TTIC 31070: Convex Optimization
Homework 4 Solutions

Professor Zhiyuan Li

Spring 2026

Problem 1

(a) Since $\nabla f(x) = Ax - b = A(x - x^\star) = Ae_t$, the update gives

$$e_{t+1} = e_t - \alpha A e_t + \beta(e_t - e_{t-1}) = (I - \alpha A + \beta I)e_t - \beta e_{t-1}.$$

Write $A = Q\Lambda Q^\top$, where Q is orthogonal and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. In coordinates $\widetilde{e}_t = Q^\top e_t$,

$$\widetilde{e}_{t+1} = (I - \alpha\Lambda + \beta I)\widetilde{e}_t - \beta\widetilde{e}_{t-1}.$$

Thus in an eigen-direction with eigenvalue λ,

$$e_{t+1}^{(\lambda)} = (1 + \beta - \alpha\lambda)e_t^{(\lambda)} - \beta e_{t-1}^{(\lambda)}.$$

Trying $e_t^{(\lambda)} = r^t$ gives

$$r^2 - (1 + \beta - \alpha\lambda)r + \beta = 0.$$

(b) Let $s := \sqrt{\kappa}$ and choose

$$\alpha = \frac{1}{L}, \qquad \beta = \left(1 - \frac{1}{s}\right)^2.$$

We prove the bound with

$$\theta := 1 - \frac{1}{2s}.$$

Fix $\lambda \in [\mu, L]$, and write $t := \lambda/L$. Then

$$t \in [1/s^2, 1].$$

The characteristic polynomial in this eigendirection is

$$q_\lambda(r) = r^2 - c_\lambda r + \beta, \qquad c_\lambda := 1 + \beta - \alpha\lambda = 1 + \beta - t.$$

We first bound c_λ. Since $t \le 1$,

$$c_\lambda \ge 1 + \beta - 1 = \beta \ge 0.$$

Since $t \ge 1/s^2$,

$$c_\lambda \le 1 + \left(1 - \frac{1}{s}\right)^2 - \frac{1}{s^2} = 2 - \frac{2}{s}.$$

Thus $0 \le c_\lambda \le 2 - 2/s = 2(1 - 1/s)$.

Let r_1, r_2 be the two roots of q_λ. The discriminant satisfies

$$c_\lambda^2 - 4\beta \le 4\left(1 - \frac{1}{s}\right)^2 - 4\left(1 - \frac{1}{s}\right)^2 = 0.$$

Therefore the square-root term in the quadratic formula is never positive inside:

$$r_1, r_2 = \frac{c_\lambda \pm i\sqrt{4\beta - c_\lambda^2}}{2},$$

allowing the square root to be zero. Hence

$$|r_1|^2 = |r_2|^2 = \frac{c_\lambda^2 + (4\beta - c_\lambda^2)}{4} = \beta.$$

Thus

$$|r_1| = |r_2| = \sqrt{\beta} = 1 - \frac{1}{s} \le \theta.$$

We have shown that $\rho_\lambda(\alpha, \beta) \le \theta$ for this arbitrary $\lambda \in [\mu, L]$. Taking the supremum over $\lambda \in [\mu, L]$ gives

$$\sup_{\lambda \in [\mu, L]} \rho_\lambda(\alpha, \beta) \le 1 - \frac{1}{2\sqrt{\kappa}}.$$

This is a spectral factor $1 - \Theta(1/\sqrt{\kappa})$, giving the iteration scale $O(\sqrt{\kappa}\log(1/\varepsilon))$.

(c) Gradient descent with the best fixed stepsize $\alpha = 2/(L + \mu)$ has contraction factor

$$\frac{L - \mu}{L + \mu} = \frac{\kappa - 1}{\kappa + 1}.$$

The choice in part (b) gives the accelerated spectral factor

$$1 - \Theta(1/\sqrt{\kappa}),$$

whereas the best fixed-stepsize gradient descent factor is $1 - \Theta(1/\kappa)$. This proof is special to quadratics because the error dynamics are a fixed linear recurrence that diagonalizes in the eigenbasis of one fixed matrix A. For a general smooth strongly convex function, the Hessian changes along the trajectory and the momentum dynamics do not decouple into scalar eigenmodes.

