Some notes on Linear Algebra

Mark Schmidt

September 10, 2009

References

- Linear Algebra and Its Applications. Strang, 1988.
- Practical Optimization. Gill, Murray, Wright, 1982.
- Matrix Computations. Golub and van Loan, 1996.
- Scientific Computing. Heath, 2002.
- Linear Algebra and Its Applications. Lay, 2002.

- The material in these notes is from the first two references, the outline roughly follows the second one. Some figures/examples taken directly from these sources.

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Vectors, Matrices

- Scalar (1 by 1): α
- Column Vector (m by 1): $a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$
- Row Vector (1 by n): $a^T = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$
- Matrix (m by n):

$$A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

- Matrix transpose (n by m):

$$(A^T)_{ij} = (A)_{ji} \qquad A^T = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

- A matrix is symmetric if $A = A^T$

Addition and Scalar Multiplication

- Vector Addition:

$$a + b = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}$$

- Scalar Multiplication:

$$\alpha b = \alpha \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \alpha b_1 \\ \alpha b_2 \end{bmatrix}$$

- These are associative and commutative:

$$A + (B + C) = (A + B) + C$$

$$A + B = B + A$$

- Applying them to a set of vectors is called a linear combination:

$$c = \alpha_1 b_1 + \alpha_2 b_2 + \ldots \alpha_n b_n$$

Inner Product

- The inner product between vectors of the same length is:

$$a^T b = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \gamma$$

- The inner product is a scalar:

$$(a^T b)^{-1} = 1/(a^T b)$$

- It is commutative and distributive across addition:

$$a^T b = b^T a$$
$$a^T (b + c) = a^T b + a^T c$$

- In general it is not associative (result is not a scalar):

$$a^T (b^T c) \neq (a^T b)^T c$$

- Inner product of non-zero vectors can be zero:

$$a^T b = 0$$ Here, a and b are called orthogonal

Matrix Multiplication

- We can 'post-mulitply' a matrix by a column vector:

$$Ax = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1^T x_1 \\ a_2^T x_2 \\ a_3^T x_3 \end{bmatrix}$$

- We can 'pre-multiply' a matrix by a row vector:

$$x^T A = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} x^T a_1 & x^T a_2 & x^T a_3 \end{bmatrix}$$

- In general, we can multiply matrices A and B when the number of columns in A matches the number of rows in B:

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & a_1^T b_3 \\ a_2^T b_1 & a_2^T b_2 & a_2^T b_3 \\ a_3^T b_1 & a_3^T b_2 & a_3^T b_3 \end{bmatrix}$$

Matrix Multiplication

- Matrix multiplication is associative and distributive across (+):

$$A(BC) = (AB)C$$

$$A(B + C) = AB + AC$$

- In general it is not commutative:

$$AB \neq BA$$

- Transposing product reverses the order (think about dimensions):

$$(AB)^T = B^T A^T$$

- Matrix-vector multiplication always yields a vector:

$$x^T Ay = x^T (Ay) = \gamma = (Ay)^T x = y^T A^T x$$

- Matrix powers don't change the order: $(AB)^2 = ABAB$

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Identity Matrix

- The identity matrix has 1's on the diagonal and 0's otherwise:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Multiplication by the identity matrix of the appropriate size yields the original matrix:

$$I_m A = A = A I_n$$

- Columns of the identity matrix are called elementary vectors:

$$e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Triangular/Tridiagonal

- A diagonal matrix has the form:

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \qquad D = diag(d)$$

- An upper triangular matrix has the form:

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

- 'Triangularity' is closed under multiplication
- A tridiagonal matrix has the form:

$$T = \begin{bmatrix} t_{11} & t_{12} & 0 & 0 \\ t_{21} & t_{22} & t_{23} & 0 \\ 0 & t_{32} & t_{33} & t_{34} \\ 0 & 0 & t_{43} & t_{44} \end{bmatrix}$$

- 'Tridiagonality' is lost under multiplication

Rank-1, Elementary Matrix

- The inner product between vectors is a scalar, the outer product between vectors is a rank-1 matrix:

$$uv^T = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{bmatrix}$$

- The identity plus a rank-1 matrix is a called an elementary matrix:

$$E = I + \alpha uv^T$$

- These are 'simple' modifications of the identity matrix

Orthogonal Matrices

- A set of vectors is orthogonal if:

$$q_i^T q_j = 0, i \neq j$$

- A set of orthogonal vectors is orthonormal if:

$$q_i^T q_i = 1$$

- A matrix with orthonormal columns is called orthogonal
- Square orthogonal matrices have a very useful property:

$$Q^T Q = I = Q Q^T$$

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Linear Combinations

- Given k vectors, a linear combination of the vectors is:

$$c = \alpha_1 b_1 + \alpha_2 b_2 + \dots \alpha_n b_n$$

- If all alpha$_i$=0, the linear combination is trivial
- This can be re-written as a matrix-vector product:

$$c = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

- Conversely, any matrix-vector product is a linear combination of the columns

$$\begin{bmatrix} 1 & 0 \\ 5 & 4 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \rightarrow u \begin{bmatrix} 1 \\ 5 \\ 2 \end{bmatrix} + v \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Linear Dependence

- A vector is linearly dependent on a set of vectors if it can be written as a linear combination of them:

$$c = \alpha_1 b_1 + \alpha_2 b_2 + \dots \alpha_n b_n$$

- We say that c is 'linearly dependent' on $\{b_1, b_2, \dots, b_3\}$, and that the set $\{c, b_1, b_2, \dots, b_3\}$ is 'linearly dependent'

- A set is linearly dependent iff the zero vector can be written as a non-trivial combination:

$$\exists_{\alpha \neq 0}, \;\; s.t. \;\; 0 = \alpha_1 b_1 + \alpha_2 b_2 + \dots \alpha_n b_n \Rightarrow \{b_1, b_2, \dots, b_n\} \;\; \text{dependent}$$

Linear Independence

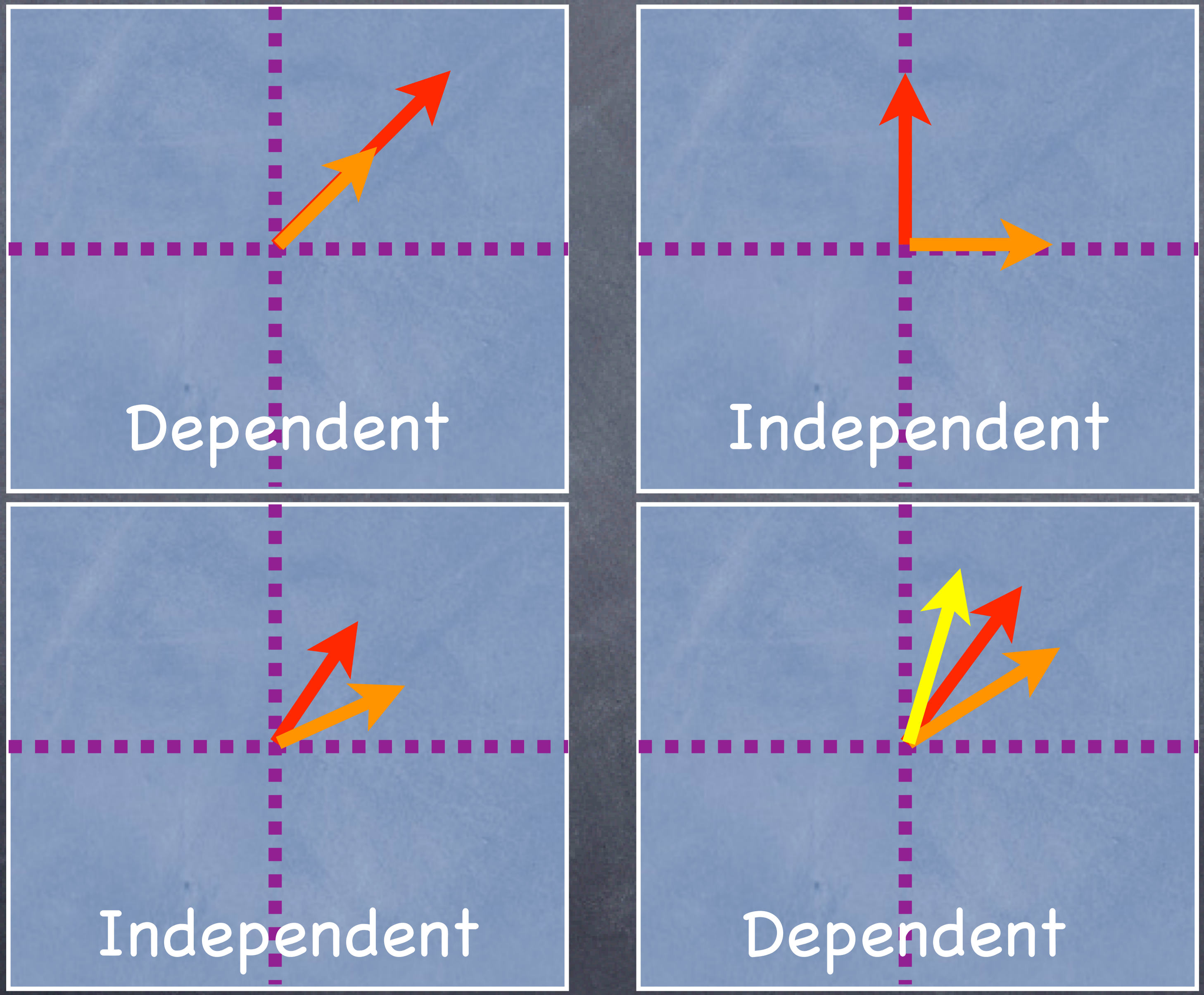
- If a set of vectors is not linearly dependent, we say it is linearly independent
- The zero vector cannot be written as a non-trivial combination of independent vectors:

