Essential Prerequisites Review

Linear Algebra and Calculus Fundamentals

Navish Kumar

February 24, 2025

Vector Norms: Definition Recap

Definition (Vector Norm): A function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a norm if it satisfies, for all $x, y \in \mathbb{R}^n$ and scalar $c \in \mathbb{R}$:

1. **Non-negativity:** $\|x\| \geq 0$, and $\|x\| = 0 \iff x = \mathbf{0}$.
2. **Homogeneity:** $\|c\,x\| = |c|\,\|x\|$.
3. **Triangle Inequality:** $\|x + y\| \leq \|x\| + \|y\|$.

Examples:

- L_1 norm: $\|x\|_1 = \sum_i |x_i|$.
- L_2 norm: $\|x\|_2 = \sqrt{\sum_i x_i^2}$.
- L_∞ norm: $\|x\|_\infty = \max_i |x_i|$.

We will see how these measure "size" of vectors in different ways.

Exercise (30s): Vector Norm Properties

Exercise:

1. Let $x = (3, -4, 1)$. Compute $\|x\|_1$, $\|x\|_2$, and $\|x\|_\infty$.
2. Prove that $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$ for $x \in \mathbb{R}^n$. (Hint: Use Cauchy–Schwarz inequality for the second part.)
3. (Optional Challenge) Show that $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$.

Think (30s): Recall definitions and inequalities (Cauchy–Schwarz, etc.).

Solution: Vector Norm Properties I

Solution Outline:

1. **Compute norms for $x = (3, -4, 1)$:**

$$\|x\|_1 = |3| + |-4| + |1| = 3 + 4 + 1 = 8.$$

$$\|x\|_2 = \sqrt{3^2 + (-4)^2 + 1^2} = \sqrt{9 + 16 + 1} = \sqrt{26}.$$

$$\|x\|_\infty = \max\{3, 4, 1\} = 4.$$

2. **Inequality $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\,\|x\|_2$**

LHS:

$$||x||_2^2 = \sum_{i=1}^{N} |x_i|^2 \leq \left(\sum_{i=1}^{N} |x_i|^2 + 2 \cdot \sum_{i,j,i<j} |x_i||x_j| \right) = ||x||_1^2 \qquad (1)$$

RHS: You can prove $\|x\|_1 \leq \sqrt{n}\,\|x\|_2$ via Cauchy–Schwarz:

$$\|x\|_1 = \sum_i |x_i| = \mathbf{1}^T|x| \leq \|\mathbf{1}\|_2\,\||x|\|_2 = \sqrt{n}\,\|x\|_2,$$

Solution: Vector Norm Properties II

where $\mathbf{1}$ is the all-ones vector.
3. $\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$**:** - Similar logic. For $\|x\|_2 \le \sqrt{n}\|x\|_\infty$, note $x_i^2 \le (\|x\|_\infty)^2$, so $\sum x_i^2 \le n(\|x\|_\infty)^2$. - Also, $\max_i |x_i| \le \sqrt{\sum x_i^2}$ because if you pick the largest component, it's at most the Euclidean norm.

Matrix Norms: Definition Recap

Definition (Matrix Norm): A function $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ is a norm if it satisfies analogous properties (non-negativity, homogeneity, triangle inequality).

Examples:

- **Frobenius Norm:**

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\mathrm{trace}(A^\top A)}.$$

- **Operator Norm (L_2 or Spectral Norm):**

$$\|A\|_2 = \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_{\max}(A),$$

  where $\sigma_{\max}(A)$ is the largest singular value of A.

We now practice using these definitions.

Exercise (1 min): Matrix Norms

Exercise:

1. Compute the Frobenius norm of

$$A = \begin{pmatrix} 2 & 1 \\ 0 & -3 \end{pmatrix}.$$

2. Argue why $\|A\|_F \geq \|A\|_2$. (Hint: largest singular value is at most the square root of the sum of the singular values squared.)
3. (Advanced) For $B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, find $\|B\|_2$ by explicitly computing $\sigma_{\max}(B)$ (or equivalently the largest eigenvalue of B if B is symmetric).

