SC-646
Distributed Optimization and Machine Learning
Homework - 1

Name: Medhansh Sharma
Roll no.: 22b1287

February 2026

1 Primer on Strong Convexity

Proof. Let

$$\phi_x(z) = f(z) - \nabla f(x)^\top z \tag{1}$$

Take any $z_1, z_2 \in \mathbb{R}^d$ and $\lambda \in (0, 1)$. Then

$$\phi_x(\lambda z_1 + (1-\lambda)z_2) = f(\lambda z_1 + (1-\lambda)z_2) - \nabla f(x)^\top(\lambda z_1 + (1-\lambda)z_2) \tag{2}$$

Since f is μ-strongly convex,

$$f(\lambda z_1 + (1-\lambda)z_2) \le \lambda f(z_1) + (1-\lambda)f(z_2) - \frac{\mu}{2}\lambda(1-\lambda)\|z_1 - z_2\|^2 \tag{3}$$

Substituting (3) into (2) and rearranging gives

$$\phi_x(\lambda z_1 + (1-\lambda)z_2) \le \lambda \phi_x(z_1) + (1-\lambda)\phi_x(z_2) - \frac{\mu}{2}\lambda(1-\lambda)\|z_1 - z_2\|^2 \tag{4}$$

Hence, ϕ_x is μ-strongly convex. A μ-strongly convex function satisfies the Polyak–Łojasiewicz inequality:

$$\frac{1}{2}\|\nabla \phi_x(y)\|^2 \ge \mu\big(\phi_x(y) - \phi_x^*\big) \tag{5}$$

where $\phi_x^* = \min_z \phi_x(z)$.
We compute

$$\nabla \phi_x(y) = \nabla f(y) - \nabla f(x) \tag{6}$$

At the minimizer y^* of ϕ_x, we have $\nabla \phi_x(y^*) = 0$, so

$$\nabla f(y^*) = \nabla f(x) \tag{7}$$

Since f is strongly convex, its gradient is injective, implying

$$y^* = x \tag{8}$$

Therefore,

$$\phi_x^* = \phi_x(x) = f(x) - \nabla f(x)^\top x \tag{9}$$

Substituting (6) and (9) into (5) yields

$$\frac{1}{2}\|\nabla f(y) - \nabla f(x)\|^2 \ge \mu\Big(f(y) - \nabla f(x)^\top y - f(x) + \nabla f(x)^\top x\Big) \tag{10}$$

Rearranging gives

$$f(y) \le f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2\mu}\|\nabla f(y) - \nabla f(x)\|^2 \tag{11}$$

This completes the proof.

2 AGD vs HB

Proof. We are given

$$f(x) = \log\Big(x + \sqrt{1+x^2}\Big) + x^2 \tag{1}$$

Computing the first derivative,

$$\nabla f(x) = \frac{1}{\sqrt{1+x^2}} + 2x \tag{2}$$

The second derivative is

$$\nabla^2 f(x) = -\frac{x}{(1+x^2)^{3/2}} + 2 \tag{3}$$

Let

$$g(x) = \frac{x}{(1+x^2)^{3/2}} \tag{4}$$

Then

$$\nabla^2 f(x) = 2 - g(x) \tag{5}$$

Differentiating $g(x)$,

$$g'(x) = \frac{1-2x^2}{(1+x^2)^{5/2}} \tag{6}$$

Setting $g'(x) = 0$, we obtain the critical points

$$x = \pm\frac{1}{\sqrt{2}} \tag{7}$$

Evaluating $g(x)$ at these points,

$$\max g(x) = \frac{2}{3\sqrt{3}}, \qquad \min g(x) = -\frac{2}{3\sqrt{3}} \tag{8}$$

Hence,

$$\min \nabla^2 f(x) = 2 - \frac{2}{3\sqrt{3}} \tag{9}$$

Therefore f is μ-strongly convex with

$$\mu = 2 - \frac{2}{3\sqrt{3}} \approx 1.615 \tag{10}$$

For Lipschitz continuity of the gradient, we require

$$|\nabla^2 f(x)| \leq L \tag{11}$$

Using (10),

$$\max \nabla^2 f(x) = 2 + \frac{2}{3\sqrt{3}} \tag{12}$$

Thus ∇f is L-Lipschitz with

$$L = 2 + \frac{2}{3\sqrt{3}} \tag{13}$$

Hence, the function is strongly convex and admits a Lipschitz continuous gradient. Here's the link to the jupyter notebook implementing Accelerated Gradient Descent and Heavy-Ball Method: Code for AGD and HB

3 Contraction Coefficient

Proof. Consider gradient descent

$$x^{k+1} = G_\alpha(x^k) = x^k - \alpha\nabla f(x^k) \tag{1}$$

Let x^* denote the minimizer of f, so that

$$G_\alpha(x^*) = x^* \tag{2}$$

(a) Linear convergence under contraction
Assume G_α is a contraction:

$$\|G_\alpha(x) - G_\alpha(y)\| \leq L_G\|x-y\|, \qquad L_G < 1 \tag{3}$$

Setting $y = x^*$ and using (2),

$$\|x^{k+1} - x^*\| = \|G_\alpha(x^k) - G_\alpha(x^*)\| \leq L_G\|x^k - x^*\| \tag{4}$$

By recursion,

$$\|x^k - x^*\| \leq L_G^k\|x^0 - x^*\| \tag{5}$$

Hence gradient descent converges linearly to x^*.

