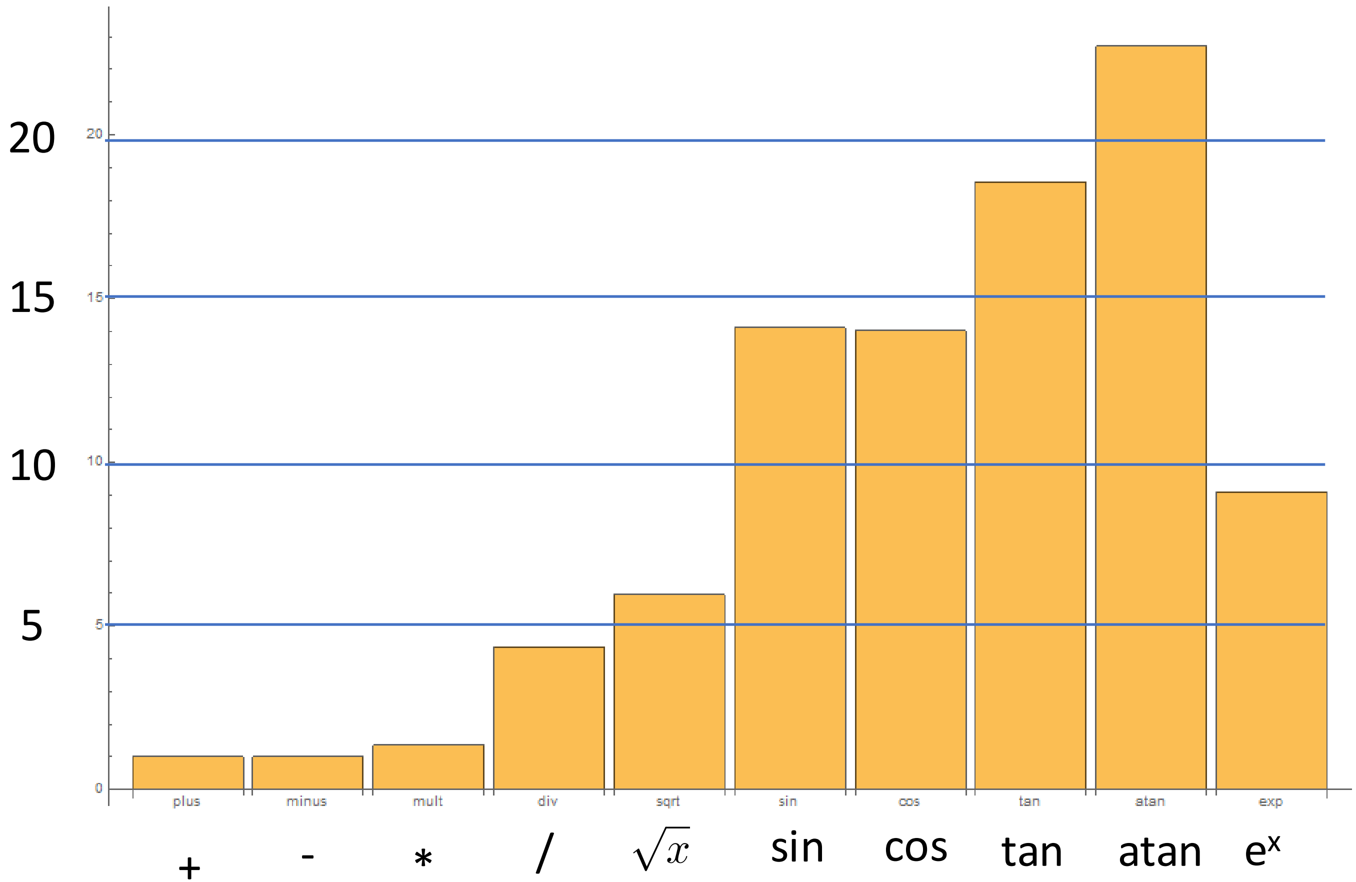
Cost of operations

https://latkin.org/blog/2014/11/09/a-simple-benchmark-of-various-math-operations/



