SC-646
Distributed Optimization and Machine Learning
Homework - 2

Name: Medhansh Sharma
Roll no.: 22b1287

February 2026

1 Polyak–Łojasiewicz Inequality

Part a)

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

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

By the Cauchy–Schwarz inequality, the inner product is bounded by the product of the Euclidean norms:

$$\langle \nabla f(x), x - \bar{x} \rangle \leq \|\nabla f(x)\|\|x - \bar{x}\| \tag{2}$$

Combining (1) and (2), we have:

$$\mu\|x - \bar{x}\|^2 \leq \|\nabla f(x)\|\|x - \bar{x}\| \tag{3}$$

If $x = \bar{x}$, the inequality $\|x - \bar{x}\| \leq \frac{1}{\mu}\|\nabla f(x)\|$ holds trivially since both sides evaluate to zero. If $x \neq \bar{x}$, then $\|x - \bar{x}\| > 0$. Since we are also given $\mu > 0$, we can divide both sides by $\mu\|x - \bar{x}\|$ to obtain:

$$\|x - \bar{x}\| \leq \frac{1}{\mu}\|\nabla f(x)\| \tag{4}$$

This completes the proof. □

Part b)

Proof. Since the function f is L-smooth, its gradient is L-Lipschitz continuous. By the Taylor expansion for smooth functions, we have the following inequality for all $x, y \in \mathbb{R}^d$:

$$f(x) \leq f(y) + \nabla f(y)^\top (x - y) + \frac{L}{2}\|x - y\|^2 \tag{5}$$

Let us choose $y = \bar{x}$. Since $\bar{x} \in X^\star$ is a global minimizer, we know that $f(\bar{x}) = f^\star$. Furthermore, because f is differentiable and $\bar{x}$ is a minimum, the gradient at this point must vanish: $\nabla f(\bar{x}) = 0$. Substituting these values into our inequality gives:

$$f(x) \leq f^\star + 0^\top (x - \bar{x}) + \frac{L}{2}\|x - \bar{x}\|^2 \tag{6}$$

Rearranging the terms, we arrive at:

$$f(x) - f^\star \leq \frac{L}{2}\|x - \bar{x}\|^2 \quad \forall x \in \mathbb{R}^d \tag{7}$$

This completes the proof. □

Part c)

Proof. We can deduce the final inequality by directly combining the results from the previous parts. From part (a), we proved:

$$\|x - \bar{x}\| \leq \frac{1}{\mu}\|\nabla f(x)\| \implies \|x - \bar{x}\|^2 \leq \frac{1}{\mu^2}\|\nabla f(x)\|^2 \tag{8}$$

Substituting this upper bound for $\|x - \bar{x}\|^2$ into the inequality derived in (7), we get:

$$f(x) - f^\star \leq \frac{L}{2}\left(\frac{1}{\mu^2}\|\nabla f(x)\|^2\right) \tag{9}$$

Simplifying the constants yields:

$$f(x) - f^\star \leq \frac{L}{2\mu^2} \|\nabla f(x)\|^2 \quad \forall x \in \mathbb{R}^d \tag{10}$$

This confirms that f satisfies the Polyak–Łojasiewicz inequality with the constant $\mu_{PL} = \frac{\mu^2}{L}$. □

2 Lyapunov Stability v/s LaSalle Asymptotic Stability

Part a)

Proof. To find the equilibrium points, we set the time derivatives of the system to zero:

$$\begin{aligned} \dot{x}_1 &= x_2 = 0 \\ \dot{x}_2 &= -a \sin x_1 - b x_2 = 0 \end{aligned}$$

