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

TTIC 31070 / CMSC 35470 / BUSF 36903 / STAT 31015

Convex Optimization

Prof. Zhiyuan Li

Spring 2026

From Lecture 14 to Self-Concordance

Lecture 14. Newton's method with the Newton direction d_f and decrement λ_f, analyzed under a Euclidean Hessian-Lipschitz assumption, gives local quadratic convergence.

Where this hypothesis breaks down. When the natural scale of $\nabla^2 f$ depends on position — e.g., its eigenvalues grow as x approaches the boundary of an open domain — no finite global Euclidean Hessian-Lipschitz constant ρ exists. A *fixed* background metric is the wrong yardstick.

New idea (today). Measure steps in the Hessian metric at the current point. Ask that this metric not change too fast in its own local units. **Self-concordance** is precisely this affine-invariant Hessian-stability condition.

Local Hessian Norms (Recap from Lecture 14)

At $x \in U$ with $\nabla^2 f(x) \succ 0$:

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

Newton direction and decrement. Same as Lecture 14:

$$d_f(x) = -(\nabla^2 f(x))^{-1} \nabla f(x), \qquad \lambda_f(x) = \|\nabla f(x)\|_{x,*}.$$

New interpretation. $\lambda_f(x)$ is also the *local Hessian length of the full Newton step*:

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

The decrement is not a new object — it is the Newton step measured in the moving metric.

Toward the Definition: The Target

Target slogan.

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

Differential condition. Fix u, move along $x_t := x + tu$, track $q(t) := \|u\|_{x_t}$.

Bounded logarithmic speed per unit local distance. With $ds = q(t)\,dt$:

$$\left|\frac{d}{ds}\log q(t)\right| \leq 1 \quad \Longleftrightarrow \quad |q'(t)| \leq q(t)^2.$$

Convert to a third-derivative condition. $q(t)^2 = D^2 f(x_t)[u, u]$, so $2qq' = D^3 f[u, u, u]$. Then $|q'| \leq q^2$ becomes the pointwise condition

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

Self-Concordance: The Definition

Definition 15.1 (Self-concordant function). $f : U \to \mathbb{R}$ with $U \subseteq E$ open convex is *self-concordant* if:

1. f is convex and C^3 on U;
2. for every $x \in U$ and every $h \in E$,

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

(Standard normalization with constant 2.)

In local-norm notation: $|D^3 f(x)[h, h, h]| \leq 2\|h\|_x^3$.

Integrated form coming up. The Hessian-comparison theorem (Theorem 15.5 below) is the formal integrated version of the target slogan from the previous slide.

Two Remarks on the Definition

Remark 1 (degenerate Hessians allowed). The definition does *not* require $\nabla^2 f \succ 0$. Affine functions ($\nabla^2 f = 0$, $D^3 f = 0$) are self-concordant. (Matters when we add linear terms to a self-concordant function later.)

Remark 2 (fixed normalization vs. barrier parameter). The constant 2 is a fixed normalization, not a complexity parameter. A separate parameter ν (the barrier parameter, Lecture 16) controls path-following complexity:

$$\text{self-concordance} \;\to\; \text{local Newton stability}, \quad \nu \;\to\; \text{global path complexity}.$$

Example: The Logarithmic Model

Example 15.1. For $\phi(t) = -\log t$ on $(0, \infty)$:

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

Hence

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

The standard normalization saturates exactly on $-\log t$.

Interpretation. $-\log t$ is the *tight one-dimensional model* for self-concordance — the prototype the theory was designed to capture.

Sanity check. The naive Euclidean Hessian-Lipschitz constant $\sup_t |\phi'''(t)| = \sup_t 2/t^3$ is infinite on $(0, \infty)$. Yet the dimensionless ratio $|\phi'''|/(\phi'')^{3/2} = 2$ is bounded.

Self-concordance is the right scaling. Hessian variation is bounded in f's own units, even when no Euclidean Hessian-Lipschitz constant exists.

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.

Proposition 15.2 (Sums + affine perturbations). If $f_1, \ldots, f_m : U \to \mathbb{R}$ are self-concordant on the same U, then $\sum_{i=1}^{m} f_i$ is self-concordant on U. Adding an affine function preserves self-concordance.

