CENG 382 - Analysis of Dynamic Systems
Spring 2024
Take Home Exam 2
Student's Solution

Ulutaş, Berk
e2522084@ceng.metu.edu.tr

May 31, 2025

---

1. Check that $V(x)$ is always positive except at the origin
   Since $x_1^2 \geq 0$ and $x_2^2 \geq 0$, we have:

$$V(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2^2) > 0 \quad \text{for all } x \neq 0$$

   and

$$V(0,0) = 0$$

   So $V(x)$ satisfies the first condition

   Compute $\dot{V}(x)$ along the system
   We use:

$$\dot{V} = \frac{\partial V}{\partial x_1}\dot{x}_1 + \frac{\partial V}{\partial x_2}\dot{x}_2$$

   Compute the partial derivatives:

$$\frac{\partial V}{\partial x_1} = x_1, \quad \frac{\partial V}{\partial x_2} = x_2$$

   Then substitute the system dynamics:

$$\dot{V} = x_1(-x_2) + x_2(x_1 - x_2^3) = -x_1x_2 + x_1x_2 - x_2^4 = -x_2^4$$

   So we get:

$$\dot{V}(x_1, x_2) = -x_2^4$$

   Since $x_2^4 \geq 0$ for all x_2, we have:

$$\dot{V}(x_1, x_2) = -x_2^4 \leq 0 \quad \text{for all } x$$

   and

$$\dot{V}(x_1, x_2) = 0 \quad \text{only when } x_2 = 0$$

   We found that:

   - $V(x) > 0$ for all $x \neq 0$, and $V(0) = 0$
   - $\dot{V}(x) \leq 0$ for all x

   So this $V(x)$ is a valid Lyapunov function and we conclude that the origin is a stable fixed point

2. We can clearly see $V(x)$ is positive, since $x_1^2 \geq 0$ and $x_2^2 \geq 0$, we have:

$$V(x_1, x_2) > 0 \quad \text{for all } x \neq 0, \quad \text{and} \quad V(0,0) = 0$$

   So $V(x)$ satisfies the first condition

   Let's calculate derivative of V

$$\dot{V} = \frac{\partial V}{\partial x_1}\dot{x}_1 + \frac{\partial V}{\partial x_2}\dot{x}_2 = 2x_1\dot{x}_1 + 2x_2\dot{x}_2$$

Substitute the expressions for $\dot{x}_1$ and $\dot{x}_2$:

$$\dot{V} = 2x_1(-4x_2) + 2x_2(Ax_2 + 4x_1 - 3x_1^2 x_2)$$

$$\dot{V} = -8x_1x_2 + 2Ax_2^2 + 8x_1x_2 - 6x_1^2x_2^2$$

Simplifying:

$$\dot{V}(x_1, x_2) = 2Ax_2^2 - 6x_1^2x_2^2 = 2x_2^2(A - 3x_1^2)$$

- If $A < 0$: then $A - 3x_1^2 < 0$ always, so

  $$\dot{V} < 0 \text{ for all } x \neq 0$$

  In this case, the system keeps losing energy, so trajectories spiral inward toward the origin. The origin is a stable fixed point.
- If $A = 0$: then $\dot{V} = -6x_1^2x_2^2 \leq 0$ Energy still decreases (except when $x_1 = 0$ or $x_2 = 0$). So again, we expect the origin to be stable
- If $A > 0$: the sign of $\dot{V}$ depends on the region:

  $$\dot{V} > 0 \text{ when } x_1^2 < \frac{A}{3}, \quad \dot{V} < 0 \text{ when } x_1^2 > \frac{A}{3}$$

  So the energy increases in some regions and decreases in others. This means trajectories are not always moving inward, they may get trapped in a region where they can't escape
  This is the kind of behavior that leads to a periodic orbit. Since the system is two dimensional and smooth, and there is a bounded trapping region, the Poincare Bendixson theorem says that trajectories must eventually settle into a closed orbit a limit cycle

3. (a) Fixed points satisfy:

$$x = x^2 - 1 \Rightarrow x^2 - x - 1 = 0 \Rightarrow x = \frac{1 \pm \sqrt{5}}{2}$$

So the fixed points are:

$$x_1 = \frac{1+\sqrt{5}}{2} \approx 1.618, \quad x_2 = \frac{1-\sqrt{5}}{2} \approx -0.618$$

We check their stability using the derivative of $f(x) = x^2 - 1$:

$$f'(x) = 2x$$

- At x_1, we have $f'(x_1) = 2x_1 = 1 + \sqrt{5} > 1 \Rightarrow$ unstable
- At x_2, we have $f'(x_2) = 1 - \sqrt{5} \approx -1.236 \Rightarrow |f'(x_2)| > 1 \Rightarrow$ unstable

(b) We want $f^2(x) = x$ but $f(x) \neq x$
First compute $f^2(x)$:

$$f(x) = x^2 - 1, \quad f(f(x)) = (x^2-1)^2 - 1 = x^4 - 2x^2 \Rightarrow f^2(x) = x^4 - 2x^2$$

Now solve:

$$f^2(x) = x \Rightarrow x^4 - 2x^2 = x \Rightarrow x^4 - 2x^2 - x = 0 \Rightarrow x(x^3 - 2x - 1) = 0$$

This gives one root at $x = 0$ Check:

$$f(0) = -1, \quad f(-1) = 0 \Rightarrow f^2(0) = 0, \quad f^2(-1) = -1$$

So:

$$\{0, -1\} \text{ is a period 2 orbit}$$

And it's prime period 2, since neither point is fixed.

