A Review on Matrix Computations
(Part 1)

Dr. José Ernesto Rayas-Sánchez

1

Outline

- Notation
- Definitions and basic operations
- Norms
- Elements of vector calculus

Dr. J. E. Rayas-Sánchez

2

Notation

- Scalars: a, b, c
- Vectors: $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$
- Matrices: $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$
- Handwriting…
- Identity matrix: $\boldsymbol{I}$
- Examples:

$$a \in \Re \qquad \boldsymbol{b} \in \Re^n \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \qquad \boldsymbol{C} \in \Re^{n\times m} \qquad \boldsymbol{C} = \begin{bmatrix} c_{11} & \dots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \dots & c_{nm} \end{bmatrix}$$

Some Definitions and Basic Operations

- Transposition

$$\boldsymbol{b} \in \Re^n \qquad \boldsymbol{b}^T = [b_1 \quad \dots \quad b_n]$$

$$\boldsymbol{C} \in \Re^{n\times m} \qquad \boldsymbol{C} = [\boldsymbol{c}_1 \quad \dots \quad \boldsymbol{c}_m] \qquad \boldsymbol{c}_i \in \Re^n \qquad \boldsymbol{C}^T = \begin{bmatrix} \boldsymbol{c}_1^T \\ \vdots \\ \boldsymbol{c}_m^T \end{bmatrix}$$

- A square matrix $\boldsymbol{A}$ is symmetric if $\boldsymbol{A}^T = \boldsymbol{A}$
- Transposition of a matrix multiplication: $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
- Inversion of a product of matrices: $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Some Definitions and Basic Operations (cont)

- Inner product or dot product

$$\boldsymbol{a}, \boldsymbol{b} \in \Re^n \qquad \langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^T \boldsymbol{b} = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n = \boldsymbol{b}^T \boldsymbol{a}$$

- $\boldsymbol{a}$ and $\boldsymbol{b}$ are orthogonal vectors if $\boldsymbol{a}^T\boldsymbol{b} = 0$
- A set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is orthogonal if $\boldsymbol{a}_i^T\boldsymbol{a}_j = 0$ for all $i \neq j$
- A set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is orthonormal if $\boldsymbol{a}_i^T\boldsymbol{a}_j = 0$ for all $i \neq j$ and $\boldsymbol{a}_i^T\boldsymbol{a}_i = 1$
- An orthogonal matrix is a square matrix $\boldsymbol{Q}$ satisfying $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I} = \boldsymbol{Q}\boldsymbol{Q}^T$ or $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$

Dr. J. E. Rayas-Sánchez 5

Some Definitions and Basic Operations (cont)

- Outer product or dyadic product

$$\boldsymbol{a} \in \Re^n \qquad \boldsymbol{b} \in \Re^m \qquad \boldsymbol{a}\boldsymbol{b}^T = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \ldots & a_1 b_m \\ a_2 b_1 & a_2 b_2 & \ldots & a_2 b_m \\ \vdots & & \ddots & \\ a_n b_1 & a_n b_2 & \ldots & a_n b_m \end{bmatrix}$$

- A set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ is linearly independent if the equality $b_1\boldsymbol{a}_1+b_2\boldsymbol{a}_2+\ldots+b_n\boldsymbol{a}_n = \boldsymbol{0}$ implies $b_i = 0$ for $i = 1\ldots n$, or
- A set of vectors is linearly dependent if at least one of the vectors can be expressed as a linear combination of the remaining vectors

Dr. J. E. Rayas-Sánchez 6

Some Definitions and Basic Operations (cont)

- If $\boldsymbol{A}$ is a matrix, then $\boldsymbol{b} = \boldsymbol{Ax}$ is a linear transformation of vector $\boldsymbol{x}$, or, $\boldsymbol{b}$ is the linear combination of the columns of $\boldsymbol{A}$
- A matrix $\boldsymbol{A}$ is nonsingular if $\boldsymbol{Ax} = \boldsymbol{0}$ implies $\boldsymbol{x} = \boldsymbol{0}$; $\boldsymbol{A}$ must be square to be nonsingular
- The rank of a matrix is the minimum between the maximum number of linearly independent columns and the maximum number of linearly independent rows
- A nonsingular matrix is a full rank matrix, that is, all its columns and all its rows are linearly independent

