1 Boundedness, attainment, and critical points

For each function below:

1. Determine whether it is bounded below;
2. Compute its infimum;
3. Determine whether the infimum is attained;
4. Find every first-order critical point; and
5. State all global minimizers, if any.

Rigorously justify each conclusion.

Worked example 1.1

Consider

$$f(x) = e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Solution. (1) Since $f(x) > 0$ for every finite x and $f(x) \to 0$ as $|x| \to \infty$, the function is bounded below. (2) $\inf_{x\in\mathbb{R}} f(x) = 0$. (3) The infimum is not attained. (4) Since

$$f'(x) = -xe^{-x^2/2}, \qquad f''(x) = (x^2 - 1)e^{-x^2/2}.$$

The only critical point is $x = 0$. (5) Since the infimum is not attained, the global minimizer does not exist.

1. Let $f(x) = 4x - 1$ for $x \in \mathbb{R}$. In addition, for a fixed stepsize $\alpha > 0$, derive the gradient-descent iterates in closed form and describe the behavior of x_k and $f(x_k)$ as $k \to \infty$.
2. Let $f(x) = \frac{1}{2}(x - 3)^2 + 2$ for $x \in \mathbb{R}$. In addition, for a fixed stepsize $\alpha > 0$, derive the gradient-descent iterates in closed form and describe the behavior of x_k and $f(x_k)$ as $k \to \infty$.
3. $f(x) = e^x$ for $x \in \mathbb{R}$.
4. $f(x) = x^3 - 3x$ for $x \in \mathbb{R}$.
5. $f(x) = x^4 - 2x^2$ for $x \in \mathbb{R}$.
6. $f(x, y) = x^2 - y^2$ for $(x, y) \in \mathbb{R}^2$.
7. **(G)** $f(x, y) = x^2 + e^y$ for $(x, y) \in \mathbb{R}^2$.
8. **(G)** $f(x, y) = x^4 + y^4 - 2x^2 - 4y^2$ for $(x, y) \in \mathbb{R}^2$.
9. **(G)** $f(x, y) = (xy - 1)^2 + x^2$ for $(x, y) \in \mathbb{R}^2$.

2 Convex minimization from first-order optimality

For each convex objective below:

1. Compute the gradient and Hessian;
2. Identify all first-order critical points;
3. Verify convexity to guarantee global optimality at critical points.

Rigorously justify each conclusion.

Worked example 2.1

Minimize the log-barrier objective $f(x) = \frac{1}{2}(x-2)^2 - \log x$ for $x > 0$.
Solution. (1) The derivatives are

$$f'(x) = x - 2 - \frac{1}{x}, \qquad f''(x) = 1 + \frac{1}{x^2} > 0.$$

(2) Solving $f'(x) = 0$ gives $x^2 - 2x - 1 = 0$. Only $x^\star = 1 + \sqrt{2}$ lies in the domain $x > 0$. (3) f is convex because $f''(x) \geq 1$ for all $x > 0$. Therefore $x^\star = 1 + \sqrt{2}$ is globally optimal.

1. Minimize $f(x) = \frac{1}{2}(x-4)^2 + \frac{1}{2}(2x+1)^2$ for $x \in \mathbb{R}$.
2. Minimize
$$f(x_1, x_2) = \tfrac{1}{2}(x_1 - 1)^2 + \tfrac{1}{2}(x_2 + 2)^2 + \tfrac{1}{2}(x_1 + x_2)^2.$$
3. Minimize $f(x) = \frac{1}{2}(x-3)^2 - 2\log x$ for $x > 0$.
4. Minimize
$$f(x_1, x_2) = \tfrac{1}{2}(x_1 - 2)^2 + \tfrac{1}{2}(x_2 - 1)^2 - \log x_1 - 2\log x_2,$$
on the domain $x_1 > 0$, $x_2 > 0$.

3 Global minimization from second-order optimality

In class, we learned that every *twice-differentiable coercive function attains a global minimum, and every global minimizer is a second-order critical point.*

Use this theorem to globally optimize each nonconvex objective f below:

1. Compute the gradient and Hessian;
2. Identify all second-order critical points;
3. Prove that f is not convex;
4. State the global minimizers by directly examining the second-order critical points. (You do not have to prove that f is coercive.)

Rigorously justify each conclusion.

Worked example 3.1

Minimize $f(x, y) = x^2 + y^4 - 2y^2$ over $\mathbb{R}^2$.
Solution. (1) The gradient and Hessian are

$$\nabla f(x,y) = \begin{bmatrix} 2x \\ 4y(y^2-1) \end{bmatrix}, \qquad \nabla^2 f(x,y) = \begin{bmatrix} 2 & 0 \\ 0 & 12y^2 - 4 \end{bmatrix}.$$

(2) The first-order critical points are $(0, 0)$ and $(0, \pm 1)$. Of these,

$$\nabla^2 f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix}, \qquad \nabla^2 f(0,\pm 1) = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix},$$

so $(0, \pm 1)$ are the only second-order critical points. (3) Since $\nabla^2 f(0, 0) \not\succeq 0$, f is not convex. (4) Both second-order critical points have the same objective $f(0, \pm 1) = -1$. Since f must attain its minimum at a second-order critical point, these are exactly the global minimizers, with minimum value -1.