(c) We need to check the derivative of $(f^2)'(x)$:

$$(f(f(x)))' = f'(f(x)) \cdot f'(x)$$

At $x = 0$, we have:

$$f(0) = -1, \quad f'(0) = 0, \quad f'(-1) = -2 \Rightarrow (f^2)'(0) = (-2)(0) = 0$$

At $x = -1$, we have:

$$f(-1) = 0, \quad f'(-1) = -2, \quad f'(0) = 0 \Rightarrow (f^2)'(-1) = 0$$

Since the product is 0 in both cases:

$$|(f^2)'(x)| < 1 \Rightarrow \text{period 2 orbit is stable}$$

4. Let's show that strong convexity implies strict convexity

   By the definition of strong convexity, for all $x, y \in \text{dom}(f)$, and for any $\lambda \in [0, 1]$, we have:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$$

   Now, if $x \neq y$, then $\|x-y\|^2 > 0$ and since $\mu > 0$, the right-hand side becomes strictly less than $\lambda f(x) + (1-\lambda)f(y)$ So we get:

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y)$$

   This is exactly the definition of strict convexity. So we've shown that strong convexity $\Rightarrow$ strict convexity.

   Now we want to prove that the function can't have two different global minimizers. So we do this by contradiction. Suppose for contradiction that f has two different global minimizers, say a and b, where $a \neq b$, and:

$$f(a) = f(b) = \min f$$

   Let us define a point between them:

$$z = \lambda a + (1-\lambda)b, \quad \text{for } \lambda \in (0, 1)$$

   Since we already showed that f is strictly convex, we know:

$$f(z) < \lambda f(a) + (1-\lambda)f(b)$$

   But since both $f(a)$ and $f(b)$ are equal to the minimum value of f, we can write:

$$f(z) < \lambda \min f + (1-\lambda)\min f = \min f$$

   Now we have reached a contradiction: We found a point z such that $f(z) < \min f$, which is impossible, since $\min f$ is supposed to be the smallest value f can take.

   Therefore, our assumption must be false. So f cannot have two different global minimizers.

   - Strong convexity implies strict convexity,
   - Strict convexity implies uniqueness of the global minimum

   So, if f is strongly convex, it must have a unique global minimizer

5. Let's call this approximation function $h(x)$

$$h(x) = f(x_t) + \nabla f(x_t)^\top (x - x_t) + \frac{1}{2}(x - x_t)^\top \nabla^2 f(x_t)(x - x_t)$$

   To find the value of x that minimizes $h(x)$, we compute the gradient of $h(x)$ with respect to x, set it equal to zero, and solve.

   - The gradient of $f(x_t)$ is zero, since it's constant.
   - The gradient of $\nabla f(x_t)^\top (x - x_t)$ is $\nabla f(x_t)$
   - The gradient of the quadratic term is $\nabla^2 f(x_t)(x - x_t)$ (using notion that $\nabla(\frac{1}{2}v^T A v) = Av$)

   So:

$$\nabla h(x) = \nabla f(x_t) + \nabla^2 f(x_t)(x - x_t)$$

   Set this equal to 0:

$$\nabla f(x_t) + \nabla^2 f(x_t)(x - x_t) = 0$$

   Solve for x: (questions says that $\nabla^2 f(x_t)$ invertible)

$$\nabla^2 f(x_t)(x - x_t) = -\nabla f(x_t)$$

$$x - x_t = -\nabla^2 f(x_t)^{-1} \nabla f(x_t)$$

$$x = x_t - \nabla^2 f(x_t)^{-1} \nabla f(x_t)$$

   This is exactly the Newton step update rule:

$$x_{t+1} = \arg\min h(x) = x_t - \nabla^2 f(x_t)^{-1} \nabla f(x_t)$$

   So we have shown that the vector x_{t+1} from Newton step is same x minimizes the function $h(x)$

6. (a)

$$L(\theta) = \frac{1}{2N}\sum_{i=1}^{N}(h_\theta(x_i) - y_i)^2 = \frac{1}{2N}\sum_{i=1}^{N}(\theta x_i - y_i)^2$$