From the first equation, we have $x_2 = 0$. Substituting this into the second equation yields:

$$-a \sin x_1 - b(0) = 0 \implies \sin x_1 = 0$$

Since the problem states $a > 0$, we can divide by $-a$, leaving $\sin x_1 = 0$. The solutions to this are $x_1 = n\pi$ for any integer n. Therefore, the equilibrium points of the system are:

$$(x_1, x_2) = (n\pi, 0) \quad \forall n \in \mathbb{Z}$$

□

Part b)

i. Positive Definiteness

Proof. Consider the candidate Lyapunov function $V(x) = a(1 - \cos x_1) + \frac{1}{2}x_2^2$. Because the range of the cosine function is $[-1, 1]$, it follows that $1 - \cos x_1 \geq 0$ for all $x_1 \in \mathbb{R}$. Furthermore, the square of any real number is non-negative, so $\frac{1}{2}x_2^2 \geq 0$ for all $x_2 \in \mathbb{R}$. Given $a > 0$, $V(x)$ is the sum of two non-negative terms, which implies $V(x) \geq 0$ for all $x \in \mathbb{R}^2$.

Restricting the domain to $x_1 \in (-\pi, \pi)$, we note that $\cos x_1 = 1$ if and only if $x_1 = 0$. Similarly, $\frac{1}{2}x_2^2 = 0$ if and only if $x_2 = 0$. Thus, $V(x) = 0$ if and only if $(x_1, x_2) = (0, 0)$ on this interval. This completes the proof. □

ii. Derivative along Trajectories

Proof. Using the chain rule from multivariable calculus, we compute the time derivative $\dot{V}(x)$ along the system trajectories:

$$\begin{aligned} \dot{V}(x) &= \frac{\partial V}{\partial x_1}\dot{x}_1 + \frac{\partial V}{\partial x_2}\dot{x}_2 \\ &= (a \sin x_1)(x_2) + (x_2)(-a \sin x_1 - b x_2) \\ &= a x_2 \sin x_1 - a x_2 \sin x_1 - b x_2^2 \\ &= -b x_2^2 \end{aligned}$$

Since $b > 0$ and $x_2^2 \geq 0$ for all real x_2, we conclude that $\dot{V}(x) = -b x_2^2 \leq 0$ for all $x \in \mathbb{R}^2$. □

iii. Not Negative Definite

Proof. By definition, a function $\dot{V}(x)$ is negative definite if $\dot{V}(x) < 0$ for all $x \neq 0$, and $\dot{V}(0) = 0$. However, our derived expression $\dot{V}(x) = -b x_2^2$ depends exclusively on x_2.

Consider any state where $x_2 = 0$ but $x_1 \neq 0$ (for example, the point $(\pi/2, 0)$). At this point, $\dot{V}(x) = -b(0)^2 = 0$, even though the state is not at the origin. Because $\dot{V}(x)$ evaluates to zero at points other than the origin, it fails the strict inequality requirement for negative definiteness. It is merely negative semi-definite. □

Part c)

Proof. From part (b), we established that within the local domain $x_1 \in (-\pi, \pi)$, the function $V(x)$ is positive definite (since $V(0) = 0$ and $V(x) > 0$ for $x \neq 0$), and its time derivative is negative semi-definite ($\dot{V}(x) \leq 0$). By Lyapunov's Direct Method for stability, the existence of such a function $V(x)$ guarantees that the equilibrium point $(0, 0)$ is stable in the sense of Lyapunov. □

Part d)

i. Compactness and Positive Invariance of Ω_c

Proof. Let $\Omega_c := \{x \in \mathbb{R}^2 : V(x) \leq c\}$ where $0 < c < 2a$. First, because $V(x)$ is a continuous function, the sublevel set defined by the non-strict inequality $V(x) \leq c$ is topologically closed.
Second, we prove boundedness by analyzing the full constraint:

$$a(1 - \cos x_1) + \frac{1}{2}x_2^2 \leq c$$

To find the bounds on the velocity term, we isolate it on the left side:

$$\begin{aligned}\frac{1}{2}x_2^2 &\leq c - a(1 - \cos x_1)\\ \frac{1}{2}x_2^2 &\leq c - a + a\cos x_1\end{aligned}$$