Proposition 15.3 (Affine invariance + restriction). Let f be self-concordant on U, $B : F \to E$ linear, $a \in E$, $V := \{z \in F : a + Bz \in U\}$ nonempty. Then $g(z) := f(a + Bz)$ is self-concordant on V. In particular, restriction to an affine subspace is self-concordant.

Closure Proofs (Compact)

Proposition 15.1 (scaling). $D^2 g = \alpha D^2 f$, $D^3 g = \alpha D^3 f$:

$$|D^3 g| \;\leq\; 2\alpha\,(D^2 f)^{3/2} \;=\; 2\alpha^{-1/2}(D^2 g)^{3/2} \;\leq\; 2(D^2 g)^{3/2}$$

since $\alpha \geq 1 \Rightarrow \alpha^{-1/2} \leq 1$.

Proposition 15.2 (sums). Let $a_i := D^2 f_i(x)[h, h] \geq 0$. Triangle + individual SC:

$$|D^3 f[h, h, h]| \leq \sum_i |D^3 f_i[h, h, h]| \leq 2 \sum_i a_i^{3/2}.$$

Power-mean $\sum a_i^{3/2} \leq (\sum a_i)^{3/2}$ for nonneg a_i:

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

(Affine perturbations: D^2, D^3 unchanged.)

Proposition 15.3 (affine invariance). Chain rule: $D^k g(z)[h, \ldots] = D^k f(a + Bz)[Bh, \ldots]$. Apply self-concordance at $a + Bz$ with vector Bh. □

Examples Generated by Closure Rules

Example 15.2. Self-concordant functions built from $-\log t$ + closure rules:

1. **Positive orthant.** $f(x) = -\sum_{i=1}^{m} \log x_i$ on $\mathbb{R}^m_{++}$ (sum rule).
2. **Polyhedron.** On $K = \{x : \langle a_i, x\rangle < b_i,\ \forall i\}$, $f(x) = -\sum_{i=1}^{m} \log(b_i - \langle a_i, x\rangle)$ (affine inv. + sum rule).
3. **PSD cone.** $F(X) = -\log\det X$ on $\mathbb{S}^n_{++}$ (matrix analogue of $-\log t$).

Pattern. Linear pieces → logarithmic barrier → self-concordant. These reappear in Lecture 16 as the canonical interior-point barriers for linear, polyhedral, and SDP feasibility.

The $-\log\det X$ verification uses matrix derivative identities (no new self-concordant idea).

Mixed-Variable Estimate

Lemma 15.4 (Polarization of self-concordance). If T is a symmetric trilinear form on a Hilbert space with $|T[h,h,h]| \le M\|h\|^3$, then

$$|T[u,v,w]| \le M\|u\|\|v\|\|w\|.$$

In particular, if f is self-concordant and $\nabla^2 f(x) \succ 0$:

$$|D^3 f(x)[u,v,w]| \;\le\; 2\|u\|_x\|v\|_x\|w\|_x, \quad |D^3 f(x)[u,v,v]| \;\le\; 2\|u\|_x\|v\|_x^2.$$

Interpretation. The diagonal cubic bound "polarizes" to a bound with three different arguments. Lets us differentiate the Hessian along one direction while testing it against another.

Why we need it. Hessian comparison and the decrement-recursion proof differentiate things like $\langle \nabla^2 f(x+tu)v, v\rangle$ in t. The chain rule produces $D^3 f[u,v,v]$, a mixed expression.

Hessian Comparison along a Short Segment

Theorem 15.5 (Hessian comparison). f self-concordant, $x \in U$, $u \in E$ with $x + tu \in U$ and $\nabla^2 f(x + tu) \succ 0$ for $t \in [0, 1]$. Set $r := \|u\|_x$. If $r < 1$, then for every $t \in [0, 1]$:

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

Interpretation. Inside the local Hessian unit ball ($r < 1$), the Hessian only changes by a factor of $(1 - tr)^{\pm 2}$. **The metric is stable on its own unit scale.**

Caveat (domain control). The theorem assumes the segment stays in U. For a general self-concordant f, "$\|u\|_x < 1 \Rightarrow x + u \in U$" is *not* automatic — it's a separate *barrier* property (Lecture 16, Dikin inclusion).