Singular Matrices – Example

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad \boldsymbol{A} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3]$$

$$b_1\boldsymbol{a}_1 + b_2\boldsymbol{a}_2 + b_3\boldsymbol{a}_3 = 0$$

$$\boldsymbol{a}_1 = -\left(\frac{b_2}{b_1}\right)\boldsymbol{a}_2 - \left(\frac{b_3}{b_1}\right)\boldsymbol{a}_3 = \begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} = -(-2)\begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} - (1)\begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

$$\boldsymbol{a}_2 = -\left(\frac{b_1}{b_2}\right)\boldsymbol{a}_1 - \left(\frac{b_3}{b_2}\right)\boldsymbol{a}_3 = \begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix} = -(-0.5)\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} - (-0.5)\begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

$$\boldsymbol{a}_3 = -\left(\frac{b_1}{b_3}\right)\boldsymbol{a}_1 - \left(\frac{b_2}{b_3}\right)\boldsymbol{a}_2 = \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix} = -(1)\begin{bmatrix} 1 \\ 4 \\ 7 \end{bmatrix} - (-2)\begin{bmatrix} 2 \\ 5 \\ 8 \end{bmatrix}$$

Any column in $\boldsymbol{A}$ is a linear combination of the other two columns

Singular Matrices – Example (cont.)

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Any column in $\boldsymbol{A}$ is a linear combination of the other two columns

$$\text{rank}(\boldsymbol{A}) = 2 \qquad \text{cond}(\boldsymbol{A}) = 5.05 \times 10^{16}$$

$$\boldsymbol{A}^{-1} \approx 1 \times 10^{16} \begin{bmatrix} -0.31 & -0.63 & 0.31 \\ -0.63 & 1.26 & -0.63 \\ 0.31 & -0.63 & 0.31 \end{bmatrix}$$

Outer Products Yield Singular Matrices

Example

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \qquad \boldsymbol{a}\boldsymbol{b}^T = \boldsymbol{C} = \begin{bmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix}$$

Any row in $\boldsymbol{C}$ is a linear combination of just one row

$$\text{rank}(\boldsymbol{C}) = 1 \qquad \text{cond}(\boldsymbol{C}) = 1.4 \times 10^{17}$$

$$\boldsymbol{C}^{-1} \approx 1 \times 10^{15} \begin{bmatrix} -4.5 & 0 & 1.5 \\ 0 & 1.8 & -1.2 \\ 3.0 & -1.5 & 0 \end{bmatrix}$$

Non-Singular Matrices – Example

$$\boldsymbol{B} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 8 & 2 \\ 6 & 3 & 1 \end{bmatrix}$$

No column (or row) in $\boldsymbol{B}$ is a linear combination of the other two columns (or rows)

$$\text{rank}(\boldsymbol{B}) = 3 \qquad \text{cond}(\boldsymbol{B}) = 4.52$$

$$\boldsymbol{B}^{-1} = \begin{bmatrix} -0.0192 & -0.0673 & 0.1923 \\ -0.0962 & 0.1635 & -0.0385 \\ 0.4038 & -0.0865 & -0.0385 \end{bmatrix}$$

$\boldsymbol{B}$ is full rank, or well conditioned

Norms

- They are criteria to determine the "magnitude" of a vector
- The p-th norm of a vector $\boldsymbol{e} \in \Re^n$ is

$$\|\boldsymbol{e}\|_p = \left[\sum_{i=1}^{n} |e_i|^p \right]^{1/p}$$

- Manhattan norm, or l_1 norm ($p = 1$) $\quad \|\boldsymbol{e}\|_1 = \sum_{i=1}^{n} |e_i|$
- Euclidean norm, or l_2 norm ($p = 2$) $\quad \|\boldsymbol{e}\|_2 = \sqrt{\sum_{i=1}^{n} (e_i)^2}$
- Chebyshev norm, or infinite norm ($p \to \infty$) $\quad \|\boldsymbol{e}\|_\infty = \max_i |e_i|$

Norms (cont)

- The Frobenius norm for matrices is the equivalent to the Euclidean norm for vectors
- If $\boldsymbol{E} \in \Re^{m \times n}$

$$\left\| \boldsymbol{E} \right\|_F = \left[\sum_{i=1}^{m} \sum_{j=1}^{n} (e_{ij})^2 \right]^{1/2} = \sqrt{\sum_{i=1}^{n} \boldsymbol{e}_i^T \boldsymbol{e}_i}$$

Positive Definite – Quadratic Forms

- An n by n matrix $\boldsymbol{A}$ is positive definite if
  $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq 0$ and
  $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} = 0$ implies $\boldsymbol{x} = 0$
- An n by n matrix $\boldsymbol{A}$ is positive semidefinite if
  $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq 0$
- If matrix $\boldsymbol{A}$ is symmetric, diagonally dominant, and has positive diagonal elements, then $\boldsymbol{A}$ is positive definite
- If matrix $\boldsymbol{A}$ is positive definite, then $\boldsymbol{A}$ is diagonally dominant and has positive diagonal elements