(b) Bound on the contraction coefficient
We write

$$G_\alpha(x) - G_\alpha(y) = (x - y) - \alpha(\nabla f(x) - \nabla f(y)) \tag{6}$$

By the Mean Value Theorem,

$$\nabla f(x) - \nabla f(y) = H(\xi)(x - y) \tag{7}$$

for some ξ on the line segment between x and y, where $\mu I \preceq H(\xi) \preceq LI$.
Substituting (7) into (6),

$$G_\alpha(x) - G_\alpha(y) = (I - \alpha H(\xi))(x - y) \tag{8}$$

Taking norms,

$$\|G_\alpha(x) - G_\alpha(y)\| \leq \|I - \alpha H(\xi)\|\|x - y\| \tag{9}$$

Since the eigenvalues of $H(\xi)$ lie in $[\mu, L]$,

$$\|I - \alpha H(\xi)\| \leq \max\{|1 - \alpha\mu|, |1 - \alpha L|\} \tag{10}$$

Thus the contraction coefficient satisfies

$$L_G \leq \max\{|1 - \alpha\mu|, |1 - \alpha L|\} \tag{11}$$

(c) Optimal step size
We minimize

$$\rho(\alpha) = \max\{|1 - \alpha\mu|, |1 - \alpha L|\} \tag{12}$$

The minimum occurs when the two terms are equal:

$$1 - \alpha\mu = \alpha L - 1 \tag{13}$$

Solving,

$$\alpha^* = \frac{2}{L + \mu} \tag{14}$$

Substituting into (12),

$$\rho^* = \frac{L - \mu}{L + \mu} \tag{15}$$

Let $\kappa = L/\mu$ be the condition number. Then

$$\rho^* = \frac{\kappa - 1}{\kappa + 1} \tag{16}$$

Using linear convergence from part (a),

$$\|x^{k+1} - x^*\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^{k+1} \|x^0 - x^*\| \tag{17}$$

4 Kernel SVMs

Here's the link to the jupyter notebook containing the code: Code for Kernel SVMs

5 Linear Convergence

Proof. Consider the proximal gradient update

$$x^{k+1} = \arg\min_{y} \left\{ \langle \nabla f(x^k), y - x^k \rangle + \frac{L}{2}\|y - x^k\|^2 + g(y) \right\} \tag{1}$$

From the L-smoothness of f, we have

$$f(x^{k+1}) \le f(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 \tag{2}$$

Adding $g(x^{k+1})$ to both sides gives

$$F(x^{k+1}) \le F(x^k) + \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2 + g(x^{k+1}) - g(x^k) \tag{3}$$

By the definition of $D_g(x^k, L)$,

$$F(x^{k+1}) \le F(x^k) - \frac{1}{2L} D_g(x^k, L) \tag{4}$$

Using the proximal-PL inequality,

$$D_g(x^k, L) \ge 2\mu(F(x^k) - F^*) \tag{5}$$

Substituting (5) into (4) yields

$$F(x^{k+1}) - F^* \le F(x^k) - F^* - \frac{\mu}{L}(F(x^k) - F^*) \tag{6}$$

Therefore,

$$F(x^{k+1}) - F^* \le \left(1 - \frac{\mu}{L}\right)(F(x^k) - F^*) \tag{7}$$

Hence F converges linearly with contraction coefficient

$$\rho = 1 - \frac{\mu}{L} \tag{8}$$

6 Necessity and Sufficiency of KKT Conditions

Proof. Consider the convex optimization problem

$$\min_{x} f(x) \quad \text{subject to} \quad g_i(x) \le 0,\ h_j(x) = 0$$

The Lagrangian is

$$\mathcal{L}(x, \lambda, \nu) = f(x) + \sum_i \lambda_i g_i(x) + \sum_j \nu_j h_j(x) \tag{1}$$

with $\lambda_i \ge 0$

Necessity
Suppose x^* and (λ^*, ν^*) are primal and dual optimal and strong duality holds. Then

$$f(x^*) = \inf_x \mathcal{L}(x, \lambda^*, \nu^*) \tag{2}$$

Hence x^* minimizes the Lagrangian, implying stationarity:

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0 \tag{3}$$

Primal feasibility gives

$$g_i(x^*) \le 0 \quad h_j(x^*) = 0 \tag{4}$$

Dual feasibility gives

$$\lambda_i^* \ge 0 \tag{5}$$

Complementary slackness follows from

$$\lambda_i^* g_i(x^*) = 0 \tag{6}$$

Thus the KKT conditions hold.