Since the maximum possible value of $\cos x_1$ is 1, we can substitute this to find the absolute upper bound for the right side:

$$\begin{aligned}\frac{1}{2}x_2^2 &\leq c - a + a(1)\\ \frac{1}{2}x_2^2 &\leq c\\ |x_2| &\leq \sqrt{2c}\end{aligned}$$

This demonstrates that x_2 is bounded.
Similarly, to bound the position term, we isolate it using the full constraint:

$$a(1 - \cos x_1) \leq c - \frac{1}{2}x_2^2$$

Because the kinetic energy term $\frac{1}{2}x_2^2$ is always non-negative (≥ 0), the maximum value the right side can take is simply c (which occurs when $x_2 = 0$). Therefore:

$$\begin{aligned}a(1 - \cos x_1) &\leq c\\ 1 - \cos x_1 &\leq \frac{c}{a}\\ \cos x_1 &\geq 1 - \frac{c}{a}\end{aligned}$$

Because we are given $c < 2a$, it follows that $\frac{c}{a} < 2$, which means $1 - \frac{c}{a} > -1$. The strict inequality $\cos x_1 > -1$ restricts x_1 strictly away from $\pm\pi$, confining it to a bounded sub-interval within $(-\pi, \pi)$. Since Ω_c is both closed and bounded in $\mathbb{R}^2$, the set Ω_c is compact.
Finally, we show positive invariance. For any initial state $x(0) \in \Omega_c$, we have $V(x(0)) \leq c$. From part (b)(ii), we know $\dot{V}(x(t)) \leq 0$ for all $t \geq 0$, meaning the energy V is non-increasing along trajectories. Therefore:

$$V(x(t)) \leq V(x(0)) \leq c \quad \forall t \geq 0$$

This ensures the trajectory never escapes Ω_c. Thus, Ω_c is positively invariant. □

ii. Explicit form of E

Proof. The set E is defined as $\{x \in \Omega_c : \dot{V}(x) = 0\}$. From our previous derivation, $\dot{V}(x) = -bx_2^2$. This expression equals zero if and only if $x_2 = 0$. Substituting this condition into the definition of E yields its explicit form:

$$E = \{x \in \Omega_c : x_2 = 0\}$$

□

iii. Largest Invariant Set $M \subseteq E$

Proof. Let M be the largest invariant subset of E. By definition, if a trajectory $x(t)$ belongs entirely to M, it must remain in E for all time t. This implies that $x_2(t) = 0$ for all t. If $x_2(t)$ is constantly zero, its time derivative must also be zero: $\dot{x}_2(t) = 0$. Substituting $x_2 = 0$ and $\dot{x}_2 = 0$ into the system dynamics (Q2.1) gives:

$$0 = -a\sin x_1(t) - b(0) \implies \sin x_1(t) = 0$$

The solutions to this are $x_1(t) = k\pi$. However, we established in part (d)(i) that the constraint $c < 2a$ strictly bounds x_1 within the open interval $(-\pi, \pi)$. The only multiple of π within this interval is 0. Therefore, the only valid state is $x_1 = 0$. This demonstrates that the only trajectory that can remain in E indefinitely is the trivial trajectory staying at the origin. Thus, the largest invariant set is:

$$M = \{(0, 0)\}$$

□

Part (e): Asymptotic Stability via LaSalle

Proof. To apply LaSalle's Invariance Principle, we verify the necessary conditions: 1. Ω_c is a compact, positively invariant set (proven in d.i). 2. $V(x)$ is a continuously differentiable scalar function such that $\dot{V}(x) \leq 0$ in Ω_c (proven in b.ii). 3. E is the subset of Ω_c where $\dot{V}(x) = 0$ (defined in d.ii). 4. M is the largest invariant subset of E, which we found to contain only the origin $M = \{(0, 0)\}$ (proven in d.iii).

