Lecture 15: Self-Concordant Analysis of Newton's Method

Lecture 14 developed Newton's method from the local quadratic model. It introduced the Newton direction

$$d_f(x) = -(\nabla^2 f(x))^{-1}\nabla f(x)$$

and the Newton decrement

$$\lambda_f(x) = \sqrt{\langle \nabla f(x), (\nabla^2 f(x))^{-1}\nabla f(x)\rangle}.$$

It also proved local quadratic convergence under a Euclidean Hessian-Lipschitz assumption. That theorem is the right local theorem near a nondegenerate minimizer, where a fixed Euclidean background geometry gives a reasonable scale for Hessian variation. It is not the right language for the open-domain geometry that appears near constraints and barriers: there, the metric itself should adapt to the current curvature.

The new idea is therefore to measure steps in the Hessian metric at the current point and to ask that this metric not change too fast in its own local units. Self-concordance is precisely this affine-invariant Hessian-stability condition. The logarithmic model will appear below as a positive example: $-\log t$ is not an exception to the theory, but the one-dimensional prototype that the theory is designed to capture.

15.1 Local Hessian Norms

Throughout this lecture, E denotes a finite-dimensional real vector space and $U \subseteq E$ is a nonempty open convex set. We use the intrinsic Hessian convention from Lecture 14:

$$\nabla^2 f(x) : E \to E^*, \qquad \left\langle \nabla^2 f(x)u, v\right\rangle = D^2 f(x)[u, v].$$

The definition of self-concordance below allows degenerate Hessians. For a twice differentiable function f, whenever we use local norms, Newton directions, or Newton decrements, we impose the additional nondegeneracy assumption that the relevant Hessian is positive definite. At such a point, the local norm and dual local norm are those of Definition 14.1:

$$\|h\|_x := \sqrt{\langle \nabla^2 f(x)h, h\rangle}, \qquad \|g\|_{x,*} := \sqrt{\langle g, (\nabla^2 f(x))^{-1}g\rangle}.$$

Thus the Newton decrement from Lecture 14 can be read as

$$\lambda_f(x) = \|\nabla f(x)\|_{x,*} = \|d_f(x)\|_x\,.$$

So $\lambda_f(x)$ is not a new object: it is exactly the local Hessian length of the full Newton step.

15.2 Self-Concordance

The fixed Euclidean smoothness condition from Lecture 14 controls Hessian variation in a fixed background norm. For barrier-like functions, that is the wrong scale. The natural target is instead:

If one moves by local length at most one, then the Hessian should change by at most a constant factor.

Here is the differential condition that leads to this target. Fix a direction u, move along the line

$$x_t := x + tu,$$

and track the local length of this same direction:

$$q(t) := \|u\|_{x_t}.$$

If the local length changes at bounded logarithmic speed per unit of local distance, then with $ds = q(t)\,dt$ we want

$$\left|\frac{d}{ds}\log q(t)\right| \le 1.$$

Equivalently,

$$|q'(t)| \le q(t)^2.$$

Since $q(t)^2 = D^2 f(x_t)[u,u]$, formally differentiating gives

$$2q(t)q'(t) = D^3 f(x_t)[u,u,u].$$

Thus a convenient pointwise way to enforce the desired local Hessian stability is exactly

$$\left|D^3 f(x)[h,h,h]\right| \le 2\big(D^2 f(x)[h,h]\big)^{3/2}.$$

This motivates the following definition.

Definition 15.1 (Self-concordant function)**.** Let $U \subseteq E$ be open and convex. A function $f : U \to \mathbb{R}$ is called *self-concordant* if it satisfies the following conditions. We use the standard normalization with constant 2.

1. f is convex and three times continuously differentiable on U;
2. for every $x \in U$ and every $h \in E$,

$$\left|D^3 f(x)[h,h,h]\right| \le 2\big(D^2 f(x)[h,h]\big)^{3/2}.$$

Theorem 15.5 below is the formal integrated version of the target property above.

