Tutorial 1
Basic mathematical background
236873

Written By: Elad Osherov
Edited By: Tamar Rott Shaham

Oct 2017
Linear Algebra in Computer Vision

Linear algebra is frequently used in...

- Representation of points and features
- Computation of similarities and distances
- Transformations
- Finding algebraic solution
- Optimization
Vectors

Geometric object that has both a magnitude and direction \( x \in \mathbb{R}^n \)

\[ x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \]

Transpose of a vector
\[ x = (x_1, x_2, \ldots, x_n)^T \]

Magnitude of a vector
\[ \|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \]

Cardinality of a vector - the number of non zero elements
Vectors – inner product space

Definition An inner product on $\mathbb{R}^n$ is a map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ with the following properties:

1. (symmetry) $\langle x, y \rangle = \langle y, x \rangle$ for any $x, y \in \mathbb{R}^n$.
2. (additivity) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ for any $x, y, z \in \mathbb{R}^n$.
3. (homogeneity) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ for any $\lambda \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.
4. (positive definiteness) $\langle x, x \rangle \geq 0$ for any $x \in \mathbb{R}^n$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Examples

- the “dot product”

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i \text{ for any } x, y \in \mathbb{R}^n.$$

- the “weighted dot product”

$$\langle x, y \rangle_w = \sum_{i=1}^{n} w_i x_i y_i,$$

where $w \in \mathbb{R}_{++}^n$. 
Vector Norms

Definition. A norm \( \| \cdot \| \) on \( \mathbb{R}^n \) is a function \( \| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R} \) satisfying

- **(Nonnegativity)** \( \|x\| \geq 0 \) for any \( x \in \mathbb{R}^n \) and \( \|x\| = 0 \) if and only if \( x = 0 \).
- **(Positive homogeneity)** \( \|\lambda x\| = |\lambda|\|x\| \) for any \( x \in \mathbb{R}^n \) and \( \lambda \in \mathbb{R} \).
- **(Triangle inequality)** \( \|x + y\| \leq \|x\| + \|y\| \) for any \( x, y \in \mathbb{R}^n \).

One natural way to generate a norm on \( \mathbb{R}^n \) is to take any inner product \( \langle \cdot, \cdot \rangle \) defined on \( \mathbb{R}^n \), and define the associated norm

\[
\|x\| \equiv \sqrt{\langle x, x \rangle}, \text{ for all } x \in \mathbb{R}^n,
\]

- \( \|x\|_1 = \sum_{i=1}^{n} |x_i| \) \( \ell_1 \)-norm
- \( \|x\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} \) \( \ell_2 \)-norm
- \( \|x\|_{\infty} = \max(|x_1|, |x_2|, \ldots, |x_n|) \) \( \ell_{\infty} \)-norm
- \( \|x\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p} \) \( \ell_p \)-norm
Vector Norms

\( \ell_2 \)-norm

\[ \begin{align*}
\mathbf{x} &= (x_1, x_2, \ldots, x_n)^T \\
\mathbf{y} &= (y_1, y_2, \ldots, y_n)^T \\
\|\mathbf{x}\|_2^2 &= \sum_{i=1}^{n} |x_i|^2 = x_1^2 + x_2^2 + \cdots + x_n^2 = \mathbf{x}^T \mathbf{x} \\
\|\mathbf{x} - \mathbf{y}\|_2^2 &= \sum_{i=1}^{n} |x_i - y_i|^2 = (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}) = \|\mathbf{x}\|^2 - 2\mathbf{x}^T \mathbf{y} + \|\mathbf{y}\|^2
\end{align*} \]
Vector Norms in Image processing

MSE – love it or leave it? Zhou Wang and Alan C. Bovik

![Comparison of image fidelity measures for “Einstein” image altered with different types of distortions. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression. (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise).]
Linear Dependency

Given a set of vectors $X = \{x_1, x_2, \ldots, x_n\}$, Linear combination of vectors is written as

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

$x_i \in X$ is linearly dependent if it can be written as a linear combination of $X \setminus \{x_i\}$.

