Lecture 22

22.1 Path-Following Scheme

We consider the following constrained convex optimization problem,

$$\min_{x \in \bar{Q}} \langle c, x \rangle, \tag{22.1}$$

where $Q \subset \mathbb{R}^n$ is an open bounded convex set equipped with a *self-concordant barrier* $F : Q \to \mathbb{R}$. We denote by $\theta > 0$ the parameter of the barrier (see previous lecture), which describes the complexity of set Q. Note that the same set can be equipped with different self-concordant barriers, and ideally we would want to choose that one with the smallest possible barrier parameter θ.

For solving (22.1), we trace the *central path* $x_t^\star \in Q$, for $t \geq 0$, defined as the minimum of the following subproblem:

$$x_t^\star \;\; = \;\; \arg\min_{x \in \mathrm{dom}\, F} \Big[\; f_t(x) \;\; := \;\; t\langle c, x \rangle + F(x) \; \Big], \tag{22.2}$$

and when $t \to +\infty$ the central path converges to a solution $x^\star \in \partial Q$ to (22.1): $x_t^\star \to x^\star$ (we prove a convergence rate for the central path in terms of the functional residual in Section 22.1.3).

Note that for our family of perturbed objectives $f_t(\cdot)$, the gradients are different by a constant term, while all the Hessians are the same, for any $t \geq 0$ and $\Delta \geq 0$:

$$\begin{array}{rcccl} \nabla f_{t+\Delta}(x) & \equiv & \nabla f_t(x) \; + \; \Delta c & \equiv & \nabla F(x) \; + \; (t + \Delta)c, \\ \nabla^2 f_{t+\Delta}(x) & \equiv & \nabla^2 f_t(x) & \equiv & \nabla^2 F(x). \end{array}$$

Therefore, for any moment of time t, the local norm at any point $x \in Q$ remains the same for all t, and it is fully defined by the Hessian $\nabla^2 F(x)$ of the self-concordant barrier at this point.

22.1.1 Newton's Step on Perturbed Problem

We are going to trace the central path approximately, utilizing the machinery of self-concordant functions, that we developed previously.

The optimality condition for the exact optimum in (22.2) is, for any $t \geq 0$:

$$\nabla f_t(x_t^\star) \;\; = \;\; tc + \nabla F(x_t^\star) \;\; = \;\; 0. \tag{22.3}$$

Let us assume that at a current fixed moment of time $t \geq 0$, we have a point $x \in Q$ that is an approximation of the central path, $x \approx x_t^\star$, under the following inexactness condition of a small gradient norm:

$$\|\nabla f_t(x)\|_x \;\; := \;\; \langle \nabla f_t(x), \nabla^2 F(x)^{-1} \nabla f_t(x) \rangle^{1/2} \;\; \leq \;\; \delta \;\; := \;\; \tfrac{1}{10}. \tag{22.4}$$

We fix $\delta = 1/10$ for simplicity of the presentation, while tighter constants can be obtained.

Consider an update of time,

$$\boxed{t^+ \;=\; t + \Delta \;>\; 0,}$$

for some small Δ (which can be either positive, if we increase t, or negative, if we want to trace the central path backwards in time), and one Newton's step from x for the new objective $f_{t^+}(\cdot)$:

$$\boxed{x^+ \;=\; x - \nabla^2 F(x)^{-1}(\nabla F(x) + t^+ c).}$$

It appears that if Δ is small, then we can ensure that x^+ will remain close to the central path with the same inexactness condition as in (22.4).

Theorem 22.1.1. *Let x satisfy (22.4) with $\delta = \frac{1}{10}$ and let*

$$|\Delta| \;\leq\; \tfrac{\gamma}{\|c\|_x}, \tag{22.5}$$

for $\gamma = \frac{1}{10}$ and $\|c\|_x := \langle c, \nabla F(x)^{-1} c\rangle^{1/2}$ is the dual norm of the target linear form. Then:

$$\|\nabla f_{t^+}(x^+)\|_{x^+} \;\leq\; \delta. \tag{22.6}$$

Proof. Denote by $\lambda(y)$ the local norm of the gradient of the *new function* f_{t^+} at point y. Thus,

$$\lambda(x) \;=\; \|\nabla f_{t^+}(x)\|_x \;=\; \|\nabla f_t(x) + \Delta c\|_x \;\overset{(22.5)}{\leq}\; \|\nabla f_t(x)\|_x \;+\; \gamma \;\overset{(22.4)}{\leq}\; \delta + \gamma. \tag{22.7}$$