Linear Algebra Review, Norms, Condition Number

- **Norms**
  - notation
  - 3 basic properties
  - p-norms
  - unit balls
  - matrix norm, geometric interp.
- **Condition number (C)**
  - “physical” meaning
  - In terms of relative errors
  - magnitude of C effect on roundoff
- **Matrix multiplication**
  - 2 interpretations
  - 2 identities (transpose, inverse)
- **Definitions**
  - range, rank, basis, inner product, linear dep/indep, singular matrix
- Ax=b in terms of basis
  - interpretation/bais of x, b
- Eigenvalue decomposition
  - Geometry of eigenvectors
  - Why is EV decomp useful
  - How to do it.

Norms

Notation

$||x||$ for the norm of x.

Think of the norm as the vector length

Properties

$||x|| \geq 0$ and $||x|| = 0$ if and only if $x = 0$.

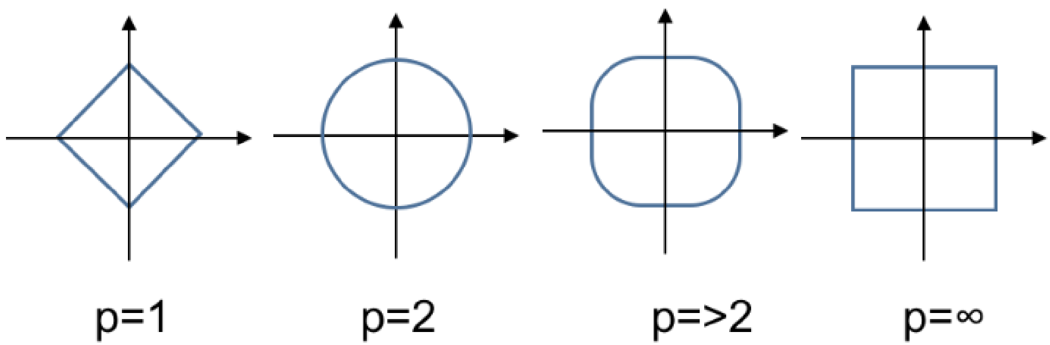
$||x + y|| \leq ||x|| + ||y||$ triangle inequality

$||\alpha x|| = |\alpha|||x||$, where α is a scalar.

p-norms, unit balls

- $\|x\|_1 = \sum |x_i|$
- $\|x\|_2 = (\sum |x_i|^2)^{1/2}$
- $\|x\|_p = (\sum |x_i|^p)^{1/p}$
- $\|x\|_\infty = \max |x_i|.$

