Recitation Material: Linear Algebra

10-606

September 2025

Matrix Calculus

1. Let $f(x) = x^T A x$, where $A \in \mathbb{R}^{n \times n}$. Compute the gradient of $f(x)$ with respect to x. What if A is symmetric? **Solution:** The gradient is:

   $$\nabla f(x) = (A + A^\top)x.$$

   To see this, write $x^\intercal A x = \sum_{i,j} x_i x_j a_{ij}$. So

   $$\frac{\partial x^\top A x}{\partial x_k} = \sum_{i \neq k} x_i a_{ik} + \sum_{j \neq k} x_j a_{kj} + 2 x_k a_{kk}, \tag{1}$$

   (consider the sums with no k's, one k and two k's) which is the k-th row of the vector $(A + A^\top)x$. If A is symmetric then $(A + A^\top)x = 2Ax$.

2. Let $A \in \mathbb{R}^{n \times n}$ and $x \in \mathbb{R}^n$. Compute the derivative of $f(x) = x^T A x + b^T x$, where $b \in \mathbb{R}^n$. **Solution:** The gradient of $f(x) = x^T A x + b^T x$ is:

   $$\nabla f(x) = (A + A^\top)x + b.$$

3. For linear regression, the loss function is $L(\beta) = \|y - X\beta\|^2$, where $X \in \mathbb{R}^{m \times n}$ is the design matrix, $\beta \in \mathbb{R}^n$ is the parameter vector, and $y \in \mathbb{R}^m$ is the target vector. Compute the gradient of $L(\beta)$ with respect to β. **Solution:** The loss function can be rewritten as:

   $$L(\beta) = (y - X\beta)^T (y - X\beta).$$

   Taking the gradient with respect to β gives:

   $$\nabla_\beta L(\beta) = -2X^T (y - X\beta).$$

4. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be

   $$f(x, y) = \begin{pmatrix} x^2 y \\ \sin(x + y) \end{pmatrix}.$$

   Compute the Jacobian matrix $J_f(x, y)$.

   Solution.

   $$J_f(x, y) = \begin{pmatrix} \frac{\partial}{\partial x}(x^2 y) & \frac{\partial}{\partial y}(x^2 y) \\ \frac{\partial}{\partial x} \sin(x + y) & \frac{\partial}{\partial y} \sin(x + y) \end{pmatrix} = \begin{pmatrix} 2xy & x^2 \\ \cos(x + y) & \cos(x + y) \end{pmatrix}.$$

5. Let $b \in \mathbb{R}^n$ and $g(x) = (b^\top x)^2$.

   (a) Compute $\nabla g(x)$.

   (b) Compute the Hessian $\nabla^2 g(x)$.

   (c) Is $\nabla^2 g(x)$ positive semidefinite?

   Solution. (a) $\nabla g(x) = 2bb^\top x$ (use question 1). (b) $\nabla^2 g(x) = 2\,bb^\top$. To see this, similarly to question 1 we have,

$$\frac{\partial (b^\top x)^2}{\partial x_k} = \sum_{i \neq k} x_i b_i b_k + \sum_{j \neq k} x_j b_k b_j + 2x_k b_k^2,$$

   so $\frac{\partial^2 (b^\top x)^2}{\partial x_\ell \partial x_k} = 2b_\ell b_k$. (c) Yes. $bb^\top$ is PSD ($y^\top (bb^\top) y = (b^\top y)^2 \geq 0$), so $2bb^\top$ is PSD.

6. Let $f : \mathbb{R}^n \to \mathbb{R}$. Let $J_x(y)$ be the Jacobian of a vector y with respect to the vector x (i.e., $J(y)_{ij} = \partial y_i / \partial x_j$). Show that the Hessian of f obeys

$$H(f) = J(\nabla_x f(x)).$$

   Recall that the Hessian of a function g satisfies $H(g)_{ij} = \partial^2 g / \partial x_i \partial x_j$. **Solution:** This is a matter of unwinding definitions. We have $\nabla f(x) = (\partial f / \partial x_1, \ldots, \partial f / \partial x_n)^\top$, so

$$J(\nabla f(x))_{ij} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j} = H(f)_{ij}.$$

Convexity and Optimization

1. Let $f(x) = x^2 + 3x + 5$. Is f convex? **Solution:** Yes. The second derivative of f is $f''(x) = 2$, which is strictly positive so f is convex.

2. Compute the Hessian of the function $f(x, y) = x^2 + y^2 + xy$. Can we conclude whether or not f is convex? **Solution:** The Hessian matrix of $f(x, y)$ is:

$$H(f) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

   The eigenvalues of $H(f)$ are 3 and 1, both positive, so $f(x, y)$ is convex.

3. Show that the Hessian of $f(x) = x^\top A x$ is $A + A^\top$. **Solution:** Take the second derivative of equation (1) above.

4. Consider the loss function in linear regression $L(\beta) = \|y - X\beta\|^2$. Prove that this loss function is convex. **Solution:** The loss function $L(\beta) = (y - X\beta)^T (y - X\beta)$ is quadratic in β, and its Hessian is $H(L) = 2X^T X$. Since $X^T X$ is positive semi-definite (why?), the Hessian is positive semi-definite, proving that the loss function is convex.

5. Perform two iterations of gradient descent for $f(x) = x^2 + 4x + 4$, starting at $x_0 = 3$, with a learning rate $\alpha = 0.1$. **Solution:** The gradient is $\nabla f(x) = 2x + 4$. At $x_0 = 3$, we have:

$$\nabla f(3) = 2(3) + 4 = 10.$$

The first iteration gives:

$$x_1 = x_0 - \alpha \nabla f(x_0) = 3 - 0.1 \times 10 = 2.$$

The gradient at $x_1 = 2$ is:

$$\nabla f(2) = 2(2) + 4 = 8.$$

The second iteration gives:

$$x_2 = x_1 - \alpha \nabla f(x_1) = 2 - 0.1 \times 8 = 1.2.$$

After two iterations, $x_2 = 1.2$.

6. Consider the function $f(x, y) = x^2 + y^2$. Starting at $(x_0, y_0) = (1, 2)$, perform one iteration of gradient descent with learning rate $\alpha = 0.1$. **Solution:**
The gradient of $f(x, y) = x^2 + y^2$ is $\nabla f = (2x, 2y)$. At $(1, 2)$, the gradient is $(2, 4)$. Updating using gradient descent:

$$x_1 = x_0 - \alpha \frac{\partial f}{\partial x} = 1 - 0.1 \times 2 = 0.8,$$

$$y_1 = y_0 - \alpha \frac{\partial f}{\partial y} = 2 - 0.1 \times 4 = 1.6.$$

Thus, the new point is $(x_1, y_1) = (0.8, 1.6)$.