

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

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

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

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

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

Question 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 $k \rightarrow \infty$.

Solution

(1) Along $x \rightarrow -\infty$, $f(x) = 4x - 1 \rightarrow -\infty$, so the function is not bounded below. (2) $\inf_{x \in \mathbb{R}} f(x) = -\infty$. (3) The infimum is not attained. (4) Since $f'(x) = 4$ for every x , there are no first-order critical points. (5) There is no global minimizer. For gradient descent, $x_{k+1} = x_k - 4\alpha$, hence

$$x_k = x_0 - 4\alpha k, \quad f(x_k) = 4x_0 - 16\alpha k - 1.$$

Therefore $x_k \rightarrow -\infty$ and $f(x_k) \rightarrow -\infty$ as $k \rightarrow \infty$.

Question 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 $k \rightarrow \infty$.

Solution

(1) Since $f(x) = \frac{1}{2}(x-3)^2 + 2 \geq 2$, the function is bounded below. (2) $\inf_{x \in \mathbb{R}} f(x) = 2$. (3) The infimum is attained at $x = 3$. (4) Since $f'(x) = x - 3$, the only first-order critical point is $x = 3$. (5) Therefore $x = 3$ is the unique global minimizer. Gradient descent satisfies

$$x_{k+1} - 3 = (1 - \alpha)(x_k - 3), \quad x_k = 3 + (1 - \alpha)^k(x_0 - 3),$$

and therefore $f(x_k) = 2 + \frac{1}{2}(1 - \alpha)^{2k}(x_0 - 3)^2$. If $0 < \alpha < 2$, then $x_k \rightarrow 3$ and $f(x_k) \rightarrow 2$. If $\alpha = 2$ and $x_0 \neq 3$, the iterates alternate between two points and the objective value remains constant. If $\alpha > 2$ and $x_0 \neq 3$, then $|x_k - 3| \rightarrow \infty$ and $f(x_k) \rightarrow \infty$. When $x_0 = 3$, the iterates

remain at the minimizer for every $\alpha > 0$.

Question 3

$f(x) = e^x$ for $x \in \mathbb{R}$.

Solution

(1) Since $e^x > 0$ for every finite x , the function is bounded below. (2) Since $e^x \rightarrow 0$ as $x \rightarrow -\infty$, $\inf_{x \in \mathbb{R}} f(x) = 0$. (3) The infimum is not attained because $e^x \neq 0$ for every finite x . (4) Since $f'(x) = e^x > 0$, there are no first-order critical points. (5) There is no global minimizer.

Question 4

$f(x) = x^3 - 3x$ for $x \in \mathbb{R}$.

Solution

(1) As $x \rightarrow -\infty$, $x^3 - 3x \rightarrow -\infty$, so the function is not bounded below. (2) $\inf_{x \in \mathbb{R}} f(x) = -\infty$. (3) The infimum is not attained. (4) Since $f'(x) = 3(x^2 - 1)$, the first-order critical points are $x = \pm 1$. (5) There is no global minimizer.

Question 5

$f(x) = x^4 - 2x^2$ for $x \in \mathbb{R}$.

Solution

(1) Completing the square gives $f(x) = (x^2 - 1)^2 - 1 \geq -1$, so the function is bounded below. (2) $\inf_{x \in \mathbb{R}} f(x) = -1$. (3) The infimum is attained at $x = \pm 1$. (4) Since $f'(x) = 4x(x^2 - 1)$, the first-order critical points are $x = 0, \pm 1$. (5) The global minimizers are exactly $x = \pm 1$.

Question 6

$f(x, y) = x^2 - y^2$ for $(x, y) \in \mathbb{R}^2$.

Solution

(1) Along $(x, y) = (0, t)$, $f(0, t) = -t^2 \rightarrow -\infty$, so the function is not bounded below. (2) $\inf_{(x, y) \in \mathbb{R}^2} f(x, y) = -\infty$. (3) The infimum is not attained. (4) Since $\nabla f(x, y) = (2x, -2y)$, the only first-order critical point is $(0, 0)$. (5) There is no global minimizer.

Question 7

(G) $f(x, y) = x^2 + e^y$ for $(x, y) \in \mathbb{R}^2$.