Proof of Theorem 15.5: Hessian Comparison

Step 1 (1D along u). Let $q(t) := \|u\|_{x+tu}$. Then $q(t)^2 = D^2 f(x+tu)[u,u]$ and

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

Self-concordance gives $|q'(t)| \le q(t)^2$, i.e., $|(1/q)'| \le 1$. With $q(0) = r$:

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

Step 2 (test against arbitrary v). Define $\psi(t) := \langle \nabla^2 f(x+tu)v, v\rangle$. By Lemma 15.4,

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

Step 3 (Integrate logarithmic ODE). $|\frac{d}{dt}\log\psi(t)| \le \frac{2r}{1-tr}$. Integrating from 0 to t:

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

Exponentiating: $(1-tr)^2\psi(0) \le \psi(t) \le (1-tr)^{-2}\psi(0)$.
Since this holds for every v, Loewner comparison follows. $\square$

Function-Value Inequalities: ω and ω^*

Setup. Define

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

Note: $\omega^*(t) = \omega(-t)$. **Both** $\approx t^2/2$ **near** 0; ω^* blows up as $t \uparrow 1$.

Theorem 15.6 (Function-value inequalities). f self-concordant, $x \in U$, $u \in E$, $r := \|u\|_x$. Segment $x + tu \in U$ with $\nabla^2 f \succ 0$ on it. Then

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

and if $r < 1$,

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

Interpretation. Self-concordant Taylor expansion: linear term + scalar function of the local length r, replacing the Euclidean $\frac{L}{2}\|u\|^2$.

Proof of Theorem 15.6: Function-Value Inequalities

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

$$\phi'(0) = \langle \nabla f(x), u \rangle, \quad \phi''(0) = r^2, \quad \phi''(t) = D^2 f(x + tu)[u, u].$$

1D self-concordance on ϕ. Apply the Hessian-comparison ODE in dim 1:

$$\frac{r^2}{(1 + tr)^2} \le \phi''(t) \le \frac{r^2}{(1 - tr)^2} \quad (0 \le t \le 1,\ r < 1 \text{ for upper}).$$

(The lower bound holds for all $t \ge 0$ regardless of r.)

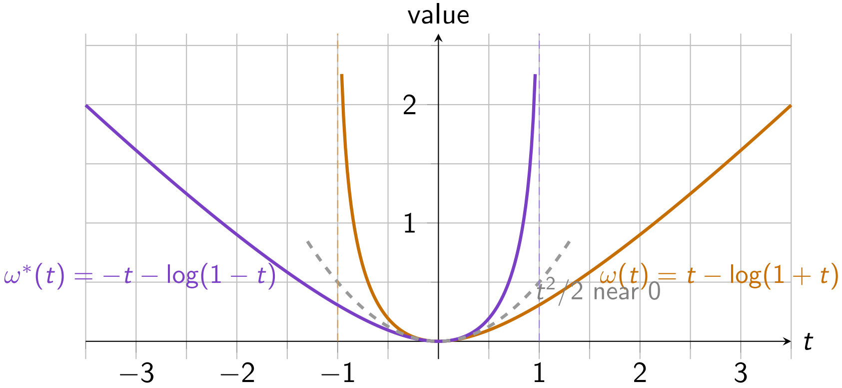
Integrate Taylor remainder. $\phi(1) - \phi(0) - \phi'(0) = \int_0^1 (1 - s)\, \phi''(s)\, ds$.

Lower bound: $\int_0^1 (1 - s) \dfrac{r^2}{(1 + sr)^2}\, ds = r - \log(1 + r) = \omega(r)$.

Upper bound ($r < 1$): $\int_0^1 (1 - s) \dfrac{r^2}{(1 - sr)^2}\, ds = -r - \log(1 - r) = \omega^*(r)$.

$\square$

ω and ω^*: Picture



$\omega^*(t) = \omega(-t)$ — mirror images. Both $\approx t^2/2$ at the origin. ω^* blows up at $t = 1$ — the local-unit-ball boundary. The certificate threshold $\lambda_f(x) < 1$ matches this blow-up.

Fenchel pair. $\sup_{r \geq 0}\{\lambda r - \omega(r)\} = \omega^*(\lambda)$ for $0 \leq \lambda < 1$ — used in the suboptimality-certificate proof.