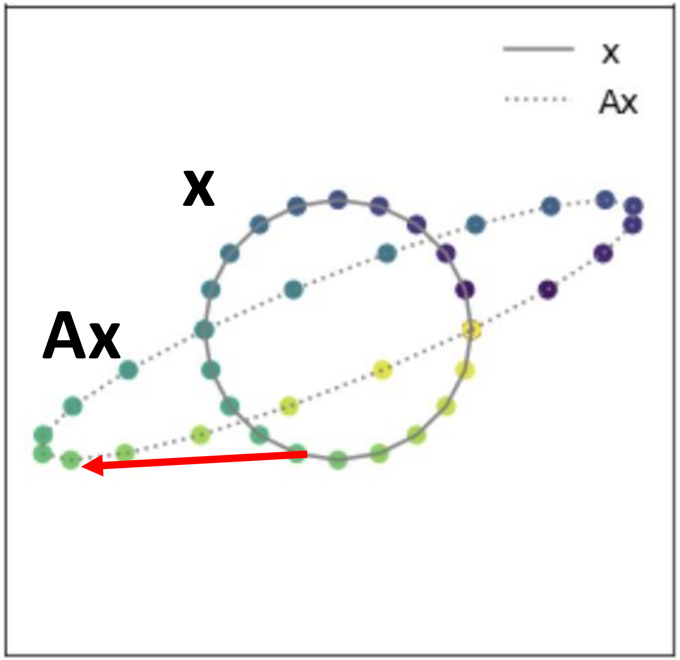
Unit *balls: norm is 1 on the line*



Induced Matrix Norm

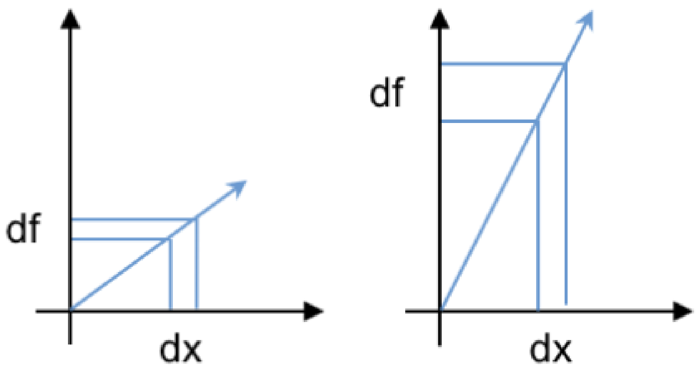
$$||A|| = \max(||Ax||/||x||) \text{ for all } x.$$

Think of this as the maximum factor that a matrix can stretch a vector



Condition Number

- Consider $f(x)$ vs. x. A given δx added to x will result in some δf added to f. That is, $x + \delta x \to f + \delta f$.
- The condition number relates sensitivity of errors in x to errors in f.



Example

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2.0001 \end{bmatrix}$$

- The solution to this is $(1, 1)$.
- But if we change the 2.0001 to 2.0000, the solution is $(2, 0)$.
- A small change in b gives a large change in x.
- The condition number of A is about 40,000.

```python
import numpy as np

A = np.array([[1,1],[1,1.0001]])
b = np.array([2,2.0000])
x = np.linalg.solve(A,b)
print('x = ', x)
print('Condition number of A = ', np.linalg.cond(A))
```

```
x =  [2. 0.]
Condition number of A =  40002.00007491187
```

Condition Number

Note, a property of norms gives $||Ax|| \leq ||A|| ||x||$

- $Ax = b \rightarrow ||b|| \leq ||A|| ||x||$.
- $A(x + \delta x) = b + \delta b \rightarrow A\delta x = \delta b$
- So, $\delta x = A^{-1}\delta b$
- $\rightarrow ||\delta x|| \leq ||A^{-1}|| ||\delta b||$.
- Combining expressions for $||b||$ and $||\delta x||$, we get:

$$||b|| ||\delta x|| \leq ||A|| ||x|| ||A^{-1}|| ||\delta b|| = ||A|| ||A^{-1}|| ||x|| ||\delta b||$$

$$\frac{||\delta x||}{||x||} \leq ||A|| ||A^{-1}|| \frac{||\delta b||}{||b||}$$

- Here, $||A|| ||A^{-1}||$ is the condition number $C(A)$.
- This equation means (the relative change in x) $\leq C(A)\cdot$ (the relative change in b).

In terms of relative error (RE):

$$RE_x = C(A) \cdot RE_b.$$

- If $C(A) = 1/\epsilon_{mach}$, and $RE_b = \epsilon_{mach}$, then $RE_x = 1$. That is RE_x = 100% error.
- **When doing matrix inversion, you can expect to lose 1 digit of accuracy for each order of magnitude of $C(A)$.**
  - So, if $C(A) = 1000$, you will lose 3 digits of accuracy (out of 16).