Solution

(1) Since $x^2 + e^y > 0$, the function is bounded below. (2) Along $(x, y) = (0, t)$ with $t \rightarrow -\infty$, $f(0, t) = e^t \rightarrow 0$, so $\inf f = 0$. (3) The infimum is not attained because $x^2 + e^y > 0$ at every finite point. (4) Since $\nabla f(x, y) = (2x, e^y)$ never vanishes, there are no first-order critical points. (5) There is no global minimizer.

Question 8

(G) $f(x, y) = x^4 + y^4 - 2x^2 - 4y^2$ for $(x, y) \in \mathbb{R}^2$.

Solution

(1) We can write

$$f(x, y) = (x^2 - 1)^2 + (y^2 - 2)^2 - 5 \geq -5.$$

Hence the function is bounded below. (2) $\inf f = -5$. (3) The infimum is attained at the four points $(\pm 1, \pm\sqrt{2})$. (4) The gradient equations are $4x(x^2 - 1) = 0$ and $4y(y^2 - 2) = 0$, so the nine first-order critical points satisfy $x \in \{0, \pm 1\}$ and $y \in \{0, \pm\sqrt{2}\}$. (5) The global minimizers are exactly $(\pm 1, \pm\sqrt{2})$.

Question 9

(G) $f(x, y) = (xy - 1)^2 + x^2$ for $(x, y) \in \mathbb{R}^2$.

Solution

(1) Since $f(x, y) = (xy - 1)^2 + x^2 \geq 0$, the function is bounded below. (2) Along $(x, y) = (\varepsilon, 1/\varepsilon)$, $f(x, y) = \varepsilon^2 \rightarrow 0$ as $\varepsilon \rightarrow 0$, so $\inf f = 0$. (3) The infimum is not attained, because value zero would require both $x = 0$ and $xy = 1$. (4) The stationarity equations are

$$2y(xy - 1) + 2x = 0, \quad 2x(xy - 1) = 0.$$

The second equation gives either $x = 0$ or $xy = 1$. The latter is incompatible with the first equation, while $x = 0$ forces $y = 0$. Hence $(0, 0)$ is the unique first-order critical point. (5) There is no global minimizer.

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

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

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

Question 1

Minimize $f(x) = \frac{1}{2}(x-4)^2 + \frac{1}{2}(2x+1)^2$ for $x \in \mathbb{R}$.

Solution

(1) The derivatives are $f'(x) = 5x - 2$ and $f''(x) = 5$. (2) Solving $f'(x) = 0$ gives the unique first-order critical point $x^* = 2/5$. (3) Since $f''(x) = 5 > 0$ for every x , the function is convex. Therefore $x^* = 2/5$ is globally optimal.

Question 2

Minimize

$$f(x_1, x_2) = \frac{1}{2}(x_1 - 1)^2 + \frac{1}{2}(x_2 + 2)^2 + \frac{1}{2}(x_1 + x_2)^2.$$

Solution

(1) The gradient and Hessian are

$$\nabla f(x_1, x_2) = \begin{bmatrix} 2x_1 + x_2 - 1 \\ x_1 + 2x_2 + 2 \end{bmatrix}, \quad \nabla^2 f(x_1, x_2) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

(2) Solving $\nabla f(x_1, x_2) = 0$ gives $x^* = (4/3, -5/3)$. (3) The Hessian is positive definite because

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}^T \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = 2z_1^2 + 2z_2^2 + 2z_1z_2 = z_1^2 + z_2^2 + (z_1 + z_2)^2 \geq 0.$$

Hence f is convex, so x^* is globally optimal.

Question 3

Minimize $f(x) = \frac{1}{2}(x-3)^2 - 2\log x$ for $x > 0$.

Solution

(1) On $x > 0$,

$$f'(x) = x - 3 - \frac{2}{x}, \quad f''(x) = 1 + \frac{2}{x^2}.$$

(2) Solving $f'(x) = 0$ gives $x^2 - 3x - 2 = 0$. Its only positive root is $x^* = (3 + \sqrt{17})/2$. (3) Since $f''(x) > 0$ for every $x > 0$, the objective is convex on its domain. Therefore x^* is globally optimal.