Cauchy and Triangle Inequalities

- Cauchy inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| = |\boldsymbol{x}^T \boldsymbol{y}| \leq \|\boldsymbol{x}\|_2 \, \|\boldsymbol{y}\|_2$$

- Triangle inequality

$$\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$$

Gradient

If u is a multidimensional scalar function,
u: $\Re^n \rightarrow \Re$
the gradient of u is

$$\nabla u(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial u}{\partial x_1} \\ \dfrac{\partial u}{\partial x_2} \\ \vdots \\ \dfrac{\partial u}{\partial x_n} \end{bmatrix}$$

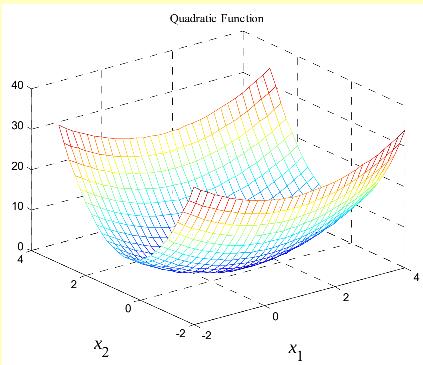
The ∇u is perpendicular to the contours of u and points uphill, in the direction of maximum increase of the function u

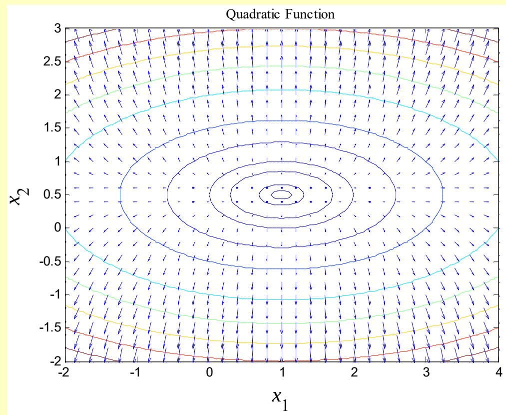
Its magnitude depends on the rate of change of u

$$\nabla u(\boldsymbol{x}) \in \Re^n$$

Gradient – Example: Quadratic Function

$$y(\boldsymbol{x}) = (x_1 - 1)^2 + (2x_2 - 1)^2 \qquad \text{where} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



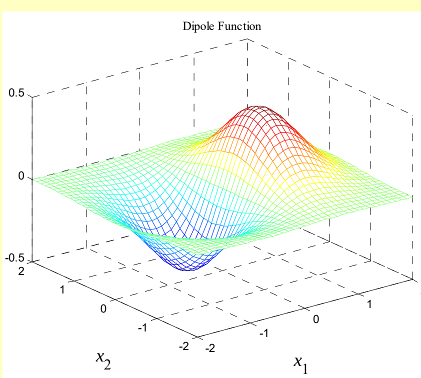


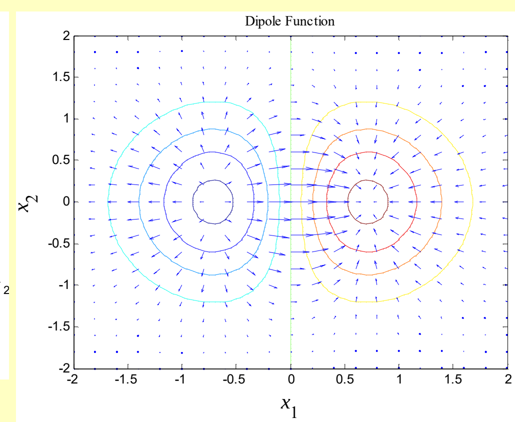
Dr. J. E. Rayas-Sánchez

17

Gradient – Example: Dipole Function

$$u(\boldsymbol{x}) = \frac{x_1}{e^{(x_1^2 + x_2^2)}} \qquad \text{where} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$





Dr. J. E. Rayas-Sánchez

18

Directional Derivative

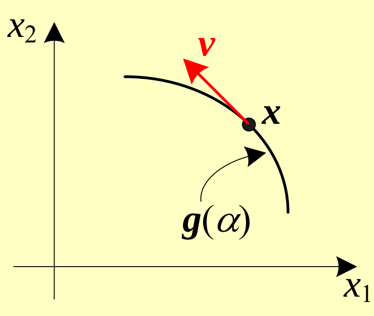
- If u is a multidimensional scalar function, u: $\Re^n \rightarrow \Re$, the directional derivative of u in the direction of $\boldsymbol{d}$, $Du|_{\boldsymbol{d}}$, is

$$Du|_{\boldsymbol{d}} = \frac{\partial u(\boldsymbol{x})}{\partial \boldsymbol{d}} = \lim_{\alpha \to 0} \frac{u(\boldsymbol{x} + \alpha \boldsymbol{d}) - u(\boldsymbol{x})}{\alpha} = \nabla u^T \boldsymbol{d}$$

