1. **[Q1. On PL inequality]**
Let $f : \mathbb{R}^d \to \mathbb{R}$ be differentiable and L-smooth (i.e., ∇f is L-Lipschitz). Assume the set of global minimizers

$$X_\star := \arg\min_{x\in\mathbb{R}^d} f(x)$$

is nonempty, and let $f_\star := \min_x f(x)$. For each $x \in \mathbb{R}^d$, let $\bar{x} := \Pi_{X_\star}(x)$ denote the Euclidean projection of x onto $X_\star$. Assume also that $\nabla f(x_\star) = 0$ for all $x_\star \in X_\star$.

Suppose there exists a constant $\mu > 0$ such that for all $x \in \mathbb{R}^d$,

$$\langle \nabla f(x),\, x - \bar{x}\rangle \;\geq\; \mu\, \|x - \bar{x}\|^2. \tag{1}$$

(a) Prove that

$$\|x - \bar{x}\| \;\leq\; \frac{1}{\mu}\, \|\nabla f(x)\| \qquad \forall x \in \mathbb{R}^d. \tag{2}$$

(b) Using L-smoothness, show that

$$f(x) - f_\star \;\leq\; \frac{L}{2}\, \|x - \bar{x}\|^2 \qquad \forall x \in \mathbb{R}^d. \tag{3}$$

(c) Deduce that f satisfies the Polyak–Łojasiewicz (PL) inequality with constant $\mu_{\mathrm{PL}} = \mu^2/L$, i.e.,

$$f(x) - f_\star \;\leq\; \frac{L}{2\mu^2}\, \|\nabla f(x)\|^2 \qquad \forall x \in \mathbb{R}^d. \tag{4}$$

2. **[Q2. Pendulum with Friction: Lyapunov Stability vs LaSalle Asymptotic Stability]**
Consider the damped pendulum model

$$\begin{aligned}\dot{x}_1 &= x_2,\\ \dot{x}_2 &= -a\sin x_1 - bx_2, \qquad a > 0,\ b > 0,\end{aligned} \tag{Q2.1}$$

where x_1 is the angle and x_2 is the angular velocity.

(a) **(Equilibria)** Find all equilibrium points of (Q2.1).

(b) **(Energy Lyapunov function)** Consider the candidate Lyapunov function

$$V(x) = a(1 - \cos x_1) + \frac{1}{2}x_2^2. \tag{Q2.2}$$

i. Show that $V(x) \geq 0$ for all x, and that $V(x) = 0$ if and only if $(x_1, x_2) = (0, 0)$ when $x_1 \in (-\pi, \pi)$.

ii. Compute $\dot{V}(x)$ along trajectories of (Q2.1) and show that

$$\dot{V}(x) = -bx_2^2 \leq 0. \tag{Q2.3}$$

iii. Explain why $\dot{V}$ is not negative definite (i.e., why $\dot{V}(x) = 0$ for all points with $x_2 = 0$ regardless of x_1).

(c) **(Stability via Lyapunov)** Using Lyapunov's direct method, conclude that the equilibrium $(0, 0)$ is (Lyapunov) stable.

(d) **(LaSalle set and largest invariant subset)** Fix a constant c satisfying $0 < c < 2a$, and define the sublevel set

$$\Omega_c := \{x \in \mathbb{R}^2 : V(x) \leq c\}. \tag{Q2.4}$$

i. Show that Ω_c is compact and positively invariant for (Q2.1).

ii. Let

$$E := \{x \in \Omega_c : \dot{V}(x) = 0\}. \tag{Q2.5}$$

Find E explicitly.

iii. Determine the *largest invariant set* $M \subseteq E$. (Hint: if $x_2(t) \equiv 0$ along a trajectory, then $\dot{x}_2(t) \equiv 0$ must also hold.)

(e) **(Asymptotic stability via LaSalle)** Use LaSalle's invariance principle to prove that every solution starting in Ω_c satisfies

$$(x_1(t), x_2(t)) \to (0,0) \quad \text{as} \quad t \to \infty, \tag{Q2.6}$$

i.e., $(0,0)$ is *asymptotically stable* with a region of attraction that contains Ω_c.

3. **[Q3. Motivating Nesterov's lower bound]**
This problem is a guided investigation that ends with the key message: *there is a fundamental speed limit* for *any* first-order (gradient-based) method on smooth convex problems.

Setup: An algorithm interacts with an unknown convex function $f : \mathbb{R}^d \to \mathbb{R}$ only through the oracle

$$\texttt{Oracle}(x) = \big(f(x),\, \nabla f(x)\big).$$

At iteration k, the algorithm chooses x_k using only past oracle replies, and satisfies the span restriction

$$x_{k+1} \in \operatorname{span}\{x_0, \nabla f(x_0), \ldots, \nabla f(x_k)\}.$$

Assume f is convex and L-smooth, and there exists a minimizer $x_\star$ with $\|x_0 - x_\star\| \le R$.

(a) **Hard instance (build it).** For $d \ge k+2$, define

$$f(x) := \frac{L}{8}\left(x_1^2 + \sum_{i=1}^{d-1}(x_{i+1} - x_i)^2\right), \tag{H}$$

and take $x_0 = (R, 0, 0, \ldots, 0)$.

i. Show that f is convex and L-smooth.

ii. Find (or argue) that a minimizer is $x_\star = 0$ and compute $f_\star$.

(b) **Local-reveal property.** Let $\mathcal{S}_k := \operatorname{span}\{\nabla f(x_0), \ldots, \nabla f(x_k)\}$. Show the following zero-pattern phenomenon for (H):

$$\text{If } (x_k)_{k+2} = (x_k)_{k+3} = \cdots = (x_k)_d = 0, \ \ \text{then } (\nabla f(x_k))_{k+3} = \cdots = (\nabla f(x_k))_d = 0.$$

Interpret in one line: each gradient call can "activate" at most one new coordinate direction.