Remark 15.1 (Degenerate functions are allowed). Convexity implies $\nabla^2 f(x) \succeq 0$ when f is C^2, but Definition 15.1 does not require $\nabla^2 f(x) \succ 0$. Thus affine functions are self-concordant:

$$\nabla^2 f = 0, \qquad D^3 f = 0.$$

This convention is important later because barrier objectives have the form

$$t\,\langle c, x\rangle + \Phi(x),$$

where $c \in E^*$, the linear term contributes no curvature, and the barrier Φ provides the positive Hessian used by Newton's method.

Remark 15.2 (Fixed normalization versus barrier complexity). The constant 2 in Definition 15.1 is a fixed normalization for self-concordance. It is not a complexity parameter. The separate parameter ν appears only in Lecture 16, when a self-concordant function is also used as a barrier for a feasible set. In short:

self-concordance controls local Newton stability,

while

the barrier parameter ν controls path-following complexity and accuracy.

Example 15.1 (The logarithmic model)**.** For $\phi(t) = -\log t$ on $(0, \infty)$,

$$\phi''(t) = \frac{1}{t^2}, \qquad \phi'''(t) = -\frac{2}{t^3}.$$

Hence

$$|\phi'''(t)| = 2\big(\phi''(t)\big)^{3/2}.$$

Thus $-\log t$ is the tight one-dimensional model for the standard self-concordance normalization.

15.3 Basic Closure Properties

Proposition 15.1 (Scaling)**.** *If $f : U \to \mathbb{R}$ is self-concordant and $\alpha \geq 1$, then αf is self-concordant on U.*

Proof of Proposition 15.1. Let $g = \alpha f$. Then

$$D^2 g = \alpha D^2 f, \qquad D^3 g = \alpha D^3 f.$$

Therefore

$$\Big|D^3 g(x)[h,h,h]\Big| \leq 2\alpha\big(D^2 f(x)[h,h]\big)^{3/2} = 2\alpha^{-1/2}\big(D^2 g(x)[h,h]\big)^{3/2} \leq 2\big(D^2 g(x)[h,h]\big)^{3/2}.$$

□

Proposition 15.2 (Sums and affine perturbations)**.** *Let $f_1, \ldots, f_m : U \to \mathbb{R}$ be self-concordant on the same open convex set U. Then $f := \sum_{i=1}^m f_i$ is self-concordant on U. Moreover, adding an affine function to a self-concordant function preserves self-concordance.*

Proof of Proposition 15.2. Fix $x \in U$ and $h \in E$. Set

$$a_i := D^2 f_i(x)[h,h] \geq 0.$$

Then

$$\Big|D^3 f(x)[h,h,h]\Big| \leq \sum_{i=1}^m \Big|D^3 f_i(x)[h,h,h]\Big| \leq 2\sum_{i=1}^m a_i^{3/2}.$$

Since $a_i \geq 0$,

$$\sum_{i=1}^{m} a_i^{3/2} \leq \left(\sum_{i=1}^{m} a_i\right)^{3/2}.$$

Thus

$$\left|D^3 f(x)[h,h,h]\right| \leq 2\big(D^2 f(x)[h,h]\big)^{3/2}.$$

Convexity is immediate from the assumptions. Adding an affine function does not change the second or third derivative. □

Proposition 15.3 (Affine invariance and affine restrictions)**.** *Let $f : U \to \mathbb{R}$ be self-concordant. Let F be a finite-dimensional real vector space, let $B : F \to E$ be linear, and let $a \in E$. Define*

$$V := \{z \in F : a + Bz \in U\}, \qquad g(z) := f(a + Bz).$$

Assume V is nonempty. Then g is self-concordant on V. In particular, the restriction of a self-concordant function to an affine subspace is self-concordant on that affine subspace.

Proof of Proposition 15.3. The set V is open and convex because it is an affine inverse image of U. For every $z \in V$ and every $h \in F$,

$$D^2 g(z)[h,h] = D^2 f(a + Bz)[Bh, Bh],$$

and

$$D^3 g(z)[h,h,h] = D^3 f(a + Bz)[Bh, Bh, Bh].$$

The self-concordance inequality for f gives

$$\left|D^3 g(z)[h,h,h]\right| \leq 2\big(D^2 f(a+Bz)[Bh,Bh]\big)^{3/2} = 2\big(D^2 g(z)[h,h]\big)^{3/2}.$$

Convexity follows by composition with an affine map. □

Example 15.2 (Examples generated by closure rules)**.** The logarithmic model in Example 15.1, together with the closure rules above, gives several basic self-concordant functions.

