Linear Algebra for Computer Vision - part 2

CMSC 828 D

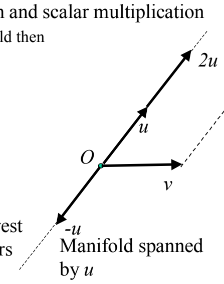
Outline

- Background and potpourri
- Summation Convention
- Eigenvalues and Eigenvectors
- Rank and Degeneracy
- Gram Schmidt Orthogonalization
- Fredholm Alternative Theorem
- Least Squares Formulation
- Singular Value Decomposition
- Applications

Summary: Linear Spaces

- *n* dimensional points in a vector space.
  - Length, distance, angles
  - Dot product (inner product)
- Linear dependence of a set of vectors
- Basis : a collection of n independent vectors so that any vector can be expressed as a sum of these vectors
- Orthogonality $<\mathbf{a},\mathbf{b}>=0$
- Orthogonal basis: basis vectors satisfy $<\mathbf{b}_i,\mathbf{b}_j> = 0$
- Vector is represented in a particular basis (coordinate system) $<\mathbf{u},\mathbf{b}_i> = u_i$

- Linear Manifolds (M): linear spaces that are subsets of the space that are closed under vector addition and scalar multiplication
  - If vectors **u** and **v** belong to the manifold then so do $\alpha_1 \mathbf{u} + \alpha_2 \mathbf{v}$
  - Manifold must contain zero vector
  - Essentially a full linear space of smaller dimension.
- Span of a set of vectors: set of all vectors that can be created by scalar multiplication and addition.
- Vectors in the space that are in the rest of the space are orthogonal to vectors in M. ($M^\perp$)



- Projection Theorem: any vector in the space X can be written only one way in terms of a vector in M and a vector in $M^\perp$.

Gram Schmidt Orthogonalization

- Given a set of basis vectors ($\mathbf{b}_1,\mathbf{b}_2, \ldots, \mathbf{b}_n$) construct an orthonormal basis ($\mathbf{e}_1,\mathbf{e}_2, \ldots, \mathbf{e}_n$) from it.
  - Set $\mathbf{e}_1 = \mathbf{b}_1/\|\mathbf{b}_1\|$
  - $\mathbf{g}_2 = \mathbf{b}_2 - < \mathbf{b}_2, \mathbf{e}_1> \mathbf{e}_1$, $\quad \mathbf{e}_2 = \mathbf{g}_2/\|\mathbf{g}_2\|$
  - For k=3, …,n
    $\mathbf{g}_k = \mathbf{b}_k - \Sigma_j < \mathbf{b}_k, \mathbf{e}_j> \mathbf{e}_j$, $\quad \mathbf{e}_k = \mathbf{g}_k/\|\mathbf{g}_k\|$

Euclidean 3D

- Three directions with basis vectors **i**,**j**,**k** or $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$, with $\mathbf{e}_i . \mathbf{e}_j = \delta_{ij}$
- Distance between two vectors **u** and **v** is $\|\mathbf{u}-\mathbf{v}\|$
- Dot product of two vectors **u** and **v** is $\|\mathbf{u}\|.\|\mathbf{v}\| \cos\theta$
- Cross product of two vectors is $\mathbf{u}\times\mathbf{v}$
  - magnitude equal to the area of the parallelogram formed by **u** and **v.**
  - Magnitude is $\|u\|\,\|v\| \sin\theta$
  - Direction is perpendicular to **u** and **v** so that the three vectors form a right handed system

$$\mathbf{u}\times\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

- Is also written using the permutation symbol ε_{ijk}

Summation Convention

- Boldface, transpose symbol and summation signs are tiresome.
  - Especially if you have to do things such as differentiation
  - Vectors can be written in terms of unit basis vectors

$$\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}= a_1\mathbf{e}_1+a_2\mathbf{e}_2+a_3\mathbf{e}_3$$

