Problem 1

Consider the problem of minimizing

$$f(x) = 4x - \ln(x), \quad x > 0,$$

by using the Newton's method with the step-size $\gamma = 1$.

(a) What is the minimizer x^* of f?

(b) Find an interval $I \in \mathbb{R}$ such that for any $x^0 \in I$, the Newton's method converges to x^*. Show that the order of convergence is at least quadratic.

Solution 1

(a) The first- and second-order derivatives of f are given by

$$f'(x) = 4 - \frac{1}{x}, \qquad f''(x) = \frac{1}{x^2} > 0.$$

We note that $f'(1/4) = 0$ and $f''(x) > 0$ for all $x > 0$. We directly conclude that $x^* = 1/4$.

(b) Newton's update is given by

$$x^+ = x - \frac{f'(x)}{f''(x)} = x - x^2 \cdot (4 - \frac{1}{x}) = -4x^2 + 2x.$$

Consider

$$\delta(x) = \frac{|x^+ - x^*|}{|x - x^*|} = \frac{|-4x^2 + 2x - \frac{1}{4}|}{|x - \frac{1}{4}|} = \frac{4\,(x - \frac{1}{4})^2}{|x - \frac{1}{4}|} = |4x - 1|.$$

Note that if $0 < x < \frac{1}{2}$ then $\delta \in (0, 1)$, which means that the method brings the next iterate closer to the solution $x^* = \frac{1}{4}$. Therefore, $I = (0, \frac{1}{2})$. Consider

$$\beta = \limsup_{t\to\infty} \frac{|x^{t+1} - x^*|}{|x^t - x^*|^2} = \limsup_{t\to\infty} \frac{4(x^t - \frac{1}{4})^2}{(x^t - \frac{1}{4})^2} = 4 < \infty.$$

We conclude that the order of convergence is at least *quadratic*.

Problem 2

Consider a set of $m+1$, possibly nonconvex, functions $\{f_0(\boldsymbol{x}), f_1(\boldsymbol{x}), \dots, f_m(\boldsymbol{x})\}$ defined over the whole $\mathbb{R}^n$ and the function $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined as

$$L(\boldsymbol{x}, \boldsymbol{\theta}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \theta_i f_i(\boldsymbol{x}).$$

Show that the following function over $\boldsymbol{\theta} \in \mathbb{R}^m$ is always convex

$$g(\boldsymbol{\theta}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{L(\boldsymbol{x}, \boldsymbol{\theta})\}.$$

Hint: You may rely on the definition of convexity and/or the properties of convex functions.

Solution 2

First note that $L(\boldsymbol{x}, \boldsymbol{\theta})$ is linear in $\boldsymbol{\theta} \in \mathbb{R}^m$, which means that it is convex along $\boldsymbol{\theta}$. Hence, for $\alpha \in [0, 1]$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^m$, we have

$$\begin{aligned}
g(\alpha\boldsymbol{\theta}_1 + (1-\alpha)\boldsymbol{\theta}_2) &= \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{L(\boldsymbol{x}, \alpha\boldsymbol{\theta}_1 + (1-\alpha)\boldsymbol{\theta}_2)\} \\
&\leq \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{\alpha L(\boldsymbol{x}, \boldsymbol{\theta}_1) + (1-\alpha) L(\boldsymbol{x}, \boldsymbol{\theta}_2)\} \\
&\leq \alpha \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{L(\boldsymbol{x}, \boldsymbol{\theta}_1)\} + (1-\alpha) \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{L(\boldsymbol{x}, \boldsymbol{\theta}_2)\} \\
&\leq \alpha g(\boldsymbol{\theta}_1) + (1-\alpha) g(\boldsymbol{\theta}_2),
\end{aligned}$$

where in the first inequality we used the convexity of L along $\boldsymbol{\theta}$ and in the second the fact that the supremum gets only bigger if it is taken separately. Note that this is a general rule that pointwise supremum or maximum preserves convexity.

Problem 3

Consider the problem of minimizing the following function over $\mathbb{R}^3$

$$f(\boldsymbol{x}) = \frac{1}{2}\exp\left(\frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d\right), \quad \text{where} \quad \boldsymbol{Q} = \begin{bmatrix} 2 & -1 & \theta \\ -1 & 2 & -1 \\ \theta & -1 & 2 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad d = 5.$$

(a) For what values of $\theta \in \mathbb{R}$ is the problem strictly convex?

(b) Is $\boldsymbol{g} = \boldsymbol{c} - \boldsymbol{Q}\boldsymbol{x}$ a descent direction for f at any $\boldsymbol{x} \in \mathbb{R}^3$?

(c) What is the minimizer $\boldsymbol{x}^*$ of f for $\theta = 0$?