Think (1 min): Apply definitions and recall that for a symmetric matrix B, $\|B\|_2 =$ max eigenvalue in absolute value.

Solution: Matrix Norms

Solution Outline:

1. **Frobenius norm of** $A = \begin{pmatrix} 2 & 1 \\ 0 & -3 \end{pmatrix}$**:**

$$\|A\|_F = \sqrt{2^2 + 1^2 + 0^2 + (-3)^2} = \sqrt{4+1+0+9} = \sqrt{14}.$$

2. **Why** $\|A\|_F \geq \|A\|_2$**:** - Fact: $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \dots$ (sum of all singular values squared). - $\|A\|_2 = \sigma_{\max}(A)$ is the largest singular value. So $\sigma_{\max}(A)^2 \leq \sigma_1^2 + \sigma_2^2 + \cdots = \|A\|_F^2$. Hence $\sigma_{\max}(A) \leq \|A\|_F$.

3. $\|B\|_2$ **for** $B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$**:** - B is symmetric, so $\|B\|_2$ is the maximum absolute eigenvalue. - Solve $\det(B - \lambda I) = 0$:

$$\det \begin{pmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = 0.$$

The roots are $\lambda = 3$ or $\lambda = -1$. - The largest eigenvalue in magnitude is 3. Thus $\|B\|_2 = 3$.

Gradients: Definition Recap

For a scalar function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient at x is

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}.$$

Key property: $\nabla f(x)$ points in the direction of steepest increase of f at x. A critical point is where $\nabla f(x) = 0$.
We now look at a few exercises with multivariable functions.

Exercise (30s): Compute Gradients

Exercise:

1. Let $f(x, y) = x^2 + xy + y^2$. Find $\nabla f(x, y)$.
2. Let $g(x, y, z) = e^{x+y^2-2z}$. Compute $\nabla g(x, y, z)$.
3. Identify any critical points of f in part (1). (Hint: set gradient to zero.)

Think (30s): Apply partial derivatives carefully.

Solution: Gradients I

Solution Outline:

1. **For** $f(x,y) = x^2 + xy + y^2$**:**

$$\frac{\partial f}{\partial x} = 2x + y, \quad \frac{\partial f}{\partial y} = x + 2y.$$

Therefore,

$$\nabla f(x,y) = \begin{bmatrix} 2x + y \\ x + 2y \end{bmatrix}.$$

2. **For** $g(x,y,z) = e^{x+y^2-2z}$**:**

$$\frac{\partial g}{\partial x} = e^{x+y^2-2z}, \quad \frac{\partial g}{\partial y} = 2y\, e^{x+y^2-2z}, \quad \frac{\partial g}{\partial z} = -2\, e^{x+y^2-2z}.$$

So

$$\nabla g(x,y,z) = e^{x+y^2-2z} \begin{bmatrix} 1 \\ 2y \\ -2 \end{bmatrix}.$$

Solution: Gradients II

3. **Critical points of f (part 1):** Solve

$$2x + y = 0 \quad \text{and} \quad x + 2y = 0.$$

From the first equation $y = -2x$. Plug into second: $x + 2(-2x) = x - 4x = -3x = 0$. Hence $x = 0$, which implies $y = 0$. The only critical point is $(0, 0)$.

Chain Rule: Definition Recap

For $f(x) = g\big(h(x)\big)$ with $h : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}$, we have:

$$\nabla f(x) = \left[\frac{\partial h}{\partial x}(x)\right]^\top \nabla g\big(h(x)\big),$$

where $\frac{\partial h}{\partial x}(x)$ is the Jacobian of h at x, and $\nabla g(u)$ is the gradient of g at u.
We'll now apply this to a composite function.

Exercise (30s): Chain Rule in 2 Steps

Exercise:
Let
$$h(x,y) = \begin{bmatrix} x^2 + y \\ x + e^y \end{bmatrix}, \quad g(u,v) = u^2 + 3v.$$
Define $f(x,y) = g\big(h(x,y)\big) = (x^2+y)^2 + 3(x+e^y)$.