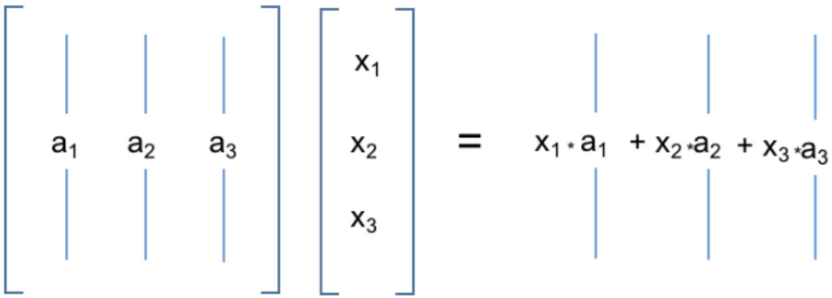
Linear systems

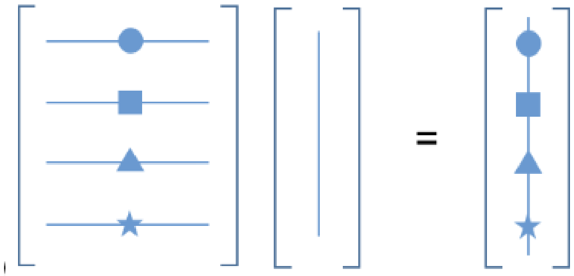
Linear systems

- Write as $Ax = b$, where A is a matrix and x, b are vectors.
  - A is an $m \times n$ matrix: m rows and n columns. Usually, $m = n$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \dots \\ b_m \end{bmatrix}$$

Matrix Multiplication: 2 views





x acts on A to produce b

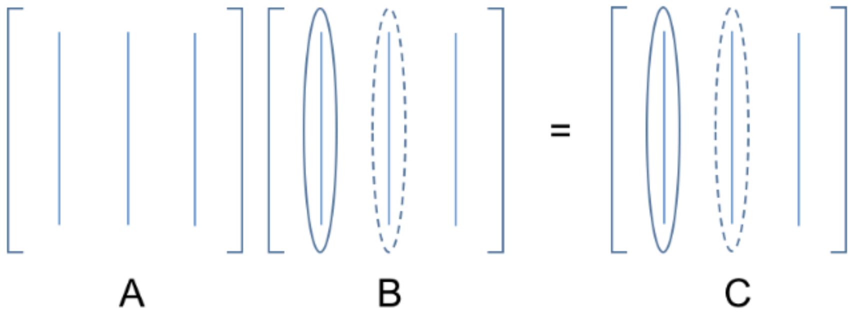
- b is a linear combination of columns of A
- Coordinates of x define scalar multipliers in the combination of the columns of A
- Columns of A are vectors: we combine them through x to get a new vector b
- Elements of x are like coordinates in the columns of A, instead of our usual coordinates i, j, k

A acts on x to produce b

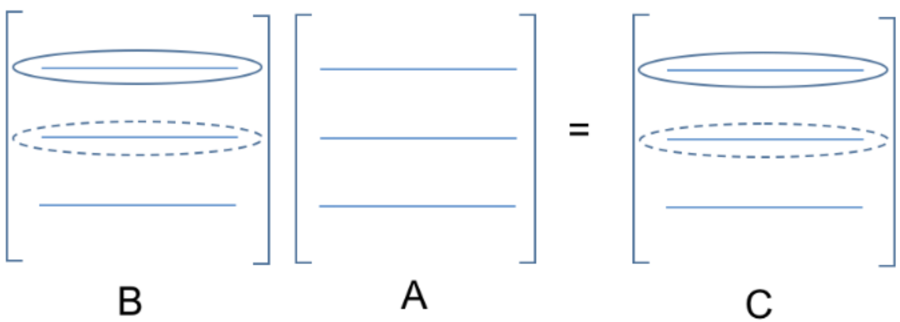
- elements of b are dot product of corresponding row of A and x

$$b_i = \sum_j a_{i,j} x_j$$

Matrix-Matrix Multiplication



Here, the columns of C are linear combinations of the columns of A, where elements of B are the multipliers in those linear combinations.



Here, the rows of C are linear combinations of the rows of A, where elements of B are the multipliers in those linear combinations.

