DSC 243 — Homework 1

Convex Quadratics, Chebyshev Acceleration, and Krylov Methods

Instructions. Use the notation from the notes. Unless explicitly stated otherwise, assume

$$\begin{aligned}
&f(x) = \frac{1}{2}x^\top Ax - b^\top x,\\
&A = A^\top \succeq 0,\\
&b \in \text{range}(A),\\
&\alpha = \lambda_{\min}(A),\\
&\beta = \lambda_{\max}(A),\\
&\text{and when } \alpha > 0,\ \kappa = \beta/\alpha.
\end{aligned}$$

You may quote results proved in class or in the notes, but every nontrivial step must be justified.

Part I: Theory

T1. Gradient descent lives in the Krylov subspace. Let A be positive definite and consider gradient descent with *arbitrary* (possibly iteration-dependent) stepsizes $\eta_0, \eta_1, \eta_2, \dots$:

$$x_{k+1} = x_k - \eta_k \nabla f(x_k) = x_k - \eta_k(Ax_k - b).$$

(a) Show by induction that for every $k \geq 1$, we have

$$x_k \in x_0 + \mathcal{K}_k(A, r_0),$$

where $r_0 = b - Ax_0$ is the initial residual.

(b) Deduce that the Krylov method (which minimizes f over $x_0 + \mathcal{K}_k(A, r_0)$) produces iterates that are at least as good as gradient descent with *any* choice of stepsizes:

$$f(x_k^{\text{Krylov}}) \leq f(x_k^{\text{GD}}).$$

T2. CG as coordinate descent in the A-orthogonal basis. Let A be positive definite on $\mathbb{R}^d$. Assume that CG does not terminate before step $d-1$, so that the search directions $p_0, \dots, p_{d-1}$ are well-defined and form an A-orthogonal basis of $\mathbb{R}^d$. Define the change of coordinates

$$x = x_0 + \sum_{j=0}^{d-1} \alpha_j \, p_j,$$

and consider the objective rewritten in the new coordinates:

$$g(\alpha_0, \dots, \alpha_{d-1}) := f\Big(x_0 + \sum_{j=0}^{d-1} \alpha_j \, p_j\Big).$$

(a) First expand g and show that

$$g(\alpha_0, \dots, \alpha_{d-1}) = f(x_0) + \sum_{j=0}^{d-1} \Big[\frac{1}{2}(p_j^\top Ap_j)\, \alpha_j^2 - (r_0^\top p_j)\, \alpha_j\Big] + \sum_{0 \leq i < j \leq d-1} (p_j^\top Ap_i)\, \alpha_i \alpha_j.$$

Use A-orthogonality to conclude that the cross terms vanish, and hence g decouples into a sum of one-dimensional quadratics.

(b) Deduce that minimizing g over $\alpha_0, \dots, \alpha_{d-1}$ simultaneously is equivalent to minimizing each α_j independently. Write the closed-form minimizer α_j^*.

(c) The CG algorithm performs a line search along p_k at step k, obtaining

$$x_{k+1} = x_k + \eta_k p_k, \qquad \eta_k = \frac{r_k^\top p_k}{p_k^\top A p_k}.$$

Show that

$$r_k^\top p_k = r_0^\top p_k,$$

and deduce that $\eta_k = \alpha_k^*$ from part (b), so that CG is exactly coordinate descent in the $\{p_j\}$ basis, processing one coordinate per iteration.

(d) Explain why this viewpoint gives an alternative proof that CG terminates in at most d iterations.

T3. Support-dependent finite termination. Let $A \succeq 0$, and let $\lambda_1, \dots, \lambda_m$ be the distinct nonzero eigenvalues of A with corresponding orthonormal eigenvectors $v_1, \dots, v_m$. Suppose the initial residual satisfies

$$r_0 \in \operatorname{span}\{v_{i_1}, \dots, v_{i_r}\}$$

for some subset of r distinct eigendirections. Show that the Krylov method terminates in at most r iterations.

[**Hint:** Use that every point in $x_0 + \mathcal{K}_k(A, r_0)$ can be written as $x_0 + q(A)r_0$ for some polynomial q of degree at most $k-1$.]