Problem 2

(a) Put

$$\tau_i(x) := a_i^\top x + b_i, \qquad u_i(x) := A_i x + p_i.$$

The fixed-t barrier objective is

$$F_t(x) = t\, c^\top x - \sum_{i=1}^m \log\left((a_i^\top x + b_i)^2 - \|A_i x + p_i\|_2^2\right).$$

Its strict-feasibility domain is

$$a_i^\top x + b_i > \|A_i x + p_i\|_2 \qquad i = 1, \dots, m.$$

(b) Let $w = (\tau, z)$, $J = \begin{bmatrix} 1 & 0 \\ 0 & -I_k \end{bmatrix}$, and $\delta = w^\top J w = \tau^2 - \|z\|_2^2$. Since $\nabla\delta(w) = 2Jw$ and $\nabla^2\delta(w) = 2J$,

$$\nabla\Phi(w) = -\frac{\nabla\delta(w)}{\delta(w)} = -\frac{2Jw}{\delta(w)}.$$

Equivalently,

$$\nabla\Phi(\tau, z) = \frac{1}{\tau^2 - \|z\|_2^2} \begin{bmatrix} -2\tau \\ 2z \end{bmatrix}.$$

Differentiating once more,

$$\nabla^2\Phi(w) = \frac{4(Jw)(Jw)^\top}{\delta(w)^2} - \frac{2J}{\delta(w)}.$$

In block form this is

$$\nabla^2\Phi(\tau, z) = \begin{bmatrix} \dfrac{2(\tau^2 + \|z\|_2^2)}{\delta^2} & -\dfrac{4\tau z^\top}{\delta^2} \\ -\dfrac{4\tau z}{\delta^2} & \dfrac{2}{\delta} I_k + \dfrac{4zz^\top}{\delta^2} \end{bmatrix}, \qquad \delta = \tau^2 - \|z\|_2^2.$$

(c) The affine block is

$$z_i(x) = B_i x + d_i, \qquad B_i = \begin{bmatrix} a_i^\top \\ A_i \end{bmatrix}, \qquad d_i = \begin{bmatrix} b_i \\ p_i \end{bmatrix}.$$

Since $F_t(x) = t\,c^\top x + \sum_{i=1}^m \Phi(z_i(x))$, the chain rule gives

$$\nabla F_t(x) = t\,c + \sum_{i=1}^m B_i^\top \nabla\Phi(z_i(x)),$$

and

$$\nabla^2 F_t(x) = \sum_{i=1}^m B_i^\top \nabla^2\Phi(z_i(x)) B_i.$$

At a strictly feasible point with $\nabla^2 F_t(x) \succ 0$, the Newton direction solves

$$\nabla^2 F_t(x)\Delta x = -\nabla F_t(x).$$

(d) The line-search step must keep every cone block in the interior:

$$a_i^\top(x + \alpha\Delta x) + b_i > \|A_i(x + \alpha\Delta x) + p_i\|_2, \qquad i = 1, \ldots, m.$$

If we only treat

$$\|A_i x + p_i\|_2^2 - (a_i^\top x + b_i)^2 \le 0$$

as a scalar inequality, the sign condition $a_i^\top x + b_i \ge 0$ is easy to lose, and the dual object is only a scalar multiplier. The conic formulation keeps the full vector $(a_i^\top x + b_i,\ A_i x + p_i)$ inside $\mathcal{Q}^{k_i+1}$, so a primal-dual method attaches a slack vector in the dual second-order cone and preserves the geometry of the quadratic constraint.

Problem 3

(a) Let $s = \langle r, x \rangle$. Since $r \geq 0$, $r \neq 0$, and $x \in \mathrm{ri}(\Delta_d)$, we have $s > 0$, and

$$\nabla \ell_r(x) = -\frac{r}{s}.$$

For Φ_2, $\nabla^2 \Phi_2(x) = I$, so

$$\|\nabla \ell_r(x)\|^2_{\Phi_2,x,*} = \frac{\|r\|_2^2}{s^2}.$$

For entropy, $\nabla^2 \Phi_{\mathrm{ent}}(x) = \mathrm{diag}(1/x_i)$, hence

$$\|\nabla \ell_r(x)\|^2_{\Phi_{\mathrm{ent}},x,*} = \sum_{i=1}^d x_i \frac{r_i^2}{s^2}.$$