Identities

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

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

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

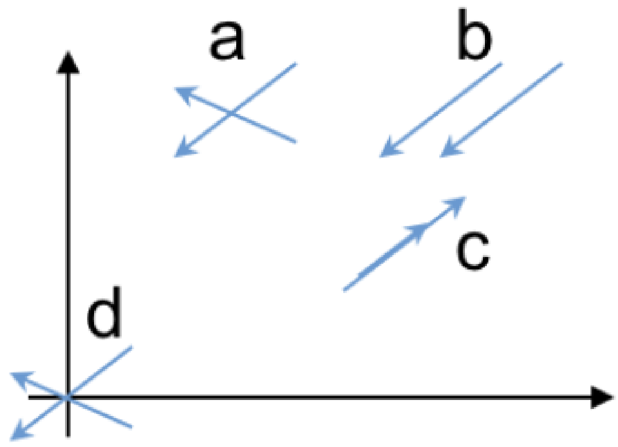
$$(AB)^T \qquad\qquad B^T A^T$$

$$\left(\begin{pmatrix} - & a_1 & - \\ - & a_2 & - \end{pmatrix} \begin{pmatrix} | & | \\ b_1 & b_2 \\ | & | \end{pmatrix} \right)^T = \begin{pmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{pmatrix}^T = \begin{pmatrix} b_1 a_1 & b_1 a_2 \\ b_2 a_1 & b_2 a_2 \end{pmatrix} = \begin{pmatrix} - & b_1 & - \\ - & b_2 & - \end{pmatrix} \begin{pmatrix} | & | \\ a_1 & a_2 \\ | & | \end{pmatrix}$$

Definitions

- **Range:** the set of vectors (b) that can be written as Ax=b for some x. This is the space spanned by the columns of A since b is a linear combination of the columns of A.
- **Rank:** the number of linearly independent rows or columns.
    - for m x n with full rank with $m \leq n$ → rank is m.
- **Basis:** a basis of vectors spans the space and is linearly independent.
- **Inner product:** $x^T x$→s where s is a scalar.
- **Outer product:** $x x^T$→M where M is a matrix. (What's the rank?)
- **Linearly dependent** set of vectors: at least one is a linear combo of the others
- **Nonsingular** matrices are invertable and have solutions to Ax=b.
- **Singular matrices** are not invertible.
    - Have less than full rank
    - Have linearly dependent columns (or rows)