1. On the positive orthant,
$$x \mapsto -\sum_{i=1}^{m} \log x_i, \qquad x \in \mathbb{R}^m_{++},$$
is self-concordant by the sum rule.

2. If a convex set is described by linear inequalities
$$K = \{x \in E : \langle a_i, x\rangle < b_i,\ i = 1, \ldots, m\},$$
then
$$x \mapsto -\sum_{i=1}^{m} \log\big(b_i - \langle a_i, x\rangle\big)$$
is self-concordant on K by the sum rule and affine invariance.

3. On the positive definite cone,

$$X \mapsto -\log\det X, \qquad X \in \mathbb{S}^n_{++},$$

is the matrix analogue of $-\log t$. We will use this example later; the verification uses matrix derivative identities rather than new ideas.

15.4 Hessian Comparison and Value Inequalities

Lemma 15.4 (Absolute-value estimate for symmetric multilinear forms)**.** *Let H be a finite-dimensional real Hilbert space with Euclidean norm $\|\cdot\|$. Let $T : H^k \to \mathbb{R}$ be a symmetric k-linear form, where $k \geq 1$. Suppose there is $M \geq 0$ such that*

$$|T[h,\dots,h]| \leq M\,\|h\|^k \qquad \forall h \in H.$$

Then

$$|T[u_1,\dots,u_k]| \leq M \prod_{i=1}^{k} \|u_i\| \qquad \forall u_1,\dots,u_k \in H.$$

Consequently, if $f : U \to \mathbb{R}$ is self-concordant and $\nabla^2 f(x) \succ 0$, then

$$\left|D^3 f(x)[u,v,w]\right| \leq 2\,\|u\|_x\,\|v\|_x\,\|w\|_x\,,$$

and in particular

$$\left|D^3 f(x)[u,v,v]\right| \leq 2\,\|u\|_x\,\|v\|_x^2\,.$$

Proof of Lemma 15.4. The first statement is the classical absolute-value estimate for symmetric multilinear forms over real Euclidean spaces; see [Wat90]. Apply it with the local Hilbert norm $\|h\|_x = \left(\langle \nabla^2 f(x)h, h\rangle\right)^{1/2}$, the symmetric trilinear form $T = D^3 f(x)$, and $M = 2$. The required diagonal bound is exactly Definition 15.1. □

Theorem 15.5 (Hessian comparison along a short segment)**.** *Let $f : U \to \mathbb{R}$ be self-concordant, let $x \in U$, and let $u \in E$. Assume that $x + tu \in U$ and $\nabla^2 f(x+tu) \succ 0$ for every $t \in [0,1]$, and set*

$$r := \|u\|_x\,.$$

If $r < 1$, then for every $t \in [0,1]$,

$$(1-tr)^2 \nabla^2 f(x) \preceq \nabla^2 f(x+tu) \preceq \frac{1}{(1-tr)^2}\nabla^2 f(x).$$

In particular, at $t = 1$,

$$(1-r)^2 \nabla^2 f(x) \preceq \nabla^2 f(x+u) \preceq \frac{1}{(1-r)^2}\nabla^2 f(x).$$

Proof of Theorem 15.5. First apply the one-dimensional argument to the direction u. Define

$$q(t) := \|u\|_{x+tu} = \sqrt{\langle \nabla^2 f(x+tu)u, u\rangle}.$$

Differentiating $q(t)^2 = D^2 f(x+tu)[u,u]$ gives

$$2q(t)q'(t) = D^3 f(x+tu)[u,u,u].$$

By self-concordance along the line,

$$|q'(t)| \le q(t)^2.$$

Equivalently,

$$-1 \le \frac{d}{dt}\frac{1}{q(t)} \le 1.$$

Since $q(0) = r$, this implies

$$q(t) \le \frac{r}{1-tr} \qquad 0 \le t \le 1.$$

Now fix $v \in E$ and define

$$\psi(t) := \left\langle \nabla^2 f(x+tu)v, v \right\rangle.$$

By Lemma 15.4,

$$|\psi'(t)| = \left| D^3 f(x+tu)[u,v,v] \right| \le 2\,\|u\|_{x+tu}\,\psi(t) \le \frac{2r}{1-tr}\psi(t).$$