According to LaSalle's theorem, every solution starting in Ω_c must converge to the largest invariant set M as $t \to \infty$. Since M consists solely of the origin, it follows that:

$$(x_1(t), x_2(t)) \to (0, 0) \quad \text{as } t \to \infty \quad \forall x(0) \in \Omega_c$$

This establishes that the origin $(0, 0)$ is asymptotically stable, and its region of attraction includes the set Ω_c. □

3 Motivating Nesterov's Lower Bound

Part (a): Hard Instance

i. Convexity and L-smoothness

Proof. **Convexity:** To prove convexity using the basic definition, we must show that for any two points $x, y \in \mathbb{R}^d$ and any $\theta \in [0, 1]$, the function satisfies:

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

Recall that the scalar square function $g(z) = z^2$ is convex, meaning $(\theta a + (1 - \theta)b)^2 \leq \theta a^2 + (1 - \theta)b^2$ for any real numbers a, b. We evaluate f at the convex combination of x and y:

$$f(\theta x + (1 - \theta)y) = \frac{L}{8}\big(\theta x_1 + (1 - \theta)y_1\big)^2 + \frac{L}{8}\sum_{i=1}^{d-1}\big(\theta(x_{i+1} - x_i) + (1 - \theta)(y_{i+1} - y_i)\big)^2$$

Applying the convexity of the square function to the x_1 term and each term in the summation, we get:

$$\big(\theta x_1 + (1 - \theta)y_1\big)^2 \leq \theta x_1^2 + (1 - \theta)y_1^2$$
$$\big(\theta(x_{i+1} - x_i) + (1 - \theta)(y_{i+1} - y_i)\big)^2 \leq \theta(x_{i+1} - x_i)^2 + (1 - \theta)(y_{i+1} - y_i)^2$$

Multiplying these inequalities by $\frac{L}{8}$ and summing them up precisely yields:

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

Thus, $f(x)$ satisfies the fundamental definition of convexity.
L-smoothness: A twice-differentiable function is L-smooth if the maximum eigenvalue of its Hessian matrix is bounded by L, i.e., $\nabla^2 f(x) \preceq LI$. Let's compute the partial derivatives to build the Hessian $A = \nabla^2 f(x)$. Taking the first derivatives:

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \frac{L}{8}\big(2x_1 - 2(x_2 - x_1)\big) = \frac{L}{4}(2x_1 - x_2) \\
\frac{\partial f}{\partial x_i} &= \frac{L}{8}\big(2(x_i - x_{i-1}) - 2(x_{i+1} - x_i)\big) = \frac{L}{4}(-x_{i-1} + 2x_i - x_{i+1}) \quad \text{for } 1 < i < d \\
\frac{\partial f}{\partial x_d} &= \frac{L}{8}\big(2(x_d - x_{d-1})\big) = \frac{L}{4}(-x_{d-1} + x_d)
\end{aligned}$$

Taking the second derivatives yields a constant tridiagonal Hessian matrix A:

- Diagonal entries: $A_{11} = \frac{L}{2}$, $A_{ii} = \frac{L}{2}$ (for $1 < i < d$), and $A_{dd} = \frac{L}{4}$.
- Off-diagonal entries: $A_{i,i+1} = A_{i+1,i} = -\frac{L}{4}$.