- The rate of increase of u in the direction of $\boldsymbol{d}$ is $\nabla u^T \boldsymbol{U}_d$, where $\boldsymbol{U}_d$ is the unit vector in the direction of $\boldsymbol{d}$,

$$\boldsymbol{U}_d = \frac{\boldsymbol{d}}{\|\boldsymbol{d}\|_2}$$

Gradient is Perpendicular to the Contour – Proof



Function $\boldsymbol{g}(\alpha)$: $\Re \rightarrow \Re^n$ generates the points on the contour of level c,

$$u(\boldsymbol{g}(\alpha)) = c$$

Let $\boldsymbol{v}$ be a vector tangential to the contour at point $\boldsymbol{x}$,

$$\boldsymbol{g}(\alpha) = \boldsymbol{x} + \alpha \boldsymbol{v} \quad \text{for small } \alpha$$

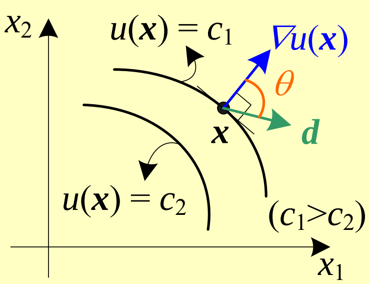
Since

$$u'(\alpha) = 0 = Du|_{\boldsymbol{v}} = \nabla u^T \boldsymbol{v}$$

Hence,

$$\nabla u \perp \boldsymbol{v}$$

Gradient is the Max Increase Direction – Proof



Since the rate of increase of u in the direction $\boldsymbol{d}$ is given by

$$\nabla u^T \boldsymbol{U}_d = \nabla u^T \frac{\boldsymbol{d}}{\|\boldsymbol{d}\|_2} = \|\nabla u\|_2 \cos\theta$$

Then, the rate of increase of u at point $\boldsymbol{x}$ is maximum when $\theta = 0$ ($\boldsymbol{d}$ has the same direction as ∇u)

Jacobian

If $\boldsymbol{f}$ is a multidimensional vector function,
$\boldsymbol{f}: \Re^n \to \Re^m$
the Jacobian of $\boldsymbol{f}$ is

$$\boldsymbol{J}(\boldsymbol{f}(\boldsymbol{x})) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & & \ddots & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \equiv \boldsymbol{f}'(\boldsymbol{x})$$

$$\boldsymbol{J}(\boldsymbol{f}(\boldsymbol{x})) \in \Re^{m \times n}$$

Hessian

If u is a multidimensional scalar function,
u: $\Re^n \rightarrow \Re$
the Hessian of u is

$$\boldsymbol{H}(u(\boldsymbol{x})) = \begin{bmatrix} \frac{\partial^2 u}{\partial x_1{}^2} & \frac{\partial^2 u}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 u}{\partial x_1 \partial x_n} \\ \frac{\partial^2 u}{\partial x_2 \partial x_1} & \frac{\partial^2 u}{\partial x_2{}^2} & \cdots & \frac{\partial^2 u}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \\ \frac{\partial^2 u}{\partial x_n \partial x_1} & \frac{\partial^2 u}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 u}{\partial x_n{}^2} \end{bmatrix} = \boldsymbol{J}(\nabla u(\boldsymbol{x}))$$

$$\boldsymbol{H}(u(\boldsymbol{x})) \in \Re^{n \times n}$$

Taylor Series Expansion

- Taylor Series approximation of a unidimensional scalar function u around point x_0, with $\Delta x = x - x_0$

$$u(x) = u(x_0 + \Delta x) \approx u(x_0) + \Delta x u'(x_0) + \frac{(\Delta x)^2}{2!} u''(x_0) + \ldots$$

- Taylor Series approximation of a multidimensional scalar function u around point $\boldsymbol{x}_0$, with $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$

$$u(\boldsymbol{x}) = u(\boldsymbol{x}_0 + \Delta\boldsymbol{x}) \approx u(\boldsymbol{x}_0) + \Delta\boldsymbol{x}^T \nabla u(\boldsymbol{x}_0) + \frac{1}{2} \Delta\boldsymbol{x}^T \boldsymbol{H}(u(\boldsymbol{x}_0)) \Delta\boldsymbol{x} + \ldots$$