- However, even this is clumsy. E.g., in 10 dimensions

$$\mathbf{a}= a_1\mathbf{e}_1+a_2\mathbf{e}_2+\ldots+a_{10}\mathbf{e}_{10} =\Sigma_{i=1}^{10} a_i\mathbf{e}_i$$

- Notice that index *i* occurs twice in the expression.
  - Einstein noticed this always occurred, so whenever index was repeated twice he avoided writing Σ_i
  - instead of writing $\Sigma_i a_i b_i$, write $a_i b_i$ with the Σ_i implied

Permutation Symbol

- Permutation symbol ε_{ijk}
  - If i, j and k are in cyclic order $\varepsilon_{ijk}=1$
    - Cyclic => (1,2,3) or (2,3,1) or (3,1,2)
  - If in anticyclic order $\varepsilon_{ijk}=-1$
    - Anticyclic => (3,2,1) or (2,1,3) or (1,3,2)
  - Else, $\varepsilon_{ijk}=0$
    - (1,1,2), (2,3,3), …
- $\mathbf{c} = \mathbf{a}\times\mathbf{b}$ => $c_i = \varepsilon_{ijk}a_j b_k$
- ε δ identity $\varepsilon_{ijk}\varepsilon_{irs} = \delta_{jr}\,\delta_{ks} - \delta_{js}\,\delta_{kr}$
  - Very useful in proving vector identities
- Indicial notation is also essential for working with tensors
  - Tensors are essentially linear operators (matrices or their generalizations to higher dimensions)

Examples

- $A_i\, B_i$ in 2 dimensions: $A_1B_1+A_2B_2$
- $A_{ij}\, B_{jk}$ in 3D? We have 3 indices here (*i*,*j*,*k*), but only *j* is repeated twice and so it is $A_{i1}B_{1k} + A_{i2}B_{2k} + A_{i3}B_{3k}$
- Matrix vector product

  $\mathbf{Ax} = A_{ij}x_{j,}$ $\qquad \mathbf{A}^t\mathbf{x} = A_{ij}x_i$

- $(\mathbf{a} \times \mathbf{b}).\ \mathbf{c}= \varepsilon_{ijk}a_j b_k c_i$
  - Using indicial notation can easily show

$$\mathbf{a}\ .(\mathbf{b}\times\mathbf{c}) = (\mathbf{a} \times \mathbf{b}).\ \mathbf{c}$$

- Homework: show $\mathbf{a} \times(\mathbf{b} \times\mathbf{c})=\mathbf{b}(\mathbf{a}.\mathbf{c})-\mathbf{c}(\mathbf{a}.\mathbf{b})$

Operators / Matrices

- Linear Operator $\mathbf{A}(\alpha_1\,\mathbf{u} + \alpha_2\mathbf{v}) = \alpha_1\,\mathbf{Au} + \alpha_2\,\mathbf{Av}$
- maps one vector to another

$$\mathbf{Ax}=\mathbf{b}$$

  - $m\times n$ dimensional matrix **A** multiplying a *n* dimensional vector **x** to produce a *m* dimensional vector **b** in the dual space
- Square matrix of dimension *n* by *n* takes vector to another vector in the same space.
- Matrix entries are representations of the matrix using basis vectors $A_{ij} = \langle \mathbf{Ab}_j,\mathbf{b}_i\rangle$
- Eigenvectors are characteristic directions of the matrix.
- Matrix decomposition is a factorization of a matrix into matrices with specific properties.

