Norms and Inner Products

Szymon Rusinkiewicz
COS 302, Fall 2020

PRINCETON UNIVERSITY

Concrete vs. Abstract

More Concrete	More Abstract
Matrices	Linear mappings
Rank of a matrix	Dimension of the image of a map
Full-rank matrices	Injective (one-to-one) linear maps
Gaussian elimination	When do systems have solutions

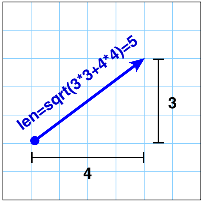
Concrete vs. Abstract

More Concrete	**More Abstract**
Matrices	Linear mappings
Rank of a matrix	Dimension of the image of a map
Full-rank matrices	Injective (one-to-one) linear maps
Gaussian elimination	When do systems have solutions
Euclidean distance in $\mathbb{R}^n$	General norms
Dot products	General inner products

In Euclidean Space…

Length of a vector $\boldsymbol{x}$: $\|\boldsymbol{x}\| = \sqrt{\sum_i x_i^2}$

Distance between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$: $\|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\sum_i (x_i - y_i)^2}$



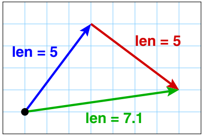
These let us talk about lengths and distances in $\mathbb{R}^n$.

Dot product between vectors $\boldsymbol{x}$ and $\boldsymbol{y}$: $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_i x_i y_i$

This lets us talk about angles and perpendicularity (orthogonality).

Properties of Euclidean Length

- Real-valued function on vectors
- "Positive definite":
  - Non-negative: $\|\boldsymbol{x}\| \geq 0$
  - Positive except for $\mathbf{0}$ vector
- Absolutely homogeneous: $\|\lambda \boldsymbol{x}\| = |\lambda| \, \|\boldsymbol{x}\|$
- Obeys triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
- Induces a distance metric between vectors: $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$



All of these can apply to more general notions of "length."

Properties of Dot Product

- Real-valued function on *pairs* of vectors
- Symmetric: $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$
- Bilinear: $(\lambda \boldsymbol{x} + \psi \boldsymbol{y}) \cdot \boldsymbol{z} = \lambda(\boldsymbol{x} \cdot \boldsymbol{z}) + \psi(\boldsymbol{y} \cdot \boldsymbol{z})$

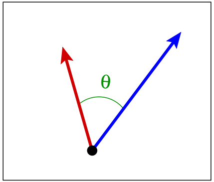
  $\boldsymbol{x} \cdot (\lambda \boldsymbol{y} + \psi \boldsymbol{z}) = \lambda(\boldsymbol{x} \cdot \boldsymbol{y}) + \psi(\boldsymbol{x} \cdot \boldsymbol{z})$
- Positive definite: $\boldsymbol{x} \cdot \boldsymbol{x} > 0$ unless $\boldsymbol{x} = \boldsymbol{0}$
- Induces a norm (in this case, standard Euclidean norm): $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$

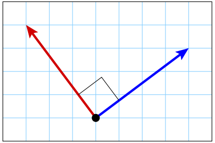
All of these can apply to more general notions of "product."

More Properties of Dot Product

Before we generalize, two more properties:

- Relation to matrix product: $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{y} = \boldsymbol{y}^\top \boldsymbol{x}$, and so $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$
- Relation to angles: $\boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos \theta$
  - Important special case: for nonzero $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{x} \cdot \boldsymbol{y} = 0$ iff $\boldsymbol{x}$ and $\boldsymbol{y}$ are perpendicular





Generalizing Dot Product: Inner Product

Suppose you apply a linear mapping to both vectors, then take a dot product in the new space.

- Will this be the same as the original dot product?

Generalizing Dot Product: Inner Product

Suppose you apply a linear mapping to both vectors, then take a dot product in the new space.

- Will this be the same as the original dot product? No.
- Will this always produce a "valid" dot product?

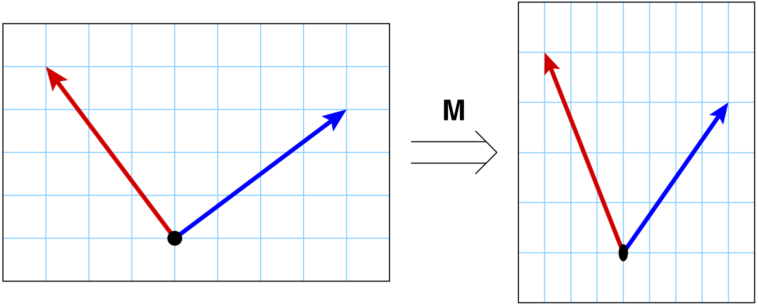
Generalizing Dot Product: Inner Product

Suppose you apply a linear mapping to both vectors,
then take a dot product in the new space.