Question 4

Minimize

$$f(x_1, x_2) = \frac{1}{2}(x_1 - 2)^2 + \frac{1}{2}(x_2 - 1)^2 - \log x_1 - 2 \log x_2,$$

on the domain $x_1 > 0, x_2 > 0$.

Solution

(1) The gradient and Hessian are

$$\nabla f(x_1, x_2) = \begin{bmatrix} x_1 - 2 - x_1^{-1} \\ x_2 - 1 - 2x_2^{-1} \end{bmatrix}, \quad \nabla^2 f(x_1, x_2) = \begin{bmatrix} 1 + x_1^{-2} & 0 \\ 0 & 1 + 2x_2^{-1} \end{bmatrix} \text{diag}(1 + x_1^{-2}, 1 + 2x_2^{-2}).$$

(2) The first-order equations are $x_1^2 - 2x_1 - 1 = 0$ and $x_2^2 - x_2 - 2 = 0$. Their positive solutions are $x_1^* = 1 + \sqrt{2}$ and $x_2^* = 2$. (3) The Hessian is positive definite everywhere on $x_1 > 0, x_2 > 0$, because

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}^T \begin{bmatrix} 1 + x_1^{-2} & 0 \\ 0 & 1 + 2x_2^{-1} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = (1 + x_1^{-2})z_1^2 + (1 + 2x_2^{-2})z_2^2 \geq 0,$$

so the objective is convex. Hence $x^* = (1 + \sqrt{2}, 2)$ is globally optimal.

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

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

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

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

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 .

Question 1

Minimize $f(x) = x^4 + 2x^3$ for $x \in \mathbb{R}$.

Solution

(1) The derivatives are

$$f'(x) = 4x^3 + 6x^2, \quad f''(x) = 12x^2 + 12x.$$

(2) The first-order critical points are $x = 0$ and $x = -3/2$. Since $f''(0) = 0$ and $f''(-3/2) = 9 > 0$, both are second-order critical points. (3) Since $f''(-1/2) = -3 < 0$, the function is not convex. (4) The objective is coercive, so a global minimizer must be a second-order critical point. Their objective values are $f(0) = 0$ and $f(-3/2) = -27/16$. Therefore the unique global minimizer is $x^* = -3/2$, with minimum value $-27/16$.

Question 2

Minimize $f(x) = \frac{1}{4}\|x\|_2^4 - \frac{1}{2}\|x\|_2^2$ for $x \in \mathbb{R}^2$.

Solution

(1) The gradient and Hessian are

$$\nabla f(x) = (\|x\|_2^2 - 1)x, \quad \nabla^2 f(x) = (\|x\|_2^2 - 1)I + 2xx^T.$$

(2) The first-order critical points are $x = 0$ and every x with $\|x\|_2 = 1$. At the origin, $\nabla^2 f(0) = -I \not\succeq 0$. On the unit circle, $\nabla^2 f(x) = 2xx^T \succeq 0$. Hence the second-order critical points are exactly the unit circle. (3) Since $\nabla^2 f(0) = -I \not\succeq 0$, the function is not convex. (4) Every second-order critical point has objective value $-1/4$. Since the objective is coercive, the entire unit circle is the set of global minimizers, and the minimum value is $-1/4$.

Question 3

Minimize $f(x, y) = (x^2 - 1)^2 + y^2$ for $(x, y) \in \mathbb{R}^2$.

Solution

(1) The gradient and Hessian are

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

(2) The first-order critical points are $(0, 0)$ and $(\pm 1, 0)$. At the origin, the Hessian is $\begin{bmatrix} -4 & 0 \\ 0 & 2 \end{bmatrix} \not\succeq 0$; at $(\pm 1, 0)$, it is $\begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix} \succ 0$. Hence the only second-order critical points are $(\pm 1, 0)$. (3) The Hessian at the origin is not positive semidefinite, so f is not convex. (4) Both second-order critical points have objective value zero. Since the objective is coercive, $(\pm 1, 0)$ are exactly the global minimizers, with minimum value zero.