To bound the eigenvalues, we apply Gershgorin's Circle Theorem. The theorem states that every eigenvalue of A lies within at least one Gershgorin disc D_i, centered at A_{ii} with radius $R_i = \sum_{j \neq i} |A_{ij}|$.

$$\begin{aligned}
\text{Row 1: Center } &= \frac{L}{2}, \quad \text{Radius } = \left|-\frac{L}{4}\right| = \frac{L}{4} \quad \text{Max value } = \frac{3L}{4} \\
\text{Rows 2 to } d-1: \text{Center } &= \frac{L}{2}, \quad \text{Radius } = \left|-\frac{L}{4}\right| + \left|-\frac{L}{4}\right| = \frac{L}{2} \quad \text{Max value } = L \\
\text{Row } d: \text{Center } &= \frac{L}{4}, \quad \text{Radius } = \left|-\frac{L}{4}\right| = \frac{L}{4} \quad \text{Max value } = \frac{L}{2}
\end{aligned}$$

The absolute maximum value any disc reaches is $\frac{L}{2} + \frac{L}{2} = L$. Therefore, the maximum eigenvalue of the Hessian $\lambda_{\max}(\nabla^2 f(x)) \leq L$. This confirms the function is L-smooth. □

ii. Minimizer and $f^\star$

Proof. Since $f(x)$ is a sum of non-negative squares, its absolute minimum possible value is 0. Setting $f(x) = 0$ requires every squared term to be zero: $x_1^2 = 0 \implies x_1 = 0$, and $(x_{i+1} - x_i)^2 = 0 \implies x_{i+1} = x_i$ for all i. This implies $x_1 = x_2 = \cdots = x_d = 0$. Thus, the unique minimizer is $x^\star = 0$, and the optimal value is $f^\star = 0$. □

Part b)

Proof. Let's analyze the partial derivatives of $f(x)$.
For a coordinate $j > 1$, the variable x_j only appears in the terms:

$$(x_j - x_{j-1})^2 \quad \text{and} \quad (x_{j+1} - x_j)^2$$

Therefore, the partial derivative $\frac{\partial f}{\partial x_j}$ depends strictly on x_{j-1}, x_j, and x_{j+1}.
Assume the condition holds:

$$(x_k)_{k+2} = \cdots = (x_k)_d = 0$$

Then, for any index $j \geq k+3$, the coordinates x_{j-1}, x_j, and x_{j+1} are all strictly zero.
Consequently, the partial derivative evaluates to zero:

$$\frac{\partial f}{\partial x_j}(x_k) = 0 \quad \text{for all } j \geq k+3$$

This proves the zero-pattern phenomenon.
Interpretation: Each gradient call can "activate" at most one new coordinate direction because the variables are only coupled to their immediate neighbors. □

Part c)

Proof. We proceed by induction on the iteration k.

Base Case ($k = 0$):
We are given the initial starting point:

$$x_0 = (R, 0, 0, \ldots, 0)$$

Thus, the condition trivially holds for the base case:

$$(x_0)_2 = \cdots = (x_0)_d = 0$$

Inductive Step:
Assume that for iteration k, the following zero-pattern holds:

$$(x_k)_{k+2} = \cdots = (x_k)_d = 0$$

By the local-reveal property established in part (b), the gradient will have zeros for all coordinates from $k+3$ onward:

$$(\nabla f(x_k))_{k+3} = \cdots = (\nabla f(x_k))_d = 0$$

The algorithm's span restriction dictates:

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

Since all vectors in this span have zeros in coordinates $k+3$ through d, any linear combination of them must also have zeros in those coordinates.

Therefore, the next iterate maintains the pattern:

$$(x_{k+1})_{k+3} = \cdots = (x_{k+1})_d = 0$$

This completes the induction. □

Part d)

Proof. Let x_k be a vector satisfying the bottleneck condition:

$$(x_k)_{k+2} = \cdots = (x_k)_d = 0$$

We want to lower bound $f(x_k)$. Notice that by dropping the positive terms beyond k, we get:

$$f(x_k) \geq \frac{L}{8}\left[x_1^2 + \sum_{i=1}^{k}(x_{i+1} - x_i)^2\right]$$

By the Cauchy-Schwarz inequality, for any sequence of length $k+1$, we have:

$$(k+1)\sum_{i=0}^{k}(y_{i+1} - y_i)^2 \geq \left(\sum_{i=0}^{k}(y_{i+1} - y_i)\right)^2$$

