Continuous Optimization: Problem Set 3

April 16, 2026

1. Recall the definition of α-self-concordance and barrier parameter θ:

$$\forall u, v, w : |D^3\phi(x)[u, v, w]| \le \frac{2}{\sqrt{\alpha}}\|u\|_x\|v\|_x\|w\|_x, \qquad \theta := \sup_{x\in D} \|(\nabla^2\phi(x))^{-1}\nabla\phi(x)\|_x.$$

   1. Assume ϕ_k are each α_k-self-concordant with barrier parameter θ_k for domain D. Compute the self-concordance and barrier parameter for $\sum_k \phi_k$.
   2. Show $-\log\det(X)$ is self-concordant for $D := \{X \succ 0\}$, and compute its barrier parameter.

Solution:

1. We use the remark from the notes that to show the self-concordance property, it is enough to bound all univariate restrictions $h(t) := \sum_k h_k(t) := \sum_k \phi_k(x+tu)$. Then we can bound

$$|h'''(t)| \le \sum_k |h_k'''(t)| \le \sum_k \frac{2}{\sqrt{\alpha_k}}(h_k''(t))^{3/2},$$

where we plugged in $h = \sum_k h_k$ and used the self-concordance property for each term. The terms $\{h_k''(t)\}$ have a fixed sum $\sum_k h_k''(t) = h''(t)$, and we can notice the above expression is a convex function of $\{h_k''(t)\}$, so the maximum is achieved at the boundary of the simplex, where only one term is non-zero:

$$|h'''(t)| \le \sum_k \frac{2}{\sqrt{\alpha_k}}(h_k''(t))^{3/2} \le \max_k \frac{2}{\sqrt{\alpha_k}}(h''(t))^{3/2},$$

which therefore verifies ϕ satisfies $\min_k \alpha_k$-self-concordance.

This bound is tight for e.g. the sum of two log-barriers for orthogonal hyperplanes: $\log(x_1) + \log(x_2)$.

For the barrier parameter, we use the following slick trick: it can be verified (by Schur complement) that ϕ has barrier parameter θ for domain D iff

$$\forall x \in D^\circ : \quad \|\nabla\phi(x)\|_{x,*} \le \theta \iff \begin{pmatrix} \nabla^2\phi(x) & \nabla\phi(x) \\ (\nabla\phi(x))^* & \theta^2 \end{pmatrix} \succeq 0.$$

Therefore, if ϕ_k has barrier parameter θ_k, then we can bound

$$\sum_k \begin{pmatrix} \nabla^2 \phi_k(x) & \nabla \phi_k(x) \\ (\nabla \phi_k(x))^* & \theta_k^2 \end{pmatrix} \succeq 0$$

which then implies $\phi := \sum_k \phi_k$ has barrier parameter $\leq \sqrt{\sum_k \theta_k^2}$.

2. We use the following standard trick for positive definite matrix $P \succ 0$ and symmetric matrix X:

$$\log\det(P + tX) = \log\det(P) + \log\det(I + P^{-1/2} X P^{-1/2}),$$

where we used the multiplicative property of the determinant. Now let symmetric matrix $P^{-1/2} X P^{-1/2}$ have eigenvalues $\{\lambda_i\}$, then we have

$$\log\det(P + tX) = \log\det(P) + \sum_i \log(1 + t\lambda_i).$$

With this we can use the result from the previous part, as this barrier is the sum of univariate barriers $\log(1+t\lambda_i)$. Each of these satsifies 1-self-concordance with barrier parameter 1 by affine invariance. Therefore $-\log\det$ satisfies α-self-concordance for $\alpha = \min_i 1 = 1$, with barrier parameter $\theta := \sqrt{\sum_i 1^2} = \sqrt{n}$.

2. (Theorem 4.1.3 in Nesterov) We made the assumption that $\nabla^2 \succ 0$ everywhere. In this question we remove the assumption. Show if there is some $x \in D, v \in \mathbb{R}^n$ such that $\langle v, \nabla^2 \phi(x) v \rangle = 0$ and ϕ is *self-concordant*, then in fact

$$\forall y \in \mathrm{dom}(\phi) : \langle v, \nabla^2 \phi(y) v \rangle = 0.$$

What does this say about the domain in this direction?

