Homework 2

Ilia Kamyshev

February 12, 2021

Contents

1 Minimum

Given

$$f(x) = ax^2 + bx + c$$

This is a convex function, so the optimal solution is global and unique:

$$\frac{d}{dx}f(x) = 2ax + b = 0$$

$$x^* = -\frac{b}{2a}$$

The optimal value of $f(x)$ is as follows:

$$f(x^*) = \frac{b^2}{4a} - \frac{b}{a} + c = \frac{b}{a}(\frac{b}{4} - 1) + c$$

For the minimum following holds $\frac{d^2}{dx^2}f(x) > 0$:

$$\frac{d^2}{dx^2}f(x) = 2a > 0 \implies a > 0$$

Answer: the minimum is at $x^* = -\frac{b}{2a}$ and its value is $f(x^*) = \frac{b}{a}(\frac{b}{4} - 1) + c$ for $a > 0$, $b \in \mathbb{R}$ and $c \in \mathbb{R}$.

2 Gradient dimension

Given

$$h(x) = f(Ax),\ f : \mathbb{R}^m \to \mathbb{R},\ A \in \mathbb{R}^{m\times k}$$

Let's assign $y = Ax$ then $h(x) = (f \circ y)(x)$. The total derivative $Dh(\mathbf{x}) = Df(y(\mathbf{x}))Dy(\mathbf{x})$. The gradient is $\nabla(\circ) = (D(\circ))^T$. Thus,

$$Dh(\mathbf{x}) = Df(y(\mathbf{x}))\mathbf{A}$$

$$\nabla_{\mathbf{x}} h(\mathbf{x}) = \mathbf{A}^T (Df(y(\mathbf{x})))^T = \mathbf{A}^T \nabla_y f(y)$$

Since $f : \mathbb{R}^m \to \mathbb{R}$, then the dimension of $\nabla_y f(y)$ is $m \times 1$ (*denominator-layout*, the gradient is a column vector), and from formula above we can conclude that $(k \times m) \times (m \times 1) = k \times 1$

Answer: $k \times 1$

3 Gradient and Hessian

Given

$$f(x) = (x, c)^2, \; x \in \mathbb{R}^m$$

The inner product $(x, c)^2$ can be rewritten as $(x^T c)^2$. Let's assign $y = x^T c$, $g = y^2$ then $f(x) = g(y(x))$ or $f(x) = (g \circ y)(x)$. Thus, applying same technique as before:

1. $Df(x) = Dg(y(x))Dy(x)$. The gradient $\nabla_x f(x) = (Dg(y(x))Dy(x))^T = (Dy(x))^T(Dg(y(x)))^T$. Here $Dg(y(x)) = D(y^2(x)) = 2y(x) = 2x^T c$ and $Dy(x) = c^T$, continuing $(c^T)^T(2x^T c)^T = 2c(c^T x)$

2. The Hessian can be derived with use of Jacobian $H = J(\nabla_x(2c(c^T x)) = 2cc^T$

4 Hessian matrix

Given

$$f(x) = g(Ax + b), \; g : \mathbb{R}^m \to \mathbb{R}, \; \mathbb{A} \in \mathbb{R}^{m \times n}, \; b \in \mathbb{R}^m, \; x \in \mathbb{R}^n$$

Let's assign $y = Ax + b$, then $f(x) = (g \circ y)(x)$. The total derivative is $Df(x) = Dg(y(x))Dy(x) = Dg(y(x))A$. Then, the gradient is $\nabla_x f(x) = A^T(Dg(y(x)))^T$. The Hessian is as before $H = J(\nabla_x f(x)) = A^T A D^2 g(y(x)) = A^T A H(g)$.