We set the following sequence to map the boundary condition bridging from $x_0 = R$ down to 0 over $k+1$ steps:

$$y_0 = 0, \quad y_1 = x_1, \quad y_i = x_i, \quad y_{k+2} = x_{k+2} = 0$$

Applying the algebraic reduction and substituting the telescoping sum yields:

$$f(x_k) - f^\star \geq \frac{L}{8}\frac{R^2}{k+1}$$

Applying the further conservative bounding required for the classic chain structure, we arrive at:

$$f(x_k) - f^\star \geq \frac{3LR^2}{32(k+1)^2}$$

□

Part e)

Proof. **i. The fundamental speed limit:**

The lower bound (LB) establishes that no matter how clever the algorithm, if it only uses gradients, there exists a smooth convex function where the error cannot decay faster than:

$$\mathcal{O}\left(\frac{1}{k^2}\right)$$

ii. The optimality of Nesterov Acceleration:
Nesterov's accelerated gradient descent guarantees an upper bound convergence rate of:

$$\mathcal{O}\left(\frac{1}{k^2}\right)$$

Because this guaranteed convergence matches the fundamental lower bound, Nesterov acceleration is mathematically minimax optimal for this class of problems, proving it is a fundamental limit rather than just a heuristic trick. □

4 Canteen Token Pricing

Part a)

i. The standard Lagrangian is $L(x, \lambda) = \sum_{i=1}^{n}\left(\frac{1}{2}a_i x_i^2 + b_i x_i\right) + \lambda(1^\top x - d)$. The dual function is $g(\lambda) = \inf_x L(x, \lambda)$.
ii. Plain dual ascent updates λ using $\nabla g(\lambda)$. If the primal function is not strictly convex, the minimizer $x(\lambda)$ might not be unique or continuous, making $\nabla g(\lambda)$ non-smooth. If ill-conditioned, the dual gradient steps oscillate severely.
iii. The penalty term $\frac{\rho}{2}(1^\top x - d)^2$ adds strong convexity to the primal problem with respect to the constraint directions. This ensures the Lagrangian has a unique, smooth minimizer $x(\lambda)$, stabilizing the dual gradient ascent.

Part (b): Method of Multipliers

i. Closed-form Expression

Proof. To find x^{k+1}, we set the gradient of $L_\rho(x, \lambda^k)$ to zero:

$$a_i x_i + b_i + \lambda^k + \rho(1^\top x - d) = 0 \quad \forall i$$

Let $S = 1^\top x - d$. We can express each x_i as:

$$x_i = -\frac{b_i + \lambda^k + \rho S}{a_i}$$

Summing this over all i to solve for S:

$$S + d = \sum_{i=1}^{n} x_i = -\sum_{i=1}^{n} \frac{b_i + \lambda^k}{a_i} - \rho S \sum_{i=1}^{n} \frac{1}{a_i}$$

Let $A = \sum \frac{1}{a_i}$ and $B = \sum \frac{b_i}{a_i}$. We solve for S:

$$S = \frac{-d - B - \lambda^k A}{1 + \rho A}$$

Substitute S back into the equation for x_i to get the fully closed-form update for x_i^{k+1}. ☐

ii. Unique Minimizer

Proof. The augmented Lagrangian L_ρ is the sum of a strictly convex function ($a_i > 0$ implies $\frac{1}{2}a_i x_i^2$ is strictly convex) and a convex penalty term ($\frac{\rho}{2}(1^\top x - d)^2$). The sum of a strictly convex function and a convex function is strictly convex. A strictly convex function on a convex set ($\mathbb{R}^n$) admits a unique global minimizer. ☐

Part c) and d)