Solution: We first show the claim for the Dikin ellipsoid of x, namely for all y such that $r := \|y - x\|_x < 1$:

$$0 \leq \langle v, \nabla^2 \phi(y) v \rangle \leq \frac{\langle v, \nabla^2 \phi(x) v \rangle}{(1 - r)^2} = 0,$$

where the first inequality is by convexity of ϕ, in the second we used the Hessian approximation shown in the notes for $r < 1$, and the final step was by assumption. Therefore all inequalities are tight and we have $\langle v, \nabla^2 \phi(y) v \rangle = 0$ for $\|y - x\|_x < 1$.

Iterating this procedures gives the result for all $y \in D$.

Further, integrating this result twice tells us that the function is fixed, and therefore the domain must be unbounded in the u direction.

3. 1. Prove affine invariance of the Newton step: for convex f let $g(y) := f(Ay)$ so for $y = A^{-1}x$ $g(y) = f(x)$. Compute the gradient and Hessian and Newton step and show $y_{t+1} = A^{-1}x_{t+1}$ if $y_t = A^{-1}x_t$.
   2. Prove affine invariance of the definition of self-concordance and barrier parameter.

Solution: The calculation follows from the chain rule

$$\nabla g(y) = \nabla f(Ay) = A^T \nabla f(Ay)$$
$$\nabla^2 g(y) = \nabla^2 f(Ay) = A^T \nabla^2 f(Ay) A.$$

The Newton step can be seen to be invariant by plugging in these calculations:

$$(\nabla^2 g(y))^{-1} \nabla g(y) = (A^T \nabla^2 f(Ay) A)^{-1} A^T \nabla f(Ay) = A^{-1} (\nabla^2 f(Ay))^{-1} \nabla f(Ay).$$

Therefore if $x_t = Ay_t$ then

$$x_{t+1} - x_t = (\nabla^2 f(x_t))^{-1} \nabla f(x_t) = AA^{-1} (\nabla^2 f(Ay_t))^{-1} \nabla f(Ay_t) = A(y_{t+1} - y_t).$$

4. (Exercise 9.10 in BV): Newton's method with fixed step size $\lambda = 1$ can diverge if the initial point is not close to x^*. In this problem we consider two examples.

   a. $f(x) = \log(e^x + e^{-x})$ has unique minimizer $x^* = 0$. Run Newton's method with fixed step size $t = 1$, starting at $x^0 = 1$ and at $x^0 = 1.1$.
   b. $f(x) = -\log(x) + x$ has unique minimizer $x^* = 1$. Run Newton's method with fixed step size $\lambda = 1$, starting at $x^0 = 3$.

Solution: Newton's method: $x_{t+1} = x_t - \lambda f'(x_t)/f''(x_t)$.

a.
$$f'(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad f''(x) = \frac{(e^x + e^{-x})^2 - (e^x - e^{-x})^2}{(e^x + e^{-x})^2} = \frac{4}{(e^x + e^{-x})^2}$$

$$\begin{aligned} x_{t+1} &= x_t - f'(x_t)/f''(x_t) = x_t - \frac{(e^{x_t} - e^{-x_t})/(e^{x_t} + e^{-x_t})}{4/(e^{x_t} + e^{-x_t})^2} \\ &= x_t - \frac{(e^{x_t} - e^{-x_t})(e^{x_t} + e^{-x_t})}{4} = x_t - \frac{1}{4}(e^{2x_t} - e^{-2x_t}) \end{aligned}$$

For $x_0 = 1$, we have

$$x_1 = 1 - (e^2 - e^{-2})/4 \approx -0.8; x_2 = x_1 - (e^{2x_1} - e^{-2x_1})/4 \approx 0.4; ...$$

This sequence alternates signs and is strictly decreasing in magnitude, so it will eventually converge to the optimizer $x^* = 0$.