T4. Repeated Chebyshev cycles. In this problem, assume A is positive definite and $\sigma(A) \subset [\alpha, \beta] \subset (0, \infty)$. Fix an integer $s \geq 1$, and let $\eta_1, \dots, \eta_s$ be the degree-s Chebyshev stepsizes from Theorem 3.1. Define the polynomial

$$p_s(\lambda) = \prod_{j=1}^{s} (1 - \eta_j \lambda).$$

Repeat this same s-step schedule for M epochs, so the total number of iterations is $N = Ms$.

(a) Show that the error after N iterations is

$$e_N = p_s(A)^M e_0.$$

(b) Prove that

$$f(x_N) - f^* \leq \left(\max_{\lambda \in [\alpha, \beta]} |p_s(\lambda)|^2 \right)^M \big(f(x_0) - f^*\big).$$

(c) Deduce the bound

$$f(x_N) - f^* \leq \frac{1}{T_s(\sigma)^{2M}} \big(f(x_0) - f^*\big) \leq 4^M \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \right)^{2N} \big(f(x_0) - f^*\big).$$

(d) Using the exact bound

$$f(x_N) - f^* \leq \frac{1}{T_s(\sigma)^{2M}} \big(f(x_0) - f^*\big),$$

derive an iteration-complexity guarantee for the restarted scheme: how large must $N = Ms$ be in order to ensure

$$f(x_N) - f^* \leq \varepsilon?$$

Express your answer in terms of κ, s, ε, and $f(x_0) - f^*$.

(e) The one-shot degree-N Chebyshev schedule satisfies

$$f(x_N) - f^* \le \frac{1}{T_N(\sigma)^2}\big(f(x_0) - f^*\big).$$

Derive the corresponding iteration-complexity estimate for this one-shot scheme.

(f) Compare the two complexity estimates. In particular, identify explicitly where the restarted method loses efficiency relative to the one-shot degree-N schedule.

T5. (Optional) Krylov convergence under clustered eigenvalues. In this problem, assume A is positive definite and that its eigenvalues lie in a union of r disjoint intervals

$$[\ell_1, u_1],\ [\ell_2, u_2],\ \dots,\ [\ell_r, u_r] \subset [\alpha, \beta].$$

Define the cluster centers and half-widths by

$$c_j := \frac{\ell_j + u_j}{2}, \qquad \delta_j := \frac{u_j - \ell_j}{2}, \qquad j = 1, \dots, r.$$

Also define

$$C_{\text{clust}} := \max_{1\le j\le r} \left(\frac{\delta_j}{c_j} \prod_{i\ne j} \left(1 + \frac{u_j}{c_i}\right)\right).$$

(a) Consider the degree-r polynomial

$$\pi_r(\lambda) := \prod_{j=1}^{r} \left(1 - \frac{\lambda}{c_j}\right).$$

Show that $\pi_r(0) = 1$ and that, for every $\lambda \in [\ell_j, u_j]$,

$$|\pi_r(\lambda)| \le \frac{\delta_j}{c_j} \prod_{i\ne j}\left(1 + \frac{u_j}{c_i}\right).$$

(b) Deduce that

$$\max_{\lambda\in\sigma(A)} |\pi_r(\lambda)| \le C_{\text{clust}}.$$

(c) Let q_t be any degree-t polynomial with $q_t(0) = 1$, and define

$$p_{r+t}(\lambda) := \pi_r(\lambda)\, q_t(\lambda).$$

Explain why $p_{r+t}(0) = 1$ and why the Krylov method after $r + t$ iterations satisfies

$$f(x_{r+t}) - f^* \le \left(\max_{\lambda\in\sigma(A)} |\pi_r(\lambda)|^2\right)\left(\max_{\lambda\in[\alpha,\beta]} |q_t(\lambda)|^2\right)\big(f(x_0) - f^*\big).$$

(d) Now choose q_t to be the degree-t shifted Chebyshev polynomial from Section 3. Deduce the bound

$$f(x_{r+t}) - f^* \le \frac{C_{\text{clust}}^2}{T_t(\sigma)^2}\big(f(x_0) - f^*\big) \le 4C_{\text{clust}}^2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^{2t} (f(x_0) - f^*).$$

(e) Interpret the result. What happens when the clusters collapse to points, i.e. $\delta_j \to 0$ with the centers c_j fixed? How does this relate to the exact finite-termination theorem for r distinct eigenvalues?