$$0 = \alpha_1 b_1 + \alpha_2 b_2 + \dots \alpha_n b_n \Rightarrow \alpha_i = 0 \;\; \forall_i$$

- A matrix with independent columns has full column rank
- In this case, Ax=0 implies that x=0

Linear [In]Dependence

- Independence in R^2:



Vector Space

- A **vector space** is a set of objects called 'vectors', with closed operations 'addition' and 'scalar multiplication' satisfying certain axioms:

  1. $x + y = y + x$
  2. $x + (y + z) = (x + y) + z$
  3. exists a "zero-vector" 0 s.t. $\forall_x, x + 0 = x$
  4. $\forall_x$, exists an 'additive inverse' $-x$, s.t. $x + (-x) = 0$
  5. $1x = x$
  6. $(c_1 c_2)x = c_1(c_2 x)$
  7. $c(x + y) = cx + cy$
  8. $(c_1 + c_2)x = c_1 x + c_2 x$

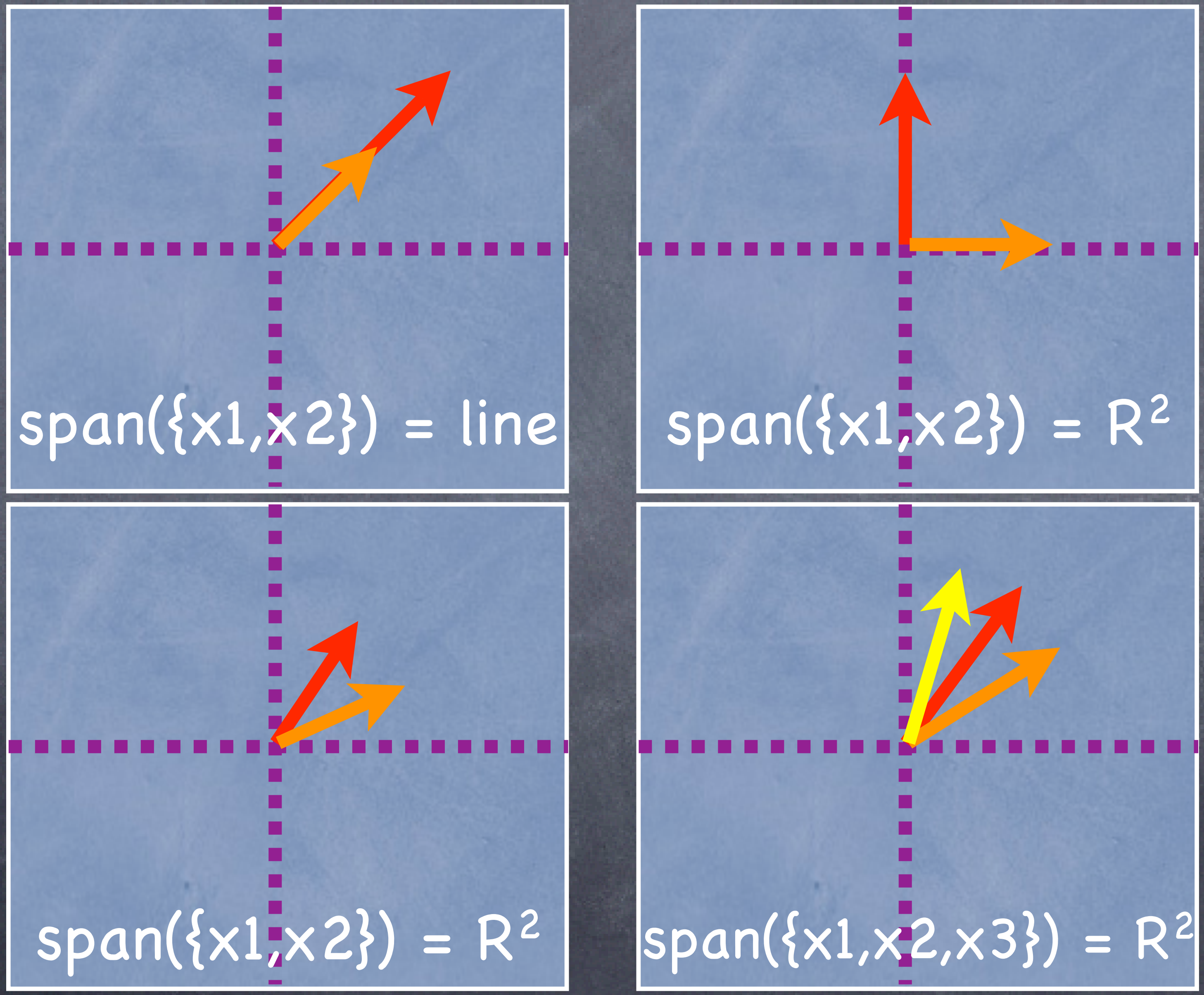
- Examples: $\mathbb{R}, \mathbb{R}^2, \mathbb{R}^n, \mathbb{R}^{mn}$

Subspace

- A (non-empty) subset of a vector space that is closed under addition and scalar multiplication is a subspace
- Possible subspaces of R^3:
  - 0 vector (smallest subspace and in all subspaces)
  - any line or plane through origin
  - All of R^3
- All linear combinations of a set of vectors {a1,a2,...,an} define a subspace
- We say that the vectors generate or span the subspace, or that their range is the subspace