Note that f_{t^+} is a standard self-concordant functions (the parameter of self-concordance is $M = 2$). Hence, by our choice of δ and γ, the point x lies in the region of local convergence of Newton's method: $\lambda(x) \leq \frac{1}{5} < \frac{1}{M} = \frac{1}{2}$ (see Theorem 20.1.2 in Lecture 20), and thus, after one Newton's, step we have

$$\lambda(x^+) \;=\; \|\nabla f_{t^+}(x^+)\|_{x^+} \;\leq\; 2\lambda(x)^2 \;\leq\; \tfrac{2}{25} \;<\; \delta,$$

which proves the required statement. □

Therefore, starting from a point x_0 within proximity to the analytic center: $\|\nabla f_0(x_0)\|_{x_0} = \|\nabla F(x_0)\|_{x_0} \leq \delta = \frac{1}{10}$, we can remain close the central path, ensuring the invariant (22.4), as soon as time t is updated not very fast.

22.1.2 The Rate of Updating Time

Note that to prove the previous theorem, we did not use the third *barrier property* of F, which is the Lipschitzness of F with respect to the local norm. This property is crucial to establish a fast *linear rate* with which we are able to update the time.

Lemma 22.1.2. *Let $\|\nabla f_t(x)\|_x \leq \delta = \frac{1}{10}$. Then,*

$$\|c\|_x \;\leq\; \tfrac{1}{t}\Big(\tfrac{1}{10} + \sqrt{\theta}\Big). \tag{22.8}$$

Consequently, ensuring (22.5), we can update time with the linear rate:

$$t^+ \;:=\; t + \tfrac{1}{10\|c\|_x} \;\overset{(22.8)}{\geq}\; t \cdot \Big(1 + \tfrac{1}{1+10\sqrt{\theta}}\Big) \;\geq\; t \cdot \exp\Big(\tfrac{1}{2(1+10\sqrt{\theta})}\Big). \tag{22.9}$$

Proof. Indeed, we have

$$t\|c\|_x \;=\; \|\nabla f_t(x) - \nabla F(x)\|_x \;\leq\; \|\nabla f_t(x)\|_x + \|\nabla F(x)\|_x \;\leq\; \delta + \sqrt{\theta},$$

which proves (22.8). □

A similar rate to (22.9) holds for tracing the central path backwards, replacing '+' by '−'.

22.1.3 Convergence Rate of Central Path

We are ready to prove the global complexity for the method which traces the central path.

First, let us show that the exact central path $x_t^\star$ indeed convergences to the optimal solution in terms of the objective function value. For that, we observe that according to optimality condition (22.3), for any positive moment of time $t > 0$, it holds:

$$c \overset{(22.3)}{=} -\tfrac{1}{t}\nabla F(x_t^\star). \tag{22.10}$$

At the same time, by the set-limitedness of F (see Lemma 21.2.8 in Lecture 21), for any $x, y \in Q$, we have

$$\langle \nabla F(x), y - x \rangle \;\leq\; \theta. \tag{22.11}$$

Therefore,

$$\langle c, x_t^\star \rangle - \langle c, x^\star \rangle \;=\; \langle c, x_t^\star - x^\star \rangle \overset{(22.10)}{=} \tfrac{1}{t}\langle \nabla F(x_t^\star), x^\star - x_t^\star \rangle \overset{(22.11)}{\leq} \tfrac{\theta}{t}. \tag{22.12}$$

Now, let $x \approx x_t^\star$ be an approximate point that satisfies condition (22.4). Note that the small gradient norm implies that the distance between points is also small. In particular, using Proposition 20.2.1 from Lecture 20, under condition (22.4): $\|\nabla f_t(x)\|_x \leq \frac{1}{\delta} = \frac{1}{10} < \frac{1}{4} = \frac{1}{2M}$, we have[1]

$$\|x - x_t^\star\|_x \;\leq\; 3\|\nabla f_t(x)\|_x \;\leq\; \tfrac{3}{10}. \tag{22.13}$$

Thus,

$$\langle c, x - x_t^\star \rangle \overset{(22.13)}{\leq} \tfrac{3}{10}\|c\|_x \overset{(22.8)}{\leq} \tfrac{3}{10t}\Big(\tfrac{1}{10} + \sqrt{\theta}\Big), \tag{22.14}$$

and we obtain the bound on the functional residual at point x, as follows:

$$\langle c, x \rangle - \langle c, x^\star \rangle \overset{(22.14),(22.12)}{\leq} \tfrac{1}{t}\Big[\theta + \tfrac{3}{10}\Big(\tfrac{1}{10} + \sqrt{\theta}\Big)\Big] \;\leq\; \tfrac{2}{t}\Big[\theta + \tfrac{3}{100}\Big]. \tag{22.15}$$