linearly dependent

linearly independent
Basis

A basis is a linearly independent set of vectors that spans the “whole space”

We can write every vector in our space as linear combination of the vectors in basis

Standard basis (a.k.a. unit vectors) \( \{ \mathbf{e}_i \in \mathbb{R}^n \mid \mathbf{e}_i = (0, \ldots, 0,1, \ldots, 0)^T \} \)

Projection of a vector: \( \mathbf{x} \cdot \mathbf{e}_i = \mathbf{x}^T \mathbf{e}_i = \mathbf{e}_i^T \mathbf{x} \)

\[
x^T = (3,2,5)^T = 3e_1^T + 2e_2^T + 5e_3^T
\]

\( \mathbf{e}_i^T \mathbf{e}_j = 0 \quad \mathbf{e}_i^T \mathbf{e}_i = 1 \)

Orthogonal + Normalized \( \equiv \) Orthonormal
Matrix

Rectangular (2D) array of numbers

\[ A \in \mathbb{R}^{m \times n} \quad A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad A_{ij} = a_{ij} \]

Special matrices

- Square matrix: matrix with (# of rows) = (# of columns)
- Identity matrix: matrix with diagonal elements ones and non-diagonal elements zeros

\[ AI = IA = A \]
Matrix Operations

Addition

Commutative: \[ A + B = B + A \]
Associative: \[ (A + B) + C = A + (B + C) \]

Multiplication

Associative: \[ A(BC) = A(BC) \]
Distributive: \[ A(B + C) = AB + AC \]
Non-commutative: \[ AB \neq BA \]

Transpose

\[ (A^T)_{ij} = A_{ji} \hspace{2cm} (A^T)^T = A \hspace{2cm} (AB)^T = B^T A^T \]
Matrix Rank

The rank of a matrix is the maximal number of linearly independent column or rows of a matrix.

\[
\text{rank}(A) = \text{rank}(A^T)
\]

\[
\text{rank}(A^T A) = \text{rank}(A)
\]

\[
\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)
\]

\[
\text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B))
\]

\[
A \in \mathbb{R}^{m \times n}, \text{ rank}(A) \leq \min(m, n)
\]

if \(\text{rank}(A) = \min(m, n)\), \(A\) is said to be full rank

**Singular matrix** – Has depended rows (and at least one zero Eigen-value)
Range and Nullspace

The range of a matrix is the span of the columns of the matrix, denoted by the set

\[ \mathcal{R}(A) = \{ y \mid y = Ax \} \]

The nullspace of a matrix, is the set of vectors that when multiplied by the matrix result in 0, given by the set

\[ \mathcal{N}(A) = \{ x \mid Ax = 0 \} \]
Determinant

\[
\begin{vmatrix}
  x_1 & y_1 & z_1 \\
  x_2 & y_2 & z_2 \\
  x_3 & y_3 & z_3 \\
\end{vmatrix} = x_1 \begin{vmatrix}
  x_2 & z_2 \\
  y_3 & z_3 \\
\end{vmatrix} - x_2 \begin{vmatrix}
  y_1 & z_1 \\
  y_3 & z_3 \\
\end{vmatrix} + x_3 \begin{vmatrix}
  y_1 & z_1 \\
  y_2 & z_2 \\
\end{vmatrix},
\]

\[\det(A) = 0 \quad \text{iff } A \text{ is singular}\]

\[
\det(AB) = \det(A)\det(B) \\
\det(A^{-1}) = \det(A)^{-1} \\
\det(\lambda A) = \lambda^n \det(A)
\]
Solving linear equation analytically

Finding the exact solution \[ Ax = b \] Where \( A \) is an \( nxn \) matrix, and \( x \) and \( b \) are \( nx1 \) vectors

\[ x = A^{-1}b \]

Finding the least square solution where \( A \) is an \( m \times n \) matrix, \( x \) is an \( n \times 1 \) vector, and \( b \) is an \( m \times 1 \) vector.