Thus

$$-\frac{2r}{1-tr} \le \frac{d}{dt}\log\psi(t) \le \frac{2r}{1-tr}.$$

Integrating from 0 to t gives

$$2\log(1-tr) \le \log\frac{\psi(t)}{\psi(0)} \le -2\log(1-tr).$$

Exponentiating yields

$$(1-tr)^2\psi(0) \le \psi(t) \le \frac{1}{(1-tr)^2}\psi(0).$$

Since this holds for every v, the Loewner comparison follows. ☐

Remark 15.3 (Domain control is not automatic). Theorem 15.5 assumes the segment stays inside U. For an arbitrary self-concordant function this cannot be omitted. The missing statement

$$\|u\|_x < 1 \quad \Longrightarrow \quad x+u \in \operatorname{int}(K)$$

is a barrier property, not a property of self-concordance alone. Lecture 16 proves this Dikin-domain statement for self-concordant barriers.

Theorem 15.6 (Self-concordant function-value inequalities)**.** *Let $f : U \to \mathbb{R}$ be self-concordant, let $x \in U$, and let $u \in E$. Assume that $x+tu \in U$ and $\nabla^2 f(x+tu) \succ 0$ for every $t \in [0,1]$, and* *set*

$$r := \|u\|_x .$$

Define the scalar functions

$$\omega(t) := t - \log(1+t) \qquad (t > -1),$$

and

$$\omega^*(t) := -t - \log(1-t) = \omega(-t) \qquad (t < 1).$$

Then

$$f(x+u) \geq f(x) + \langle \nabla f(x), u \rangle + \omega(r).$$

If $r < 1$, then

$$f(x+u) \leq f(x) + \langle \nabla f(x), u \rangle + \omega^*(r).$$

In the rest of this lecture, ω and ω^* refer to the scalar functions introduced in Theorem 15.6. The notation ω^* is consistent with the Fenchel conjugate identity

$$\sup_{r \geq 0} \{\lambda r - \omega(r)\} = \omega^*(\lambda), \qquad 0 \leq \lambda < 1,$$

which is verified in the exercises.

Proof of Theorem 15.6. Define

$$\phi(t) := f(x+tu), \qquad t \in [0,1].$$

Then

$$\phi'(0) = \langle \nabla f(x), u \rangle, \qquad \phi''(0) = r^2.$$

The one-dimensional self-concordance argument gives

$$\phi''(t) \geq \frac{r^2}{(1+tr)^2} \qquad 0 \leq t \leq 1,$$

and, if $r < 1$,

$$\phi''(t) \leq \frac{r^2}{(1-tr)^2} \qquad 0 \leq t \leq 1.$$

Using

$$\phi(1) - \phi(0) - \phi'(0) = \int_0^1 (1-s)\phi''(s)\, ds,$$

we get

$$\phi(1) - \phi(0) - \phi'(0) \geq \int_0^1 (1-s) \frac{r^2}{(1+sr)^2}\, ds = r - \log(1+r) = \omega(r),$$

and, for $r < 1$,

$$\phi(1) - \phi(0) - \phi'(0) \leq \int_0^1 (1-s) \frac{r^2}{(1-sr)^2}\, ds = -r - \log(1-r) = \omega^*(r).$$