Sufficiency

Suppose x^* and (λ^*, ν^*) satisfy KKT conditions. By convexity of f and g_i and linearity of h_j,

$$f(x) \geq f(x^*) + \sum_i \lambda_i^* g_i(x) + \sum_j \nu_j^* h_j(x) = \mathcal{L}(x, \lambda^*, \nu^*) \tag{7}$$

Taking infimum over x gives

$$p^* \geq d^*. \tag{8}$$

Since weak duality always gives $d^* \leq p^*$ and strong duality holds, we conclude

$$p^* = d^* = f(x^*) \tag{9}$$

Hence x^* and (λ^*, ν^*) are optimal.
Therefore, KKT conditions are necessary and sufficient for optimality.

7 KKT conditions in action

Solution. Consider the optimization problem

$$\max_{x,y \geq 0} xy \quad \text{subject to} \quad x + y^2 \leq 2$$

We convert it to a minimization problem:

$$\min_{x,y \geq 0} -xy \quad \text{subject to} \quad g_1(x,y) = x + y^2 - 2 \leq 0,$$

$$g_2(x,y) = -x \leq 0, \qquad g_3(x,y) = -y \leq 0$$

The Lagrangian is

$$\mathcal{L}(x, y, \lambda) = -xy + \lambda_1(x + y^2 - 2) - \lambda_2 x - \lambda_3 y \tag{1}$$

where $\lambda_i \geq 0$.

KKT stationarity conditions

$$\frac{\partial \mathcal{L}}{\partial x} = -y + \lambda_1 - \lambda_2 = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial y} = -x + 2\lambda_1 y - \lambda_3 = 0 \tag{3}$$

Complementary slackness

$$\lambda_1(x + y^2 - 2) = 0, \quad \lambda_2 x = 0, \quad \lambda_3 y = 0 \tag{4}$$

Since we are maximizing xy, the optimal solution must satisfy $x > 0$ and $y > 0$. Hence from complementary slackness,

$$\lambda_2 = \lambda_3 = 0 \tag{5}$$

From (2), we obtain

$$\lambda_1 = y \tag{6}$$

Substituting into (3) gives

$$-x + 2y^2 = 0 \quad \Rightarrow \quad x = 2y^2 \tag{7}$$

Because $\lambda_1 > 0$, the first constraint is active:

$$x + y^2 = 2 \tag{8}$$

Substituting $x = 2y^2$ into (8),

$$2y^2 + y^2 = 2 \quad \Rightarrow \quad 3y^2 = 2 \quad \Rightarrow \quad y = \sqrt{\frac{2}{3}} \tag{9}$$

Hence

$$x = 2y^2 = \frac{4}{3} \tag{10}$$

The optimal value is

$$f^* = xy = \frac{4}{3}\sqrt{\frac{2}{3}} \tag{11}$$

Therefore, the optimal solution is

$$(x^*, y^*) = \left(\frac{4}{3}, \sqrt{\frac{2}{3}}\right)$$

8 Condition for Convexity

Proof. We are given for all $x, y \in \mathbb{R}^d$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \tag{1}$$

Let $\lambda \in (0, 1)$ and define

$$z = \lambda x + (1 - \lambda) y \tag{2}$$

Applying (1) with (x, z) replaced by (z, y) gives

$$f(y) \geq f(z) + \nabla f(z)^\top (y - z) \tag{3}$$

Similarly, applying (1) with (x, z) replaced by (z, x) yields

$$f(x) \geq f(z) + \nabla f(z)^\top (x - z) \tag{4}$$

Multiplying (3) by $(1 - \lambda)$ and (4) by λ, we obtain

$$(1 - \lambda) f(y) \geq (1 - \lambda) f(z) + (1 - \lambda) \nabla f(z)^\top (y - z) \tag{5}$$

$$\lambda f(x) \geq \lambda f(z) + \lambda \nabla f(z)^\top (x - z) \tag{6}$$

Adding (5) and (6) gives

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) + \nabla f(z)^\top \big(\lambda (x - z) + (1 - \lambda)(y - z)\big) \tag{7}$$

Using the definition of z in (2), we have

$$\lambda (x - z) + (1 - \lambda)(y - z) = 0 \tag{8}$$

Substituting (8) into (7) yields

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) \tag{9}$$

Finally, using (2) again,

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y) \tag{10}$$

Therefore, f is convex.