Lecture 19

Self-concordant functions

Instructor: Prof. Gabriele Farina (✉ gfarina@mit.edu)*

In Lecture 17, we have taken a look at the standard analysis of Newton's method. The key result we established was that the method converges doubly exponentially fast (in distance) to a local minimum of a function f if the Hessian of f is M-Lipschitz continuous and the curvature of the function around the minimum is sufficiently large (say, $\nabla^2 f(x_\star) \succeq \mu I$). However, this analysis has two shortcomings.

■ **Inability to handle log functions.** The first shortcoming is practical. The analysis we gave in Lecture 17 breaks down for functions that "shoot to infinity" at the boundary of their domain. This is the case—for example—of functions such as $f(x) = x - \log(x)$, $f(x) = -\log(x) - \log(1 - x)$, *et cetera*. These functions have enormous applications in optimization, but their Hessian is *not* Lipschitz continuous. Yet, as we show next, Newton's method converges to the minimum at a double exponential rate.

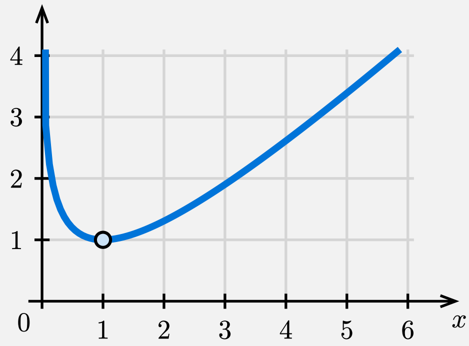
Example L19.1.

Consider the function $f : (0, \infty) \to \mathbb{R}$

$$f(x) = x - \log(x),$$

whose minimum is at $x = 1$. The update of Newton's method is

$$\begin{aligned} x_{t+1} &= x_t - [f''(x_t)]^{-1} f'(x_t) \\ &= x_t - x_t^2 \cdot \left(1 - \frac{1}{x_t}\right) = 2x_t - x_t^2. \end{aligned}$$



So, we have $1 - x_{t+1} = 1 + x_t^2 - 2x_t = (1 - x_t)^2$.

This shows that not only does Newton's method converge to the minimum at a quadratic rate, but it does so with *equality*. One can check very easily that changing the parameterization from x to kx also rescales the radius of convergence by k.

The above hints to the fact that requiring Lipschitz continuity of the Hessian is not the most natural condition for studying the quadratic convergence of Newton's method. Of course, if

*These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

Lipschitz continuity has to go, one has to wonder: "*what other condition ensuring smoothness can we impose?*" We will propose one later today, called *self-concordance*.

■ **Lack of affine invariance in the radius of fast convergence.** The second shortcoming is *conceptual*. For $\eta = 1$, Newton's method repeatedly moves to the minimum of the quadratic approximation of f around the current iterate. This point is independent of the choice of coordinates used to parameterize the space, and only depends on the function itself. Hence, the radius of fast convergence is also independent of the choice of coordinates.

However, the radius of fast convergence we gave in Lecture 17 was *not* affine invariant. For example, consider a function $f(x)$. If we now make a transformation of coordinates and consider the function $f(Ax)$, the constants M and μ change, and in general we arrive at a different radius of fast convergence.

Part of the issue is that even the requirement $\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\|_s \le M\|x-y\|_2$ is *not* affine invariant, in that reparameterization of x and y lead to a different constant. Ideally, we would like to impose a condition that is fully affine invariant. Self-concordance satisfies this *desideratum*.

L19.1 Self-concordant functions

With these two shortcomings in mind, we introduce the concept of *self-concordant functions*, which are a class of strictly convex functions (in particular, with positive definite Hessians) that are smooth and have a well-behaved Hessian. The definition provides an affine-invariant condition focused on *bounding the approximation error of the second-order expansion of the function around each point*.

In other words, instead of focusing on the Lipschitz continuity of the Hessian, we focus on the *quality* of the approximation of the function by the second-order Taylor expansion around each point.