For the log barrier, $\nabla^2 \Phi_{\log}(x) = \mathrm{diag}(1/x_i^2)$, hence

$$\|\nabla \ell_r(x)\|^2_{\Phi_{\log},x,*} = \sum_{i=1}^d x_i^2 \frac{r_i^2}{s^2}.$$

Define $p_i := x_i r_i / s$. Then $p_i \geq 0$ and $\sum_i p_i = 1$, so

$$\|\nabla \ell_r(x)\|^2_{\Phi_{\log},x,*} = \sum_i p_i^2 \leq 1.$$

Thus the log-barrier local dual norm is uniformly bounded by 1. The Euclidean and entropy quantities are not uniformly bounded: take $r = e_j$ and let $x_j = p \downarrow 0$. Then the Euclidean squared local dual norm is $1/p^2$, and the entropy squared local dual norm is $1/p$.

(b) Since

$$u_i^\rho = (1-\rho)u_i + \frac{\rho}{d} \geq \frac{\rho}{d},$$

we have $u^\rho \in \Delta_d^\rho$. Also,

$$\langle r_t, u^\rho \rangle = (1-\rho)\langle r_t, u \rangle + \frac{\rho}{d}\langle r_t, \mathbf{1} \rangle \geq (1-\rho)\langle r_t, u \rangle.$$

Therefore

$$\ell_t(u^\rho) \leq \ell_t(u) + \log\frac{1}{1-\rho},$$

and summing over t gives the smoothing penalty.

For the initial potentials, with $x_1 = \mathbf{1}/d$,

$$D_{\Phi_2}(u^\rho, x_1) = \frac{1}{2}\|u^\rho - x_1\|_2^2 \leq 1.$$

For entropy,

$$D_{\Phi_{\mathrm{ent}}}(u^\rho, x_1) = \sum_i u_i^\rho \log(du_i^\rho) \leq \log d.$$

For the log barrier,

$$D_{\Phi_{\log}}(u^\rho, x_1) = \sum_i \left[-\log(du_i^\rho) + du_i^\rho - 1\right].$$

Since $du_i^\rho \geq \rho$, the first term is at most $\log(1/\rho)$, and $\sum_i (du_i^\rho - 1) = 0$. Hence

$$D_{\Phi_{\log}}(u^\rho, x_1) \leq d \log\frac{1}{\rho}.$$

(c) The Euclidean update on Δ_d^ρ is

$$x_{t+1} = \Pi_{\Delta_d^\rho}(x_t - \eta g_t) = \Pi_{\Delta_d^\rho}\left(x_t + \eta \frac{r_t}{\langle r_t, x_t\rangle}\right).$$

Writing $\ell_\rho := \rho/d$ and

$$v_t := x_t - \eta g_t = x_t + \eta \frac{r_t}{\langle r_t, x_t\rangle},$$

this projection has the threshold form

$$x_{t+1,i} = \max\{\ell_\rho,\ v_{t,i} - \theta_t\}, \qquad \sum_{i=1}^d \max\{\ell_\rho,\ v_{t,i} - \theta_t\} = 1,$$

where $\theta_t \in \mathbb{R}$ is uniquely determined by the simplex constraint.

The entropy update is

$$x_{t+1} \in \arg\min_{x \in \Delta_d^\rho} \left\{ \eta \langle g_t, x - x_t\rangle + \sum_i x_i \log \frac{x_i}{x_{t,i}} \right\}.$$

Before imposing the floor constraint, the multiplicative-weights point is

$$\widetilde{x}_{t+1,i} = \frac{x_{t,i} \exp(-\eta g_{t,i})}{\sum_j x_{t,j} \exp(-\eta g_{t,j})} = \frac{x_{t,i} \exp(\eta r_{t,i}/\langle r_t, x_t\rangle)}{\sum_j x_{t,j} \exp(\eta r_{t,j}/\langle r_t, x_t\rangle)}.$$

The constrained entropy update is the entropy Bregman projection of $\widetilde{x}_{t+1}$ onto Δ_d^ρ, equivalently

$$x_{t+1,i} = \max\{\ell_\rho,\ c_t \widetilde{x}_{t+1,i}\}, \qquad \sum_{i=1}^d \max\{\ell_\rho,\ c_t \widetilde{x}_{t+1,i}\} = 1,$$

where $c_t > 0$ is uniquely determined.

