In this lecture, we cover the following topics

- Properties of Self-Concordant Functions
- Minimization of Self-Concordant Functions
- (Damped) Newton Method of Self-Concordant Functions

Reference:
Nesterov, Introductory Lectures on Convex Optimization, 2004, Chapter 4.1.4 - 4.1.5

19.1 Properties of Self-Concordant Functions

Let \( f(x) \) be a standard self-concordant function, i.e. \( \forall x \in \text{dom}(f), h \in \mathbb{R}^n \)

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

i.e.

\[
\left| \frac{d^3}{dt^3} \big|_{t=0} f(x + th) \right| \leq 2 \left( \frac{d^2}{dt^2} \big|_{t=0} f(x + th) \right)^{3/2}
\]

**Definition 19.1** We define the local norm of \( h \) at \( x \in \text{dom}(f) \) as

\[
\| h \|_x = \sqrt{h^T \nabla^2 f(x) h}
\]

We state below a basic inequality without proof, for standard self-concordant function \( f \), it holds that

\[
\left| D^3 f(x)[h_1, h_2, h_3] \right| < 2 \| h_1 \|_x \cdot \| h_2 \|_x \cdot \| h_3 \|_x
\]

**Remark** (“Lipschitz continuity”) at a high level,

\[
\left| \frac{d}{dt} \big|_{t=0} D^2 f(x + t\delta)[h, h] \right| \leq 2 \| \delta \|_x D^2 f(x)[h, h]
\]

The second derivative is relatively Lipschitz continuous w.r.t. the local norm defined by \( f \).
For instance, when \( f \) is self-concordant on \( \mathbb{R} \) and strictly convex,

\[
\frac{|f'''(x)|}{|f''(x)|^{3/2}} \leq 1 \implies \left| \frac{d}{dx} \left[ f''(x)^{-1/2} \right] \right| \leq 1
\]

\[
\implies -y \leq \int_0^y \frac{d}{dx} \left[ f''(x) \right]^{-1/2} dx \leq y
\]

\[
\implies -y \leq \frac{1}{\sqrt{f''(y)}} - \frac{1}{\sqrt{f''(0)}} \leq y
\]

Simplifying the above terms, we arrive at

\[
\frac{f''(0)}{(1 + y\sqrt{f''(0)})^2} \leq f''(y) \leq \frac{f''(0)}{(1 - y\sqrt{f''(0)})^2}, \quad \forall 0 \leq y < \sqrt{f''(0)}
\]

Note that the Hessian is nearly proportional to \( f''(0) \) around a neighborhood of \( x = 0 \).

Moreover, if we integrate (\( \ast \)) and integrate again on both sides (e.g. lower bound):

\[
\frac{|f'''(x)|}{|f''(x)|^{3/2}} \leq 1 \implies \sqrt{f''(0)} - \frac{\sqrt{f''(0)}}{1 + y\sqrt{f''(0)}} \leq f'(y) - f'(0)
\]

\[
\implies y\sqrt{f''(0)} - \ln(1 + y\sqrt{f''(0)}) \leq f(y) - f(0) - f'(0)y
\]

Rewriting the above equations and considering both sides, we arrive at: \( \forall 0 \leq y < \sqrt{f''(0)} \)

\[
\frac{(y\sqrt{f''(0)})^2}{1 + y\sqrt{f''(0)}} \leq y(f'(y) - f'(0)) \leq \frac{(y\sqrt{f''(0)})^2}{1 - y\sqrt{f''(0)}}
\]

and

\[
y\sqrt{f''(0)} - \ln(1 + y\sqrt{f''(0)}) \leq f(y) - f(0) - f'(0)y \leq -y\sqrt{f''(0)} - \ln(1 - y\sqrt{f''(0)})
\]

More generally, for self-concordant functions in \( \mathbb{R}^n \), similar results hold (see, Nesterov, 2004 Theorem 4.1.6-4.1.8). We provide the results below without providing the proofs.