L19.1.1 Intrinsic norms

Before we can give a meaningful quantification of the approximation error of the second-order expansion in our setting, special care must be taken to describe the radius around which the approximation is good in an affine-invariant way. Furthermore, one needs to use care close to the boundary of the domain. In the theory of self-concordant functions, both of these issues are elegantly resolved by moving away from the Euclidean norm and using instead the notion of *intrinsic norm at a point x* in the domain of f, defined as.

$$\boxed{\|v\|_x := \sqrt{\langle v, \nabla^2 f(x) v\rangle}}\,.$$

The name *intrinsic* norm comes from the following two important facts:

- the intrinsic norm is affine-invariant, in that if all points x are now replaced with $x = Ax'$, the vectors v with Av', and the function $g(x') := f(Ax')$ is introduced, then

$$\|v'\|_{x'} = \sqrt{\langle v', \nabla^2 g(x') v'\rangle} = \sqrt{\langle v', A^\top \nabla^2 f(x) A v'\rangle} = \sqrt{\langle Av', \nabla^2 f(x) A v'\rangle} = \|v\|_x.$$

- the intrinsic norm is insensitive to the choice of reference inner product $\langle \cdot, \cdot \rangle$ for the space. This is a consequence of the previous point, since all inner products are equivalent up to change of coordinates.

L19.1.2 Definition of self-concordance

We are now ready to define self-concordance. For this lecture, we will in particular focus on the notion of *strong nondegenerate* self-concordance, which is a stronger version of the definition, and is the definition that plays the central role in the theory of interior point methods.

Definition L19.1. Let $\Omega \subseteq \mathbb{R}^n$ be open and convex. A twice-differentiable function $f : \Omega \to \mathbb{R}$ is said to be *strongly nondegenerate self-concordant* if:

1. The Hessian of f is positive definite everywhere on Ω;
2. The ellipsoid $W(x) := \{y \in \mathbb{R}^n : \|y - x\|_x^2 < 1\}$ is contained in Ω for all $x \in \Omega$; and
3. Inside of the ellipsoid $W(x)$, the function f is *almost quadratic*:

$$(1 - \|y - x\|_x)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq \frac{1}{(1 - \|y - x\|_x)^2} \nabla^2 f(x) \qquad \forall x \in \Omega, y \in W(x),$$

which is equivalent to the statement

$$(1 - \|y - x\|_x)\|v\|_x \leq \|v\|_y \leq \frac{\|v\|_x}{1 - \|y - x\|_x} \qquad \forall x \in \Omega, y \in W(x), v \in \mathbb{R}^n.$$

The following equivalent characterization is often used too.

Theorem L19.1. If $f : \Omega \to \mathbb{R}$ is three-time differentiable with positive definite Hessian everywhere on Ω, then strong nondegenerate self-concordance is equivalent to f satisfying the following two properties:

1. for any $x_0 \in \Omega$ and $d \in \mathbb{R}^n$, the restriction $\varphi(\gamma) := f(x_0 + \gamma d)$ of f to the segment $\{\gamma : x_0 + \gamma d \in \Omega\}$ satisfies
$$\varphi'''(\gamma) \leq 2\varphi''(\gamma)^{3/2}; \qquad \text{and}$$
2. any sequence $\{x_k\}$ converging to a point on the boundary of Ω is such that $f(x_k) \to +\infty$.

Remark L19.1. In this lecture, we will use the term *self-concordant* to mean *strongly nondegenerate self-concordant*. In different contexts, self-concordance without qualifications refers to only condition 1 in Theorem L19.1.

While not a proof, it is easy to see how the definition of self-concordance in Definition L19.1, for univariate functions, implies the first item of Theorem L19.1. In particular, for a univariate function Definition L19.1 implies that for any $t > 0$ sufficiently small

$$\phi''(x+t) \leq \frac{1}{\left(1-\sqrt{\phi''(x)}t\right)^2}\phi''(x).$$

Subtracting $\phi''(x)$ from both sides and dividing by t yields

$$\frac{\phi''(x+t)-\phi''(x)}{t} \leq \frac{2(\phi''(x))^{3/2}-(\phi''(x))^2 t}{\left(1-\sqrt{\phi''(x)}t\right)^2} \leq 2(\phi''(x))^{3/2}.$$

Taking a limit as $t \downarrow 0$ then yields $\phi'''(x) \leq 2(\phi''(x))^{3/2}$ as claimed. The other direction is slighlty more involved but still elementary.