1. Compute $\nabla f(x,y)$ *directly* by partial derivatives.
2. Compute $\nabla f(x,y)$ *using the chain rule formula*, i.e., find $J_h(x,y)$ and $\nabla g(u,v)$, then compose.
3. Check consistency of both results.

Think (30s): It's mechanical but good practice.

Solution: Chain Rule in 2 Steps I

Solution Sketch:

1. **Direct partial derivatives:**

$$f(x,y) = (x^2 + y)^2 + 3(x + e^y).$$

$$\frac{\partial f}{\partial x} = 2(x^2+y)\cdot\frac{\partial}{\partial x}(x^2+y)+3\cdot\frac{\partial}{\partial x}(x+e^y) = 2(x^2+y)\cdot 2x+3\cdot 1 = 4x(x^2+y$$

$$\frac{\partial f}{\partial y} = 2(x^2+y)\cdot\frac{\partial}{\partial y}(x^2+y)+3\cdot\frac{\partial}{\partial y}(x+e^y) = 2(x^2+y)\cdot 1+3\cdot e^y = 2(x^2+y)$$

Hence

$$\nabla f(x,y) = \begin{bmatrix} 4x(x^2+y)+3 \\ 2(x^2+y)+3e^y \end{bmatrix}.$$

2. **Using chain rule form:**

$$h(x,y) = \begin{bmatrix} x^2 + y \\ x + e^y \end{bmatrix}, \quad g(u,v) = u^2 + 3v.$$

Solution: Chain Rule in 2 Steps II

- Jacobian of h:

$$J_h(x,y) = \begin{bmatrix} \frac{\partial}{\partial x}(x^2+y) & \frac{\partial}{\partial y}(x^2+y) \\ \frac{\partial}{\partial x}(x+e^y) & \frac{\partial}{\partial y}(x+e^y) \end{bmatrix} = \begin{bmatrix} 2x & 1 \\ 1 & e^y \end{bmatrix}.$$

- Gradient of $g(u,v)$:

$$\nabla g(u,v) = \begin{bmatrix} \frac{\partial}{\partial u}(u^2+3v) \\ \frac{\partial}{\partial v}(u^2+3v) \end{bmatrix} = \begin{bmatrix} 2u \\ 3 \end{bmatrix}.$$

- Then

$$\nabla f(x,y) = J_h(x,y)^\top \nabla g\big(h(x,y)\big).$$

So

$$\nabla g(h(x,y)) = \begin{bmatrix} 2(x^2+y) \\ 3 \end{bmatrix}, \quad J_h(x,y)^\top = \begin{bmatrix} 2x & 1 \\ 1 & e^y \end{bmatrix}.$$

Solution: Chain Rule in 2 Steps III

Multiply:

$$\begin{bmatrix} 2x & 1 \\ 1 & e^y \end{bmatrix} \begin{bmatrix} 2(x^2+y) \\ 3 \end{bmatrix} = \begin{bmatrix} 2x \cdot 2(x^2+y) + 1 \cdot 3 \\ 1 \cdot 2(x^2+y) + e^y \cdot 3 \end{bmatrix} = \begin{bmatrix} 4x(x^2+y) + 3 \\ 2(x^2+y) + 3e^y \end{bmatrix}.$$

- Same as the direct computation result.

Taylor Expansion: Definition Recap

First-order Taylor Expansion (multivariate): For $f : \mathbb{R}^n \to \mathbb{R}$ differentiable at a,

$$f(x) \approx f(a) + \nabla f(a)^\top (x - a).$$

Higher-order expansions include Hessians, etc. We'll do a second-order example.

Exercise (30s): Second-Order Taylor Expansion

Exercise: Consider

$$f(x, y) = x^2 + 3xy + 2y^2.$$

1. Write the second-order Taylor expansion around the point $(a, b) = (1, 1)$ up to (and including) the quadratic terms.
2. Evaluate this quadratic approximation at $(x, y) = (1.1, 0.9)$ to estimate $f(1.1, 0.9)$.
3. Compare with the actual value of $f(1.1, 0.9)$ (just to see how close it is).