For $x_0 = 1.1$, we have

$$x_1 = 1.1 - \frac{(e^{2.2} - e^{-2.2})}{4} \approx 1.1 - 2.23 = -1.13; x_2 = x_1 - \frac{(e^{2x_1} - e^{-2x_1})}{4} \approx -1.13 + 2.37 = 1.2...$$

Newton's method diverges as the iterates alternate signs but are strictly increasing in magnitude.

b.

$$f'(x) = -\frac{1}{x} + 1 = \frac{x-1}{x}, \quad f''(x) = \frac{1}{x^2}$$

$$x_{t+1} = x_t - f'(x_t)/f''(x_t) = x_t - \frac{(x_t - 1)/x_t}{1/x_t^2} = x_t - x_t(x_t - 1) = x_t(2 - x_t)$$

Starting at $x_0 = 3$ we have

$$x_1 = (3)(-1) = -3, x_2 = (-3)(5) = -15, x_3 = (-15)(17) = -255, ...$$

Newton's method diverges as the iterates strictly increase in magnitude.

5. (Daniel Dadush Course, Ex 9.11) Let $\mathsf{A} \in \mathbb{R}^{m \times n}$, $\text{rank}(\mathsf{A}) = m$, $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$. For $\mu > 0$, examine the optimization problems

$$\min_{\mathbf{x}} \{\mathbf{c}^\mathsf{T}\mathbf{x} - \mu \sum_{i=1}^n \ln x_i : \mathsf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} > 0\} \tag{BP}$$

$$\max_{\mathbf{s},\mathbf{y}} \{\mathbf{b}^\mathsf{T}\mathbf{y} + \mu \sum_{i=1}^n \ln s_i : \mathsf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} = \mathbf{c}, \mathbf{s} > 0\}. \tag{BD}$$

Assume that the feasible regions for both programs are non-empty (i.e., the primal and dual LPs are both strictly feasible).

1. Show that the Lagrangian dual of BP is equivalent to BD and prove that strong duality holds.
2. Conclude that $\mathbf{x}_\mu$ is an optimal to solution to (BP) iff $\exists$ a solution $\mathbf{s}_\mu, \mathbf{y}_\mu$ to (BD) satisfying $x_{\mu,i} s_{\mu,i} = \mu$, $\forall i \in [n]$.

Solution:

1. Let $\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} > 0\}$, and let us write the Lagrangian function for BP:

$$L(\mathbf{x}, \mathbf{y}) = \mathbf{c}^\mathsf{T}\mathbf{x} - \mu \sum_{i=1}^n \ln x_i + \mathbf{y}^\mathsf{T}(\mathbf{b} - \mathsf{A}\mathbf{x}) = \mathbf{b}^\mathsf{T}\mathbf{y} - \mu \sum_{i=1}^n \ln x_i + (\mathbf{c} - \mathsf{A}^\mathsf{T}\mathbf{y})^\mathsf{T}\mathbf{x}$$

$$g(\mathbf{y}) = \inf_{\mathbf{x}\in\mathcal{C}} L(\mathbf{x},\mathbf{y})$$

$L(\mathbf{x},\mathbf{y})$ can be minimized coordinate by coordinate: for arbitrary $\mathbf{y}$, let $s_i = (\mathbf{c} - \mathsf{A}^\mathsf{T}\mathbf{y})_i$. If any $s_i \leq 0$, the infimum is $-\infty$ so $L(\mathbf{x},\mathbf{y}) = -\infty$ in that case. Let us now assume $s_i > 0\ \forall i$. The note that

$$\arg\min_{x_i>0} s_i x_i - \mu \ln x_i = \frac{\mu}{s_i},$$

noting that the derivative at $\frac{\mu}{s_i}$ is $s_i - \mu/x_i = 0$. So $g(\mathbf{y}) = \mathbf{b}^\mathsf{T}\mathbf{y} - \mu\sum_{i=1}^{n}\ln(\frac{\mu}{s_i}) + \sum_{i=1}^{n} s_i\frac{\mu}{s_i}$ whenever $s > 0$ or more simply

$$g(\mathbf{y}) = \mathbf{b}^\mathsf{T}\mathbf{y} - n\mu\ln\mu + \mu\sum_{i=1}^{n}\ln s_i + n\mu$$