L19.1.3 Notable example: the log function

As a sanity check, let's show that the function $f(x) = -\log(x)$ on the domain $\Omega := (0, \infty)$ is self-concordant. We will show the multidimensional version of this result.

Example L19.2. The function $f(x) = -\sum_{i=1}^n \log(x_i)$ on the domain $\Omega := \mathbb{R}^n_{>0}$ is self-concordant.

Solution. The function f is twice differentiable with positive definite Hessian

$$\nabla^2 f(x) = \operatorname{diag}\left(\frac{1}{x_1^2}, ..., \frac{1}{x_n^2}\right) \succ 0 \qquad \forall x \in \Omega = \mathbb{R}^n_{>0}.$$

The ellipsoid $W(x)$ is therefore equivalent to

$$W(x) = \left\{y \in \mathbb{R}^n : \sum_{i=1}^n \left(\frac{y_i - x_i}{x_i}\right)^2 < 1\right\} = \left\{y \in \mathbb{R}^n : \sum_{i=1}^n \left(\frac{y_i}{x_i} - 1\right)^2 < 1\right\}.$$

It is immediate to check that the condition implies that $y_i > 0$ for all $i = 1, ..., n$. So, $W(x) \subseteq \Omega$. We now check the third condition. We will try to relate $\|v\|_y^2$ to $\|v\|_x^2$, by observing that

$$\|v\|_y^2 = \sum_{i=1}^n \left(\frac{v_i}{y_i}\right)^2 = \sum_{i=1}^n \left(\frac{v_i}{x_i}\right)^2 \left(\frac{x_i}{y_i}\right)^2.$$

In particular, the previous equality implies immediately that

$$\|v\|_x^2 \min_i \left(\frac{x_i}{y_i}\right)^2 \leq \|v\|_y^2 \leq \|v\|_x^2 \max_i \left(\frac{x_i}{y_i}\right)^2. \tag{1}$$

We will now use the triangle inequality to bound the minimum and maximum of $\frac{x_i}{y_i}$. Specifically, we have that, for all $i \in \{1, ..., n\}$,

$$\frac{y_i}{x_i} \le 1 + \left|\frac{y_i}{x_i} - 1\right| \le 1 + \sqrt{\sum_{i=1}^{n}\left(\frac{y_i}{x_i} - 1\right)^2} = 1 + \|y - x\|_x$$
$$\frac{y_i}{x_i} \ge 1 - \left|\frac{y_i}{x_i} - 1\right| \ge 1 - \sqrt{\sum_{i=1}^{n}\left(\frac{y_i}{x_i} - 1\right)^2} = 1 - \|y - x\|_x.$$

Taking a minimum over $i = 1, ..., n$, we then get

$$\min_i \left(\frac{x_i}{y_i}\right)^2 \ge \frac{1}{(1 + \|y - x\|_x)^2}, \qquad \max_i \left(\frac{x_i}{y_i}\right)^2 \le \frac{1}{(1 - \|y - x\|_x)^2}.$$

Plugging the previous bounds into (1), we find

$$\frac{\|v\|_x^2}{(1 + \|y - x\|_x)^2} \le \|v\|_y^2 \le \frac{\|v\|_x^2}{(1 - \|y - x\|_x)^2}.$$

Finally, using the fact that $\frac{1}{1+z} \ge 1 - z$ (valid for all $z > -1$), and taking square roots,

$$(1 - \|y - x\|_x)\|v\|_x \le \|v\|_y \le \frac{\|v\|_x}{1 - \|y - x\|_x}.$$

□

L19.1.4 Other examples

Other examples of self-concordant functions include the following:

- $-\log\det X$ on the set of positive definite matrices X
- $-\log(1 - \|Ax\|_2^2)$ on the ellipsoid defined by $A \succ 0$
- Quadratic functions $x^\top Ax$, with $A \succ 0$, on $\mathbb{R}^n$

L19.1.5 Composition properties of self-concordant functions

In general, checking whether a function is self-concordant is a nontrivial task. Usually, one does not use the definition; rather, the following composition rules are used.

i) (Sum of self-concordant function) The set of self-concordant functions is closed under addition.