Damped Newton Decrease

Domain-safety assumption (DS). Whenever $\nabla^2 f(x) \succ 0$:

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

Automatic if $U = E$; otherwise a barrier property (Lec 16 Dikin inclusion).

Theorem 15.7 (Damped Newton decrease). Assume self-concordance, $\nabla^2 f \succ 0$ on U, and (DS). At $x \in U$, let $\lambda := \lambda_f(x)$ and

$$x^+ := x + \frac{1}{1+\lambda}\, d_f(x).$$

Then

$$\boxed{f(x^+) \;\le\; f(x) - \omega(\lambda).}$$

Why the step size $1/(1+\lambda)$? It is the *exact* minimizer of the self-concordant upper model along the Newton direction. Larger $\lambda \Rightarrow$ shorter step.

Proof of Theorem 15.7: Damped Newton Decrease

Trial step. For $\alpha \in [0, 1/\lambda)$, consider $x + \alpha d$. Apply Theorem 15.6 upper bound with $u = \alpha d$, $r = \alpha\lambda$:

$$f(x + \alpha d) - f(x) \le \alpha\langle \nabla f(x), d\rangle + \omega^*(\alpha\lambda).$$

Plug in Newton identity. $\langle \nabla f(x), d\rangle = -\lambda^2$ (Lemma 14.2), so

$$f(x + \alpha d) - f(x) \le -\alpha\lambda^2 + \omega^*(\alpha\lambda).$$

Minimize over α. Let $r := \alpha\lambda$. Minimize $-\lambda r + \omega^*(r)$ over $r \in [0, 1)$:

$$(\omega^*)'(r) = \frac{r}{1-r} = \lambda \implies r = \frac{\lambda}{1+\lambda},\ \alpha = \frac{1}{1+\lambda}.$$

Plug back. At $r = \lambda/(1 + \lambda)$:

$$\begin{aligned} \omega^*\Big(\tfrac{\lambda}{1+\lambda}\Big) - \tfrac{\lambda^2}{1+\lambda} &= -\tfrac{\lambda}{1+\lambda} + \log(1+\lambda) - \tfrac{\lambda^2}{1+\lambda} \\ &= -\lambda + \log(1+\lambda) = -\omega(\lambda). \end{aligned}$$

□

Decrement Certifies Suboptimality

Theorem 15.8 (Suboptimality certificate). Assume self-concordance, $\nabla^2 f \succ 0$ on U, and f attains its min at $x^\star \in U$. If $\lambda_f(x) < 1$, then

$$\boxed{f(x) - f(x^\star) \;\le\; \omega^*(\lambda_f(x)).}$$

Takeaway. When $\lambda < 1$, the decrement is a *true* suboptimality certificate (no Lecture 14 caveat). $\omega^*(\lambda) \approx \lambda^2/2$ for small λ.

Reminder (Lecture 14). Without self-concordance, $\lambda^2/2$ was only the *model gap*, not a true certificate. Self-concordance is what turns the local quadratic model into a controlled bound on the original objective.

Proof of Theorem 15.8: Suboptimality Certificate

Lower bound at any y. Theorem 15.6 (lower) on the segment $[x, y]$:

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

Bound the linear term. Cauchy-Schwarz in local norm:

$$\langle \nabla f(x), y - x \rangle \;\geq\; -\|\nabla f(x)\|_{x,*} \|y - x\|_x \;=\; -\lambda_f(x)\, \|y - x\|_x.$$

Combine. With $r := \|y - x\|_x$ and $\lambda := \lambda_f(x)$:

$$f(y) \geq f(x) + \omega(r) - \lambda r, \quad \forall y \in U,\; r = \|y - x\|_x.$$

Take sup over $r \geq 0$. Fenchel identity $\sup_{r \geq 0}\{\lambda r - \omega(r)\} = \omega^*(\lambda)$ for $0 \leq \lambda < 1$:

$$f(y) \geq f(x) - \omega^*(\lambda).$$

Set $y = x^\star$. □

Heart. Self-concordance makes the lower model coercive in r; the gap is bounded by the Legendre dual of ω.

Full-Step Decrement Recursion

Theorem 15.9 (Full-step recursion). Assume self-concordance, $\nabla^2 f \succ 0$ on U, and (DS). At $x \in U$ with $\lambda := \lambda_f(x) < 1$, set $x^+ := x + d_f(x)$. Then

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

In particular, $\lambda \le 1/4 \Rightarrow \lambda_f(x^+) \le 2\lambda^2$.