Subspace

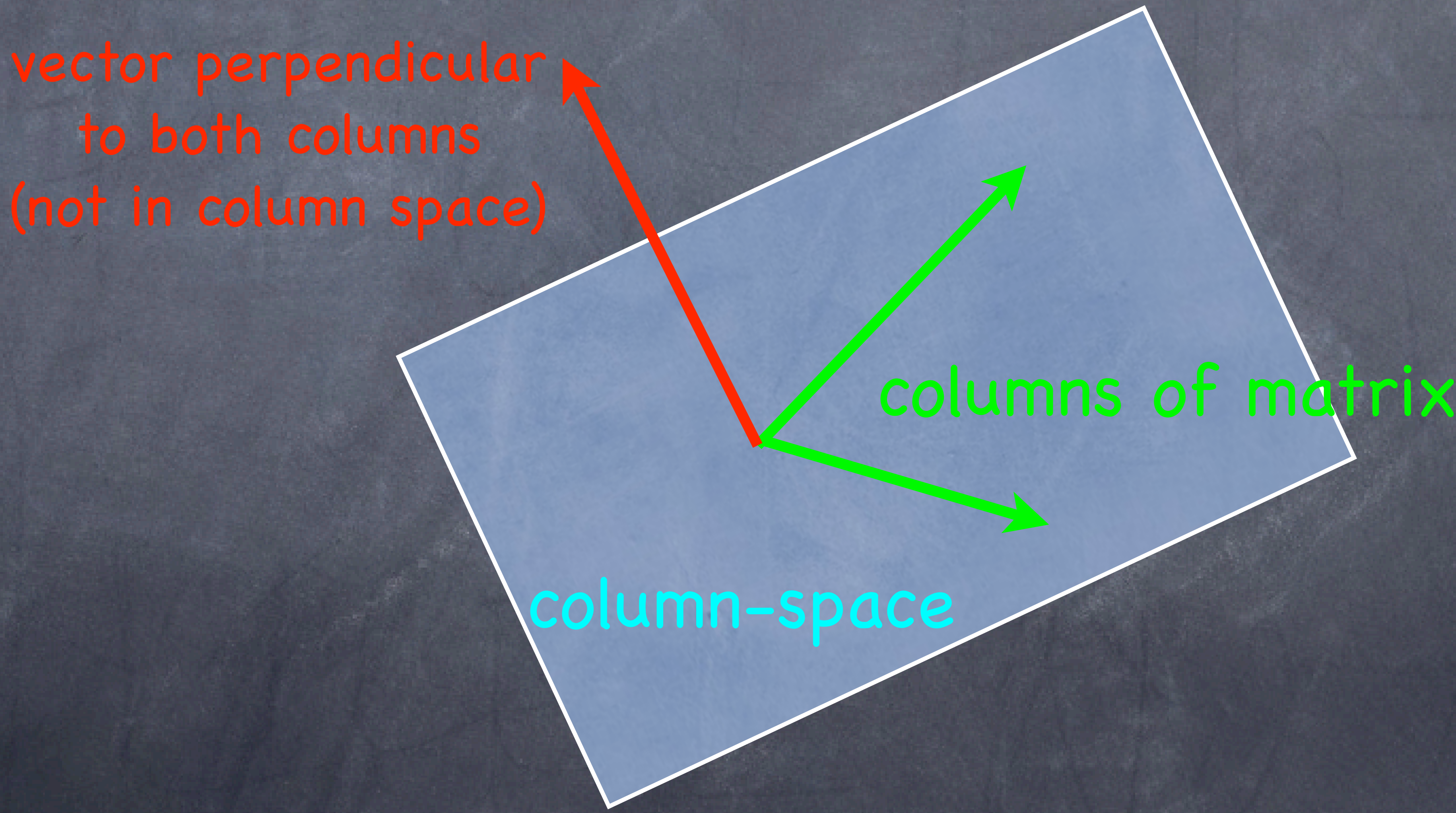
- Subspaces generated in R^2:



Column-Space

- The column-space (or range) of a matrix is the subspace spanned by its columns:

$$\mathcal{R}(A) = \{\text{All } b \text{ such that } Ax = b\}$$



- The system Ax=b is solvable iff b is in A's column-space

Column-Space

- The column-space (or range) of a matrix is the subspace spanned by its columns:

$$\mathcal{R}(A) = \{\text{All } b \text{ such that } Ax = b\}$$

- The system Ax=b is solvable iff b is in A's column-space
- Any product Ax (and all columns of any product AB) must be in the column space of A
- A non-singular square matrix will have $\Re(A) = R^m$
- We analogously define the row-space:

$$\mathcal{R}(A^T) = \{\text{All } b \text{ such that } x^T A = b^T\}$$

Dimension, Basis

- The vectors that span a subspace are not unique
- However, the minimum number of vectors needed to span a subspace is unique
- This number is called the dimension or rank of the subspace
- A minimal set of vectors that span a space is called a basis for the space
- The vectors in a basis must be linearly independent (otherwise, we could remove one and still span space)

Orthogonal Basis

- Any vector in the subspace can be represented uniquely as a linear combination of the basis
- If the basis is orthogonal, finding the unique coefficients is easy:

$$c = \alpha_1 b_1 + \alpha_2 b_2 + \ldots \alpha_n b_n$$

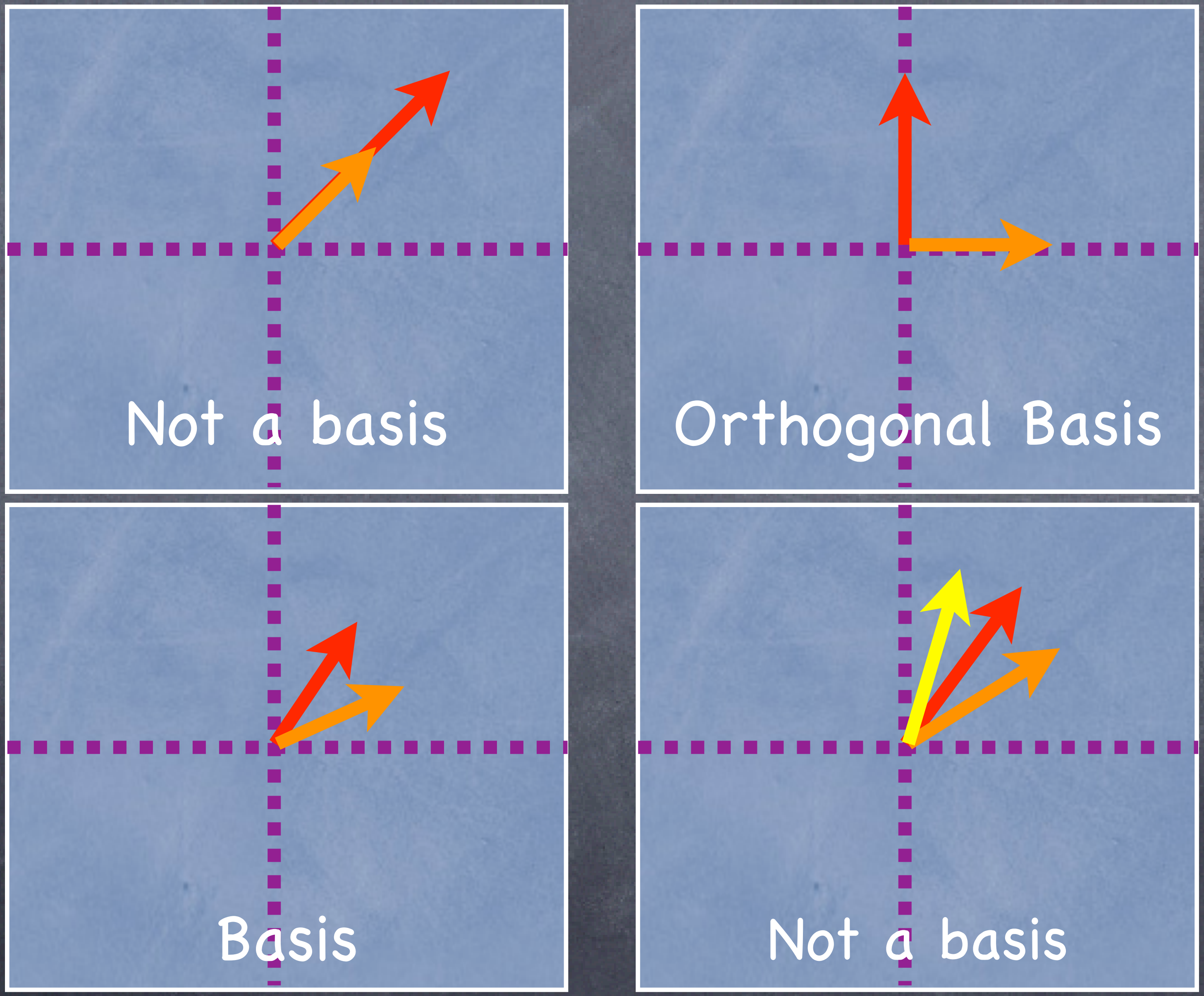
$$b_1^T c = \alpha_1 b_1^T b_1 + \alpha_2 b_1^T b_2 + \ldots \alpha_n b_1^T b_n$$
$$= \alpha b_1^T b_1$$

$$\alpha_1 = b_1^T c / b_1^T b_1$$

- The Gram-Schmidt procedure is a way to construct an orthonormal basis.