> **Theorem L19.2.** Let $f_1 : \Omega_1 \to \mathbb{R}$ and $f_2 : \Omega_2 \to \mathbb{R}$ be self-concordant functions whose domains satisfy $\Omega_1 \cap \Omega_2 \neq \emptyset$. Then, the function $f + g : \Omega_1 \cap \Omega_2 \to \mathbb{R}$ is self-concordant.

[▷ You should try to prove this!]

ii) (Addition of an affine function) Addition of an affine function to a self-concordant functions does not affect the self-concordance property, since self-concordance depends only on the Hessian of the function, and the addition of affine functions does not affect the Hessian.

Theorem L19.3. Let $f : \Omega \to \mathbb{R}$ be self-concordant function. Then, the function $g(x) := f(x) + \langle a, x\rangle + b$ is self-concordant on Ω.

iii) (Affine transformation) The composition of a self-concordant function with an injective affine transformation preserves the self-concordance.

Theorem L19.4. Let $f : \Omega \to \mathbb{R}$ be self-concordant and $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ represent an injective transformation.[1] Then, assuming $\Omega' := \{x \in \mathbb{R}^m : Ax + b \in \Omega\} \neq \emptyset$, the affinely-transformed function $g(x) := f(Ax + b)$ is self-concordant on the domain Ω'.

iv) (Consequence) Putting together the previous result together with Example L19.2, we obtain the following important corollary.

Corollary L19.1. Let $\Omega := \left\{x \in \mathbb{R}^n : a_i^\top x > b_i \text{ for all } i\right\}$ be a nonempty open polyhedral set containing no lines, where $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}$. A function of the form

$$f(x) = c^\top x - \sum_{i=1}^{m} \log(a_i^\top x - b_i),$$

where $c \in \mathbb{R}^n$, is self-concordant on Ω.

L19.2 Properties of self-concordant functions

Self-concordant functions enjoy a number of nice properties. We now scratch the surface.

L19.2.1 The "near-quadratic" nature of self-concordant functions

In point 3. of Definition L19.1, we wrote that self-concordance is effectively asking that "inside of the ellipsoid $W(x)$, the function f is *almost quadratic*". We will need to make this statement more precise.

Theorem L19.5 (Quadratic approximation bound)**.** Let $f : \Omega \to \mathbb{R}$ be self-concordant, and let q_x be the second-order Taylor expansion of f around x, that is,

$$q_x(y) := f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle y - x, \nabla^2 f(x)(y - x)\rangle.$$

Then, for all $x \in \Omega$ and $y \in W(x)$, we have

$$|f(y) - q_x(y)| \leq \frac{\|y - x\|_x^3}{3(1 - \|y - x\|_x)}.$$

[1] Injectivity is necessary to preserve the positive definiteness of the Hessian.

[2] In particular, for a univariate function $\phi(t)$ we have

Proof. We use the fundamental theorem of calculus[2] to bound the remainder of the second-order Taylor expansion as

$$f(y) - q_x(y) = \int_0^1 \int_0^a \Big\langle y - x, (\nabla^2 f(x + b(y - x)) - \nabla^2 f(x))(y - x)\Big\rangle \,\mathrm{d}b \,\mathrm{d}a.$$