- Will this be the same as the original dot product? No.
- Will this always produce a "valid" dot product? No.
  - For positive definiteness, need the linear mapping *not* to collapse dimensions

Generalizing Dot Product: Inner Product

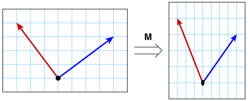
Suppose you apply an injective linear mapping to both vectors, then take a dot product in the new space.



Generalizing Dot Product: Inner Product

Suppose you apply a injective linear mapping to both vectors, then take a dot product in the new space.

Let the transformation be representable by matrix $\boldsymbol{M}$.



New inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = (\boldsymbol{M}\boldsymbol{x})^\top(\boldsymbol{M}\boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{M}^\top \boldsymbol{M}\,\boldsymbol{y} = \boldsymbol{x}^\top \boldsymbol{A}\,\boldsymbol{y}$, where the matrix $\boldsymbol{A}$ is square, symmetric, and *positive definite.*

Aside: Quadratic Forms

You'll often see notation such as $\boldsymbol{x}^\top \boldsymbol{A}\,\boldsymbol{x}$, with square and symmetric $\boldsymbol{A}$. This is a *quadratic form*: a second-order polynomial in the elements of $\boldsymbol{x}$:

Let $\boldsymbol{x} = (x_1, x_2, x_3, \ldots)$. Then,

$$\boldsymbol{x}^\top \boldsymbol{A}\,\boldsymbol{x} = ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 + \ldots$$

Also: a *bilinear form* $\boldsymbol{x}^\top \boldsymbol{A}\,\boldsymbol{y}$ is a function from two vectors to a scalar that is linear in both $\boldsymbol{x}$ and $\boldsymbol{y}$.

Generalizing Dot Product: Inner Product

The generalized inner product, applied to the same vector twice, gives

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle_A = \boldsymbol{x}^\top \boldsymbol{A} \, \boldsymbol{x}$$

for some square, symmetric $\boldsymbol{A}$.

- If $\boldsymbol{A}$ is diagonal, then we just have scaled versions of x_1^2, x_2^2, etc.
  - Application: the "weight" or "importance" of each dimension is different.
- If $\boldsymbol{A}$ is not diagonal, also have "mixed" quadratic terms: $x_1 x_2$, $x_2 x_3$, etc.
  - Application: accounting for "correlation" between dimensions.

    Example: $\boldsymbol{A} = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$, so $\boldsymbol{x}^\top \boldsymbol{A} \, \boldsymbol{x} = x_1^2 + x_2^2 - x_1 x_2$.

    The norm induced by this inner product *downweights* correlation in x_1 and x_2.

Generalizing Dot Product: Inner Product

The generalized inner product, applied to the same vector twice, gives

$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle_A = \boldsymbol{x}^\top \boldsymbol{A}\, \boldsymbol{x}$$

for some square, symmetric $\boldsymbol{A}$.

- But we know that this has to be > 0 (unless $\boldsymbol{x}$ is the $\boldsymbol{0}$ vector), because it came from a dot product in some (transformed) space.
- So we say that $\boldsymbol{A}$ is "symmetric positive definite" or SPD.
- Key duality between SPD matrices, generalized inner products, and norms on linearly-transformed vector spaces.
  - $\boldsymbol{A}$ is SPD iff it can be written as $\boldsymbol{A} = \boldsymbol{M}^\top \boldsymbol{M}$

Generalizing Norm: L^p Spaces

Not all norms come from inner products. We can have vector spaces with *valid norms* but *no well-defined inner products.*

Most important example: ℓ_p norm

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

for $p \geq 1$. (For $p < 1$, does not satisfy triangle inequality, so not a valid norm.)

Generalizing Norm: L^p Spaces

$$\|\boldsymbol{x}\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

- $p = 2$: Good ol' Euclidean norm.
- $p = 1$: Manhattan or taxicab norm.
  $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$ = sum of North-South and East-West distances, when restricted to city-block grid. Often used in *robust estimation*.
- $p = \infty$: Infinity norm or *max* norm.
  $\|\boldsymbol{x}\|_\infty = \max_i |x_i|$