Doubling-precision-per-step in the local Hessian metric. Below the threshold $\lambda \le 1/4$, the decrement squares (up to a constant 2).

Affine-invariant. Unlike Lecture 14's quadratic convergence rate (ρ/μ), this rate involves *no* explicit constants: the local Hessian metric carries all the information.

Proof of Theorem 15.9 (1/2): Residual Bound

Setup. $g_+ := \nabla f(x^+)$, $H_s := \nabla^2 f(x + sd)$. Newton: $\nabla f(x) + H_0 d = 0$.

Express g_+ as a double integral of $D^3 f$. For any $v \in E$:

$$\langle g_+, v\rangle = \int_0^1 \langle (H_s - H_0)d, v\rangle \, ds = \int_0^1 (1-\tau)\, D^3 f(x+\tau d)[d,d,v]\, d\tau.$$

Apply Lemma 15.4 + Theorem 15.5. With $\|d\|_x = \lambda$:

$$\|d\|_{x+\tau d} \le \tfrac{\lambda}{1-\tau\lambda}, \quad \|v\|_{x+\tau d} \le \tfrac{\|v\|_x}{1-\tau\lambda}, \quad |D^3 f[d,d,v]| \le 2\|d\|_{x+\tau d}^2 \|v\|_{x+\tau d}.$$

Bound $|\langle g_+, v\rangle|$.

$$|\langle g_+, v\rangle| \le 2\lambda^2 \|v\|_x \int_0^1 \tfrac{1-\tau}{(1-\tau\lambda)^3}\, d\tau.$$

Proof of Theorem 15.9 (2/2): Closing the Bound

Compute the scalar integral.

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

Therefore

$$|\langle g_+, v\rangle| \le \frac{\lambda^2}{1-\lambda}\|v\|_x \quad \forall v, \quad \text{i.e.,} \quad \|g_+\|_{x,*} \le \frac{\lambda^2}{1-\lambda}.$$

Convert dual norm at x to dual norm at x^+. Theorem 15.5 with $t = 1$ gives

$$H_1 \succeq (1-\lambda)^2 H_0 \implies H_1^{-1} \preceq \frac{1}{(1-\lambda)^2} H_0^{-1}.$$

Hence

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

Threshold $\lambda \le 1/4$. $(\lambda/(1-\lambda))^2 \le (4/3)^2\lambda^2 < 2\lambda^2$. □

Self-Concordant Newton Synthesis

Theorem 15.10 (Two-phase synthesis). Assume self-concordance, $\nabla^2 f \succ 0$, (DS), f attains min at $x^\star$. Run

$$x_{k+1} = \begin{cases} x_k + \dfrac{d_k}{1+\lambda_k} & \lambda_k > 1/4 \quad \text{(damped)} \\ x_k + d_k & \lambda_k \le 1/4 \quad \text{(full step)} \end{cases}$$

Set $\Delta_0 := f(x_0) - f(x^\star)$ and $N_{\mathrm{damp}} := \lceil \Delta_0/\omega(1/4) \rceil$.

- Some $K \le N_{\mathrm{damp}}$ achieves $\lambda_K \le 1/4$.
- For $k \ge K$: $\lambda_{k+1} \le 2\lambda_k^2$. Hence $\lambda_{K+q} \le 2^{-2^q-1}$.
- Certificate: $f(x_{K+q}) - f(x^\star) \le \omega^*(\lambda_{K+q})$.

Total iteration count. $O(\Delta_0/\omega(1/4))$ damped + $O(\log\log(1/\varepsilon))$ full-step. **Affine-invariant, barrier-ready.**

Synthesis Proof (1/2): Damped Phase

Setup. $\bar{\lambda} := 1/4$, $c := \omega(\bar{\lambda}) > 0$. ω is increasing on $[0, \infty)$ (since $\omega'(t) = t/(1+t) \geq 0$).

Each damped iteration drops f by at least c. Whenever $\lambda_k > \bar{\lambda}$, Theorem 15.7 gives

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

Finite damped phase. Suppose, for contradiction, all of $x_0, \ldots, x_{N_{\mathrm{damp}}}$ have $\lambda_k > \bar{\lambda}$. Then

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

contradicting optimality.