Assumptions-
1. \( A \) has a full column rank: \( \text{rank}(A) = n \)
2. \( m > n \) the system is inconsistent (more data than unknowns)

\[
\begin{align*}
  x_1 + 2x_2 &= 0 \\
  2x_1 + x_2 &= 1 \\
  3x_1 + 2x_2 &= 1
\end{align*}
\]

\[ x = (A^T A)^{-1} A^T b \]

Derivation on the whiteboard
Solving linear equation analytically

\[ x = \left( A^T A \right)^{-1} A^T b \]

Derivation on the whiteboard
Solving linear equation non analytically

- Simplex
- Conjugate gradient
- Optimization
- Gradient descent
- Interior point
- Newton’s method
Eigen Decomposition
Eigenvalues and Eigenvectors

Given a matrix, $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n$ are said to be an eigenvalue and the corresponding eigenvector of the matrix if

$$Ax = \lambda x, \quad x \neq 0$$

We can solve for the eigenvalues by solving for the roots of the polynomial generated by

$$|(\lambda I - A)| = 0$$
Eigenvalues properties

- \( \det(A) = \prod_{i=1}^{n} \lambda_i = \lambda_1 \lambda_2 \cdots \lambda_n \)

- \( \text{tr}(A) = \sum_{i=1}^{n} A_{ii} = \sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n \)

- The rank of a matrix is equal to the number of its non-zero eigenvalues

- Eigenvalues of a diagonal matrix, are simply the diagonal entries

- A matrix is said to be diagonalizable if we can write \( A = X \Lambda X^{-1} \)
Eigenvalues of symmetric matrixes

- Eigenvalues of symmetric matrices are real
- Eigenvectors of (real) symmetric matrices are orthonormal

Consider the optimization problem involving the symmetric matrix $A$

\[
\begin{align*}
\text{maximize} & \quad x^T A x \\
\text{subject to} & \quad \|x\|_2 = 1
\end{align*}
\]

the maximizing $x$ is the eigenvector corresponding to the largest eigenvalue
Eigen Decomposition

In the mathematical discipline of linear algebra, eigen-decomposition or sometimes spectral decomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvalues and eigenvectors. Only diagonalizable matrices can be factorized in this way.

Let $A$ be a square $(N \times N)$ and symmetric matrix with, then $A$ can be factorized as:

$$A = Q \Lambda Q^{-1}$$
Eigen Decomposition

What if A is non-square?
The SVD of matrix is given by

\[ A = UV^T \]

Where \( u_i \) are the columns of \( U \) and are called the left singular vectors.

\( v_i \) are the columns of \( V \), and are called the right singular vectors.

\( \Sigma \) is a diagonal matrix whose values are \( \sigma_i \), and called the singular values.
Singular Value Decomposition (SVD)

The left-singular vectors of $U$ are eigenvectors of $AA^T$
The right-singular vectors of $V$ are eigenvectors of $A^TA$

Singular values of $A$ are the square root of the non-zero eigenvalues of $AA^T$ or $A^TA$

$AA^T, A^TA \rightarrow \lambda_i$

$\sum_{ii} = \sigma_i = \sqrt{\lambda_i}$
Singular Value Decomposition (SVD) - MATLAB

e = eig(A)
[V,D] = eig(A)

s = svd(A)
[U,S,V] = svd(A)
Matrix Calculus -- Gradient

Let \( f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R} \). then the gradient is given by:

\[
(\nabla_A f(A))_{ij} = \frac{\partial f(A)}{\partial A_{ij}}
\]

- \( \nabla_A f(A) \) is always the same size as \( A \), thus if we just have a vector the gradient is simply

\[
\nabla_x f(x) = \begin{bmatrix}
\frac{\partial f(x)}{x_1} \\
\frac{\partial f(x)}{x_2} \\
\vdots \\
\frac{\partial f(x)}{x_n}
\end{bmatrix}
\]
Matrix Calculus -- Gradient

From partial derivatives:

\[ \nabla_x (f(x) + g(x)) = \nabla_x f(x) + \nabla_x g(x) \]

\[ \nabla_x (tf(x)) = t \nabla_x f(x) \]