Norm of a matrix

- $\|\mathbf{A}\| \geq 0 \qquad \|\mathbf{Ax}\|\leq\|\mathbf{A}\|\ \|\mathbf{x}\|$
- $\|\mathbf{A}\|_F =[a_{ij}\, a_{ij}]^{1/2}$ Froebenius norm. If **A** is diagonal $\|\mathbf{A}\|_F=[a_{11}^2+ a_{22}^2+\ldots+ a_{nn}^2\ ]^{1/2}$
- $\|\mathbf{A}\|_2 = \max_\mathbf{x} \|\mathbf{Ax}\|_2/\|\mathbf{x}\|_2$.
  Can show 2 norm = square root of largest eigenvalue of $\mathbf{A}^t\mathbf{A}$

Rank and Null Space

- Range of a $m\times n$ dimensional matrix **A**

  Range $(\mathbf{A}) = \{\mathbf{y}\in\mathbb{R}^m\mathbf{:}\ \mathbf{y}=\mathbf{Ax} \text{ for some } \mathbf{x} \in\mathbb{R}^n\}$
- Null space of **A** is the set of vectors which it takes to zero.

  Null$(\mathbf{A}) = \{\mathbf{x} \in\mathbb{R}^n : \mathbf{Ax} = \mathbf{0}\}$
- Rank of a matrix is the dimension of its range.
  Rank $(\mathbf{A})$ = Rank $(\mathbf{A}^t)$
  - Maximal number of independent rows **or** columns
- Dimension of Null$(\mathbf{A})$+Rank$(\mathbf{A}) = n$

Orthogonality

- Two vectors are orthogonal if $\langle\mathbf{u},\mathbf{v}\rangle=0$
- Orthogonal matrix is composed of orthogonal vectors as columns.

  $\mathbf{p}.\mathbf{q}=0$

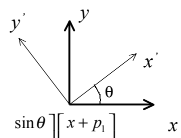
$$\begin{bmatrix} q_1 & p_1 & \cdots & r_1 \\ q_2 & p_2 & \vdots & r_2 \\ \vdots & \vdots & \ddots & \vdots \\ q_n & p_n & \cdots & r_n \end{bmatrix}$$

- Usually represented as **Q**
- By definition $\mathbf{QQ}^t=\mathbf{I}$
- Matrices that rotate coordinate axes are orthogonal matrices

Rotation in 2D and 3D

- Rotation through an angle θ

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$



- Rotation + translation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \qquad \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x + p_1 \\ y + p_2 \end{bmatrix}$$

- Rotation in 3D
  - ϕ about z axis, θ about new x axis, ψ about new y axis.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Rotation matrix

- Rotates a vector represented in one orthogonal coordinate system into a vector in another coordinate system.
  - Since length of vector should not change
  - $\|\mathbf{Qx}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$
  - Since Q will not change a vector along coordinate directions $\mathbf{QQ}^t = \mathbf{I}$
  - Columns of Q are its eigenvectors.
  - Eigenvalues are all 1.

Similarity Transforms

- Transforms vector represented in one basis to vector in another basis
- Let X={$\mathbf{x}_1,\ldots,\mathbf{x}_n$} and Y={$\mathbf{y}_1,\ldots,\mathbf{y}_n$} be two bases in a n dimensional space
  - There exists a transformation **A** which takes a vector expressed in X to one expressed in Y,
  - Inverse transformation $\mathbf{A}^{-1}$ from Y to X also exists.

$$\mathbf{u} = \alpha_i \mathbf{x}_i \quad \text{and} \quad \mathbf{u} = \beta_i \mathbf{y}_i = \beta_i A_{ij} \mathbf{x}_j$$

- Let **B** and **C** be two matrices. Then if

$$\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}\,\mathbf{A}$$

B and **C** represent the same matrix transformation with respect to different bases and are called Similar Matrices.

- If **A** is orthogonal then $\mathbf{C} = \mathbf{A}^t\mathbf{B}\mathbf{A}$

Eigenvalue problem

1. $x \neq 0$,
2. $Ax = \lambda x$.