Question 4

(G) Let $C = \text{diag}(1, 3)$ and minimize

$$f(x) = x^T C x + (\|x\|_2^2 - 1)^2, \quad x \in \mathbb{R}^2.$$

Solution

(1) The gradient and Hessian are

$$\nabla f(x) = 2Cx + 4(\|x\|_2^2 - 1)x, \quad \nabla^2 f(x) = 2C + 4(\|x\|_2^2 - 1)I + 8xx^T.$$

(2) The stationarity equations are $(4\|x\|_2^2 - 2)x_1 = 0$ and $(4\|x\|_2^2 + 2)x_2 = 0$. Thus the first-order critical points are 0 and $\pm(1/\sqrt{2}, 0)$. At the origin, $\nabla^2 f(0) = \text{diag}(-2, 2) \not\succeq 0$. At the other two points, $\nabla^2 f = 4I \succ 0$. Hence the second-order critical points are exactly $\pm(1/\sqrt{2}, 0)$. (3) Since the Hessian at the origin is not positive semidefinite, the objective is not convex. (4) Both second-order critical points have objective value $3/4$. Since the objective is coercive, they are exactly the global minimizers, with minimum value $3/4$.

Question 5

(G) Fix a nonzero vector $z \in \mathbb{R}^n$ with $n \geq 2$ and minimize

$$f(x) = \|xx^T - zz^T\|_F^2, \quad x \in \mathbb{R}^n.$$

Hint: first show that

$$f(x) = \|x\|_2^4 - 2(x^T z)^2 + \|z\|_2^4.$$

Solution

(1) Expanding the objective gives

$$f(x) = \|x\|_2^4 - 2(x^T z)^2 + \|z\|_2^4,$$

and therefore

$$\nabla f(x) = 4(\|x\|_2^2 x - (x^T z)z), \quad \nabla^2 f(x) = 4(2xx^T + \|x\|_2^2 I - zz^T).$$

(2) If a nonzero critical point had $x^T z = 0$, stationarity would force $x = 0$, a contradiction. Hence $x^T z \neq 0$, and the stationarity equation shows that x is parallel to z . The first-order critical points are therefore 0 and $\pm z$. At the origin, $z^T \nabla^2 f(0)z = -4\|z\|_2^4 < 0$, so it is not second-order critical. At $x = \pm z$,

$$\nabla^2 f(x) = 4(\|z\|_2^2 I + zz^T) \succ 0.$$

Thus the second-order critical points are exactly $\pm z$. (3) The negative curvature at the origin proves that f is not convex. (4) Both second-order critical points have objective value zero. Since the objective is coercive, $\pm z$ are exactly the global minimizers, with minimum value zero.

Question 6

(G) Let $C = C^T \succeq 0$ have eigenvalues $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n \geq 0$ and a unit eigenvector u_1 for λ_1 . Minimize

$$f(x) = \|xx^T - C\|_F^2, \quad x \in \mathbb{R}^n.$$

Find all critical points as well as all global minimizers. Hint: diagonalize C and use

$$\nabla f(x) = 4(\|x\|_2^2 x - Cx).$$

Solution

(1) Expanding the objective gives

$$f(x) = \|x\|_2^4 - 2x^T Cx + \|C\|_F^2, \quad \nabla f(x) = 4(\|x\|_2^2 x - Cx),$$

and

$$\nabla^2 f(x) = 4(2xx^T + \|x\|_2^2 I - C).$$

(2) The stationarity equation is $Cx = \|x\|_2^2 x$. Hence the first-order critical points are the origin and, for every positive eigenvalue λ of C , every vector in its eigenspace with norm $\sqrt{\lambda}$. At the origin, $u_1^T \nabla^2 f(0)u_1 = -4\lambda_1 < 0$. At a critical point associated with $\lambda < \lambda_1$, orthogonality of distinct eigenspaces gives

$$u_1^T \nabla^2 f(x)u_1 = 4(\lambda - \lambda_1) < 0.$$

For $x = \pm\sqrt{\lambda_1}u_1$,

$$\nabla^2 f(x) = 4(2\lambda_1 u_1 u_1^T + \lambda_1 I - C) \succ 0,$$

because $\lambda_1 > \lambda_2$. Thus the only second-order critical points are $\pm\sqrt{\lambda_1}u_1$. (3) The negative curvature at the origin proves that the objective is not convex. (4) At either second-order critical point, $f(x) = \|C\|_F^2 - \lambda_1^2$. Since the objective is coercive, the global minimizers are exactly $\pm\sqrt{\lambda_1}u_1$, with minimum value $\|C\|_F^2 - \lambda_1^2$.