Singular Matrices

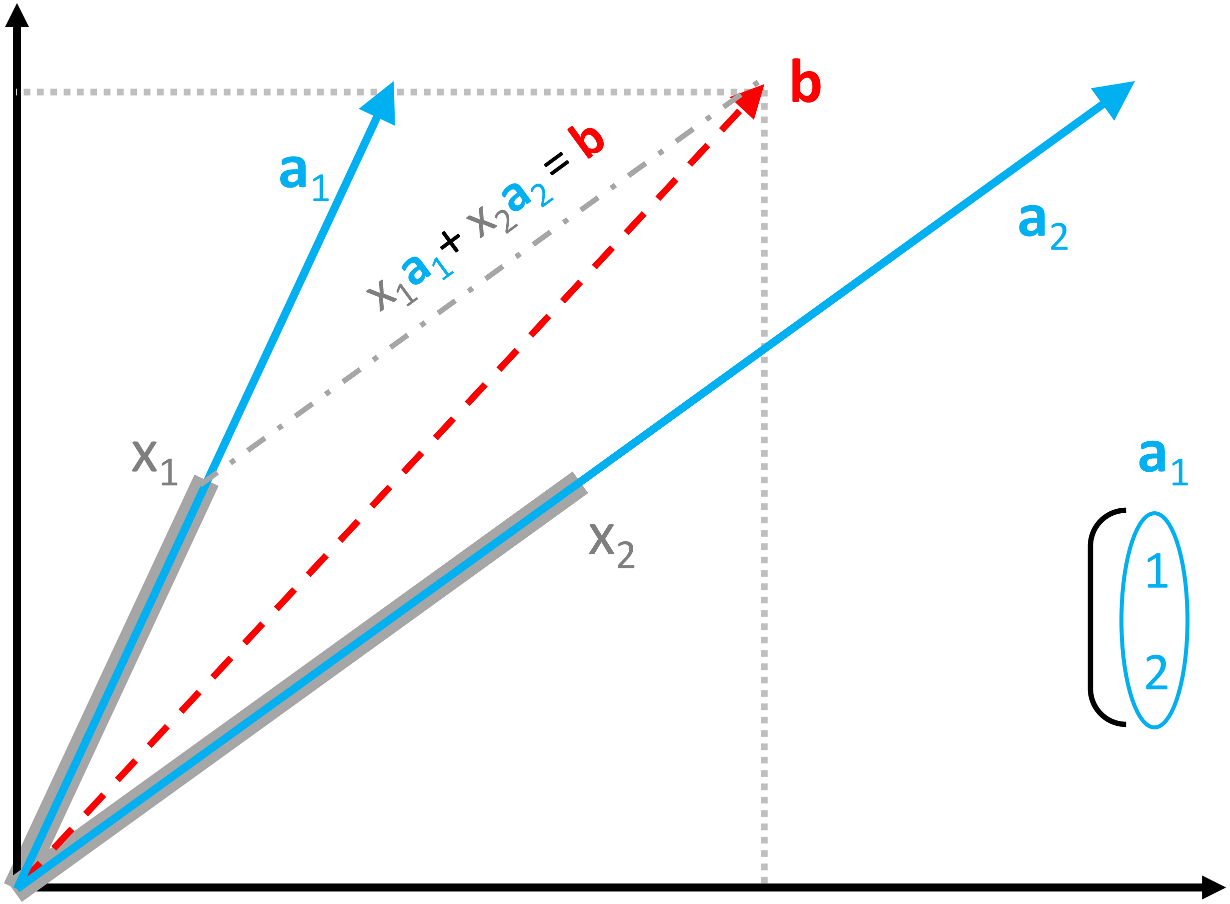


- Case a is a unique solution.
- Case b has no solution.
- Case c has ∞ solutions
- Case d is a trivial solution

Numerically singular matrices are almost singular. That is, they may be singular to within roundoff error, or the near singularity may result in inaccurate results.

(What does this say about their condition number?)

Ax=b Geometry

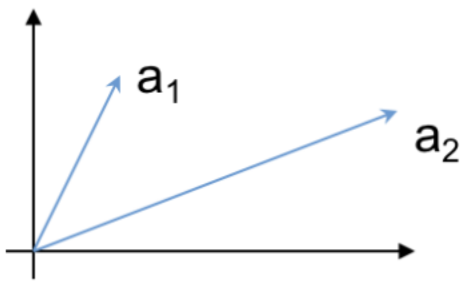


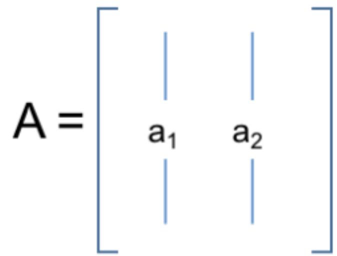
$$\begin{matrix} \mathbf{a}_1 & \mathbf{a}_2 \end{matrix} \quad \mathbf{x} \quad \mathbf{b}$$

$$\begin{pmatrix} 1 & 3 \\ 2 & 2 \end{pmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Basis, Coordinate Systems