- λ is an eigenvalue and x is an eigenvector.
- If $y^H A = \lambda y^H$, then (λ, y) is a left eigenpair
- If $Ax = \lambda x$, then $(\lambda I - A)x = 0$. Hence $(\lambda I - A)$ is singular.
- The eigenvalues of A are the roots of the *characteristic equation*

$$p(\lambda) \equiv \det(\lambda I - A) = 0.$$

- No distinction between left and right eigenvalues.
- The *characteristic polynomial* p can be factored in the form

$$p(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k},$$

where the numbers λ_i are distinct and

$$m_1 + m_2 + \cdots + m_k = n.$$

- m_i is the *algebraic multiplicity* of λ_i.

Remarks: Eigenvalues and Eigenvectors

- Eigenvalues and Eigenvectors of a real symmetric matrix are real.
- In general since eigenvalues are determined by solving a polynomial equation, they can be complex.
- Further roots can be repeated ➔ multiple eigenvectors correspond to a single eigenvalue.
- Transforming matrix into eigenbasis yields a diagonal matrix.

$\mathbf{Q}^t\mathbf{AQ} = \mathbf{\Lambda}$ $\quad$ $\mathbf{\Lambda}$ is a matrix of eigenvalues

  - Knowing the eigenvectors we can solve an equation $\mathbf{Ax} = \mathbf{b}$. Rewrite it as

$$\mathbf{Q}^t\mathbf{AQQ}^t\mathbf{x} = \mathbf{Q}^t\mathbf{b} \quad \mathbf{\Lambda y} = \mathbf{f}$$

  - Where $\mathbf{y} = \mathbf{Q}^t\mathbf{x}$ and $\mathbf{f} = \mathbf{Q}^t\mathbf{b}$
  - Can get **x** from **y** $\quad \mathbf{x} = (\mathbf{Q}^t)^{-1}\mathbf{y} = \mathbf{Qy}$
- Determinant is unchanged by an orthogonal transformation.
- Determinant: Det(**A**) = $\lambda_1 \lambda_2 \ldots \lambda_n$

When is Ax=b Solvable?

- When does the equation **Ax**=**b** have a solution?
  - Usual answer is if **A** is invertible
  - However in many situations where **A** is singular there still may be a meaningful solution.
- Fredholm Alternative Theorem.
  - Look at the homogeneous systems

$\mathbf{Ax} = 0 \quad (1) \qquad \mathbf{A}^*\mathbf{y} = 0 \quad (2)$

  - If (1) has only the trivial solution then so does (2). This occurs only if det(**A**) ≠ 0 (if **A** is invertible). Then **Ax**=**b** has a unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$
  - If (1) has nontrivial solutions then det(**A**)=0.
    - This means rows of **A** have interdependencies. In this case **b** must reflect those dependencies

- If 2nd row of A is a sum of the 1st and 3rd rows, then $b_2=b_1+b_3$
- If there are k independent solutions to equation (1) then **A** has a k dimensional ***nullspace***.
- $\mathbf{A}^*$ also has a k dimensional nullspace (but with different solutions).
  - Let these solutions be $\mathbf{n}_{*1}, \mathbf{n}_{*2}, \ldots, \mathbf{n}_{*k}$
- For **Ax=b** can have solutions iff

$$\langle \mathbf{b},\mathbf{n}_{*j} \rangle = 0 \qquad j = 1,\ldots,k$$

- **b** must be orthogonal to the nullspace of $\mathbf{A}^*$.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 1 & -2 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \text{ or } \mathbf{Ax} = \mathbf{b} \qquad \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & -2 \\ 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- Any solution with $y_2=-y_1$ and $y_3=y_1$ satisfies the adjoint equation or the nullspace of $\mathbf{A}^*$ is $\alpha\,[1,-1,1]^t$
- Here $\langle \mathbf{b},\mathbf{n}_{*1} \rangle = -1 (\neq 0)$. So equation has no solution.
- However if $\mathbf{b}=[1, 2, 1]^t$ we would have a solution
- General solution is $\mathbf{x} = \mathbf{x}^{\sim} + c_k \mathbf{n}_{*k}$ where $\mathbf{x}^{\sim}$ is a particular solution.