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable and satisfies the preconditioned smoothness condition

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

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

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

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

$$\frac{1}{2} \|\nabla f(x)\|_B^2 \geq \mu(f(x) - f^*), \quad (\text{PL})$$

then preconditioned gradient descent converges exponentially.

Question 1

Descent lemma. Using the fundamental theorem of calculus and the weighted Cauchy–Schwarz inequality $|a^T b| \leq \|a\|_B \|b\|_{B^{-1}}$, prove the quadratic upper-bound

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|_{B^{-1}}^2.$$

Prove (DL) by substituting $x_{k+1} = x_k - \alpha B \nabla f(x_k)$.

Solution

The fundamental theorem of calculus gives

$$f(y) - f(x) - \nabla f(x)^T (y - x) = \int_0^1 (\nabla f(x + t(y - x)) - \nabla f(x))^T (y - x) dt.$$

Weighted Cauchy–Schwarz and smoothness imply

$$|(\nabla f(x + t(y - x)) - \nabla f(x))^T (y - x)| \leq Lt \|y - x\|_{B^{-1}}^2.$$

Integrating over $t \in [0, 1]$ proves

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{L}{2} \|y - x\|_{B^{-1}}^2.$$

Now substitute $y = x_k - \alpha B \nabla f(x_k)$. Since $\|B \nabla f(x_k)\|_{B^{-1}}^2 = \|\nabla f(x_k)\|_B^2$,

$$f(x_{k+1}) \leq f(x_k) - \alpha \left(1 - \frac{L\alpha}{2}\right) \|\nabla f(x_k)\|_B^2,$$

which is (DL).

Question 2

Gradient convergence. Starting from (DL), prove that the step-size $\alpha = \frac{1}{L}$ guarantees the following first-order complexity bound

$$\min_{0 \leq k < N} \|\nabla f(x_k)\|_B^2 \leq \frac{2L(f(x_0) - f^*)}{N}.$$

Solution

With $\alpha = 1/L$, (DL) gives

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|_B^2.$$

Summing from $k = 0$ to $N - 1$ yields

$$\frac{1}{2L} \sum_{k=0}^{N-1} \|\nabla f(x_k)\|_B^2 \leq f(x_0) - f(x_N) \leq f(x_0) - f^*.$$

The smallest of the first N squared gradient norms is no larger than their average, so

$$\min_{0 \leq k < N} \|\nabla f(x_k)\|_B^2 \leq \frac{2L(f(x_0) - f^*)}{N}.$$

Question 3

PL inequality. Assume that f is μ -strongly convex and attains its minimum at $f(x^*) = \min_x f(x)$. Starting from the preconditioned strong-convexity condition

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \frac{\mu}{2} \|y - x\|_{B^{-1}}^2,$$

prove that f satisfies (PL) with the same μ .

Solution

Apply strong convexity with $y = x^*$:

$$f(x) - f^* \leq \nabla f(x)^T(x - x^*) - \frac{\mu}{2} \|x - x^*\|_{B^{-1}}^2.$$

Weighted Cauchy--Schwarz and the scalar inequality $ab - \frac{\mu}{2}b^2 \leq a^2/(2\mu)$ give

$$f(x) - f^* \leq \|\nabla f(x)\|_B \|x - x^*\|_{B^{-1}} - \frac{\mu}{2} \|x - x^*\|_{B^{-1}}^2 \leq \frac{1}{2\mu} \|\nabla f(x)\|_B^2.$$

Rearranging proves $\frac{1}{2} \|\nabla f(x)\|_B^2 \geq \mu(f(x) - f^*)$, which is (PL) with the same μ .

Question 4

Linear convergence. Suppose that f satisfies (PL) for some $\mu > 0$. Use (DL) to prove that gradient descent converges exponentially

$$f(x_{k+1}) - f^* \leq \left(1 - \frac{\mu}{L}\right) (f(x_k) - f^*),$$

with a suitable choice of stepsize $\alpha > 0$. State what that “suitable choice” of α is.