□

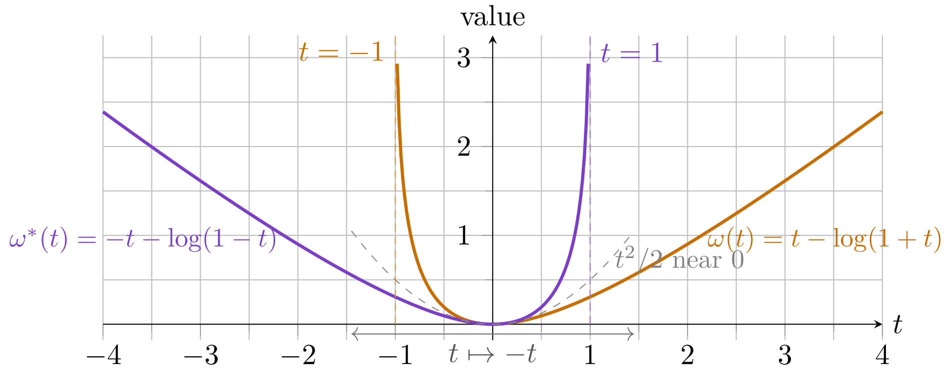

Figure 1: The scalar functions used in self-concordant Newton analysis, drawn on their full domains. They are mirror images because $\omega^*(t) = \omega(-t)$. Near the origin, both satisfy $\omega(t) = t^2/2 + O(t^3)$ and $\omega^*(t) = t^2/2 + O(t^3)$. The blow-up of ω^* as $t \uparrow 1$ matches the certificate condition $\lambda_f(x) < 1$.

15.5 Self-Concordant Analysis of Newton's Method

This subsection upgrades the Lecture 14 Newton analysis. The Newton direction and decrement are exactly the Lecture 14 objects; the difference is that the proof now uses self-concordant local geometry rather than a Euclidean Hessian-Lipschitz constant.

For the Newton update results, we use the following *domain-safety assumption*: whenever $x \in U$ and $\nabla^2 f(x) \succ 0$,

$$\{x + u : \|u\|_x < 1\} \subseteq U. \tag{DS}$$

This assumption is automatic when $U = E$. It is not a consequence of self-concordance alone; Lecture 16 will prove it for self-concordant barriers as the Dikin-ellipsoid inclusion. Each individual Newton theorem below only needs this condition at the point where the update is taken; the phase-switch theorem assumes it along the generated iterates.

The damped Newton step in Theorem 15.7 is not chosen arbitrarily. It is the minimizer of the self-concordant upper model along the Newton direction.

Theorem 15.7 (Damped Newton decrease)**.** *Let $f : U \to \mathbb{R}$ be self-concordant and assume $\nabla^2 f(z) \succ 0$ for every $z \in U$. Assume* (DS). *Let $x \in U$. Let*

$$d := d_f(x), \qquad \lambda := \lambda_f(x)$$

be the Newton direction and decrement from Definition 14.2. Define

$$x^+ := x + \frac{1}{1+\lambda} d.$$

Then

$$f(x^+) \le f(x) - \omega(\lambda).$$

Proof of Theorem 15.7. If $\lambda = 0$, then $d = 0$, $x^+ = x$, and the claim is immediate. Assume $\lambda > 0$. First consider a trial damped step $x + \alpha d$, where $0 \le \alpha\lambda < 1$. By (DS), the segment $x + s\alpha d$ lies in U for every $0 \le s \le 1$. Hence Theorem 15.6 gives

$$\begin{aligned} f(x+\alpha d) - f(x) &\le \alpha \langle \nabla f(x), d\rangle + \omega^*(\alpha\lambda) \\ &= -\alpha\lambda^2 + \omega^*(\alpha\lambda). \end{aligned}$$

Here $\langle \nabla f(x), d\rangle = -\lambda^2$ because d is the Newton direction. Write $r := \alpha\lambda$. The upper bound becomes

$$-\lambda r + \omega^*(r), \qquad 0 \le r < 1.$$

Since ω^* is convex and

$$(\omega^*)'(r) = \frac{r}{1-r},$$

the minimizing value satisfies

$$r = \frac{\lambda}{1+\lambda}, \qquad \alpha = \frac{1}{1+\lambda}.$$

Thus the damped Newton step $x^+ = x + \alpha d$ is the minimizer of this one-dimensional upper bound. Substituting this value gives

$$f(x^+) - f(x) \le \omega^*\left(\frac{\lambda}{1+\lambda}\right) - \frac{\lambda^2}{1+\lambda} = -\frac{\lambda}{1+\lambda} + \log(1+\lambda) - \frac{\lambda^2}{1+\lambda} = -\lambda + \log(1+\lambda) = -\omega(\lambda).$$