The key point of Item 3 of Definition L19.1 is that the inner difference of Hessians is bounded quite nicely for self-concordant functions. In particular, since by hypothesis $y \in W(x)$, we can write

$$\begin{aligned}\nabla^2 f(x + b(y - x)) - \nabla^2 f(x) &\preceq \frac{1}{(1 - b\|y - x\|_x)^2}\nabla^2 f(x) - \nabla^2 f(x)\\ &= \left(\frac{1}{(1 - b\|y - x\|_x)^2} - 1\right)\nabla^2 f(x).\end{aligned}$$

Hence, taking absolute values, we find

$$\begin{aligned}|f(y) - q_x(y)| &\le \int_0^1 \int_0^a \left(\frac{1}{(1 - b\|y - x\|_x)^2} - 1\right)\left|\langle y - x, \nabla^2 f(x)(y - x)\rangle\right| \mathrm{d}b \,\mathrm{d}a\\ &= \|y - x\|_x^2 \int_0^1 \int_0^a \left(\frac{1}{(1 - b\|y - x\|_x)^2} - 1\right) \mathrm{d}b \,\mathrm{d}a\\ &= \|y - x\|_x^2 \int_0^1 \left(\frac{1}{\|y - x\|_x}\left[\frac{1}{1 - b\|y - x\|_x}\right]_0^a - a\right) \mathrm{d}a\\ &= \|y - x\|_x^2 \int_0^1 \left(\frac{a}{1 - a\|y - x\|_x} - a\right) \mathrm{d}a\\ &= \|y - x\|_x^3 \int_0^1 \frac{a^2}{1 - a\|y - x\|_x} \,\mathrm{d}a\end{aligned}$$

At this point, we can just bound the integral according to

$$\int_0^1 \frac{a^2}{1 - a\|y - x\|_x} \,\mathrm{d}a \le \int_0^1 \frac{a^2}{1 - \|y - x\|_x} \,\mathrm{d}a \le \frac{1}{3(1 - \|y - x\|_x)},$$

yielding the final result. □

L19.2.2 Proximity to a minimum

A neat property of self-concordant functions is that it is possible to bound the distance from the minimum of the function simply by looking at the (intrinsic) norm of the Newton direction at any point x, which for notational convenience we denote as

---

[2]
$$\phi(1) - \phi(0) - \phi'(0) - \frac{1}{2}\phi''(0) = \int_0^1 \int_0^a \phi''(b) - \phi''(0) \,\mathrm{d}b \,\mathrm{d}a.$$

It is easy to obtain the multivariate version of the theorem by simply considering the restriction of a generic function f onto the line spanned by $y - x$, that is, $\phi(t) := f(x + t(y - x))$.

$$n(x) := -[\nabla^2 f(x)]^{-1} \nabla f(x).$$

Remark L19.2. The intrinsic norm of the descent direction $\|n(x)\|_x$ at any point x is a crucial quantity to study Newton's method for self-concordant functions. It is sometimes called the *Newton decrement*, and indicated as $\lambda(x) := \|n(x)\|_x$. We will avoid the notation $\lambda(x)$ to minimize the set of notation.

In particular, if the norm is sufficiently small, then the minimum must be near. Formally, we have the following result.

Theorem L19.6. Let $f : \Omega \to \mathbb{R}$ be self-concordant. If a point $x \in \Omega$ is such that $\|n(x)\|_x \leq 1/9$, then there exists a minimizer z of f within distance

$$\|z - x\|_x \leq 3 \cdot \|n(x)\|_x.$$

Proof. For positive definite quadratics, the theorem is immediate to prove. The idea is to now use the near-quadratic nature of self-concordant functions to establish a compact region around x where a minimum must exist, and then invoke the Weierstrass theorem.