Think (30s): You'll need partial derivatives (1st and 2nd), evaluate them at $(1, 1)$, then plug in $(1.1, 0.9)$.

Solution: Second-Order Taylor Expansion I

Solution Outline:

1. **Partial derivatives and Hessian:**

- $f(x, y) = x^2 + 3xy + 2y^2$. - First-order derivatives:

$$f_x = 2x + 3y, \quad f_y = 3x + 4y.$$

- Second-order derivatives (Hessian H_f):

$$f_{xx} = 2, \quad f_{xy} = 3, \quad f_{yy} = 4.$$

(Also $f_{yx} = 3$.) - Evaluate at $(1, 1)$:

$$f(1, 1) = 1^2 + 3(1)(1) + 2(1)^2 = 1 + 3 + 2 = 6,$$

$$f_x(1, 1) = 2 \cdot 1 + 3 \cdot 1 = 5, \quad f_y(1, 1) = 3 \cdot 1 + 4 \cdot 1 = 7,$$

$$f_{xx}(1, 1) = 2,\ f_{xy}(1, 1) = 3,\ f_{yy}(1, 1) = 4.$$

Solution: Second-Order Taylor Expansion II

2. **Second-order Taylor expansion around $(1, 1)$:** For (x, y) near $(1, 1)$,

$$f(x, y) \approx f(1, 1) + \nabla f(1, 1)^\top \begin{pmatrix} x-1 \\ y-1 \end{pmatrix} + \tfrac{1}{2} \begin{pmatrix} x-1 & y-1 \end{pmatrix} \times H_f(1, 1) \begin{pmatrix} x-1 \\ y-1 \end{pmatrix}.$$

So

$$f(x, y) \approx 6 + \begin{bmatrix} 5 & 7 \end{bmatrix} \begin{pmatrix} x-1 \\ y-1 \end{pmatrix} + \tfrac{1}{2} \begin{pmatrix} x-1 & y-1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x-1 \\ y-1 \end{pmatrix}.$$

3. **Plug in $(x, y) = (1.1, 0.9)$:** - Let $\Delta x = 0.1$, $\Delta y = -0.1$. - First-order term:

Solution: Second-Order Taylor Expansion III

$[5\ 7]\begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix} = 5(0.1) + 7(-0.1) = 0.5 - 0.7 = -0.2.$ - Quadratic term:

$$\frac{1}{2}\begin{pmatrix} 0.1 & -0.1 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix}.$$

Multiply inside:

$$\begin{aligned} \begin{pmatrix} 0.1 & -0.1 \end{pmatrix}\begin{pmatrix} 2\cdot 0.1 + 3\cdot(-0.1) \\ 3\cdot 0.1 + 4\cdot(-0.1) \end{pmatrix} &= \begin{pmatrix} 0.1 & -0.1 \end{pmatrix}\begin{pmatrix} 0.2 - 0.3 \\ 0.3 - 0.4 \end{pmatrix} \\ &= \begin{pmatrix} 0.1 & -0.1 \end{pmatrix}\begin{pmatrix} -0.1 \\ -0.1 \end{pmatrix}. \end{aligned}$$

$$= 0.1(-0.1) + (-0.1)(-0.1) = -0.01 + 0.01 = 0.$$

Then multiply by $\frac{1}{2}$, which is still 0. - So $f(1.1, 0.9) \approx 6 + (-0.2) + 0 = 5.8$ (using up-to-second-order expansion).

Solution: Second-Order Taylor Expansion IV

4. **Actual value:**

$$\begin{aligned} f(1.1, 0.9) &= (1.1)^2 + 3(1.1)(0.9) + 2(0.9)^2 = 1.21 + 2.97 + 2(0.81) \\ &= 1.21 + 2.97 + 1.62 = 5.80. \end{aligned}$$

Exactly 5.80. The second-order approximation in this case gives the exact result (coincidentally) because cross-terms canceled neatly.