Solution 3

(a) Consider the function $\varphi(t) = \frac{1}{2}\mathrm{e}^t$, which is strictly increasing and convex. Hence, if the function $h(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d$ is strictly convex, then the composite function $f = \varphi \circ h$ will also be strictly convex. In order to apply the *Sylvester's criterion*, we compute the principal minors of the Hessian $\mathsf{H}h(\boldsymbol{x}) = \boldsymbol{Q}$ as follows

$$\Delta_1 = 2, \quad \Delta_2 = \det\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 4 - 1 = 3, \quad \text{as well as}$$
$$\Delta_3 = \det \boldsymbol{Q} = 2(4-1) + 1(-2+\theta) + \theta(1-2\theta) = -2(\theta-2)(\theta+1).$$

We have $\Delta_1 > 0$ and $\Delta_2 > 0$. Hence, all one needs for strong convexity is for $\Delta_3 > 0$, which yields $\theta \in (-1, 2)$.

(b) The gradient of f is given by

$$\nabla f(\boldsymbol{x}) = \frac{1}{2}\mathrm{e}^{\frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d}(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}).$$

Hence, one can verify that

$$\nabla f(\boldsymbol{x})^\mathsf{T}\boldsymbol{g} = \frac{1}{2}\mathrm{e}^{\frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d}(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c})^\mathsf{T}\boldsymbol{g} = -\frac{1}{2}\mathrm{e}^{\frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d}\|\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}\|^2 < 0,$$

which means that $\boldsymbol{g}$ is indeed a descent direction.

(c) Note that since $\varphi(t) = \frac{1}{2}\mathrm{e}^t$ is strictly increasing, the minimizer of f can be found by directly computing the minimizers of the quadratic function $h(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d$. We write down and solve the *normal equations* for the quadratic

$$\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{c} = \boldsymbol{0} \quad \Rightarrow \quad \begin{cases} 2x_1 - x_2 = 1 \\ -x_1 + 2x_2 - x_3 = 0 \\ -x_2 + 2x_3 = 0 \end{cases} \quad \Rightarrow \quad \boldsymbol{x}^* = \left(\frac{3}{4}, \frac{1}{2}, \frac{1}{4}\right).$$

Problem 4

Consider the problem of minimizing the following function over $\mathbb{R}^n$:

$$f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \quad \text{where} \quad \boldsymbol{A} \in \mathbb{R}^{m \times n} \quad \text{and} \quad \boldsymbol{b} \in \mathbb{R}^m.$$

(a) Provide the expressions for the gradient and the Hessian of f. Comment on the convexity of f.

(b) Find the minimizer $\boldsymbol{x}^* \in \mathbb{R}^n$ of f, when $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$ is nonsingular.

(c) Assume that $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$ is nonsingular. Find a matrix $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ such that for any $\boldsymbol{x}^0 \in \mathbb{R}^n$ the following single iteration

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 - \boldsymbol{M}\nabla f(\boldsymbol{x}^0),$$

converges to the minimizer $\boldsymbol{x}^* \in \mathbb{R}^n$.

Hint: Note this result must be valid for any starting point $\boldsymbol{x}^0 \in \mathbb{R}^n$.

Solution 4

(a) The expressions are as follows

$$\nabla f(\boldsymbol{x}) = \boldsymbol{A}^\mathsf{T}(\boldsymbol{A} - \boldsymbol{b}) \quad \text{and} \quad \mathsf{H}f(\boldsymbol{x}) = \boldsymbol{A}^\mathsf{T}\boldsymbol{A}.$$

Note that for any $\boldsymbol{x} \in \mathbb{R}^n$, we have that

$$\boldsymbol{x}^\mathsf{T}\mathsf{H}f(\boldsymbol{x})\boldsymbol{x} = \boldsymbol{x}^\mathsf{T}\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^\mathsf{T}(\boldsymbol{A}\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x}\|^2 \geq 0.$$

The function is hence at least convex.

(b) The minimizer $\boldsymbol{x}^*$ of f must satisfy the normal equations

$$\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}^\mathsf{T}\boldsymbol{b},$$

Since function f is convex and $\boldsymbol{A}^\mathsf{T}\boldsymbol{A}$ is nonsingular, we have the unique solution

$$\boldsymbol{x}^* = [\boldsymbol{A}^\mathsf{T}\boldsymbol{A}]^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{b}$$

(c) We expand the expression by plugging in the gradient from (a)

$$\boldsymbol{x}^1 = \boldsymbol{x}^0 - \boldsymbol{M}\nabla f(\boldsymbol{x}^0) = \boldsymbol{x}^0 - \boldsymbol{M}\boldsymbol{A}^\mathsf{T}\boldsymbol{A}\boldsymbol{x}^0 + \boldsymbol{M}\boldsymbol{A}^\mathsf{T}\boldsymbol{b} = (\boldsymbol{I} - \boldsymbol{M}\boldsymbol{A}^\mathsf{T}\boldsymbol{A})\boldsymbol{x}^0 + \boldsymbol{M}\boldsymbol{A}^\mathsf{T}\boldsymbol{b} = [\boldsymbol{A}^\mathsf{T}\boldsymbol{A}]^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{b}.$$

Since the result must hold for any $\boldsymbol{x}^0 \in \mathbb{R}^n$, we must have

$$\boldsymbol{I} = \boldsymbol{M}\boldsymbol{A}^\mathsf{T}\boldsymbol{A} \quad \Leftrightarrow \quad \boldsymbol{M} = [\boldsymbol{A}^\mathsf{T}\boldsymbol{A}]^{-1}.$$

Bonus Problem

Consider a function $\boldsymbol{y} = \mathsf{G}(\boldsymbol{x})$ that takes in a vector $\boldsymbol{x} \in \mathbb{R}^n$ and produces another vector $\boldsymbol{y} \in \mathbb{R}^n$.