In particular, for any y in $W(x)$, we know from Theorem L19.5 that

$$\begin{aligned}
f(y) &\geq q_x(y) - \frac{\|y-x\|_x^3}{3(1-\|y-x\|_x)} \\
&= f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2}\langle y - x, \nabla^2 f(x)(y-x)\rangle - \frac{\|y-x\|_x^3}{3(1-\|y-x\|_x)} \\
&= f(x) + \left\langle [\nabla^2 f(x)]^{-1/2}\nabla f(x), [\nabla^2 f(x)]^{1/2}(y-x)\right\rangle + \frac{1}{2}\|y-x\|_x^2 - \frac{\|y-x\|_x^3}{3(1-\|y-x\|_x)}
\end{aligned}$$

Hence, using Cauchy-Schwarz,

$$f(y) \geq f(x) - \|n(x)\|_x \|y-x\|_x + \frac{1}{2}\|y-x\|_x^2 - \frac{\|y-x\|_x^3}{3(1-\|y-x\|_x)}.$$

It is easy to see by direct substitution that for any $y \in W(x)$ with $\|y - x\|_x > 3\|n(x)\|_x$, then $f(y) > f(x)$. Hence, for any we can restrict the function f to the closed ball of radius $\|y - x\|_x \leq 1/3$ (which is a subset of $W(x)$) and find a minimum z there; such a minimum is guaranteed to be a minimum of f on $W(x)$. However, since f is convex and z is in the interior of $W(x)$, then $\nabla f(z) = 0$, and therefore z is a minimizer of f on Ω as well. $\square$

The proof of Theorem L19.6 was a bit loose in our bounds. In fact, the result can be further strengthened to a larger intrinsic radius of 1/4 (instead of 1/9).

Remark L19.3. With a bit of additional work, the result in Theorem L19.6 can be strengthened as follows: If a point $x \in \Omega$ is such that $\|n(x)\|_x \leq 1/4$, then there exists a minimum z of f within distance

$$\|z - x\|_x \le \|n(x)\|_x + \frac{3\|n(x)\|_x^2}{(1 - \|n(x)\|_x)^3}.$$

L19.2.3 Existence and uniqueness of the minimum

In addition, we mention the following property.

Theorem L19.7. Let $f : \Omega \to \mathbb{R}$ be self-concordant and lower bounded. Then, f attains a unique minimum.

The "uniqueness" part of the theorem is implied directly by the strict convexity of the self-concordant functions, which is guaranteed by Item 1 of Definition L19.1. Existence on the other hand is not immediate. One pretty direct proof is to show that there must exist at least one point with $\|n(x)\|_x \le 1/9$, and then use Theorem L19.6.

L19.3 Newton's method applied to self-concordant functions

In this section, we discuss how the guarantees of Newton's method can be extended to self-concordant functions, while gaining affine invariance at the same time.

L19.3.1 Recovering the quadratic convergence rate

For self-concordant functions, the following affine-invariant guarantee can be established.

Theorem L19.8. Let $f : \Omega \to \mathbb{R}$ be self-concordant. If a point $x_t \in \Omega$ is such that $\|n(x_t)\|_{x_t} < 1$, then

$$\|n(x_{t+1})\|_{x_{t+1}} \le \left(\frac{\|n(x_t)\|_{x_t}}{1 - \|n(x_t)\|_{x_t}}\right)^2.$$

The proof involves a couple of steps but it is not particularly difficult. The key idea is to use the self-concordance property to bound the Hessian at the next iterate in terms of the Hessian at the current iterate. You can find a detailed proof in the references cited below.

Combined with the result in the previous subsection, this gives a quadratic convergence rate.

L19.4 Further readings

The short book by Renegar, J. [Ren01] and the monograph by Nesterov, Y. [Nes18] (Chapter 5) provide a comprehensive introduction to self-concordant functions and their applications in optimization.

I especially recommend the book by Renegar, J. [Ren01] for a concise yet rigorous account.

[Ren01] Renegar, J. (2001). *A Mathematical View of Interior-point Methods in Convex Optimization*. SIAM. https://doi.org/10.1137/1.9780898718812

[Nes18] Nesterov, Y. (2018). *Lectures on Convex Optimization*. Springer International Publishing. https://link.springer.com/book/10.1007/978-3-319-91578-4

Changelog

- Apr 29, 2025: Fixed a few typos (thanks Jonathan Huang!)
- May 3, 2025: Fixed two typos
- May 11, 2025: Fixed a few typos (thanks Khizer!)