Lemma 6.4 *For any $t \geq 0$ and $\tau \in [0,1)$ we have:*

$$\omega'(\omega_*'(\tau)) = \tau, \quad \omega_*'(\omega'(t)) = t,$$

$$\omega(t) = \max_{0 \leq \xi < 1} [\xi t - \omega_*(\xi)],$$

$$\omega_*(\tau) = \max_{\xi \geq 0} [\xi \tau - \omega(\xi)]$$

$$\omega(t) + \omega_*(\tau) \geq \tau t,$$

$$\omega_*(\tau) = \tau \omega_*'(\tau) - \omega(\omega_*'(\tau)),$$

$$\omega(t) = t\omega'(t) - \omega_*(\omega'(t)).$$

Proof:

Prove that as an exercise. □

Note:

The reason for the relations is that functions $\omega(t)$ and $\omega_*(\tau)$ are *conjugate*.

Minimizing the self-concordant function

Consider the problem:

$$\min\{f(x) \mid \ x \in \mathrm{dom}\, f\}. \tag{6.9}$$

Theorem 6.9 *Let* $\lambda_f(x) < 1$ *for some* $x \in \mathrm{dom}\, f$.

Then the solution of problem (6.9), x_f^**, exists and unique.*

Proof:
Indeed, in view of (6.6), for any $y \in \mathrm{dom}\, f$ we have:

$$f(y) \geq f(x) + \langle f'(x), y - x\rangle + \omega(\| y - x \|_x)$$

$$\geq f(x) - \| f'(x) \|_x^* \cdot \| y - x \|_x + \omega(\| y - x \|_x)$$

$$= f(x) - \lambda_f(x) \cdot \| y - x \|_x + \omega(\| y - x \|_x).$$

Therefore $\forall y \in \mathcal{L}_f(f(x)) = \{y \in R^n \mid \ f(y) \leq f(x)\}$ we have:

$$\frac{1}{\|y-x\|_x}\omega(\| y - x \|_x) \leq \lambda_f(x) < 1.$$

Hence, $\| y - x \|_x \leq \bar{t}$, where $\bar{t}$ is a root of the equation:

$$(1 - \lambda_f(x))t = \ln(1 + t), \quad t > 0.$$

Thus, $\mathcal{L}_f(f(x))$ is bounded and therefore x_f^* exists.

It is unique since in view of (6.6)

$$f(y) \geq f(x_f^*) + \omega(\| y - x_f^* \|_{x_f^*}), \quad \forall y \in \mathrm{dom}\, f.$$

□

Damped Newton Method

Consider the following scheme:

0. Choose $x_0 \in \mathrm{dom}\, f$.
1. For $k \geq 0$ iterate

$$x_{k+1} = x_k - \frac{1}{1+\lambda_f(x_k)}[f''(x_k)]^{-1} f'(x_k). \qquad (6.10)$$

Theorem 6.10 *For any $k \geq 0$ we have:*

$$f(x_{k+1}) \leq f(x_k) - \omega(\lambda_f(x_k)). \qquad (6.11)$$

Proof:
Denote $\lambda = \lambda_f(x_k)$. Then

$$\| x_{k+1} - x_k \|_x = \tfrac{\lambda}{1+\lambda} = \omega'(\lambda).$$

Therefore, in view of (6.8), we have:

$$\begin{aligned} f(x_{k+1}) \leq\ & f(x_k) + \langle f'(x_k), x_{k+1} - x_k \rangle \\ & + \omega_*(\| x_{k+1} - x_k \|_x) \\ =\ & f(x_k) - \tfrac{\lambda^2}{1+\lambda} + \omega_*(\omega'(\lambda)) \\ =\ & f(x_k) - \lambda\omega'(\lambda) + \omega_*(\omega'(\lambda)) \\ =\ & f(x_k) - \omega(\lambda) \end{aligned}$$

(we have used Lemma 6.4). □

Remarks:

1. We have seen that a local condition $\lambda_f(x) < 1$ provides us with a global information (existence of x_f^*).

The result of Theorem 6.9 cannot be improved

(consider $f(x) = -\ln x$).

2. For all $x \in \mathrm{dom}\, f$ with $\lambda_f(x) \geq \beta > 0$ we can decrease the value of the $f(x)$ at least by the constant $\omega(\beta) > 0$ (by one step of the Damped Newton Method).

3. The result of Theorem 6.10 is *global*. It can be used to obtain a global efficiency estimate of the process.

Local convergence

Consider the scheme of the *standard* Newton Method:

0. Choose $x_0 \in \mathrm{dom}\, f$.
1. For $k \geq 0$ iterate

$$x_{k+1} = x_k - [f''(x_k)]^{-1} f'(x_k). \qquad (6.12)$$

Let us study it *local* convergence.

We can measure the convergence of the process in different ways:

- $f(x_k) - f(x_f^*)$ - functional gap;
- $\lambda_f(x_k) = \| f'(x_k) \|_{x_k}^*$ - local norm of the gradient;
- $\| x_k - x_f^* \|_{x_k}$ - local distance to the minimum with respect to x_k;

Let us prove that locally all these measures are equivalent.