Eigenvalues: Definition Recap

Eigenvalue/Eigenvector: For $A \in \mathbb{R}^{n \times n}$, a scalar λ and non-zero vector v satisfy $Av = \lambda v$ if and only if λ is an eigenvalue and v is a corresponding eigenvector.

Key steps to find eigenvalues:

1. Solve $\det(A - \lambda I) = 0$ for λ.
2. For each λ, solve $(A - \lambda I)v = 0$ for v.

Exercise (1 min): Eigenvalue Analysis

Exercise: Let

$$A = \begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

1. Find all eigenvalues of A.
2. For each eigenvalue, give a basis for the corresponding eigenspace.
3. Is A diagonalizable? (Hint: diagonalizable if it has n linearly independent eigenvectors.)

Think (1 min): Notice the block structure and that the bottom-right corner is separate.

Solution: Eigenvalue Analysis I

Solution Outline:

1. **Find eigenvalues:** The matrix is block-diagonal in form:

$$\begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

- The first 2×2 block:

$$\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}.$$

Solve $\det\left(\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} - \lambda I\right) = 0$. This is

$$\det\begin{pmatrix} 4-\lambda & -2 \\ -2 & 1-\lambda \end{pmatrix} = (4-\lambda)(1-\lambda)-(-2)(-2) = (4-\lambda)(1-\lambda)-4.$$

Solution: Eigenvalue Analysis II

Expanding:

$$= 4 - 4\lambda - \lambda + \lambda^2 - 4 = \lambda^2 - 5\lambda = \lambda(\lambda - 5).$$

So eigenvalues are $\lambda = 0$ and $\lambda = 5$. - The 3rd row/column is 3. So another eigenvalue is $\lambda = 3$.
Hence the eigenvalues are $\{0, 5, 3\}$.
2. **Eigenvectors:** - For $\lambda = 0$, solve

$$(A - 0I)v = 0 \implies \begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} v = 0.$$

The third component must be 0 to satisfy $3 \cdot v_3 = 0$. The 2×2 block for the first two components is

$$\begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0.$$

Solution: Eigenvalue Analysis III

That system is

$$4v_1 - 2v_2 = 0, \quad -2v_1 + v_2 = 0.$$

From the second equation, $v_2 = 2v_1$. Plug into the first: $4v_1 - 2(2v_1) = 4v_1 - 4v_1 = 0$, consistent. So let $v_1 = 1$ then $v_2 = 2$.

$$v^{(0)} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

is an eigenvector basis for $\lambda = 0$.
- For $\lambda = 5$, solve $(A - 5I)v = 0$. Then the first 2×2 block becomes $\begin{pmatrix} -1 & -2 \\ -2 & -4 \end{pmatrix}$, the third diagonal is $(3 - 5) = -2$. So $v_3 = 0$. The 2×2 system is:

$$-1 \cdot v_1 - 2\,v_2 = 0 \quad \text{and} \quad -2\,v_1 - 4\,v_2 = 0.$$

Solution: Eigenvalue Analysis IV

The second row is just 2 times the first row, so we only have $-v_1 = 2v_2$. Let $v_1 = 2$, then $v_2 = -1$.

$$v^{(5)} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}.$$

- For $\lambda = 3$, we see from the block structure that $v_1 = v_2 = 0$, and the third coordinate is free. E.g.,

$$v^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

3. **Diagonalizability:** - We have found *three* linearly independent eigenvectors: $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. - Because A is a 3×3 with 3 distinct eigenvalues (or simply 3 independent eigenvectors), A **is diagonalizable**.

Summary of Exercises

We covered:

- **Vector Norms**: Basic properties, inequalities, comparisons (L_1, L_2, L_∞).
- **Matrix Norms**: Frobenius vs. spectral norm, computing norms for small matrices.
- **Gradients**: Computing partial derivatives and finding critical points.
- **Chain Rule**: Composing Jacobians and gradients.
- **Taylor Expansion**: 2D second-order approximations.
- **Eigenvalues**: Finding eigenvalues/eigenvectors of a block or structured matrix.

Next Steps:

- Explore SVD.

Thank you!

Questions?