- We say that G is *Lipschitz continuous* with constant $L > 0$, when

$$\|\mathsf{G}(\boldsymbol{x}_1) - \mathsf{G}(\boldsymbol{x}_2)\| \leq L\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|, \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n.$$

  When $L < 1$, we say that G is *contractive*.

- We say that G is *strongly monotone* with constant $M > 0$, when

$$(\mathsf{G}(\boldsymbol{x}_1) - \mathsf{G}(\boldsymbol{x}_2))^{\mathsf{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2) \geq M\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2, \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$$

(a) Show that for $f \in \mathcal{S}^1_{M,L}(\mathbb{R}^n)$, the gradient $\nabla f(\boldsymbol{x})$ is strongly monotone with constant M.

Hint: Use the definition of strong convexity at two different points $\boldsymbol{x}_1 \in \mathbb{R}^n$ and $\boldsymbol{x}_2 \in \mathbb{R}^n$.

(b) Show that for $f \in \mathcal{S}^1_{L,M}(\mathbb{R}^n)$ there exists $\gamma > 0$, such that the function $\mathsf{G}(\boldsymbol{x}) = \boldsymbol{x} - \gamma\nabla f(\boldsymbol{x})$ is contractive.

Hint: Use the fact that ∇f is both Lipschitz continuous and strongly monotone.

(c) Show that when $\mathsf{G}(\boldsymbol{x}) = \boldsymbol{x} - \gamma\nabla f(\boldsymbol{x})$ is contractive with a Lipschitz constant $0 < c < 1$, the iterates generated by the gradient method converge linearly to the minimizer $\boldsymbol{x}^* \in \mathbb{R}^n$ as follows

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\| \leq c^t\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|, \quad \forall t \geq 1.$$

Bonus Solution

(a) We apply the definition of strong convexity twice as follows

$$f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}_2) + \nabla f(\boldsymbol{x}_2)^{\mathsf{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \frac{M}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2$$
$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^{\mathsf{T}}(\boldsymbol{x}_2 - \boldsymbol{x}_1) + \frac{M}{2}\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|^2,$$

where the inequalities are true for any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$. The result is obtained by first adding these two inequalities and then simplifying the result

$$0 \geq (\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2))^{\mathsf{T}}(\boldsymbol{x}_2 - \boldsymbol{x}_1) + M\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2$$
$$\Rightarrow \quad (\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2))^{\mathsf{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2) \geq M\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2.$$

Bonus Solution (cont)

(b) Consider

$$\begin{aligned}
\|\mathsf{G}(\boldsymbol{x}_1) - \mathsf{G}(\boldsymbol{x}_2)\|^2 &= \|(\boldsymbol{x}_1 - \boldsymbol{x}_2) - \gamma(\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2))\|^2 \\
&= \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2 - 2\gamma(\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2))^{\mathsf{T}}(\boldsymbol{x}_1 - \boldsymbol{x}_2) + \gamma^2\|\nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2)\|^2 \\
&\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2 - 2\gamma M\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2 + \gamma^2 L^2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2 \\
&= (1 - 2\gamma M + \gamma^2 L^2)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2,
\end{aligned}$$

where for the inequality we used both the Lipschitz continuity and the strong monotonicity of the gradient. Hence for any $\gamma > 0$ such that

$$0 < 1 - 2\gamma M + \gamma^2 L^2 < 1 \quad \Rightarrow \quad 0 < \gamma < \frac{2M}{L^2},$$

the function G is contractive.

(c) For any γ in the interval given in (b), we have that

$$c = (1 - 2\gamma M + \gamma^2 L) \in (0, 1).$$

Note that at a minimizer $\boldsymbol{x}^*$, we have that

$$\mathsf{G}(\boldsymbol{x}^*) = \boldsymbol{x}^* - \gamma\nabla f(\boldsymbol{x}^*) = \boldsymbol{x}^*,$$

where the last inequality comes from the fact that $\boldsymbol{x}^*$ is the stationary point. Now consider a minimizer $\boldsymbol{x}^*$ and an iteration $t \geq 1$ of the gradient method, then

$$\|\boldsymbol{x}^t - \boldsymbol{x}^*\| = \|\mathsf{G}(\boldsymbol{x}^{t-1}) - \mathsf{G}(\boldsymbol{x}^*)\| \leq c\|\boldsymbol{x}^{t-1} - \boldsymbol{x}^*\| \leq \cdots \leq c^t\|\boldsymbol{x}^0 - \boldsymbol{x}^*\|,$$

hence the iterates generated by the gradient method converge linearly to $\boldsymbol{x}^*$.