5004 Homework 2

RONG Shuo

November 9, 2024

Question 1:

1. For each of the following functions $f(x_1, x_2)$, find all critical points (i.e, all x_1, x_2 such that $\nabla f(x_1, x_2) = \mathbf{0}$).

(a) $f(x_1, x_2) = (4x_1^2 - x_2)^2$
(b) $f(x_1, x_2) = 2x_2^3 - 6x_2^2 + 3x_1^2 x_2$
(c) $f(x_1, x_2) = (x_1 - 2x_2)^4 + 64x_1 x_2$
(d) $f(x_1, x_2) = x_1^2 + 4x_1 x_2 + x_2^2 + x_1 - x_2$

Answer :

(a)

$$\frac{\partial f}{\partial x_1} = 2(4x_1^2 - x_2)(8x_1) = 16x_1(4x_1^2 - x_2)$$
$$\frac{\partial f}{\partial x_2} = 2(4x_1^2 - x_2)(-1) = -2(4x_1^2 - x_2)$$

set the gradient to 0:

$$16x_1(4x_1^2 - x_2) = 0 \tag{1}$$
$$-2(4x_1^2 - x_2) = 0 \tag{2}$$

if $x_1 = 0$, from (2) we can get that $x_2 = 0$ if $x_1 \neq 0$, from equation (1), we know:

$$4x_1^2 = x_2$$

this satisfied $(x_1, x_2) = (0, 0)$ Thus, we can conclude that the critical points are:

$$(x_1, x_2) = (x_1, 4x_1^2), \forall x_1 \in \mathbb{R}.$$

(b)

$$\frac{\partial f}{\partial x_1} = 6x_1 x_2$$
$$\frac{\partial f}{\partial x_2} = 6x_2^2 - 12x_2 + 3x_1^2$$

set the gradient to 0:

$$
\begin{aligned}
6x_1x_2 &= 0 \\
6x_2^2 - 12x_2 + 3x_1^2 &= 0
\end{aligned}
$$

if $x_1 = 0$,

$$
\begin{aligned}
6x_2^2 - 12x_2 &= 0 \\
6x_2(x_2 - 2) &= 0
\end{aligned}
$$

we can conclude that $(x_1, x_2) = (0, 0),$ or $(x_1, x_2) = (0, 2)$.
if $x_2 = 0$,

$$
3x_1^2 = 0x_1 \qquad\qquad = 0
$$

This gives $(x_1, x_2) = (0, 0)$
In conclusion, the critical points is $(0, 0)$ and $(0, 2)$

(c)

$$
\begin{aligned}
\frac{f}{\partial x_1} &= 4(x_1 - 2x_2)^3 + 64x_2 \\
\frac{f}{\partial x_2} &= -8(x_1 - 2x_2)^3 + 64x_1
\end{aligned}
$$

set the gradient to 0:

$$
\begin{aligned}
4(x_1 - 2x_2)^3 + 64x_2 &= 0 \\
(x_1 - 2x_2)^2 &= -16x_2 \\
-8(x_1 - 2x_2)^3 + 64x_1 &= 0 \\
(x_1 - 2x_2)^3 &= 8x_1
\end{aligned}
$$

$$
\begin{aligned}
(x_1 - 2x_2)^3 &= -16x_2 \\
(x_1 - 2x_2)^3 &= 8x_1 \\
-16x_2 &= 8x_1 \\
-2x_2 &= x_1
\end{aligned}
$$

Substituting $-2x_2 = x_1$ to $(x_1 - 2x_2)^2 = -16x_2$, we can get:

$$
\begin{aligned}
64x_2^3 + 16x_2 &= 0 \\
x_2^2 &= \frac{1}{4}
\end{aligned}
$$

Thus, the result is

$$
\begin{aligned}
(x_1, x_2) &= (-1, \frac{1}{2}) \\
(x_1, x_2) &= (1, -\frac{1}{2})
\end{aligned}
$$

(d)

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 2x_1 + 4x_1 + 1 \\
\frac{\partial f}{\partial x_2} &= 4x_2 + 2x_2 - 1
\end{aligned}
$$

set the gradient to 0:

$$2x_1 + 4x_2 + 1 = 0$$
$$4x_2 + 2x_2 - 1 = 0$$

$$x_1 = -\frac{1}{2} - 2x_2$$

Substituting this to second equation.

$$\begin{aligned} 4(-\frac{1}{2} - 2x_2) + 2x_2 - 1 = 0 \\ -2 - 8x_2 + 2x_2 - 1 = 0 \\ x_2 = -\frac{1}{2} \\ x_1 = -\frac{1}{2} + 1 = \frac{1}{2} \end{aligned}$$

Thus, the critical point is $\left(\frac{1}{2}, -\frac{1}{2}\right)$

Question 2:

2. Find the gradient of the following functions, where the space $\mathbb{R}$ and $\mathbb{R}^{n\times n}$ are equipped with the standard inner product.
(a) $f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{x}\|_2^2$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $(\lambda > 0)$ are given.
(b) $f(\boldsymbol{X}) = \boldsymbol{b}^T\boldsymbol{X}\boldsymbol{c}$, where $\boldsymbol{X} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$
(c) $f(\boldsymbol{X}) = \boldsymbol{b}\boldsymbol{X}^T\boldsymbol{X}c$, where $\boldsymbol{X} \in \boldsymbol{R}^{n\times n}$ and $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$