The log-barrier update is run on the open simplex:

$$x_{t+1} \in \arg\min_{x \in \mathrm{ri}(\Delta_d)} \left\{ \eta \langle g_t, x - x_t\rangle + D_{\Phi_{\log}}(x, x_t) \right\},$$

where

$$D_{\Phi_{\log}}(x, x_t) = \sum_i \left[-\log \frac{x_i}{x_{t,i}} + \frac{x_i}{x_{t,i}} - 1 \right].$$

The KKT conditions give a scalar multiplier $\lambda_t \in \mathbb{R}$ such that

$$\eta g_{t,i} - \frac{1}{x_{t+1,i}} + \frac{1}{x_{t,i}} + \lambda_t = 0, \qquad i = 1, \ldots, d.$$

Equivalently,

$$x_{t+1,i} = \frac{1}{x_{t,i}^{-1} + \eta g_{t,i} + \lambda_t} = \frac{1}{x_{t,i}^{-1} - \eta r_{t,i}/\langle r_t, x_t\rangle + \lambda_t},$$

where λ_t is the unique value for which all denominators are positive and

$$\sum_{i=1}^d \frac{1}{x_{t,i}^{-1} + \eta g_{t,i} + \lambda_t} = 1.$$

(d) Let $s = \langle r, x \rangle$ and $z_i = r_i/s$. If $x \in \Delta_d^\rho$, then

$$s = \sum_i x_i r_i \geq \frac{\rho}{d} \sum_i r_i,$$

so

$$\left\| \frac{r}{s} \right\|_2 \leq \frac{\|r\|_1}{s} \leq \frac{d}{\rho}.$$

Also, $z_i \leq 1/x_i \leq d/\rho$, because $\sum_j x_j z_j = 1$ and all terms are nonnegative. Therefore $z_i^2 \leq (d/\rho) z_i$, and

$$\sum_i x_i z_i^2 \leq \frac{d}{\rho} \sum_i x_i z_i = \frac{d}{\rho}.$$

For Euclidean geometry, the constrained supremum is at most the unconstrained supremum over $h = x - y$:

$$M_{\Phi_2}^{\Delta_d^\rho}(x, r; \eta) \leq \sup_h \left\{ \eta \langle -r/s, h \rangle - \frac{1}{2} \|h\|_2^2 \right\} = \frac{\eta^2}{2} \left\| \frac{r}{s} \right\|_2^2 \leq \frac{\eta^2 d^2}{2\rho^2}.$$

For entropy, the constrained supremum is at most the full-simplex supremum. The conjugate calculation for negative entropy gives

$$M_{\Phi_{\mathrm{ent}}}^{\Delta_d^\rho}(x, r; \eta) \leq \log \left(\sum_i x_i e^{\eta z_i} \right) - \eta,$$

because $\sum_i x_i z_i = 1$. If $\eta \leq \rho/d$, then $0 \leq \eta z_i \leq 1$. Using $e^a \leq 1 + a + a^2$ for $a \in [0, 1]$,

$$\sum_i x_i e^{\eta z_i} \leq 1 + \eta \sum_i x_i z_i + \eta^2 \sum_i x_i z_i^2 \leq 1 + \eta + \frac{\eta^2 d}{\rho}.$$

Thus, using $\log(1 + a) \leq a$,

$$M_{\Phi_{\mathrm{ent}}}^{\Delta_d^\rho}(x, r; \eta) \leq \log \left(1 + \eta + \frac{\eta^2 d}{\rho} \right) - \eta \leq \frac{\eta^2 d}{\rho}.$$

For the log barrier, write $q_i = y_i/x_i$ and

$$p_i := \frac{x_i r_i}{s}.$$

Then $p_i \geq 0$, $\sum_i p_i = 1$, and

$$\left\langle -\frac{r}{s}, x - y \right\rangle = \sum_i p_i (q_i - 1).$$

Moreover,

$$D_{\Phi_{\log}}(y, x) = \sum_i [-\log q_i + q_i - 1].$$

Therefore

$$\eta \left\langle -\frac{r}{s}, x - y \right\rangle - D_{\Phi_{\log}}(y, x) = \sum_i \left[\eta p_i (q_i - 1) + \log q_i - q_i + 1 \right].$$