□

Theorem 15.8 (Decrement controls objective gap)**.** *Let $f : U \to \mathbb{R}$ be self-concordant and assume $\nabla^2 f(z) \succ 0$ for every $z \in U$. Assume also that f attains its minimum at some $x^\star \in U$. Let $x \in U$ and assume $\lambda_f(x) < 1$. Then*

$$f(x) - f(x^\star) \le \omega^*(\lambda_f(x)).$$

Proof of Theorem 15.8. Let $g := \nabla f(x)$. For any $y \in U$, the segment from x to y lies in U because U is convex. The lower self-concordant inequality gives

$$f(y) \ge f(x) + \langle g, y-x\rangle + \omega(\|y-x\|_x).$$

By duality of the local norms,

$$\langle g, y-x\rangle \ge -\|g\|_{x,*}\,\|y-x\|_x = -\lambda_f(x)\,\|y-x\|_x\,.$$

Therefore, with $r := \|y-x\|_x$,

$$f(y) \ge f(x) + \omega(r) - \lambda_f(x) r.$$

For $\lambda \in [0,1)$,

$$\sup_{r \ge 0} \{\lambda r - \omega(r)\} = \omega^*(\lambda).$$

Hence

$$f(y) \ge f(x) - \omega^*(\lambda_f(x)).$$

Taking $y = x^\star$ proves the result. □

Theorem 15.9 (Full-step decrement recursion)**.** *Let $f : U \to \mathbb{R}$ be self-concordant and assume $\nabla^2 f(z) \succ 0$ for every $z \in U$. Assume (DS). Let $x \in U$, write*

$$\lambda := \lambda_f(x), \qquad d := d_f(x),$$

and assume $\lambda < 1$. Let $x^+ := x + d$. Then

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

In particular, if $\lambda \le 1/4$, then

$$\lambda_f(x^+) \le 2\lambda^2.$$

Proof of Theorem 15.9. If $\lambda = 0$, then $\nabla f(x) = 0$, so $d = 0$, $x^+ = x$, and the claim is immediate. Assume $\lambda > 0$. Since $\|d\|_x = \lambda < 1$, (DS) implies that the segment $x + sd$, $0 \le s \le 1$, lies in U. Write

$$H_s := \nabla^2 f(x + sd), \qquad H_0 := \nabla^2 f(x), \qquad g_+ := \nabla f(x^+).$$

Since d is the Newton direction,

$$\nabla f(x) + H_0 d = 0.$$

Thus, for any $v \in E$,

$$\langle g_+, v\rangle = \int_0^1 \langle (H_s - H_0)d, v\rangle \, ds.$$

Expanding $H_s - H_0$ along the same segment gives

$$\langle g_+, v\rangle = \int_0^1 \int_0^s D^3 f(x + \tau d)[d, d, v] \, d\tau \, ds = \int_0^1 (1-\tau) D^3 f(x + \tau d)[d, d, v] \, d\tau.$$

By Theorem 15.5, for $0 \le \tau \le 1$,

$$\|d\|_{x+\tau d} \le \frac{\lambda}{1 - \tau\lambda}, \qquad \|v\|_{x+\tau d} \le \frac{\|v\|_x}{1 - \tau\lambda}.$$

Using Lemma 15.4, symmetry of $D^3 f$, and the last two estimates,

$$\begin{aligned} |\langle g_+, v\rangle| &\le \int_0^1 (1-\tau) 2 \|d\|_{x+\tau d}^2 \|v\|_{x+\tau d} \, d\tau \\ &\le 2\lambda^2 \|v\|_x \int_0^1 \frac{1-\tau}{(1-\tau\lambda)^3} \, d\tau. \end{aligned}$$

A direct integration gives

$$2\int_0^1 \frac{1-\tau}{(1-\tau\lambda)^3} \, d\tau = \frac{1}{1-\lambda}.$$

Therefore

$$|\langle g_+, v\rangle| \le \frac{\lambda^2}{1-\lambda} \|v\|_x \qquad \forall v \in E,$$

which is exactly

$$\|g_+\|_{x,*} \le \frac{\lambda^2}{1-\lambda}.$$

Finally, Theorem 15.5 at $t = 1$ gives

$$H_1 \succeq (1-\lambda)^2 H_0.$$

Hence $H_1^{-1} \preceq (1-\lambda)^{-2} H_0^{-1}$, so

$$\lambda_f(x^+) = \|g_+\|_{x^+,*} \le \frac{1}{1-\lambda} \|g_+\|_{x,*} \le \left(\frac{\lambda}{1-\lambda}\right)^2.$$