(c) **Information bottleneck lemma.** Using Step 2, prove by induction that for any first-order method obeying the span restriction,

$$(x_k)_{k+2} = (x_k)_{k+3} = \cdots = (x_k)_d = 0 \qquad \text{for all } k. \tag{B}$$

(d) **Lower bound punchline.** Let x_k be any vector satisfying (B). Prove that on the hard instance (H),

$$f(x_k) - f_\star \ \ge\ \frac{3LR^2}{32(k+1)^2}. \tag{LB}$$

Hint: reduce to a discrete inequality on sequences $\{x_1, \ldots, x_{k+1}\}$ with boundary conditions $x_1 = R$, $x_{k+2} = 0$, and use Cauchy–Schwarz/Jensen.

(e) **Reflection.** In 4–6 lines:

i. What does (LB) say about the *best possible* worst-case rate any gradient-based method can guarantee for smooth convex optimization?

ii. Why does this make Nesterov acceleration feel "optimal" rather than just "clever"?

4. **[Q4. "Canteen Token Pricing" — Augmented Lagrangian / Method of Multipliers + coding]**
You are designing a pricing controller for a canteen token system. Each stall (agent) chooses its own production plan $x_i \in \mathbb{R}$, but the campus administration enforces a global resource-balance constraint (total plates must match a target demand).

Consider the constrained optimization problem

$$\min_{x\in\mathbb{R}^n} \; f(x) := \sum_{i=1}^{n} \left(\tfrac{1}{2}a_i x_i^2 + b_i x_i\right) \quad \text{s.t.} \quad \mathbf{1}^\top x = d, \tag{P}$$

where $a_i > 0$, $b_i \in \mathbb{R}$, and $d \in \mathbb{R}$ are given.

Define the Augmented Lagrangian

$$\mathcal{L}_\rho(x,\lambda) := f(x) + \lambda(\mathbf{1}^\top x - d) + \frac{\rho}{2}\,(\mathbf{1}^\top x - d)^2, \qquad \rho > 0. \tag{AL}$$

(a) **"Why augment?" (conceptual)**

i. Write the standard Lagrangian $\mathcal{L}(x,\lambda)$ and the dual function $g(\lambda) = \inf_x \mathcal{L}(x,\lambda)$.

ii. Explain why plain dual ascent can be slow/unstable when the primal minimizer is not unique or when the dual is ill-conditioned.

iii. Explain what role the quadratic penalty term $\frac{\rho}{2}(\mathbf{1}^\top x - d)^2$ plays in stabilizing the method.

(b) **Method of Multipliers updates (derivation)** The Method of Multipliers (MoM) performs:

$$x^{k+1} = \arg\min_x \mathcal{L}_\rho(x,\lambda^k), \qquad \lambda^{k+1} = \lambda^k + \rho(\mathbf{1}^\top x^{k+1} - d). \tag{MoM}$$

i. Derive a closed-form expression for x^{k+1} in terms of λ^k, ρ, and $\{a_i, b_i\}$.

ii. Prove that for any $\rho > 0$, the x-subproblem has a unique minimizer.

(c) **Compare with two baselines** Implement and compare:

(A) **Dual ascent (no augmentation):**

$$x^{k+1} = \arg\min_x \mathcal{L}(x,\lambda^k), \qquad \lambda^{k+1} = \lambda^k + \alpha(\mathbf{1}^\top x^{k+1} - d),$$

with stepsize $\alpha > 0$.

(B) **Quadratic penalty only (no multiplier update):**

$$x^{k+1} = \arg\min_x \; f(x) + \frac{\rho}{2}(\mathbf{1}^\top x - d)^2.$$

(C) **Method of Multipliers (MoM)** as in (MoM).

(d) **Coding task (required)** Write code (Python/Matlab/Julia) that:

i. Randomly generates a reproducible instance:

$$n = 50, \quad a_i \sim \text{Uniform}(0.5, 2), \quad b_i \sim \mathcal{N}(0,1), \quad d = 5.$$

ii. Runs (A), (B), (C) for at least 200 iterations, and plots vs iteration k:

$$r^k := \mathbf{1}^\top x^k - d \;\; \text{(primal feasibility residual)}, \qquad f(x^k), \qquad \lambda^k \text{ (if applicable)}.$$

iii. Uses the same initialization $x^0 = 0$, $\lambda^0 = 0$ for all methods.

iv. Includes a short paragraph reporting:

- which method reaches $|r^k| \le 10^{-6}$ fastest,

- whether $f(x^k)$ exhibits overshoot/oscillation,
- sensitivity of each method to hyperparameters (α for dual ascent, ρ for MoM/penalty).

(e) **Hyperparameter scaling law** Experimentally vary $\rho \in \{0.1, 1, 10, 100\}$. For MoM, report the iteration count needed to reach $|r^k| \leq 10^{-6}$. Give a hypothesis (2–4 lines) for how the best ρ scales with $\{a_i\}$ (conditioning).

5. [**Q5. Fixed-Time Convergent Saddle-Point Dynamics**]
Let $F(x, z)$ be a continuously-differentiable strongly convex-strongly concave function with moduli $\mu_1, \mu_2 > 0$, respectively in x and z. Design a fixed-time convergent flow for solving the following saddle-point optimization problem and establish its fixed-time convergence:

$$\max_{z \in \mathbb{R}^{n_2}} \min_{x \in \mathbb{R}^{n_1}} F(x, z)$$