Using the scalar identity with $\alpha_i = \eta p_i$,

$$\eta p_i(q_i - 1) + \log q_i - q_i + 1 \le -\eta p_i - \log(1 - \eta p_i).$$

Let $h(a) := -a - \log(1-a) = \sum_{k\ge 2} a^k/k$. Since $\sum_i p_i = 1$ and $p_i \in [0,1]$,

$$\sum_i h(\eta p_i) \le \sum_i \sum_{k\ge 2} \frac{\eta^k p_i^k}{k} \le \sum_{k\ge 2} \frac{\eta^k}{k} = -\eta - \log(1-\eta).$$

For $0 < \eta \le 1/2$,

$$-\eta - \log(1-\eta) = \sum_{k\ge 2} \frac{\eta^k}{k} \le \frac{1}{2} \sum_{k\ge 2} \eta^k \le \eta^2.$$

Hence

$$M_{\Phi_{\log}}^{\mathrm{ri}(\Delta_d)}(x, r; \eta) \le \eta^2.$$

(e) Compare first to u^ρ, then add the smoothing cost from part (b). The superscripts 2, ent, and log denote the Euclidean, entropy, and log-barrier algorithms, respectively.

For Euclidean projected gradient, parts (b) and (d) give

$$\mathrm{Reg}_T^{(2)}(u) \le \frac{1}{\eta} + \frac{\eta T d^2}{2\rho^2} + T \log \frac{1}{1-\rho}.$$

When $0 < \rho \le 1/2$, the last term is $O(\rho T)$. For fixed ρ, choose $\eta = \rho/(d\sqrt{T})$, giving

$$\mathrm{Reg}_T^{(2)}(u) \le C\left(\frac{d\sqrt{T}}{\rho} + \rho T\right)$$

for a universal constant C. Taking

$$\rho = \min\left\{\frac{1}{2}, \sqrt{d}\, T^{-1/4}\right\}$$

gives the clean large-T scaling, and in particular when $\sqrt{d}\, T^{-1/4} \le 1/2$,

$$\mathrm{Reg}_T^{(2)}(u) = O(\sqrt{d}\, T^{3/4}).$$

For smaller T, keep the same clipped ρ in the intermediate bound.

For entropy mirror descent, the bound is

$$\mathrm{Reg}_T^{(\mathrm{ent})}(u) \le \frac{\log d}{\eta} + \frac{\eta T d}{\rho} + O(\rho T),$$

with the constraint $\eta \le \rho/d$. For fixed ρ, choose

$$\eta = \min\left\{\frac{\rho}{d}, \sqrt{\frac{\rho \log d}{dT}}\right\}.$$

In the large-T regime where the square-root term is admissible, this gives

$$\mathrm{Reg}_T^{(\mathrm{ent})}(u) \le C\left(\sqrt{\frac{dT \log d}{\rho}} + \rho T\right)$$

for a universal constant C. Taking

$$\rho = \min\left\{\frac{1}{2}, \left(\frac{d\log d}{T}\right)^{1/3}\right\},$$

and using the corresponding admissible $\eta \le \rho/d$, gives in the large-T regime

$$\mathrm{Reg}_T^{(\mathrm{ent})}(u) = O\Big((d\log d)^{1/3} T^{2/3}\Big),$$

which is the stated large-T rate. For smaller T, use the intermediate bound directly with clipped choices of $\rho \le 1/2$ and $\eta \le \rho/d$.

For log-barrier mirror descent, the algorithm is not truncated, but we compare to u^ρ. The bound is

$$\mathrm{Reg}_T^{(\log)}(u) \le \frac{d\log(1/\rho)}{\eta} + \eta T + T\log\frac{1}{1-\rho}.$$

For $T \ge 2$, take

$$\rho = \min\left\{\frac{1}{2}, \frac{1}{T}\right\}.$$

Then the smoothing term is $O(1)$. Choose

$$\eta = \min\left\{\frac{1}{2}, \sqrt{\frac{d\log T}{T}}\right\}.$$

This gives

$$\mathrm{Reg}_T^{(\log)}(u) = O\Big(\sqrt{dT\log T} + d\log T\Big).$$

The geometric reason is that the log-barrier Hessian scales like $\mathrm{diag}(1/x_i^2)$, so the local dual norm of $\nabla \ell_r(x)$ remains bounded even when some coordinate x_i is tiny. Euclidean and entropy geometry do not fully match the singularity of the no-margin log loss, so their analysis needs the floor parameter ρ, and the local cost worsens as $\rho \downarrow 0$.