Solution

Choose $\alpha = 1/L$. Then (DL) and (PL) give

$$f(x_{k+1}) - f^* \leq f(x_k) - f^* - \frac{1}{2L} \|\nabla f(x_k)\|_B^2 \leq \left(1 - \frac{\mu}{L}\right) (f(x_k) - f^*).$$

Iterating the inequality yields

$$f(x_N) - f^* \leq \left(1 - \frac{\mu}{L}\right)^N (f(x_0) - f^*).$$

Question 5

(G) Convex objective convergence. Assume that f is convex and attains its minimum at $f(x^*) = \min_x f(x)$. Starting from the first-order convexity definition

$$f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

and (DL), prove that the step-size $\alpha = \frac{1}{L}$ implies the telescope identity

$$f(x_{k+1}) - f^* \leq \frac{L}{2} (\|x_k - x^*\|_{B^{-1}}^2 - \|x_{k+1} - x^*\|_{B^{-1}}^2).$$

Then, use monotonicity $f(x_0) \geq f(x_1) \geq f(x_2) \geq \dots$ to prove convergence to the optimal value

$$f(x_N) - f^* \leq \frac{L\|x_0 - x^*\|_{B^{-1}}^2}{2N}.$$

Solution

Let $g_k = \nabla f(x_k)$ and $D_k = \|x_k - x^*\|_{B^{-1}}^2$. Since $x_{k+1} = x_k - L^{-1}Bg_k$, direct expansion gives

$$D_{k+1} = D_k - \frac{2}{L}g_k^T(x_k - x^*) + \frac{1}{L^2}\|g_k\|_B^2.$$

Convexity implies $f(x_k) - f^* \leq g_k^T(x_k - x^*)$, so

$$\frac{L}{2}(D_k - D_{k+1}) \geq f(x_k) - f^* - \frac{1}{2L}\|g_k\|_B^2.$$

By (DL) with $\alpha = 1/L$, the right-hand side is at least $f(x_{k+1}) - f^*$. Hence

$$f(x_{k+1}) - f^* \leq \frac{L}{2}(D_k - D_{k+1}).$$

Summing from $k = 0$ to $N - 1$ gives

$$\sum_{k=0}^{N-1} (f(x_{k+1}) - f^*) \leq \frac{L}{2}D_0.$$

Since (DL) makes the objective values nonincreasing, each term in the sum is at least $f(x_N) - f^*$. Therefore

$$f(x_N) - f^* \leq \frac{L\|x_0 - x^*\|_{B^{-1}}^2}{2N}.$$

Question 6

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

$$f(x_k - \alpha B \nabla f(x_k)) \leq f(x_k) - c_1 \alpha \|\nabla f(x_k)\|_B^2 \quad (\text{AC})$$

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

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

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

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

Solution

By (DL), every trial step satisfies

$$f(x_k - \alpha B \nabla f(x_k)) \leq f(x_k) - \alpha \left(1 - \frac{L\alpha}{2}\right) \|\nabla f(x_k)\|_B^2.$$

Therefore (AC) is guaranteed whenever $1 - L\alpha/2 \geq c_1$, equivalently whenever $\alpha \leq 2(1 - c_1)/L$. Since the trial steps tend to zero, backtracking terminates. If $\alpha = 1$ is accepted, then the claimed lower bound is immediate. Otherwise the preceding trial α/c_2 was rejected. Because every step at most $2(1 - c_1)/L$ must be accepted, rejection implies

$$\frac{\alpha}{c_2} > \frac{2(1 - c_1)}{L}, \quad \text{so} \quad \alpha > \frac{2c_2(1 - c_1)}{L}.$$

Hence every accepted step obeys

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

Finally, (AC) and this lower bound give

$$f(x_{k+1}) \leq f(x_k) - c_1 \underline{\alpha} \|\nabla f(x_k)\|_B^2 = f(x_k) - \min \left\{ c_1, \frac{2c_1 c_2 (1 - c_1)}{L} \right\} \|\nabla f(x_k)\|_B^2.$$