Vector spaces

DS-GA 1013 / MATH-GA 2824 Optimization-based Data Analysis

http://www.cims.nyu.edu/~cfgranda/pages/OBDA_fall17/index.html

Carlos Fernandez-Granda

Vector space

Consists of:

- A set $\mathcal{V}$
- A scalar field (usually $\mathbb{R}$ or $\mathbb{C}$)
- Two operations $+$ and $\cdot$

Properties

- For any $\vec{x}, \vec{y} \in \mathcal{V}$, $\vec{x} + \vec{y}$ belongs to $\mathcal{V}$
- For any $\vec{x} \in \mathcal{V}$ and any scalar α, $\alpha \cdot \vec{x} \in \mathcal{V}$
- There exists a zero vector $\vec{0}$ such that $\vec{x} + \vec{0} = \vec{x}$ for any $\vec{x} \in \mathcal{V}$
- For any $\vec{x} \in \mathcal{V}$ there exists an additive inverse $\vec{y}$ such that $\vec{x} + \vec{y} = \vec{0}$, usually denoted by $-\vec{x}$

Properties

- The vector sum is commutative and associative, i.e. for all $\vec{x}, \vec{y}, \vec{z} \in \mathcal{V}$

$$\vec{x} + \vec{y} = \vec{y} + \vec{x}, \quad (\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$$

- Scalar multiplication is associative, for any scalars α and β and any $\vec{x} \in \mathcal{V}$

$$\alpha\,(\beta \cdot \vec{x}) = (\alpha\,\beta) \cdot \vec{x}$$

- Scalar and vector sums are both distributive, i.e. for any scalars α and β and any $\vec{x}, \vec{y} \in \mathcal{V}$

$$(\alpha + \beta) \cdot \vec{x} = \alpha \cdot \vec{x} + \beta \cdot \vec{x}, \quad \alpha \cdot (\vec{x} + \vec{y}) = \alpha \cdot \vec{x} + \alpha \cdot \vec{y}$$

Subspaces

A subspace of a vector space $\mathcal{V}$ is any subset of $\mathcal{V}$ that is *also itself a vector space*

Linear dependence/independence

A set of m vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m$ is linearly dependent if there exist m scalar coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$ which are not all equal to zero and

$$\sum_{i=1}^{m} \alpha_i \, \vec{x}_i = \vec{0}$$

Equivalently, any vector in a linearly dependent set can be expressed as a linear combination of the rest

Span

The span of $\{\vec{x}_1, \ldots, \vec{x}_m\}$ is the set of all possible linear combinations

$$\text{span}\left(\vec{x}_1, \ldots, \vec{x}_m\right) := \left\{ \vec{y} \mid \vec{y} = \sum_{i=1}^{m} \alpha_i \, \vec{x}_i \quad \text{for some scalars } \alpha_1, \alpha_2, \ldots, \alpha_m \right\}$$

The span of any set of vectors in $\mathcal{V}$ is a subspace of $\mathcal{V}$

Basis and dimension

A basis of a vector space $\mathcal{V}$ is a set of independent vectors $\{\vec{x}_1, \ldots, \vec{x}_m\}$ such that

$$\mathcal{V} = \operatorname{span}(\vec{x}_1, \ldots, \vec{x}_m)$$

If $\mathcal{V}$ has a basis with finite cardinality then every basis contains the same number of vectors

The dimension $\dim(\mathcal{V})$ of $\mathcal{V}$ is the cardinality of any of its bases

Equivalently, the dimension is the number of linearly independent vectors that span $\mathcal{V}$

Standard basis

$$\vec{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \vec{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

The dimension of $\mathbb{R}^n$ is n

Inner product

Operation $\langle \cdot, \cdot \rangle$ that maps a pair of vectors to a scalar

Properties

- If the scalar field is $\mathbb{R}$, it is symmetric. For any $\vec{x}, \vec{y} \in \mathcal{V}$

  $$\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$$

  If the scalar field is $\mathbb{C}$, then for any $\vec{x}, \vec{y} \in \mathcal{V}$

  $$\langle \vec{x}, \vec{y} \rangle = \overline{\langle \vec{y}, \vec{x} \rangle},$$

  where for any $\alpha \in \mathbb{C}$ $\overline{\alpha}$ is the complex conjugate of α

Properties

- It is linear in the first argument, i.e. for any $\alpha \in \mathbb{R}$ and any $\vec{x}, \vec{y}, \vec{z} \in \mathcal{V}$

$$\langle \alpha \, \vec{x}, \vec{y} \rangle = \alpha \, \langle \vec{x}, \vec{y} \rangle \, ,$$
$$\langle \vec{x} + \vec{y}, \vec{z} \rangle = \langle \vec{x}, \vec{z} \rangle + \langle \vec{y}, \vec{z} \rangle \, .$$

  If the scalar field is $\mathbb{R}$, it is also linear in the second argument

- It is positive definite: $\langle \vec{x}, \vec{x} \rangle$ is nonnegative for all $\vec{x} \in \mathcal{V}$ and if $\langle \vec{x}, \vec{x} \rangle = 0$ then $\vec{x} = \vec{0}$