Answer :

(a)

$$\begin{aligned} f(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{Ay} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{y}\|_2^2 \\ f(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{Ay} - \boldsymbol{Ax} + \boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{y} - \boldsymbol{x} + \boldsymbol{x}\|_2^2 \\ f(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x}) + \boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda\|(\boldsymbol{y} - \boldsymbol{x}) + \boldsymbol{x}\|_2^2 \\ f(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x})\|^2 + \langle \boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{Ax} - \boldsymbol{b}\rangle + \frac{1}{2}\|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda\langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{x}\rangle + \lambda\|\boldsymbol{x}\|_2^2 \\ f(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x})\|^2 + \langle (\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{A}^T(\boldsymbol{Ax} - \boldsymbol{b})\rangle + \frac{1}{2}\|\boldsymbol{Ax} - \boldsymbol{b}\|_2^2 + \lambda\|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda\langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{x}\rangle + \lambda\|\boldsymbol{x}\|_2^2 \\ f(\boldsymbol{y}) &= \frac{1}{2}\|\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x})\|^2 + \langle (\boldsymbol{y} - \boldsymbol{x}), \boldsymbol{A}^T(\boldsymbol{Ax} - \boldsymbol{b})\rangle + f(\boldsymbol{x}) + \lambda\|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda\langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{x}\rangle \end{aligned}$$

$$
\begin{aligned}
&\lim_{\|\boldsymbol{y}-\boldsymbol{x}\|_2\to 0} \frac{|f(\boldsymbol{y})-(f(\boldsymbol{x})+\langle \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}),\boldsymbol{y}-\boldsymbol{x}\rangle+\lambda\langle \boldsymbol{y}-\boldsymbol{x},\boldsymbol{x}\rangle)|}{\|\boldsymbol{y}-\boldsymbol{x}\|_2}\\
&=\lim_{\|\boldsymbol{y}-\boldsymbol{x}\|_2\to 0} \frac{\frac{1}{2}\|\boldsymbol{A}(\boldsymbol{y}-\boldsymbol{x})\|^2+\lambda\|\boldsymbol{y}-\boldsymbol{x}\|^2}{\|\boldsymbol{y}-\boldsymbol{x}\|_2}\\
&\le \lim_{\|\boldsymbol{y}-\boldsymbol{x}\|_2\to 0} \frac{\frac{1}{2}\|\boldsymbol{A}\|_2^2\|\boldsymbol{y}-\boldsymbol{x}\|^2+\lambda\|\boldsymbol{y}-\boldsymbol{x}\|^2}{\|\boldsymbol{y}-\boldsymbol{x}\|_2}\\
&=\lim_{\|\boldsymbol{y}-\boldsymbol{x}\|_2\to 0} \frac{1}{2}\|\boldsymbol{A}\|_2^2\|\boldsymbol{y}-\boldsymbol{x}\|+\lambda\|\boldsymbol{y}-\boldsymbol{x}\|\\
&=0
\end{aligned}
$$

In conclusion, the gradient for(a) is $\nabla f(\boldsymbol{x})=\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})+2\lambda\boldsymbol{x}$

Question 3:

3. Let $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$ be given with $\boldsymbol{x}_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$. Assume $N < n$. Consider the ridge regression

$$
\text{minimize}_{\boldsymbol{a}\in\mathbb{R}^N} \sum_{i=1}^{N} (\langle \boldsymbol{\alpha}, \boldsymbol{x}_i\rangle - y_i)^2 + \lambda\|\boldsymbol{a}\|_2^2,
$$

where $\lambda \in \mathbb{R}$ is a regularization parameter, and we set the bias $b = 0$ for simplicity.

(a) Prove that the solution must be in the form of $\boldsymbol{a} = \sum_{i=1}^N c_i\boldsymbol{x}_i$ for some $\boldsymbol{c} = [c_1, c_2, \cdots, c_N]^T \in \mathbb{R}^N$.
(*hint: similar to the proof of the representer theorem.*)

(b) Re-express the minimization in terms of $c \in \mathbb{R}^N$, which has fewer unknowns than the original formulation as $N < n$.

Answer :

Question 4:

4. Let $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} + c$, where $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a symmetric positive semidefinite matrix, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$.
(a) Prove that $\boldsymbol{x}$ is a global minimizer of f if and only if $\boldsymbol{A}\boldsymbol{x} = -\boldsymbol{b}$.
(b) Prove that f is bounded below over $\mathbb{R}^n$ if and only if $\boldsymbol{b} \in \{\boldsymbol{A}\boldsymbol{y} : \boldsymbol{y} \in \mathbb{R}^n\}$.

Answer :

Question 5:

5. We consider the following optimization problem:

$$
\text{minimize}_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) = log\left(\sum_{i=1}^{m} \exp(\boldsymbol{a}_i^T\boldsymbol{x} + b_i)\right) \tag{3}
$$

where $\boldsymbol{a}_1, \cdots \boldsymbol{a}_m \in \mathbb{R}^n$ and $b_1, \cdots b_m \in \mathbb{R}$ are given.

(a) Find the gradient of $f(\boldsymbol{x})$.

(b) If we use gradient descent to solve Problem (1), will it converge to the global minimizer? Please justify your answer.

Answer