Hence, if we want to find a point $\bar{x}$ with an ε-accuracy in terms of the target residual:

$$\langle c, \bar{x} \rangle - \langle c, x^\star \rangle \;\leq\; \varepsilon, \tag{22.16}$$

according to (22.15), it is enough to trace the central path up to the moment:

$$t \;=\; \tfrac{2}{\varepsilon}\Big[\theta + \tfrac{3}{100}\Big]. \tag{22.17}$$

Assume that we start with some $t_1 > 0$, and update discrete timestamps $(t_1, t_2, t_3, \ldots)$ with the following linear rate, as in (22.9):

$$t_{k+1} \;\geq\; t_k \exp\Big(\tfrac{1}{2(1+10\sqrt{\theta})}\Big) \;\geq\; \ldots \;\geq\; t_1 \exp\Big(\tfrac{k}{2(1+10\sqrt{\theta})}\Big) \overset{(*)}{\geq} \tfrac{2}{\varepsilon}\Big[\theta + \tfrac{3}{100}\Big],$$

where $(*)$ holds as soon as

$$k \;\geq\; 2(1 + 10\sqrt{\theta}) \cdot \log\Big(\tfrac{2}{t_1\varepsilon}\Big[\theta + \tfrac{3}{100}\Big]\Big). \tag{22.18}$$

[1] A more precise analysis can be used to show that $\|x - x_t^\star\|_x \leq \frac{\|\nabla f_t(x)\|_x}{1 - \|\nabla f_t(x)\|_x} \leq \frac{\delta}{1-\delta} = \frac{1}{9}$.

22.2 Interior-Point Algorithm

We come to the following direct algorithm for solving problem (22.1). This method requires as initialization a point x_0 that is already close to the analytic center $x_0^\star := \arg\min_{y \in Q} F(y)$. To find x_0 we can use an auxiliary path-following algorithm, which we discuss in the next section.

Algorithm 3.1: *Path-Following Interior-Point Method.*

Initialization: Fix $\delta = \gamma = \frac{1}{10}$. Choose $x_0 \in Q$ such that $\|\nabla F(x_0)\|_{x_0} \leq \delta$. Set $t_0 = 0$.

For $k \geq 0$ **iterate:**

1. Update time: $t_{k+1} = t_k + \frac{\gamma}{\|c\|_{x_k}}$, where $\|c\|_{x_k} := \langle c, \nabla^2 F(x_k)^{-1} c\rangle^{1/2}$
2. Perform Newton's step with perturbed gradient:

$$x_{k+1} \quad = \quad x_k - \nabla^2 F(x_k)^{-1}\Big(\nabla F(x_k) + t_{k+1} c\Big)$$

3. If $t_{k+1} \geq \frac{2}{\varepsilon}\Big[\theta + \frac{3}{100}\Big]$ then **return** x_k

According to previous observations, we have proved the following result:

Theorem 22.2.1. *For any $\varepsilon > 0$, Algorithm 3.1 stops and return as the result a solution:*

$$\langle c, x_k\rangle - \langle c, x^\star\rangle \quad \leq \quad \varepsilon,$$

after the following number of iterations (Newton's steps):

$$k \quad = \quad O\Big(\Big[1 + \sqrt{\theta}\Big] \cdot \log \tfrac{1+\theta}{t_1 \varepsilon}\Big). \tag{22.19}$$

22.2.1 An Auxiliary Path to the Analytic Center

To find the analytic center $x_0^\star$, we can trace an *auxiliary path* that starts from any feasible point $y_1 \in Q$ and leads to $x_0^\star$, as follows:

$$y_s^\star \quad = \quad \arg\min_y \Big[\ \bar{f}_s(y) \quad := \quad -s\langle \nabla F(y_1), y\rangle + F(y)\ \Big], \qquad 1 \geq s \geq 0. \tag{22.20}$$

The optimality condition for (22.20) is $\nabla F(y_s^\star) = s\nabla F(y_1)$. Hence, $y_1^\star = y_1$ is the starting point of the auxiliary path, which we trace and $y_0^\star = x_0^\star$ is the endpoint that we wish to approach.

Using the same reasoning as above, it is easy to show that total complexity of this auxiliary traverse is of the same order (22.19).