We now compute the gradient of $L(\theta)$ with respect to θ:

$$\nabla L(\theta) = \frac{\partial L(\theta)}{\partial \theta} = \frac{1}{2N}\sum_{i=1}^{N}\frac{\partial}{\partial \theta}(\theta x_i - y_i)^2$$

$$= \frac{1}{2N}\sum_{i=1}^{N}2(\theta x_i - y_i)\cdot x_i = \frac{1}{N}\sum_{i=1}^{N}(\theta x_i - y_i)x_i$$

So, the gradient is:

$$\nabla L(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\theta x_i - y_i)x_i$$

The batch Gradient Descent update rule is:

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla L(\theta_t)$$

Substituting in the gradient:

$$\theta_{t+1} = \theta_t - \alpha \cdot \left(\frac{1}{N}\sum_{i=1}^{N}(\theta_t x_i - y_i)x_i\right)$$

(b)

$$L_j(\theta) = \frac{1}{2}(h_\theta(x_j) - y_j)^2 = \frac{1}{2}(\theta x_j - y_j)^2$$

Now we take the gradient of $L_j(\theta)$ with respect to θ:

$$\nabla L_j(\theta) = \frac{\partial L_j(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\frac{1}{2}(\theta x_j - y_j)^2\right]$$

$$= (\theta x_j - y_j)\cdot x_j$$

So the stochastic gradient is:

$$\nabla L_j(\theta) = (\theta x_j - y_j)x_j$$

The update rule for SGD becomes:

$$\theta_{t+1} = \theta_t - \alpha \cdot \nabla L_j(\theta_t)$$

$$\theta_{t+1} = \theta_t - \alpha \cdot (\theta_t x_j - y_j)x_j$$

(c) The batch loss function is:

$$L(\theta) = \frac{1}{2N}\sum_{i=1}^{N}(\theta x_i - y_i)^2$$

For our dataset of two points ($N = 2$):

$$L(\theta) = \frac{1}{4}\left[(\theta \cdot 1 + 2)^2 + (\theta \cdot 2 - 3)^2\right]$$

Now expand the terms:

$$L(\theta) = \frac{1}{4}\left[(\theta + 2)^2 + (2\theta - 3)^2\right]$$

$$= \frac{1}{4}\left[\theta^2 + 4\theta + 4 + 4\theta^2 - 12\theta + 9\right] = \frac{1}{4}\left[5\theta^2 - 8\theta + 13\right]$$

To minimize this function, take the derivative with respect to θ and set it to zero:

$$\frac{dL}{d\theta} = \frac{1}{4}(10\theta - 8) \quad \Rightarrow \quad \frac{1}{4}(10\theta - 8) = 0$$

$$10\theta - 8 = 0 \Rightarrow \theta = \frac{4}{5}$$

$$\theta^* = \frac{4}{5}$$

So, the optimal θ that minimizes the batch loss is $\frac{4}{5}$

(d) The batch gradient descent update rule from part (a):

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{1}{N} \sum_{i=1}^{N} (\theta_t x_i - y_i) x_i$$

We are given:

$$(x_1, y_1) = (1, -2), \quad (x_2, y_2) = (2, 3), \quad \theta_0 = 0, \quad \alpha = 0.1, \quad N = 2$$

Compute the gradient at $\theta_0 = 0$:

$$\nabla L(\theta_0) = \frac{1}{2} [(0 \cdot 1 - (-2)) \cdot 1 + (0 \cdot 2 - 3) \cdot 2] = \frac{1}{2}(2 - 6) = \frac{-4}{2} = -2$$

Now apply the update rule:

$$\theta_1 = \theta_0 - \alpha \cdot \nabla L(\theta_0) = 0 - 0.1 \cdot (-2) = 0.2$$

$$\theta_1 = 0.2$$

(e) The SGD update rule from part (b):

$$\theta_{t+1} = \theta_t - \alpha \cdot (\theta_t x_j - y_j) x_j$$

Substitute the values:

$$x_1 = 1, \quad y_1 = -2, \quad \theta_0 = 0, \quad \alpha = 0.1$$

Compute the gradient using just this one data point:

$$g = (\theta_0 \cdot x_1 - y_1) x_1 = (0 - (-2)) \cdot 1 = 2$$

Now update:

$$\theta_1 = \theta_0 - \alpha \cdot g = 0 - 0.1 \cdot 2 = -0.2$$

$$\theta_1 = -0.2$$

(f)
- Batch Gradient Descent update (part d): $\theta_1 = 0.2$
- SGD update (part e): $\theta_1 = -0.2$
- Optimal value: $\theta^* = 0.8$

Now let's compare:

- GD: Moved from $\theta_0 = 0$ to $\theta_1 = 0.2$. This is *towards* the optimal value $\theta^* = 0.8$, so the direction is correct.
- SGD: Moved from $\theta_0 = 0$ to $\theta_1 = -0.2$. This is *away* from the optimal value $\theta^* = 0.8$, since it only considered one data point.

The batch GD update moved in the correct direction toward. This shows the difference in behavior, GD gives a more stable direction by averaging over all samples, where SGD may move in a noisy or even incorrect direction in a single step