Least Squares

- Number of equations and unknowns may not match
- Look for solution by maximizing $\|\mathbf{Ax} - \mathbf{b}\|$
- $(A_{ij}x_j-b_i).(A_{ik}x_k-b_i)$ with respect to x_l
- Recall $\frac{\partial x_i}{\partial x_l} = \delta_{il}$

$$\frac{\partial}{\partial x_l}(A_{ij}x_j - b_i)\bullet(A_{ik}x_k - b_i) = 0$$

$$(A_{ij}\delta_{jl})\bullet(A_{ik}x_k - b_i) + (A_{ij}x_j - b_i)\bullet(A_{ik}\delta_{kl}) = 0$$

$$A_{il}\bullet(A_{ik}x_k - b_i) + (A_{ij}x_j - b_i)\bullet A_{il} = 2\left(A_{il}A_{ik}x_k - A_{il}b_i\right) = 0$$

$$A_{il}A_{ik}x_k = A_{il}b_i$$

- Same as the solution of $\mathbf{A}^t\mathbf{Ax}=\mathbf{A}^t\mathbf{b}$
- Shows the power of the index notation
  - See again the appearance of $\mathbf{A}^t\mathbf{A}$

Singular Value Decomposition

- Chief tool for dealing with m by n systems and singular systems.
- **Singular values**: Non negative square roots of the eigenvalues of $\mathbf{A}^t\mathbf{A}$. Denoted σ_i, $i=1,\ldots,n$
  - $\mathbf{A}^t\mathbf{A}$ is symmetric → eigenvalues and singular values are real.
- SVD: If **A** is a real m by n matrix then there exist orthogonal matrices **U** ($\in\mathbb{R}^{m\times m}$) and **V** ($\in\mathbb{R}^{n\times n}$) such that $\mathbf{U}^t\mathbf{A}\mathbf{V} = \Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$ $\qquad p=\min\{m,n\}$

$$\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^t$$

- Geometrically, singular values are the lengths of the hyperellipsoid defined by $E=\{\mathbf{Ax}: \|\mathbf{x}\|_2=1\}$
- Singular values arranged in decreasing order.

Properties of the SVD

- Suppose we know the singular values of **A** and we know r are non zero

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq \sigma_{r+1} = \ldots = \sigma_p = 0$$

  - Rank(**A**) = r.
  - Null(**A**) = span$\{\mathbf{v}_{r+1},\ldots,\mathbf{v}_n\}$
  - Range(**A**)=span$\{\mathbf{u}_1,\ldots,\mathbf{u}_r\}$
- $\|A\|_F^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_p^2 \qquad \|A\|_2 = \sigma_1$
- *Numerical rank:* If k singular values of A are larger than a given number ε. Then the ε rank of A is k.
- Distance of a matrix of rank n from being a matrix of rank $k = \sigma_{k+1}$

Why is it useful?

- Square matrix may be singular due to round-off errors. Can compute a "regularized" solution
  - $\mathbf{x}=\mathbf{A}^{-1}\mathbf{b}=(\mathbf{U}\,\Sigma\,\mathbf{V}^t)^{-1}\mathbf{b} = \sum^{n} \frac{\mathbf{u}_i^t\mathbf{b}}{\;}\mathbf{v}_i$
- If σ_i is small (vanishes) the solution "blows up"
- Given a tolerance ε we can determine a solution that is "closest" to the solution of the original equation, but that does not "blow up" $\quad \mathbf{x}_r = \sum_{i=1}^{k} \frac{\mathbf{u}_i^t\mathbf{b}}{\sigma_i}\mathbf{v}_i \qquad \sigma_k > \varepsilon,\ \sigma_{k+1} \leq \varepsilon$
- Least squares solution is the x that satisfies $\mathbf{A}^t\mathbf{Ax}=\mathbf{A}^t\mathbf{b}$
- can be effectively solved using SVD