Some common gradients:

<table>
<thead>
<tr>
<th>( y )</th>
<th>( \frac{\partial y}{\partial x} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Ax )</td>
<td>( A^T )</td>
</tr>
<tr>
<td>( x^T A )</td>
<td>( A )</td>
</tr>
<tr>
<td>( x^T x )</td>
<td>( 2x )</td>
</tr>
<tr>
<td>( x^T Ax )</td>
<td>( Ax + A^T x )</td>
</tr>
</tbody>
</table>
Matrix Calculus -- Gradient

\[ \frac{\partial}{\partial \mathbf{x}} (a^T \mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T a) = a \]

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{x}^T \mathbf{A} \mathbf{x}) = (\mathbf{A}^T + \mathbf{A}) \mathbf{x} \]

\[ \frac{\partial z}{\partial \mathbf{x}} = \frac{\partial y}{\partial \mathbf{x}} \frac{\partial z}{\partial y} \]

\[ \frac{\partial}{\partial \mathbf{x}} \mathbf{A}^T \mathbf{x} = \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{A} = \mathbf{A} \]

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{A} \mathbf{x} + \mathbf{b})^T (\mathbf{A} \mathbf{x} + \mathbf{b}) = 2 \mathbf{A}^T (\mathbf{A} \mathbf{x} + \mathbf{b}) \]

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{A} \mathbf{x} + \mathbf{b})^T \mathbf{C} (\mathbf{A} \mathbf{x} + \mathbf{b}) = \mathbf{A}^T (\mathbf{C} + \mathbf{C}^T) (\mathbf{A} \mathbf{x} + \mathbf{b}) \]

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{A} \mathbf{x} + \mathbf{b})^T (\mathbf{D} \mathbf{x} + \mathbf{e}) = \mathbf{A}^T (\mathbf{D} \mathbf{x} + \mathbf{e}) + \mathbf{D}^T (\mathbf{A} \mathbf{x} + \mathbf{b}) \]

\[ \frac{\partial}{\partial \mathbf{x}} (\mathbf{A} \mathbf{x} + \mathbf{b})^T \mathbf{C} (\mathbf{D} \mathbf{x} + \mathbf{e}) = \mathbf{A}^T \mathbf{C} (\mathbf{D} \mathbf{x} + \mathbf{e}) + \mathbf{D}^T \mathbf{C}^T (\mathbf{A} \mathbf{x} + \mathbf{b}) \]
Histogram

- Density function of the image
- Statistical tool for estimation and processing

- Gray levels vs. number of occurrences
- Can be normalized → PDF
- Global, Invariant to order of pixels
Convolution

- Convolution in 1D
  
  \[(f \ast g)(t) \overset{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau) g(t - \tau) \, d\tau = \int_{-\infty}^{\infty} f(t - \tau) g(\tau) \, d\tau.\]

- Convolution in 2D

- Usage
  - Filtering
  - Edge Detection
  - Template matching
Linear Filtering

- Linear combination of image and filter
  \[
  \begin{bmatrix}
  0 & \alpha_2 & 0 \\
  \alpha_3 & \alpha_1 & \alpha_5 \\
  0 & \alpha_4 & 0
  \end{bmatrix}
  \Rightarrow
  J[m, n] = \alpha_1 \cdot I[m, n] + \alpha_2 \cdot I[m - 1, n] + \\
  + \alpha_3 \cdot I[m, n - 1] + \alpha_4 \cdot I[m + 1, n] + \alpha_5 \cdot I[m, n + 1]
  \]

- Examples
  - Averaging
  - Gaussian
  - Laplacian

\[
\begin{bmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
0 & -1 & 0 \\
-1 & 4 & -1 \\
0 & -1 & 0
\end{bmatrix}
\begin{bmatrix}
2 & 3 & 2 \\
3 & 5 & 3 \\
2 & 3 & 2
\end{bmatrix}
\]
Non-Linear Filtering

- Not all filters can be formulated as matrices
- Minimum, Maximum
- Median filter
- Frequency mixer
- Energy transfer filter
- …