Theorem 6.11 *Let* $\lambda_f(x) < 1$. *Then*

$$\omega(\lambda_f(x)) \leq f(x) - f(x_f^*) \leq \omega_*(\lambda_f(x)), \qquad (6.13)$$

$$\omega'(\lambda_f(x)) \leq \| x - x_f^* \|_x \leq \omega_*'(\lambda_f(x)), \qquad (6.14)$$

Proof:
Denote $r = \| x - x_f^* \|_x$ and $\lambda = \lambda_f(x)$.

1. LHS of (6.13) follows from Theorem 6.10. Further, in view of (6.6) we have:

$$\begin{aligned} f(x_f^*) \geq &\ f(x) + \langle f'(x), x_f^* - x \rangle + \omega(r) \\ \geq &\ f(x) - \lambda r + \omega(r) \geq f(x) - \omega_*(\lambda). \end{aligned}$$

2. In view of (6.5) we have:

$$\tfrac{r^2}{1+r} \leq \langle f'(x), x - x_f^* \rangle \leq \lambda r.$$

That is the RHS of (6.14).

If $r \geq 1$ then the LHS of (6.14) is trivial.

Suppose that $r < 1$. Then

$$\begin{aligned} f'(x) &= G(x - x_f^*), \ G = \int_0^1 f''(x_f^* + \tau(x - x_f^*)) d\tau, \\ \lambda_f^2(x) &= \langle [f''(x)]^{-1} G(x - x_f^*), G(x - x_f^*) \rangle \\ &\leq \| H \|^2 r^2, \end{aligned}$$

where $H = [f''(x)]^{-1/2} G [f''(x)]^{-1/2}$.

In view of Corollary 6.4, we have: $G \leq \frac{1}{1-r} f''(x)$.

Therefore $\| H \| \leq \frac{1}{1-r}$ and we conclude that

$$\lambda_f^2(x) \leq \frac{r^2}{(1-r)^2} = (\omega_*'(r))^2.$$

Thus, $\lambda_f(x) \leq \omega_*'(r)$. Applying $\omega'(\cdot)$ to both sides, we get LHS of (6.14). $\square$

We will estimate the local convergence in terms of $\lambda_f(x)$.

Theorem 6.12 *Let* $x \in \operatorname{dom} f$ *and* $\lambda_f(x) < 1$. *Then the point* $x_+ = x - [f''(x)]^{-1} f'(x)$ *belongs to* $\operatorname{dom} f$ *and we have*

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

Proof:
Denote $p = x_+ - x$, $\lambda = \lambda_f(x)$. Then $\| p \|_x = \lambda < 1$. Therefore $x_+ \in \operatorname{dom} f$ (see Theorem 6.5).

Note that in view of Theorem 6.6

$$\begin{aligned} \lambda_f(x_+) &= \langle [f''(x_+)]^{-1} f'(x_+), f'(x_+) \rangle^{1/2} \\ &\le \frac{1}{1-\|p\|_x} \| f'(x_+) \|_x = \frac{1}{1-\lambda} \| f'(x_+) \|_x . \end{aligned}$$

Further,

$$f'(x_+) = f'(x_+) - f'(x) - f''(x)(x_+ - x) = Gp,$$

where $G = \int_0^1 [f''(x + \tau p) - f''(x)] d\tau$. Therefore

$$\| f'(x_+) \|_x^2 = \langle [f''(x)]^{-1} Gp, Gp \rangle \le \| H \|^2 \cdot \| p \|_x^2,$$

where $H = [f''(x)]^{-1/2} G [f''(x)]^{-1/2}$. In view of Corollary 6.4,

$$(-\lambda + \tfrac{1}{3}\lambda^2) f''(x) \le G \le \frac{\lambda}{1-\lambda} f''(x).$$

Therefore

$$\| H \| \le \max \left\{ \frac{\lambda}{1-\lambda}, \lambda - \frac{1}{3}\lambda^2 \right\} = \frac{\lambda}{1-\lambda},$$

and we conclude that

$$\lambda_f^2(x_+) \le \frac{1}{(1-\lambda)^2} \| f'(x_+) \|_x^2 \le \frac{\lambda^4}{(1-\lambda)^4}.$$

$\Box$

Remarks:

- The region of quadratic convergence is as follows:

$$\lambda_f(x) < \bar{\lambda},$$

where $\bar{\lambda}$ is the root of the equation

$$\tfrac{\lambda}{(1-\lambda)^2} = 1$$

(then we can guarantee $\lambda_f(x_+) < \lambda_f(x)$). Clearly,

$$\bar{\lambda} = \tfrac{3-\sqrt{5}}{2} > \tfrac{1}{3}.$$

- We can solve the problem (6.9) as follows.

  1. First stage: $\lambda_f(x_k) \geq \beta$, where $\beta \in (0, \bar{\lambda})$.

     Apply the Damped Newton Method. At each iteration
     $$f(x_{k+1}) \leq f(x_k) - \omega(\beta).$$
     The number of steps is bounded:
     $$N \leq \tfrac{1}{\omega(\beta)}[f(x_0) - f(x_f^*)].$$

  2. Second stage: $\lambda_f(x_k) \leq \beta$.

     Apply the standard Newton Method. At this stage the process converges quadratically:
     $$\lambda_f(x_{k+1}) \leq \left(\tfrac{\lambda_f(x_k)}{1-\lambda_f(x_k)}\right)^2 \leq \tfrac{\beta\lambda_f(x_k)}{(1-\beta)^2} < \lambda_f(x_k).$$