The Lagrangian dual of BP therefore requires us to maximize $g(\mathbf{y})$ subject to the constraint $s_i > 0 \Leftrightarrow \mathbf{c} - \mathsf{A}\mathbf{y} > 0$. One may consider s_is to be independent variables and consider g to be a function of $\mathbf{y},\mathbf{s}$, with the constraint becoming $\mathsf{A}^\mathsf{T}\mathbf{y} + \mathbf{s} = \mathbf{c}, \mathbf{s} > 0$ - written this way, the Lagrangian dual of BP is equivalent to BD, since the constraints are identical and the objective functions are simply shifted by a constant value $(n\mu\ln\mu - n\mu)$.

The Slater condition holds for (BP) since the constraints $A\mathbf{x} = \mathbf{b}$ are affine linear. Finally, $\mathcal{C} = \mathbb{R}^n_{>0}$ is a convex set, so that strong duality holds.

Note that the Slater condition also holds for (BD) by same argument, and since (BD) is a shift of the Lagrangian dual of (BD), we see that both (BD) and (BP) must have lower bounded value. By Lagrangian duality, we thus get that both (BD) and (BP) both have optimal solutions, that is, (BD) has a global minimizer and (BP) has a global maximizer.

2. Assume that $\mathbf{x}_\mu$ is an optimal solution to (BP). Let $f(\mathbf{x}) = \mathbf{c}^\mathsf{T}\mathbf{x} - \mu\sum_{i=1}^{n}\ln x_i$. Since f is convex and differentiable over $\mathcal{C}$ and the Slater condition holds, we must have that $\mathbf{x}_\mu$ is a KKT point. Since $\mathcal{C}$ is open, the only active constraints at $\mathbf{x}_\mu$ are the linear equality constraints $\mathsf{A}\mathbf{x} = \mathbf{b}$. Therefore, by the KKT equations, there exists $\mathbf{y}_\mu \in \mathbb{R}^m$ such that

$$0 = \nabla f(\mathbf{x}_\mu) - \mathsf{A}^\mathsf{T}\mathbf{y}_\mu = \mathbf{c} - \mu\sum_{i=1}^{n}\mathbf{e}_i/x_{\mu,i} - \mathsf{A}^\mathsf{T}\mathbf{y}_\mu. \tag{1}$$

Let $s_{\mu,i} = c_i - (\mathsf{A}^\mathsf{T}\mathbf{y}_\mu)_i$, $i \in [n]$. Looking at equation (1) one coordinate at a time, we see that is equivalent to

$$0 = c_i - (\mathsf{A}^\mathsf{T}\mathbf{y}_\mu)_i - \mu/x_i, \forall\ i \in [n] \Leftrightarrow s_{\mu,i} = \mu/x_{\mu,i}, \forall i \in [n].$$

In particular, $s_{\mu,i}x_{\mu,i} = \mu$, $\forall i \in [n]$. From here, it is easy to see that $\mathbf{y}_\mu, \mathbf{s}_\mu$ is a valid solution to (BD) since $\mathbf{s}_\mu = \mathbf{c} - \mathsf{A}^\mathsf{T}\mathbf{y}_\mu$ and $s_{\mu,i} = \mu/x_{\mu,i} > 0$, $\forall\ i \in [n]$.

Now assume that $\mathbf{s}_\mu, \mathbf{y}_\mu$ are solutions to (BD) satisfying $x_{\mu,i}s_{\mu,i} = \mu$,$\forall\ i \in [n]$. Then, using the KKT equations from (1), it is direct to check that $\mathbf{x}_\mu$ is a KKT point with Lagrange multipliers $\mathbf{y}_\mu$. Therefore, $\mathbf{x}_\mu$ is a global minimizer, as needed.