1. Minimize $f(x) = x^4 + 2x^3$ for $x \in \mathbb{R}$.
2. Minimize $f(x) = \frac{1}{4}\|x\|_2^4 - \frac{1}{2}\|x\|_2^2$ for $x \in \mathbb{R}^2$.
3. Minimize $f(x, y) = (x^2 - 1)^2 + y^2$ for $(x, y) \in \mathbb{R}^2$.
4. **(G)** Let $C = \operatorname{diag}(1, 3)$ and minimize
$$f(x) = x^T C x + (\|x\|_2^2 - 1)^2, \qquad x \in \mathbb{R}^2.$$
5. **(G)** Fix a nonzero vector $z \in \mathbb{R}^n$ with $n \geq 2$ and minimize
$$f(x) = \|xx^T - zz^T\|_F^2, \qquad x \in \mathbb{R}^n.$$
Hint: first show that
$$f(x) = \|x\|_2^4 - 2(x^T z)^2 + \|z\|_2^4.$$
6. **(G)** Let $C = C^T \succeq 0$ have eigenvalues $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ and a unit eigenvector u_1 for λ_1. Minimize
$$f(x) = \|xx^T - C\|_F^2, \qquad x \in \mathbb{R}^n.$$
Find all critical points as well as all global minimizers. Hint: diagonalize C and use
$$\nabla f(x) = 4(\|x\|_2^2 x - Cx).$$

4 Fixed-step preconditioned gradient descent

Let $B = B^T \succ 0$ and define $\|v\|_B^2 = v^T B v$ and $\|v\|_{B^{-1}}^2 = v^T B^{-1} v$. Assume throughout that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and satisfies the preconditioned smoothness condition

$$\|\nabla f(x) - \nabla f(y)\|_B \le L\|x - y\|_{B^{-1}} \qquad \text{for all } x, y \in \mathbb{R}^n.$$

Then, preconditioned gradient descent $x_{k+1} = x_k - \alpha B \nabla f(x_k)$ satisfies the descent lemma

$$f(x_{k+1}) \le f(x_k) - \alpha \left(1 - \frac{L\alpha}{2}\right) \|\nabla f(x_k)\|_B^2. \tag{DL}$$

Moreover, if $f^\star := \inf_x f(x) > -\infty$ and f satisfies the Polyak-Lojasiewicz inequality

$$\frac{1}{2}\|\nabla f(x)\|_B^2 \ge \mu(f(x) - f^\star), \tag{PL}$$

then preconditioned gradient descent converges exponentially.

1. **Descent lemma.** Using the fundamental theorem of calculus and the weighted Cauchy–Schwarz inequality $|a^T b| \le \|a\|_B \|b\|_{B^{-1}}$, prove the quadratic upper-bound
$$f(y) \le f(x) + \nabla f(x)^T (y - x) + \frac{L}{2}\|y - x\|_{B^{-1}}^2.$$
Prove (DL) by substituting $x_{k+1} = x_k - \alpha B \nabla f(x_k)$.

2. **Gradient convergence.** Starting from (DL), prove that the step-size $\alpha = \frac{1}{L}$ guarantees the following first-order complexity bound
$$\min_{0 \le k < N} \|\nabla f(x_k)\|_B^2 \le \frac{2L(f(x_0) - f^\star)}{N}.$$

3. **PL inequality.** Assume that f is μ-strongly convex and attains its minimum at $f(x^\star) = \min_x f(x)$. Starting from the preconditioned strong-convexity condition
$$f(y) \ge f(x) + \nabla f(x)^T (y - x) + \frac{\mu}{2}\|y - x\|_{B^{-1}}^2,$$
prove that f satisfies (PL) with the same μ.

4. **Linear convergence.** Suppose that f satisfies (PL) for some $\mu > 0$. Use (DL) to prove that gradient descent converges exponentially
$$f(x_{k+1}) - f^\star \le \left(1 - \frac{\mu}{L}\right)(f(x_k) - f^\star),$$
with a suitable choice of stepsize $\alpha > 0$. State what that "suitable choice" of α is.

5. **(G) Convex objective convergence.** Assume that f is convex and attains its minimum at $f(x^\star) = \min_x f(x)$. Starting from the first-order convexity definition
$$f(y) \ge f(x) + \nabla f(x)^T (y - x)$$
and (DL), prove that the step-size $\alpha = \frac{1}{L}$ implies the telescope identity
$$f(x_{k+1}) - f^\star \le \frac{L}{2}\left(\|x_k - x^\star\|_{B^{-1}}^2 - \|x_{k+1} - x^\star\|_{B^{-1}}^2\right).$$
Then, use monotonicity $f(x_0) \ge f(x_1) \ge f(x_2) \ge \cdots$ to prove convergence to the optimal value
$$f(x_N) - f^\star \le \frac{L\|x_0 - x^\star\|_{B^{-1}}^2}{2N}.$$

6. **(G) Backtracking line search.** Given parameters $c_1, c_2 \in (0,1)$, we pick stepsize α by testing $\alpha \in \{1, c_2, c_2^2, c_2^3, \ldots\}$ until the preconditioned Armijo condition

$$f(x_k - \alpha B \nabla f(x_k)) \le f(x_k) - c_1 \alpha \|\nabla f(x_k)\|_B^2 \tag{AC}$$

is satisfied. Prove that the accepted step is bounded below by

$$\alpha \ge \underline{\alpha} := \min\left\{1, \frac{2c_2(1-c_1)}{L}\right\},$$

and therefore backtracking line search satisfies the following variant of (DL)

$$f(x_{k+1}) \le f(x_k) - \min\left\{c_1, \frac{2c_1c_2(1-c_1)}{L}\right\} \|\nabla f(x_k)\|_B^2.$$