$$Ax = b \qquad x = A^{-1}b$$





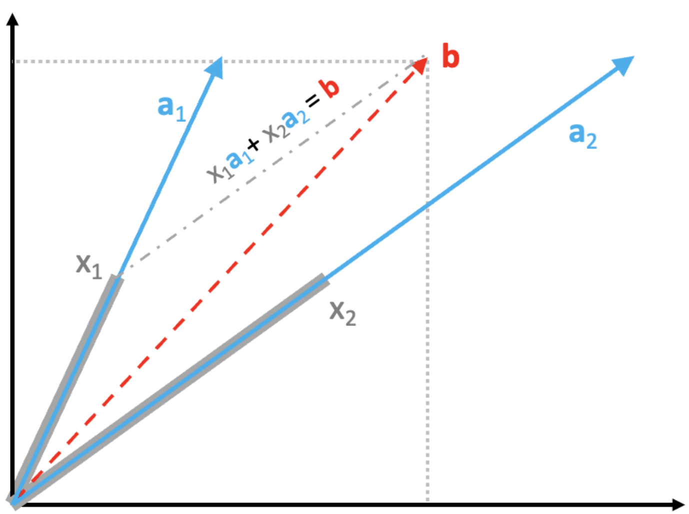
Think of A in terms of its column vectors

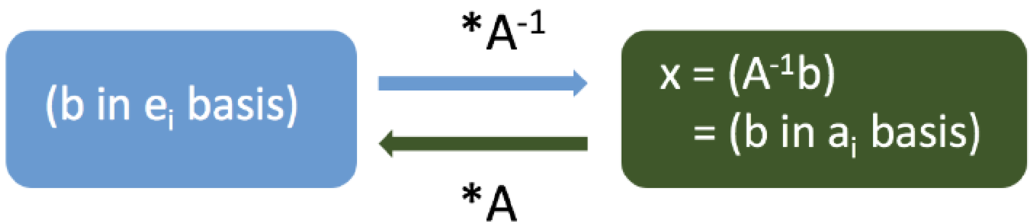
- For A·x, x are the coordinates in the basis of columns of A: a_1 and a_2.
- For given b, x is the vector of coefficients of the unique linear expansion of b in the basis of columns of A.
  - **x is b in the basis of columns of A**
  - x are the coefficients of b in the columns of A
  - coefficients of b are coordinates in columns of I

Basis, Coordinate Systems

$A^{-1}b \rightarrow$ **change basis** of b:

- Normally, when we write $x^T = (x_1, x_2)$, we mean $x = x_1\vec{i} + x_2\vec{j}$, or $x = x_1 \cdot (1,0) + x_2 \cdot (0,1)$.
- Let e_i be the coordinate vectors, so $e_1 = \vec{i}$ and $e_2 = \vec{j}$, etc.





Eigenvector Decomposition

Eigen do it if I try!

$$Av = \lambda v$$

- v is an eigenvector of A and λ is the corresponding eigenvalue.
- Note, normally for $Ax = b$, A operates on x to get a new vector b. You can think of A stretching and rotating x to get b.
- For eigenvectors, A does not rotate v, it only stretches it. And the stretching factor is λ.
  - $|\lambda| > 1$: stretch; $0 < |\lambda| < 1$: compress; $\lambda < 0$: reverse direction.

Eigenvector Decomposition

$$AV = V\Lambda.$$

- This is the matrix form. V is a matrix whose columns are the eigenvectors v of A. Λ is a diagonal matrix with the λ's of A on the diagonal.
  - Now, solve this for A:

$$A = V\Lambda V^{-1}.$$

  - Insert this into $Ax = b$:

$$Ax = b \rightarrow V\Lambda V^{-1}x = b.$$

  - Now, multiply both sides of the second equation by V^{-1}:

$$\Lambda V^{-1}x = V^{-1}b.$$

  - And group terms:

$$\Lambda(V^{-1}x) = (V^{-1}b).$$

  - Define $\hat{x} = V^{-1}x$ and $\hat{b} = V^{-1}b$.

$$\rightarrow \Lambda\hat{x} = \hat{b}$$

Eigenvector Decomposition

So, $Ax = b$ can be written as $\Lambda\hat{x} = \hat{b}$.

- Λ is diagonal, so its easy to invert.
- The second form is decoupled, meaning, each row of this second equation only involves one component of x, whereas each row in $Ax = b$ might contain every x component.
- $\hat{x}$ is x written in the basis of eigenvectors of A. Likewise for $\hat{b}$.
- Note, $x = V\hat{x}$, and $b = V\hat{b}$.
- This decoupling will be very useful later on.