For $\lambda \le 1/4$,

$$\left(\frac{\lambda}{1-\lambda}\right)^2 \le \left(\frac{4}{3}\right)^2 \lambda^2 < 2\lambda^2.$$

$\square$

Theorem 15.10 (Self-concordant Newton synthesis)**.** *Let $f : U \to \mathbb{R}$ be self-concordant and assume $\nabla^2 f(z) \succ 0$ for every $z \in U$. Assume (DS). Assume that f attains its minimum at $x^\star \in U$. Starting from $x_0 \in U$, define iterates recursively as follows. Given $x_k \in U$, write*

$$d_k := d_f(x_k), \qquad \lambda_k := \lambda_f(x_k).$$

If $\lambda_k > 1/4$, set

$$x_{k+1} := x_k + \frac{1}{1+\lambda_k} d_k.$$

If $\lambda_k \le 1/4$, set

$$x_{k+1} := x_k + d_k.$$

Then the recursive construction is well-defined for all k, and the damped phase is finite. More precisely, if

$$\Delta_0 := f(x_0) - f(x^\star), \qquad N_{\mathrm{damp}} := \left\lceil \frac{\Delta_0}{\omega(1/4)} \right\rceil,$$

then some iterate x_K, with $K \le N_{\mathrm{damp}}$, satisfies

$$\lambda_K \le \frac{1}{4}.$$

From that point on, full Newton steps satisfy

$$\lambda_{k+1} \le 2\lambda_k^2 \qquad k \ge K.$$

Consequently, for every $q \ge 0$,

$$\lambda_{K+q} \le \frac{1}{2}(2\lambda_K)^{2^q} \le 2^{-2^q-1},$$

and the objective gap is certified by

$$f(x_{K+q}) - f(x^\star) \le \omega^*(\lambda_{K+q}).$$

Proof of Theorem 15.10. Let

$$\bar{\lambda} := \frac{1}{4}, \qquad c := \omega(\bar{\lambda}) > 0.$$

Since $\omega'(t) = t/(1+t) \geq 0$ for $t \geq 0$, the function ω is increasing on $[0, \infty)$. Whenever $\lambda_k > \bar{\lambda}$, the algorithm takes the damped step, and Theorem 15.7 gives

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

Thus each such step decreases the objective by at least c.
Set $N_{\text{damp}} = \lceil \Delta_0 / c \rceil$. Suppose, for contradiction, that

$$\lambda_k > \bar{\lambda} \qquad 0 \leq k \leq N_{\text{damp}}.$$

Then the first $N_{\text{damp}} + 1$ steps are all damped, and repeated use of the previous inequality gives

$$f(x_{N_{\text{damp}}+1}) \leq f(x_0) - (N_{\text{damp}} + 1)c < f(x_0) - \Delta_0 = f(x^\star),$$

contradicting optimality of $x^\star$. Therefore some

$$K \leq N_{\text{damp}}$$

satisfies $\lambda_K \leq \bar{\lambda}$.
We now prove by induction that the method stays in the full-step phase and that the displayed quadratic recursion holds. Suppose $k \geq K$ and $\lambda_k \leq 1/4$. Then the algorithm takes a full step. The segment from x_k to x_{k+1} lies in U by (DS), so Theorem 15.9 applies and gives

$$\lambda_{k+1} \leq \left(\frac{\lambda_k}{1 - \lambda_k} \right)^2 \leq 2\lambda_k^2.$$

In particular, if $\lambda_k \leq 1/4$, then

$$\lambda_{k+1} \leq 2 \left(\frac{1}{4} \right)^2 = \frac{1}{8} \leq \frac{1}{4}.$$