Basis

- Basis in R^2:



Orthogonal Subspace

- Orthogonal subspaces: Two subspaces are orthogonal if every vector in one subspace is orthogonal to every vector in the other
- In R^3:
  - {0} is orthogonal to everything
  - Lines can be orthogonal to {0}, lines, or planes
  - Planes can be orthogonal to {0}, lines (NOT planes)
- The set of ALL vectors orthogonal to a subspace is also a subspace, called the orthogonal complement
- Together, the basis for a subspace and its orthogonal complement span R^n
- So if k is the dimension of the original subspace of R^n, then the orthogonal complement has dimension n-k

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Matrices as Transformation

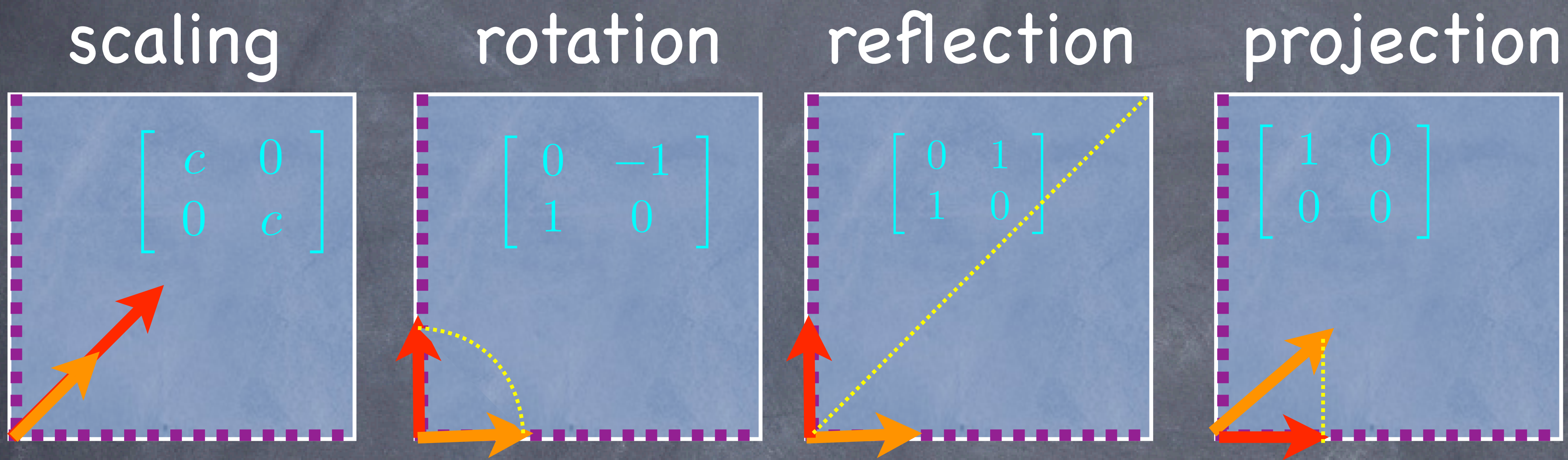
- Instead of a collection of scalars or (column/row) vectors, a matrix can also be viewed as a transformation applied to vectors:

$$T(x) = Ax$$

- The domain of the function is R^m
- The range of the function is a subspace of R^n (the column-space of A)
- If A has full column rank, the range is R^n

Matrices as Transformation

- Many transformation are possible, for example:



- The transformation must be linear:

$$T(\alpha x + \beta y) = \alpha T(x) + \alpha T(y)$$

- Any linear transformation has a matrix representation

Null-Space

- A linear transformation can't move the origin:

$$T(0) = A0 = 0$$

- But if A has linearly dependent columns, there are non-zero vectors that transform to zero:

$$\exists_{x \neq 0} \text{ s.t. } T(x) = Ax = 0$$

- A square matrix with this property is called singular

- The set of vectors that transform to zero forms a subspace called the null-space of the matrix:

$$\mathcal{N}(A) = \{\text{All } x \text{ such that } Ax = 0\}$$

Orthogonal Subspaces (again)

- The null-space:

$$\mathcal{N}(A) = \{\text{All } x \text{ such that } Ax = 0\}$$

- Recall the row-space:

$$\mathcal{R}(A^T) = \{\text{All } b \text{ such that } x^T A = b^T\}$$

- The row-Space is orthogonal to Null-Space
  - Let y be in $\Re(A^T)$, and x be in $\mathcal{N}(A)$:

$$y^T x = z^T A x = z^T (Ax) = z^T 0 = 0$$

Fundamental Theorem

- Column-space: $\mathcal{R}(A) = \{\text{All } b \text{ such that } Ax = b\}$
- Null-space: $\mathcal{N}(A) = \{\text{All } x \text{ such that } Ax = 0\}$
- Row-space: $\mathcal{R}(A^T) = \{\text{All } b \text{ such that } x^T A = b^T\}$
- The Fundamental Theorem of Linear Algebra describes the relationships between these subspaces:

$$r = dim(\mathcal{R}(A)) = dim(\mathcal{R}(A^T))$$

$$n = r + (n - r) = dim(\mathcal{R}(A)) + dim(\mathcal{N}(A))$$

- Row-space is orthogonal complement of null-space
- Full version includes results involving 'left' null-space

Inverses

- Can we undo a linear transformation from Ax to b?
- We can find the inverse iff A is square + non-singular (otherwise we either lose information to the null-space or can't get to all b vectors)
- In this case, the unique inverse matrix A^{-1} satisfies:

$$A^{-1}A = I = AA^{-1}$$

- Some useful identities regarding inverses:

$$(A^{-1})^T = (A^T)^{-1}$$

$$(\gamma A)^{-1} = \gamma^{-1}A^{-1}$$

$$(AB)^{-1} = B^{-1}A^{-1}$$

(assuming A^{-1} and B^{-1} exist)

Inverses of Special Matrices

- Diagonal matrices have diagonal inverses:

$$D = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \qquad D^{-1} = \begin{bmatrix} 1/d_1 & 0 & 0 \\ 0 & 1/d_2 & 0 \\ 0 & 0 & 1/d_3 \end{bmatrix}$$

- Triangular matrices have triangular inverses:

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}$$

- Tridiagonal matrices do not have sparse inverses

- Elementary matrices have elementary inverses (same uv^T):

$$(I + \alpha uv^T)^{-1} = I + \beta uv^T, \beta = -\alpha/(1 + \alpha u^T v)$$

- The transpose of an orthogonal matrix is its inverse:

$$Q^T Q = I = QQ^T$$

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Matrix Trace

- The trace of a square matrix is the sum of its diagonals:

$$tr(A) = \sum_{i=1}^{n} a_{ii}$$

- It is a linear transformation:

$$\gamma tr(A + B) = \gamma tr(A) + \gamma tr(B)$$

- You can reverse the order in the trace of a product:

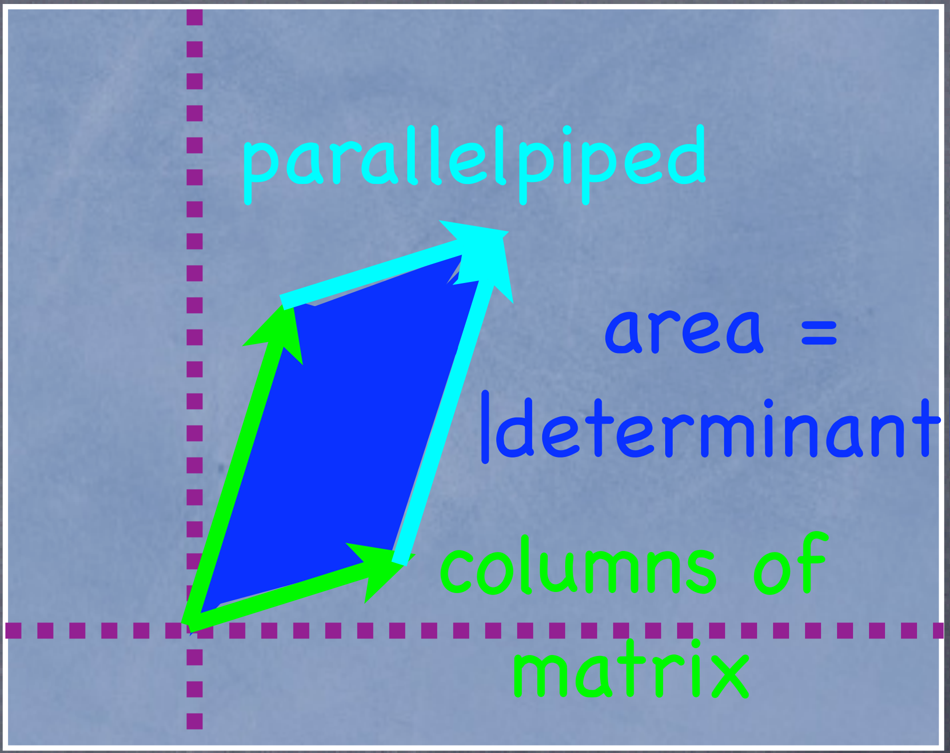
$$tr(AB) = tr(BA)$$

- More generally, it has the cyclic property:

$$tr(ABC) = tr(CAB) = tr(BCA)$$

Matrix Determinant

- The determinant of a square matrix is a scalar number associated with it that has several special properties
- Its absolute value is the volume of the parallelpiped formed from its columns
- det(A) = 0 iff A is singular
- det(AB) = det(A)det(B), det(I) = 1
- $\det(A^T) = \det(A)$, $\det(A^{-1}) = 1/\det(A)$
- exchanging rows changes sign of det(A)
- Diagonal/triangular: determinant is product(diagonals)
- determinants can be calculated from LU factorization:
  - A = PLU = det(P)det(L)det(U) = (+/-)prod(diags(U))
    (sign depends on even/odd number of row excahnges)