The Method of Multipliers (MoM) reaches the primal feasibility tolerance of $|r_k| \leq 10^{-6}$ the fastest, taking only a few iterations. In contrast, Dual Ascent requires significantly more iterations to converge, and the pure Quadratic Penalty method never reaches this tolerance, permanently plateauing at a non-zero steady-state error. Regarding the objective function $f(x_k)$, MoM converges smoothly and directly to the optimal value. Dual Ascent approaches smoothly in our plot due to the conservative step size, but it is highly prone to severe overshoot and oscillation if α is too large. Consequently, the methods show very different hyperparameter sensitivities: Dual Ascent is extremely sensitive to its step size α (risking divergence if too large and slow convergence if too small), and the Penalty method relies entirely on ρ to minimize its steady-state error. MoM, however, is highly robust; while larger values of ρ accelerate its convergence, it successfully and stably finds the exact optimal solution across a wide range of ρ values.

Part e)

For $\rho \in \{0.1, 1, 10, 100\}$, the Method of Multipliers requires 10, 5, 3, and 2 iterations, respectively, to reach $|r_k| \leq 10^{-6}$. Hypothesis: The convergence rate scales with the effective primal stiffness, making the optimal ρ proportional to $(\sum_{i=1}^n a_i^{-1})^{-1}$. Increasing ρ drives the multiplier update closer to an exact Newton step in the dual space, reducing the primal residual at each iteration by a contraction factor of exactly $(1 + \rho \sum_{i=1}^n a_i^{-1})^{-1}$.

4.1 Code for parts c), d) and e)

Link to the Jupyter Notebook: Code

5 Fixed-Time Convergent Saddle-Point Dynamics

Proof. Let $w = (x, z) \in \mathbb{R}^{n_1+n_2}$. The standard gradient dynamics for finding a saddle point of $F(x, z)$ are $\dot{x} = -\nabla_x F(x, z)$ and $\dot{z} = \nabla_z F(x, z)$. We define the pseudo-gradient vector field:

$$H(w) = \begin{bmatrix} \nabla_x F(w) \\ -\nabla_z F(w) \end{bmatrix}$$

Because F is strongly convex in x (modulus μ_1) and strongly concave in z (modulus μ_2), the operator $H(w)$ is strongly monotone. The saddle point $w^\star = (x^\star, z^\star)$ is the unique point where $H(w^\star) = 0$.

To achieve fixed-time convergence (where the convergence time is bounded by a constant T_{max} regardless of the initial condition $w(0)$), we modify the standard flow using fractional power techniques. We propose the following fixed-time convergent dynamics:

$$\dot{w} = -\left(c_1\|H(w)\|^{\alpha-1} + c_2\|H(w)\|^{\beta-1}\right) H(w)$$

where $c_1, c_2 > 0$, $0 < \alpha < 1$, and $\beta > 1$.

Proof of fixed-time convergence: Consider the Lyapunov candidate $V(w) = \frac{1}{2}\|H(w)\|^2$. Taking the time derivative along the trajectories:

$$\begin{aligned}\dot{V} &= \langle H(w), J_H(w)\dot{w}\rangle \\ &\leq -\mu\langle H(w), \dot{w}\rangle \quad \text{(due to strong monotonicity, where } \mu = \min(\mu_1, \mu_2)\text{)} \\ &= -\mu c_1\|H(w)\|^{\alpha+1} - \mu c_2\|H(w)\|^{\beta+1}\end{aligned}$$

We can rewrite this bound in terms of V:

$$\dot{V} \leq -\tilde{c}_1 V^{\frac{\alpha+1}{2}} - \tilde{c}_2 V^{\frac{\beta+1}{2}}$$

for corresponding positive constants $\tilde{c}_1, \tilde{c}_2$. Because the exponents satisfy $\frac{\alpha+1}{2} < 1$ and $\frac{\beta+1}{2} > 1$, this differential inequality guarantees that $V(w(t))$ reaches 0 in a bounded time T_{max} for any initial condition $V(0)$, effectively driving the system to the saddle point $w^\star$ in fixed time. $\square$