[**Hints:** For part (a), if $\lambda \in [\ell_j, u_j]$, then

$$\left|1 - \frac{\lambda}{c_j}\right| = \frac{|\lambda - c_j|}{c_j} \le \frac{\delta_j}{c_j}.$$

For $i \neq j$, use

$$\left|1 - \frac{\lambda}{c_i}\right| \le 1 + \frac{\lambda}{c_i} \le 1 + \frac{u_j}{c_i}.$$

For part (d), combine parts (b) and (c) with the standard Chebyshev bound on $[\alpha, \beta]$.]

Part II: Experiments

E1. CG termination and spectral support. Take $d = 120$ and let

$$A = \operatorname{diag}(\lambda_1, \ldots, \lambda_{20}, 0, \ldots, 0),$$

where $\lambda_1, \ldots, \lambda_{20}$ are 20 distinct positive numbers of your choice. Let

$$x^* = \mathbf{1} \in \mathbb{R}^d, \qquad b = Ax^*,$$

so that x^* is a minimizer.

(a) Choose x_0 so that the initial residual $r_0 = b - Ax_0 = A(x^* - x_0)$ is supported on exactly $r = 3$ eigendirections corresponding to nonzero eigenvalues, and run CG. Plot $\|r_k\|/\|r_0\|$.

(b) Repeat with r_0 supported on exactly $r = 6$ nonzero eigendirections.

(c) Repeat with a generic initial point x_0 such that the resulting residual $r_0 = b - Ax_0$ has nonzero components along all 20 nonzero eigendirections.

(d) Compare the observed termination behavior with the prediction from Problem T1.

E2. One-shot Chebyshev versus restarted Chebyshev. Take $d = 200$ and

$$A = \operatorname{diag}\left(1,\, 1 + \frac{99}{d-1},\, 1 + 2\frac{99}{d-1},\, \ldots,\, 100\right),$$

so that A is positive definite with $\kappa = 100$. Let

$$x^* = \mathbf{1} \in \mathbb{R}^d, \qquad b = Ax^*.$$

Let $N = 120$, and use the same generic random initial point x_0 for every method.

(a) Run one-shot degree-120 Chebyshev acceleration.

(b) For each cycle length $s \in \{5, 10, 20, 30\}$, run restarted Chebyshev with repeated s-step cycles until N iterations are used.

(c) Plot the relative suboptimality

$$\frac{f(x_k) - f^*}{f(x_0) - f^*}$$

of all methods on the same graph.

(d) Report which restart length performs best empirically, and compare this with the theory from Problem T2.

E3. CG under clustered spectra. Fix $d = 180$ and cluster centers

$$c_1 = 2, \qquad c_2 = 20, \qquad c_3 = 80.$$

For each $\gamma \in \{10^{-1},\, 10^{-2},\, 10^{-3}\}$, construct a diagonal positive definite matrix whose eigenvalues consist of 60 points in each interval

$$\big[c_j - \gamma c_j,\; c_j + \gamma c_j\big], \qquad j = 1, 2, 3.$$

For example, you may take the eigenvalues to be uniformly spaced within each interval. For each such matrix, set

$$x^* = \mathbf{1} \in \mathbb{R}^d, \qquad b = Ax^*,$$

and use the same generic random initial point x_0 for all runs.

(a) For each choice of γ, run CG and plot

$$\frac{f(x_k) - f^*}{f(x_0) - f^*}$$

versus k on a semilog scale.

(b) On the same plots, overlay the standard worst-case Chebyshev/Krylov bound based only on the condition number κ of that matrix.

(c) Also overlay the clustered-spectrum bound predicted by Problem T3: for $k \geq 3$, use the constant C_{clust} from T3 and the estimate

$$f(x_k) - f^* \leq 4C_{\text{clust}}^2 \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^{2(k-3)} \big(f(x_0) - f^*\big).$$

(d) Compare the three curves. Does the empirical CG behavior track the clustered-spectrum prediction more closely as the clusters become tighter?

(e) Briefly explain how the experiment illustrates the transition from generic $O(\sqrt{\kappa}\log(1/\varepsilon))$ behavior toward near finite termination.