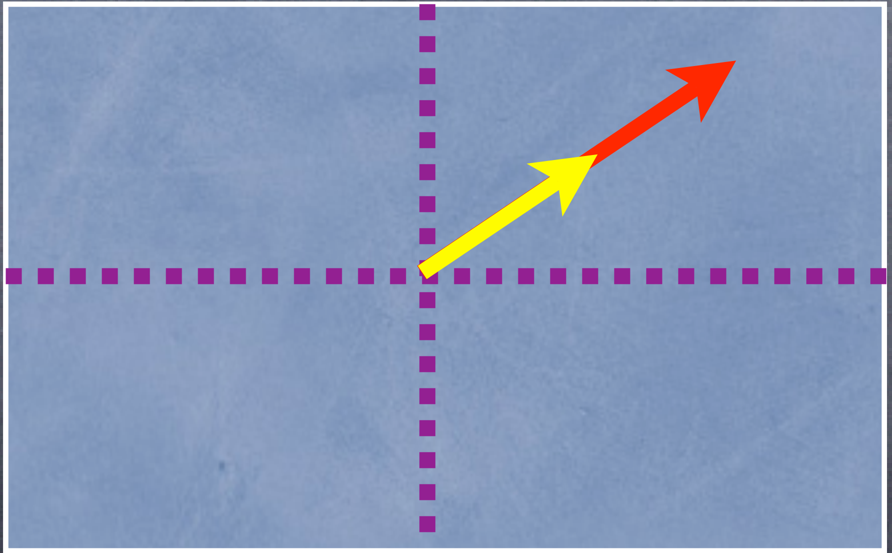


Eigenvalues

- A scalar lambda is an eigenvalue (and u is an eigenvector) of A if:

$$Au = \lambda u$$

- The eigenvectors are vectors that only change in magnitude, not direction (except sign)



- Multiplication of eigenvector by A gives exponential growth/decay (or stays in 'steady state' if lambda = 1):

$$AAAAu = \lambda AAAu = \lambda^2 AAu = \lambda^3 Au = \lambda^4 u$$

Computation (small A)

- Multiply by I, move everything to LHS:

$$Ax = \lambda x, \ (A - \lambda I)x = 0$$

- Eigenvector x is in the null-space of (A-λI)
- Eigenvalues λ make (A-λI) singular (have a Null-space)
- Computation (in principle):
  - Set up equation det(A-λI) = 0 (characteristic poly)
  - Find the roots of the polynomial (eigenvalues)
  - For each root, solve (A-λI)x = 0 (eigenvector)
- Problem: In general, no algebraic formula for roots

Eigenvalues (Properties)

- Eigenvectors are not unique (scaling)
- $\text{sum}(\lambda_i) = \text{tr}(A)$, $\text{prod}(\lambda_i) = \det(A)$, $\text{eigs}(A^{-1}) = 1/\text{eigs}(A)$
- Real matrix can have complex eigenvalues (pairs)
- Eg:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\ det(A - \lambda I) = \lambda^2 + 1,\ \lambda = i, -i$$

- If two matrices have the same eigenvalues, we say that they are similar
- For non-singular W, WAW^{-1} is similar to A:

$$Ax = \lambda x$$

$$AW^{-1}Wx = \lambda x$$

$$WAW^{-1}(Wx) = \lambda(Wx)$$

Spectral Theorem

- A matrix with n independent eigenvalues can be diagonalized by a matrix S containing its eigenvectors

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \dots \end{bmatrix}$$

- Spectral theorem: for any symmetric matrix:
  - the eigenvalues are real
  - the eigenvectors can be made orthonormal (so $S^{-1}=S^T$)
- The maximum eigenvalue satisfies: $\lambda = \max_{x \neq 0} \frac{x^T A x}{x^T x}$
- The minimum eigenvalue satisfies: $\lambda = \min_{x \neq 0} \frac{x^T A x}{x^T x}$
- The spectral radius is eigenvalue with largest absolute value

Definiteness

- A matrix is called positive definite if all eigenvalues are positive
- If this case: $\forall_{x \neq 0} \;\; x^T A x > 0$
- If the eigenvalues are non-negative, the matrix is called positive semi-definite and:

$$\forall_{x \neq 0} \;\; x^T A x \geq 0$$

- Similar definitions hold for negative [semi-]definite
- If A has positive and negative eigenvalues it is indefinite ($x^T Ax$ can be positive or negative)

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Vector Norm

- A norm is a scalar measure of a vector's length
- Norms must satisfy three properties:

$$||x|| \geq 0 \ \ \text{(with equality iff } \ x = 0)$$
$$||\gamma x|| = |\gamma| ||x||$$
$$||x + y|| \leq ||x|| + ||y|| \ \ \text{(the triangle inequality)}$$

- The most important norm is the Euclidean norm:

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \qquad ||x||_2^2 = x^T x$$

- Other important norms:

$$||x||_1 = \sum_i |x_i| \qquad ||x||_\infty = \max_i |x_i|$$

Cauchy-Scwartz

- Apply law of cosines to triangle formed from x and y:

$$||y-x||_2^2 = ||y||_2^2 + ||x||_2^2 - 2||y||_2||x||_2 \cos\theta$$

- Use: $$||y-x||_2^2 = (y-x)^T(y-x)$$

- To get relationship between lengths and angles:

$$\cos\theta = \frac{y^T x}{||x||_2||y||_2}$$

- Get Cauchy-Schwartz inequality because $|\cos(\theta)| \leq 1$:

$$|y^T x| \leq ||x||_2||y||_2$$

- A generalization is Holder's inequality:

$$|y^T x| \leq ||x||_p||y||_q \ \ (\text{for } 1/p + 1/q = 1)$$

Orthogonal Transformations

- Geometrically, an orthogonal transformation is some combination of rotations and reflections
- Orthogonal matrices preserve lengths and angles:

$$||Qx||_2^2 = x^T Q^T Q x = x^T x = ||x||_2^2$$

$$(Qx)^T (Qy) = x^T Q^T Q y = x^T y$$

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

Linear Equations

- Given A and b, we want to solve for x:

$$Ax = b \qquad \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

- This can be given several interpretations:
  - By rows: x is the intersection of hyper-planes:

$$2x - y = 1$$

$$x + y = 5$$

  - By columns: x is the linear combination that gives b:

$$x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

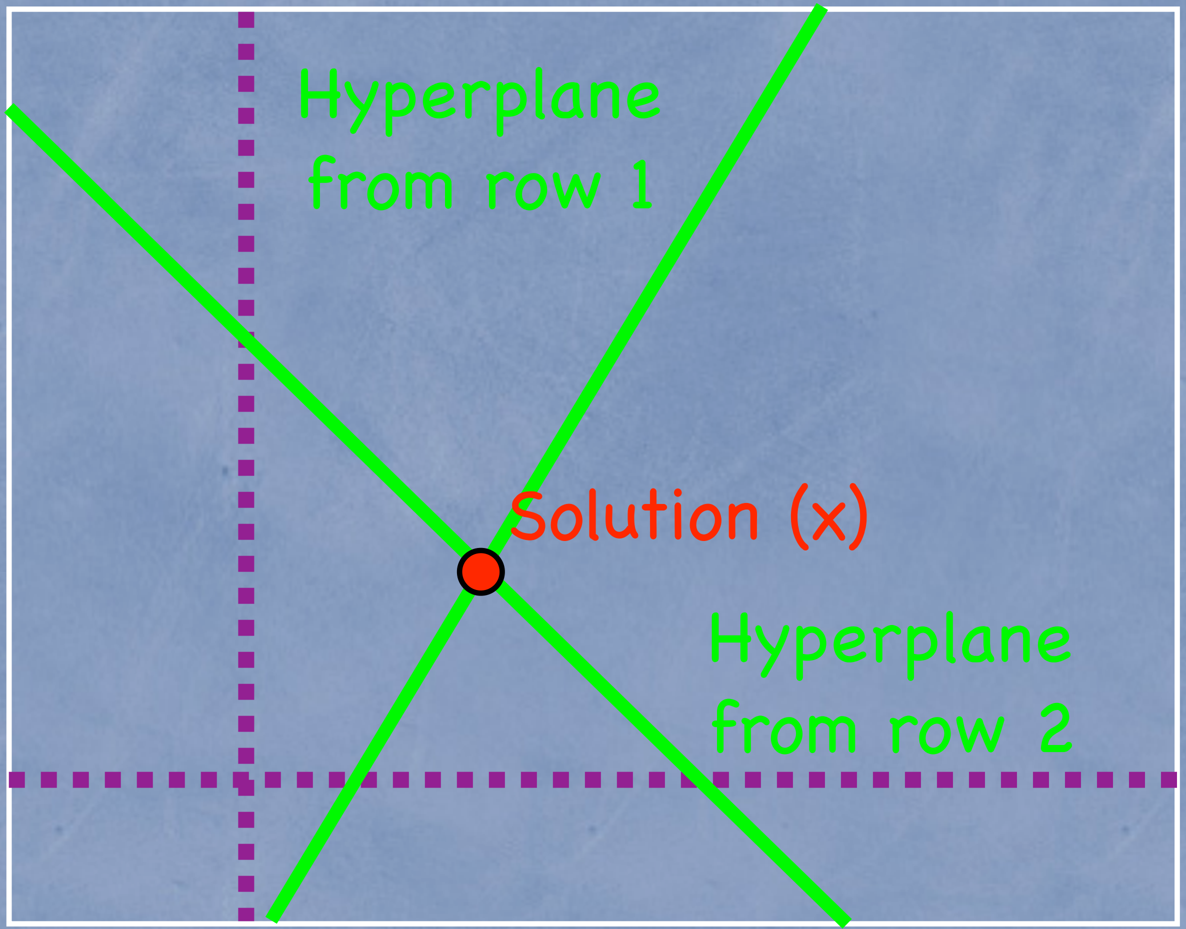
  - Transformation: x is the vector transformed to b:

$$T(x) = b$$

Geometry of Linear Equations

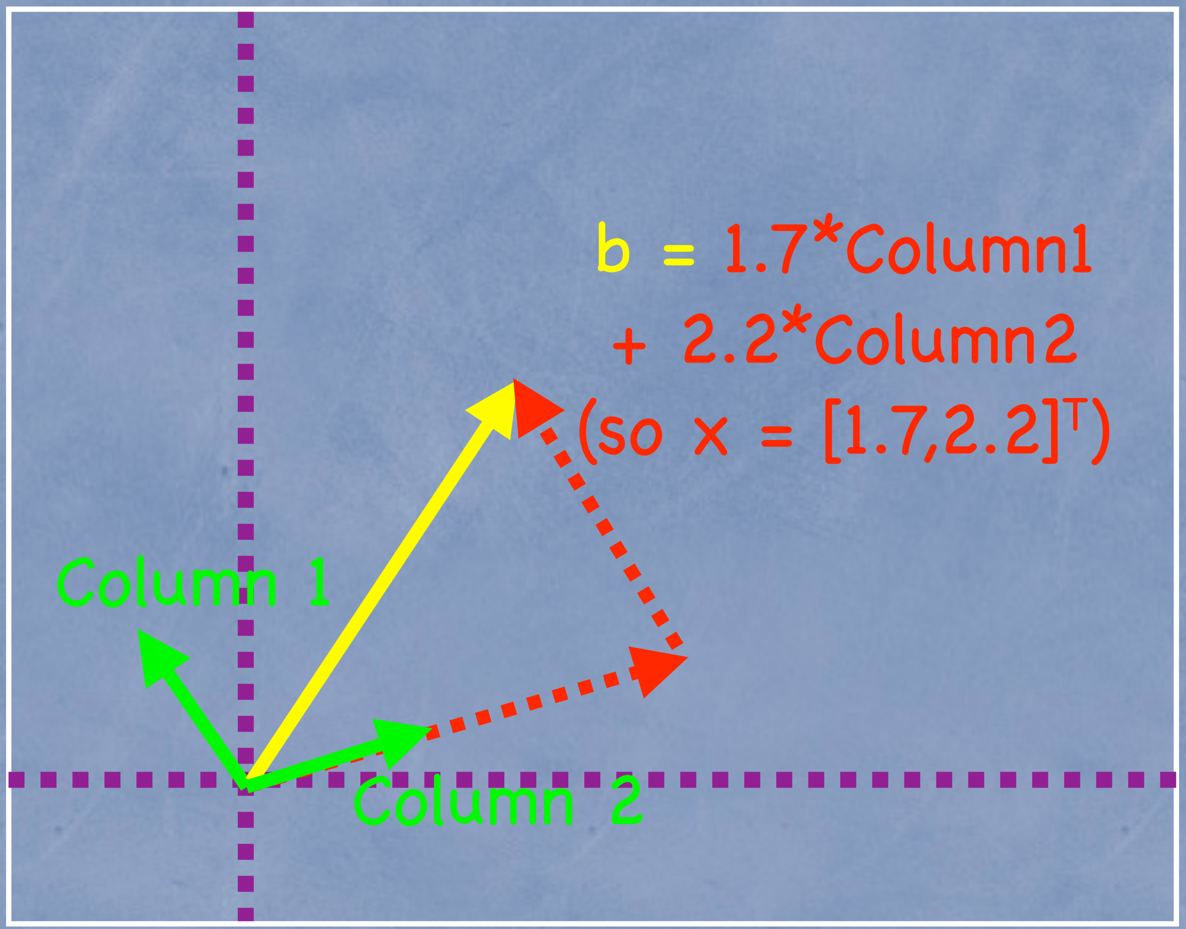
By Rows:

Find Intersection of Hyperplanes



By Columns:

Find Linear Combination of Columns

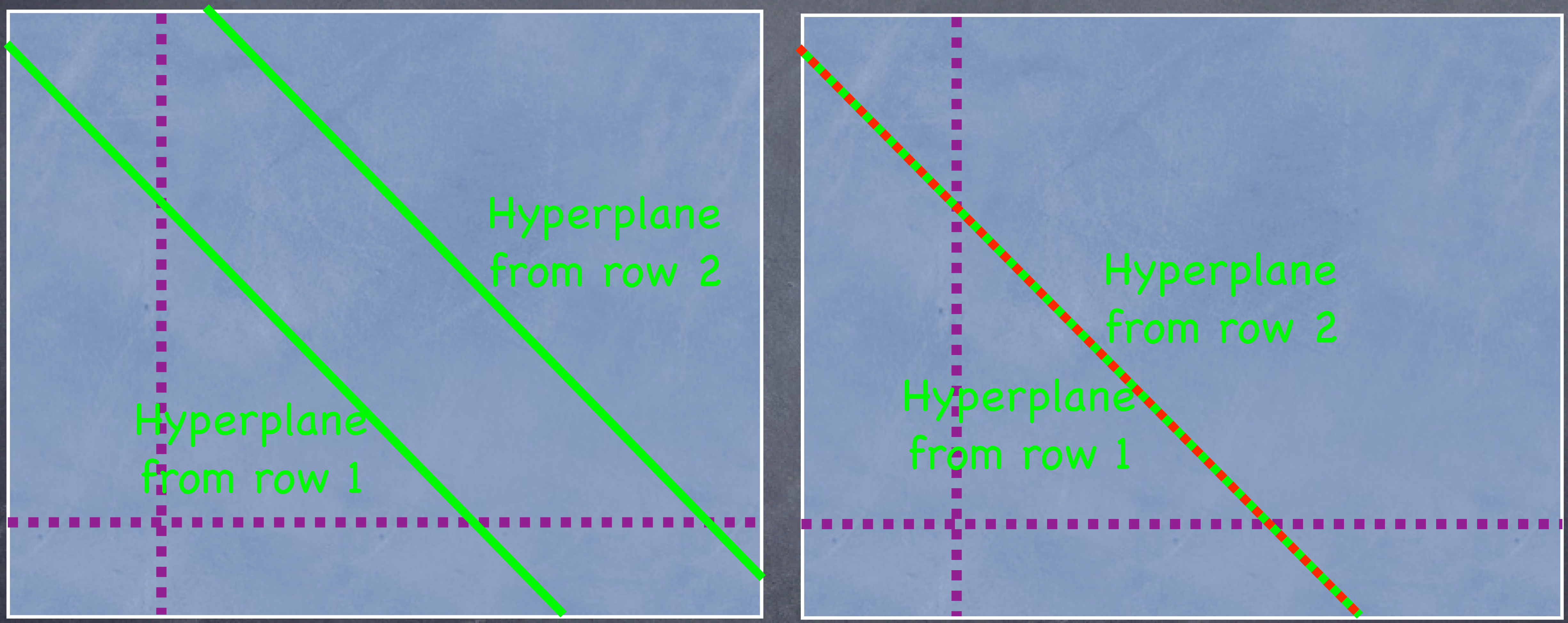


Solutions to Ax=b

- The non-singular case is easy:
  - Column-space of A is basis for R^n, so there is a unique x for every b (ie. $x = A^{-1}b$)
- In general, when does Ax=b have a solution?
  - When b is in the column-space of A

What can go wrong?