**Definition 19.2 (Dikin Ellipsoid)**

\[
W_r(x) = \{ y : \| y - x \|_2 \leq 1 \}
\]

\[
W^0_r(x) = \{ y : \| y - x \|_2 < 1 \}
\]

**Theorem 19.3** For \( x \in \text{dom}(f) \), we have \( W^0_r(x) \subseteq \text{dom}(f) \)

The Dikin ellipsoid of any point in the domain is contained in the domain. More critically, the self-concordant function behaves nicely within this Dikin ellipsoid.
Theorem 19.4 (Hessian of self-concordant function) For $x \in \text{dom}(f)$, we have

$$(1 - \| y - x \|_x)^2 \nabla^2 f(x) \preceq \nabla^2 f(y) \preceq (1 - \| y - x \|_x)^{-2} \nabla^2 f(x), \quad \forall y \in W_1^0(x)$$

Theorem 19.5 (Gradient of self-concordant function) For $x \in \text{dom}(f)$, we have

$$\frac{\| y - x \|_x^2}{1 + \| y - x \|_x} \leq \langle \nabla f(y) - \nabla f(x), y - x \rangle \leq \frac{\| y - x \|_x^2}{1 - \| y - x \|_x}, \quad \forall y \in W_1^0(x)$$

Theorem 19.6 (Linear approximation of self-concordant function)

1. $\forall x, y \in \text{dom}(f)$, we have

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

2. $\forall x \in \text{dom}(f), y \in W_2^0(1), \text{ we have}$

$$f(y) \leq f(x) + \langle f(x), y - x \rangle + \omega_s(\| y - x \|_x)$$

where $\omega(t) = t - \ln(1 + t)$ and $\omega_s(t) = -t - \ln(1 - t)$ is the conjugate

Proof: See Theorem 4.1.6, 4.1.7, 4.1.8 in (Nesterov, 2004).

19.2 Minimizing Self-Concordant Functions

Consider the unconstrained minimization

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

where $f$ is a standard self-concordant and non-degenerate (i.e. $\nabla^2 f(x) \succeq 0$) function. Note that the problem is not necessarily solvable, e.g. $f(x) = -\ln(x)$.

Definition 19.7 (Newton’s decrement) The quantity

$$\lambda_f(x) = \sqrt{\nabla f(x)[\nabla^2 f(x)]^{-1} \nabla f(x)}$$

is called Newton decrement.

Remark: The Newton decrement can be interpreted as

1. The decrease of the second order Taylor expansion after a Newton step:

$$f(x) - \min_h \left\{ f(x) + h^T \nabla f(x) + \frac{1}{2} h^T \nabla^2 f(x) h \right\} = \frac{1}{2} \lambda_f^2(x)$$
2. The conjugate local norm of $\nabla f(x)$:

$$\|\nabla f(x)\|_{x,*} = \max \{ \nabla f(x)^T y, \| y \|_x \leq 1 \} = \| \nabla^2 f(x)^{-1/2} \nabla f(x) \|_2 = \lambda_f(x)$$

3. The local norm of Newton’s direction $d(x) = -\nabla^2 f(x)^{-1} \nabla f(x)$:

$$\| d(x) \|_x = \lambda_f(x)$$

**Theorem 19.8** Assume $\lambda_f(x_0) < 1$ for some $x_0 \in \text{dom}(f)$. Then there exists a unique minimizer of $f$.

**Proof:** It suffices to show that the level set $\{ y : f(y) \leq f(x_0) \}$ is bounded. Since

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

$$\geq f(x) - \| \nabla f(x) \|_{x,*} \cdot \| y - x \|_x + \omega(\| y - x \|_x)$$

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

we have

$$f(y) \leq f(x_0) \implies \frac{\omega(\| y - x_0 \|_{x_0})}{\| y - x_0 \|_{x_0}} \leq \lambda_f(x_0) < 1$$

Notice the function $\phi(t) = \frac{\omega(t)}{t} = 1 - \frac{1}{t} \ln(1+t)$ is strictly increasing in $t \geq 0$. Hence, $\| y - x_0 \|_{x_0} \leq t^*$ for some $t^*$. This implies that the level set must be bounded.

**Remark:** Note that for self-concordant functions, local condition such as $\lambda_f(x_0) < 1$ provides some global information on $f$.

**Example:** consider the self-concordant function $f(x) = \varepsilon x - \ln(x)$ with $\text{dom}(f) := \{ x : x > 0 \}$.

$$\lambda_f(x) = \sqrt{(\varepsilon - \frac{1}{x})(\frac{1}{x^2})^{-1}(\varepsilon - \frac{1}{x})} = |1 - \varepsilon x|$$

When $\varepsilon \leq 0$, $\lambda_f(x) \geq 1$, and the function is unbounded below and there does not exist a minimizer.

When $\varepsilon > 0$, $\lambda_f(x) < 1$, for $x \in (0, \frac{1}{\varepsilon})$, there exists a unique minimizer $x^* = \frac{1}{\varepsilon}$.