Hence the next iterate is again in the full-step region, and the induction proves

$$\lambda_{k+1} \leq 2\lambda_k^2 \qquad k \geq K.$$

Define $a_q := 2\lambda_{K+q}$. The recursion gives

$$a_{q+1} = 2\lambda_{K+q+1} \leq 4\lambda_{K+q}^2 = a_q^2,$$

and $a_0 = 2\lambda_K \leq 1/2$. Therefore

$$a_q \leq a_0^{2^q} \leq 2^{-2^q},$$

which is exactly

$$\lambda_{K+q} \leq \frac{1}{2}(2\lambda_K)^{2^q} \leq 2^{-2^q - 1}.$$

Since $\lambda_{K+q} < 1$, Theorem 15.8 gives

$$f(x_{K+q}) - f(x^\star) \leq \omega^*(\lambda_{K+q}).$$

□

Remark 15.4 (Barrier-safe reading). The domain-safety assumption (DS) is explicit because self-concordance alone does not imply that the local unit ball lies inside U; see Remark 15.3. In the barrier applications of Lecture 16, Dikin inclusion supplies this assumption automatically: the damped step has local length $\lambda_k/(1+\lambda_k) < 1$, and every full step in the quadratic phase has local length at most $1/4$.

15.6 Equality-Constrained Newton as Affine-Slice Newton

Barrier subproblems in the next lecture are open-domain Newton problems, often with affine equalities. No new self-concordant convergence theory is needed: equality-constrained Newton is simply Newton on an affine slice.

Consider

$$\min f(x) \qquad \text{subject to } Ax = b,$$

where $A : E \to F$ is a surjective linear map between finite-dimensional real vector spaces and $b \in F$. If A is not surjective, replace F by $\operatorname{range}(A)$. Let

$$\mathcal{A} := \{x \in E : Ax = b\}, \qquad T := \ker A.$$

At a feasible point $x \in \mathcal{A}$, the equality-constrained Newton direction is the minimizer of the quadratic model over tangent directions:

$$d \in \operatorname*{arg\,min}_{h \in T} \left\{ \langle \nabla f(x), h \rangle + \frac{1}{2} \left\langle \nabla^2 f(x) h, h \right\rangle \right\}.$$

Equivalently, $d \in E$ and a multiplier increment $w \in F^*$ solve

$$\nabla^2 f(x) d + \nabla f(x) + A^* w = 0, \qquad Ad = 0.$$

The direction d is unique whenever $\nabla^2 f(x)$ is positive definite on T. The restricted Newton decrement is

$$\lambda_{\mathcal{A}}(x)^2 := \left\langle \nabla^2 f(x) d, d \right\rangle = \max_{h \in T \setminus \{0\}} \frac{\langle \nabla f(x), h \rangle^2}{\langle \nabla^2 f(x) h, h \rangle}.$$

All self-concordant results above apply to f restricted to $\mathcal{A}$, because self-concordance is preserved under affine restriction by Proposition 15.3.

Dependency and proof sketch

1. Definition 15.1, Remark 15.1, and Example 15.1 introduce the local third-derivative condition, explain why affine functions are allowed, and identify $-\log t$ as the motivating one-dimensional model.
2. Propositions 15.1 to 15.3 are the closure rules that turn the one-dimensional logarithmic model into the examples in Example 15.2.
3. Lemma 15.4 and Theorems 15.5 and 15.6 form the technical calculus: self-concordance controls Hessian variation in the local norm and integrates to the scalar inequalities involving ω and ω^*.
4. Theorems 15.7 to 15.10 are the self-concordant Newton analysis. They replace the Euclidean Hessian-Lipschitz constants of Lecture 14 by affine-invariant local norm control and combine into a damped phase followed by a quadratic full-step phase.

5. The affine-slice discussion explains why Lecture 16 can use the same Newton theory for barrier subproblems with linear equality constraints.

Exercises

1. Prove directly that $f(x) = -\sum_{i=1}^{m} \log x_i$ is self-concordant on $\mathbb{R}^m_{++}$.
2. Verify the identities
$$\sup_{r \geq 0} \{\lambda r - \omega(r)\} = \omega^*(\lambda) \qquad 0 \leq \lambda < 1.$$
3. Show directly that if U is open and convex and f is self-concordant on U, then the restriction of f to $U \cap \mathcal{A}$ is self-concordant as a function on the affine space $\mathcal{A}$.

References

[Wat90] William C. Waterhouse. The absolute-value estimate for symmetric multilinear forms. *Linear Algebra and its Applications*, 128:97–105, 1990.