- By Rows:

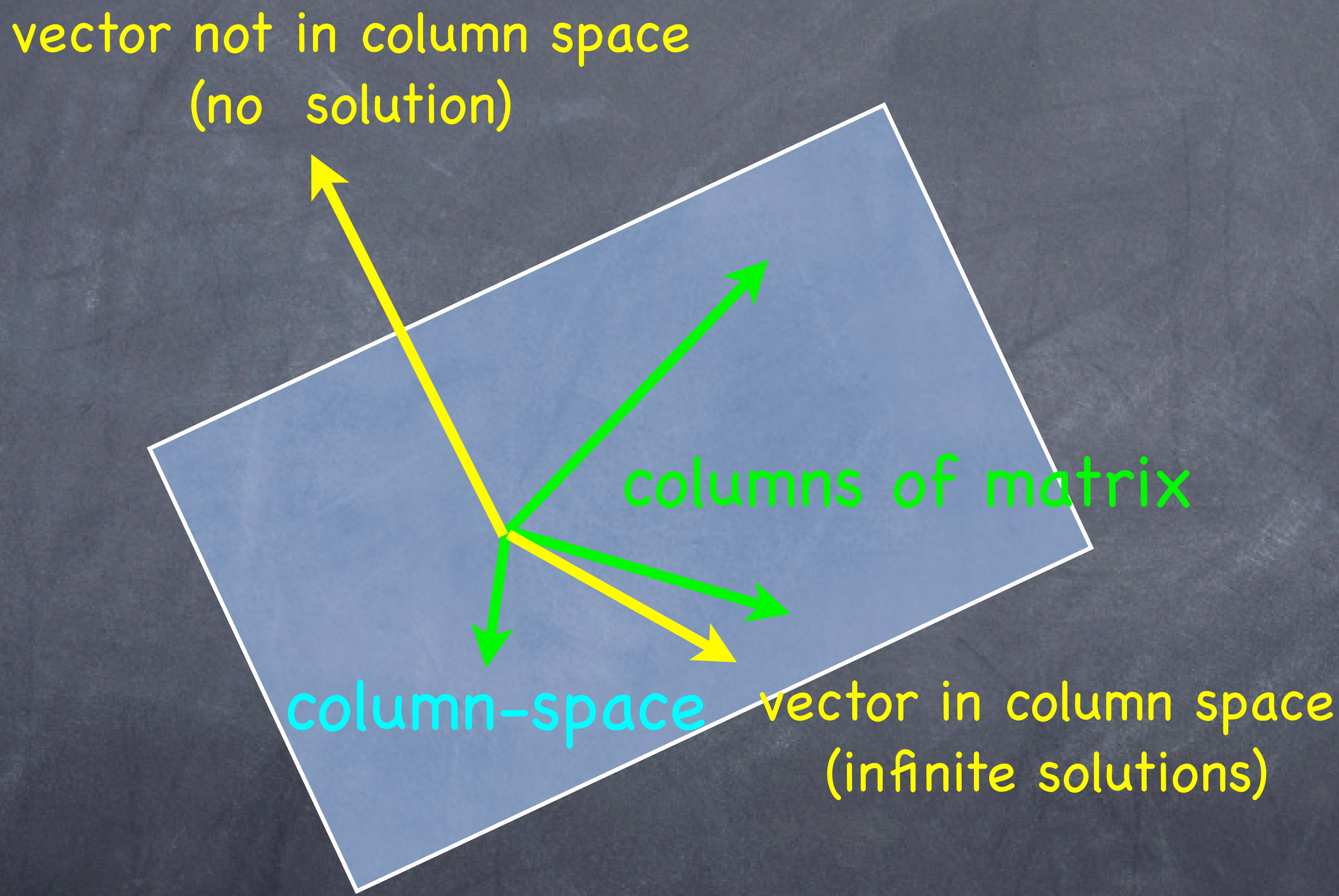


No Intersection

Infinite Intersection

What can go wrong?

- By Columns:



Solutions to Ax=b

- The non-singular case is easy:
  - Column-space of A is basis for R^n, so there is a unique x for every b (ie. $x = A^{-1}b$)
- In general, when does Ax=b have a solution?
  - When b is in the column-space of A
- In general, when does Ax=b have a unique solution?
  - When b is in the column-space of A, and the columns of A are linearly independent
  - Note: this can still happen if A is not square...

Solutions to Ax=b

- This rectangular system has a unique solution

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$$

  - b is in the column-space of A ($x_1 = 2$, $x_2 = 3$)
  - columns of A are independent (no null-space)

Characterization of Solutions

- If Ax=b has a solution, we say it is consistent
- If it is consistent, then we can find a particular solution in the column-space
- But an element of the null-space added to the particular solution will also be a solution:

$$A(x_p + y_n) = Ax_p + Ay_n = Ax_p + 0 = Ax_p = b$$

- So the general solution is:
  x = (sol'n from col-space) + (anything in null-space)
- By fundamental theorem, independent columns => trivial null-space (leading to unique solution)

Triangular Linear Systems

- Consider a square linear system with an upper triangular matrix (non-zero diagonals):

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

- We can solve this system bottom to top in O(n²)

$$u_{33}x_3 = b_3 \qquad x_1 = \frac{b_3}{u_{33}}$$

$$u_{22}x_2 + u_{23}x_3 = b_2 \qquad x_2 = \frac{b_2 - u_{23}x_2}{u_{22}}$$

$$u_{11}x_1 + u_{12}x_2 + u_{13}x_3 = b_1 \qquad x_1 = \frac{b_1 - u_{13}x_3 - u_{23}x_2}{u_{33}}$$

- This is called back-substitution
(there is an analogous method for lower triangular)

Gaussian Elimination (square)

- Gaussian elimination uses elementary row operations to transform a linear system into a triangular system:

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 5 \\ 4x_1 & + & -6x_2 & & & = & -2 \\ -2x_1 & + & 7x_2 & + & 2x_3 & = & 9 \end{array}$$

add -2 times first row to second
add 1 times first row to third

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 5 \\ & & -8x_2 & + & -2x_3 & = & -12 \\ & & 8x_2 & + & 3x_3 & = & 14 \end{array}$$

add 1 times second row to third

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 5 \\ & & -8x_2 & + & -2x_3 & = & -12 \\ & & & & x_3 & = & 2 \end{array}$$

Diagonals {2,-8,1} are called the pivots

Gaussian Elimination (square)

- Only one thing can go wrong: 0 in pivot position

Non-Singular Case

$$\begin{aligned} x_1 &+ x_2 &+ x_3 &= b_1 \\ 2x_1 &+ 2x_2 &+ 5x_3 &= b_2 \\ 4x_1 &+ 6x_2 &+ 8x_3 &= b_3 \end{aligned}$$

↓

$$\begin{aligned} x_1 + x_2 + x_3 &= \ldots \\ 3x_3 &= \ldots \\ 2x_2 + 4x_3 &= \ldots \end{aligned}$$

↓ Fix with row exchange

$$\begin{aligned} x_1 + x_2 + x_3 &= \ldots \\ 2x_2 + 4x_3 &= \ldots \\ 3x_3 &= \ldots \end{aligned}$$

Singular Case

$$\begin{aligned} x_1 &+ x_2 &+ x_3 &= \ldots \\ 2x_1 &+ 2x_2 &+ 5x_3 &= \ldots \\ 4x_1 &+ 4x_2 &+ 8x_3 &= \ldots \end{aligned}$$

↓

$$\begin{aligned} x_1 + x_2 + x_3 &= \ldots \\ 3x_3 &= \ldots \\ 4x_3 &= \ldots \end{aligned}$$

Can't make triangular...