Conclusion. Some $K \leq N_{\mathrm{damp}} = \lceil \Delta_0 / c \rceil$ has $\lambda_K \leq \bar{\lambda} = 1/4$. □

Synthesis Proof (2/2): Quadratic Phase

Inductive claim. If $\lambda_k \le 1/4$, then the full step is taken and $\lambda_{k+1} \le 2\lambda_k^2 \le 2 \cdot (1/4)^2 = 1/8 \le 1/4$. So the method stays in the full-step region.

Doubling recursion. Let $a_q := 2\lambda_{K+q}$. From $\lambda_{k+1} \le 2\lambda_k^2$:

$$a_{q+1} = 2\lambda_{K+q+1} \le 4\lambda_{K+q}^2 = a_q^2.$$

With $a_0 = 2\lambda_K \le 1/2$: $a_q \le a_0^{2^q} \le 2^{-2^q}$, hence

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

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

$$f(x_{K+q}) - f(x^\star) \le \omega^*(\lambda_{K+q}) \le \omega^*(2^{-2^q-1}).$$

Iterations to reach ε. Want $\lambda_{K+q} \le \sqrt{2\varepsilon}$ (so $\omega^* \le \varepsilon$). $q = \lceil \log_2 \log_2(1/\varepsilon) \rceil$ iterations suffice. □

Equality-Constrained Newton = Affine-Slice Newton

Problem. $\min f(x)$ subject to $Ax = b$, with $A : E \to F$ surjective (else replace F by range).

Let $\mathcal{A} := \{x : Ax = b\}$ and $T := \ker A$ (tangent space).

Restricted Newton direction. At $x \in \mathcal{A}$:

$$d \in \operatorname*{argmin}_{h \in T} \Big\{ \langle \nabla f(x), h \rangle + \tfrac{1}{2} \langle \nabla^2 f(x) h, h \rangle \Big\}.$$

Equivalently (KKT system for the quadratic model):

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

Restricted decrement. $\lambda_{\mathcal{A}}(x)^2 = \langle \nabla^2 f(x) d, d \rangle$.

All self-concordant results apply. Self-concordance is preserved under affine restriction (Proposition 15.1). So Newton on $\mathcal{A}$ inherits Theorems 15.7–15.10.

Slogan. Equality-constrained Newton is Newton on an affine slice. No new theory needed.

Summary

Self-concordance. $|D^3 f[h,h,h]| \le 2\|h\|_x^3$. Affine-invariant, holds for $-\log t$ tight, closed under sums + scaling + affine restriction.

Calculus.

- Hessian comparison: $(1-tr)^2 H \preceq H_{x+tu} \preceq (1-tr)^{-2} H$ for $r = \|u\|_x < 1$.
- Function-value: $f(x+u) - f(x) - \langle \nabla f, u \rangle$ trapped between $\omega(r)$ and $\omega^*(r)$.

Newton synthesis.

- Damped step ($\alpha = 1/(1+\lambda)$) drops f by $\omega(\lambda) \gtrsim \lambda^2/2$.
- $\lambda < 1$ is a true certificate: $f(x) - f^\star \le \omega^*(\lambda) \lesssim \lambda^2/2$.
- Full step ($\lambda \le 1/4$) squares the decrement: $\lambda_{k+1} \le 2\lambda_k^2$.

Total complexity (affine-invariant). $O(\Delta_0/\omega(1/4))$ damped + $O(\log\log(1/\varepsilon))$ full-step.

Next (Lecture 16). Self-concordant *barriers*: when (DS) holds automatically (Dikin inclusion), and the barrier parameter ν controls the path-following loop.

Bibliographic Notes

Yurii Nesterov

- **Yu. Nesterov & A. Nemirovski, Interior-Point Polynomial Algorithms in Convex Programming, SIAM (1994).** Original definition of self-concordance and the analysis used today.
- **Yu. Nesterov, Introductory Lectures on Convex Optimization, Kluwer (2004).** Chapter 4: a streamlined modern treatment; close in style to today's lecture.
- **S. Boyd & L. Vandenberghe, Convex Optimization, Cambridge (2004).** Chapter 9.6–9.7: self-concordant Newton, an accessible textbook version.