Outline

- Basic Operations
- Special Matrices
- Vector Spaces
- Transformations
- Eigenvalues
- Norms
- Linear Systems
- Matrix Factorization

LU factorization

- Each elimination step is equivalent to multiplication by a lower triangular elementary matrix:

E: add -2 times first row to second

$$EA = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix}$$

- So Gaussian elimination takes Ax=b and pre-multiplies by elementary matrices {E,F,G} until GFEA is triangular

F: add 1 times first row to third
G: add 1 times second row to third

$$GFEA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

LU factorization

- We'll use U to denote the upper triangular GFEA
- Note: $E^{-1}F^{-1}G^{-1}U = A$, we'll use L for $E^{-1}F^{-1}G^{-1}$, so A = LU
- L is lower triangular:
  - inv. of elementary is elementary w/ same vectors:

$$EE^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

  - product of lower triangular is lower triangular:

$$E^{-1}F^{-1}G^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix}$$

LU for Non-Singular

- So we have A=LU, and linear system is LUx = b
- After compute L and U, we can solve a non-singular system:
  - x = U\(L\b) (where '\' means back-substitution)
- Cost: ~(1/3)n^3 for factorization, ~n^2 for substitution
- Solve for different b': x = U\(L\b') (no re-factorization)
- If the pivot is 0 we perform a row exchange with a permutation matrix:

$$PA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 0 & 2 \end{bmatrix}$$

Notes on LU

- Diagonals of L are 1 but diagonals of U are not:
  - LDU factorization: divide pivots out of U to get diagonal matrix D (A=LDU is unique)
- If A is symmetric and positive-definite: $L=U^T$
  - Cholesky factorization ($A = LL^T$) is faster: ~$(1/6)n^3$
  - Often the fastest check that symmetric A is positive-definite
- LU is faster for band-diagonal matrices: ~w^2n (diagonal: w= 1, tri-digonal: w = 2)
- LU is not optimal, current best: $O(n^{2.376})$

QR Factorization

- LU factorization uses lower triangular elementary matrices to make A triangular
- The QR factorization uses orthogonal elementary matrices to make A triangular
- Householder transformation:

$$H = I - \frac{1}{\beta} w w^T, \beta = \frac{1}{2} ||w||_2^2$$

- Because orthogonal transformations preserve length, QR can give more numerically stable solutions

Spectral Decomposition

- Any symmetric matrix can be written as:

$$A = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

- Where U contains the orthonormal eigenvectors and Lambda is diagonal with the eigenvalues as elements

- This can be used to 'diagonalize' the matrix:

$$Q^T A Q = \Lambda$$

- It is also useful for computing powers:

$$A^3 = Q\Lambda Q^T Q\Lambda Q^T Q\Lambda Q^T = Q\Lambda\Lambda\Lambda Q^T = Q\Lambda^3 Q^T$$

Spectral Decomposition and SVD

- Any symmetric matrix can be written as:

$$A = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

- Where U contains the orthonormal eigenvectors and Lambda is diagonal with the eigenvalues as elements

- Any matrix can be written as:

$$A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

- Where U and V have orthonormal columns and Sigma is diagonal with the ‘singular’ values as elements (square roots of eigenvalues of $A^T A$)

Singular/Rectangular System

- The general solution to Ax=b is given by transforming A to echelon form:

Basic Variables (pivot)

$$U = \begin{bmatrix} \otimes & \times & \times & \times & \times & \times & \times & \times \\ 0 & \otimes & \times & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & \otimes & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \otimes \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Free Variables (no pivot)

- 1. Solve with free variables 0: x_{part} (one solution to Ax=b)
    - If this fails, b is not in the column-space
- 2. Solve with free variables e_i: $x_{hom(i)}$ (basis for nullspace)
- 3. Full set of solutions: $x = x_{part} + \Sigma\beta_i x_{hom(i)}$
(any solution) = (one solution) + (anything in null-space)

Pseudo-Inverse

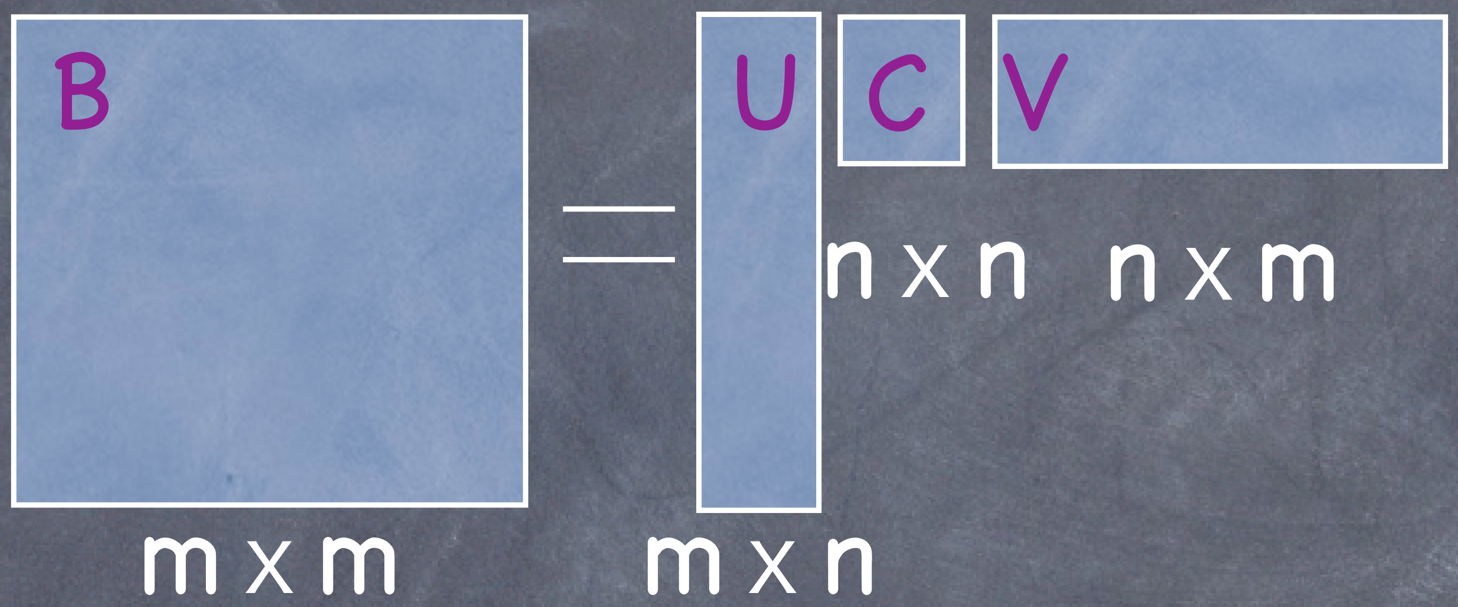
- When A is non-singular, Ax=b has the unique solution $x=A^{-1}b$
- When A is non-square or singular, the system may be incompatible, or the solution might not be unique
- The pseudo-inverse matrix A^+, is the unique matrix such that $x=A^+b$ is the vector with minimum $||x||_2$ that minimizes $||Ax-b||_2$
- It can be computed from the SVD:

$$A^+ = V\Omega U^T, \Omega = diag(\omega), \omega_i = \begin{cases} 1/\sigma_i & \text{if } \sigma_i \neq 0 \\ 0 & \text{if } \sigma_i = 0 \end{cases}$$

- If A is non-singular, $A^+ = A^{-1}$

Inversion Lemma

- Rank-1 Matrix: uv^T (all rows/cols are linearly dependent)
- Low-rank representation of m x m matrix: B = UCV



- Sherman-Morrison-Woodbury Matrix inversion Lemma:
  - $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$
  - If you know A^{-1}, invert (n x n) instead of (m x m) (ie. useful if A is diagonal or orthogonal)

Some topics not covered

- Perturbation theory, condition number, least squares
- Differentiation, quadratic functions, Wronskians
- Computing eigenvalues, Krylov subspace methods
- Determinants, general vector spaces, inner-product spaces
- Special matrices (Toeplitz, Vandermonde, DFT)
- Complex matrices (conjugate transpose, Hermitian/unitary)