Dot product

Inner product between $\vec{x}, \vec{y} \in \mathbb{R}^n$

$$\vec{x} \cdot \vec{y} := \sum_i \vec{x}[i]\ \vec{y}[i]$$

$\mathbb{R}^n$ endowed with the dot product is usually called a Euclidean space of dimension n

If $\vec{x}, \vec{y} \in \mathbb{C}^n$

$$\vec{x} \cdot \vec{y} := \sum_i \vec{x}[i]\ \overline{\vec{y}[i]}$$

Sample covariance

Quantifies joint fluctuations of two quantities or features

For a data set (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)

$$\text{cov}\left((x_1, y_1), \ldots, (x_n, y_n)\right) := \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \text{av}\left(x_1, \ldots, x_n\right)\right)\left(y_i - \text{av}\left(y_1, \ldots, y_n\right)\right)$$

where the average or sample mean is defined by

$$\text{av}\left(a_1, \ldots, a_n\right) := \frac{1}{n} \sum_{i=1}^{n} a_i$$

If $(\mathbf{x_1}, \mathbf{y_1})$, $(\mathbf{x_2}, \mathbf{y_2})$, ..., $(\mathbf{x_n}, \mathbf{y_n})$ are iid samples from $\mathbf{x}$ and $\mathbf{y}$

$$\mathrm{E}\left(\text{cov}\left((\mathbf{x_1}, \mathbf{y_1}), \ldots, (\mathbf{x_n}, \mathbf{y_n})\right)\right) = \mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right) := \mathrm{E}\left(\left(\mathbf{x} - \mathrm{E}\left(\mathbf{x}\right)\right)\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)\right)$$

Matrix inner product

The inner product between two $m \times n$ matrices A and B is

$$\begin{aligned} \langle A, B \rangle &:= \operatorname{tr}\left(A^T B\right) \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ij} \end{aligned}$$

where the trace of an $n \times n$ matrix is defined as the sum of its diagonal

$$\operatorname{tr}(M) := \sum_{i=1}^{n} M_{ii}$$

For any pair of $m \times n$ matrices A and B

$$\operatorname{tr}\left(B^T A\right) := \operatorname{tr}\left(AB^T\right)$$

Function inner product

The inner product between two complex-valued square-integrable functions f, g defined in an interval $[a, b]$ of the real line is

$$\vec{f} \cdot \vec{g} := \int_a^b f(x)\,\overline{g(x)}\,\mathrm{d}x$$

Norm

Let $\mathcal{V}$ be a vector space, a norm is a function $||\cdot||$ from $\mathcal{V}$ to $\mathbb{R}$ with the following properties

- It is homogeneous. For any scalar α and any $\vec{x} \in \mathcal{V}$

$$||\alpha \, \vec{x}|| = |\alpha| \, ||\vec{x}|| \,.$$

- It satisfies the triangle inequality

$$||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}|| \,.$$

  In particular, $||\vec{x}|| \geq 0$

- $||\vec{x}|| = 0$ implies $\vec{x} = \vec{0}$

Inner-product norm

Square root of inner product of vector with itself

$$||\vec{x}||_{\langle\cdot,\cdot\rangle} := \sqrt{\langle\vec{x},\vec{x}\rangle}$$

Inner-product norm

- Vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$: ℓ_2 norm

$$||\vec{x}||_2 := \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{\sum_{i=1}^{n} \vec{x}[i]^2}$$

- Matrices in $\mathbb{R}^{m \times n}$ or $\mathbb{C}^{m \times n}$: Frobenius norm

$$||A||_{\mathrm{F}} := \sqrt{\operatorname{tr}\left(A^T A\right)} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$$

- Square-integrable complex-valued functions: $\mathcal{L}_2$ norm

$$||f||_{\mathcal{L}_2} := \sqrt{\langle f, f \rangle} = \sqrt{\int_a^b |f(x)|^2 \, \mathrm{d}x}$$

Cauchy-Schwarz inequality

For any two vectors $\vec{x}$ and $\vec{y}$ in an inner-product space

$$|\langle \vec{x}, \vec{y} \rangle| \leq ||\vec{x}||_{\langle \cdot, \cdot \rangle} \, ||\vec{y}||_{\langle \cdot, \cdot \rangle}$$

Assume $||\vec{x}||_{\langle \cdot, \cdot \rangle} \neq 0$, then

$$\langle \vec{x}, \vec{y} \rangle = - ||\vec{x}||_{\langle \cdot, \cdot \rangle} \, ||\vec{y}||_{\langle \cdot, \cdot \rangle} \iff \vec{y} = - \frac{||\vec{y}||_{\langle \cdot, \cdot \rangle}}{||\vec{x}||_{\langle \cdot, \cdot \rangle}} \vec{x}$$

$$\langle \vec{x}, \vec{y} \rangle = ||\vec{x}||_{\langle \cdot, \cdot \rangle} \, ||\vec{y}||_{\langle \cdot, \cdot \rangle} \iff \vec{y} = \frac{||\vec{y}||_{\langle \cdot, \cdot \rangle}}{||\vec{x}||_{\langle \cdot, \cdot \rangle}} \vec{x}$$

Sample variance and standard deviation

The sample variance quantifies fluctuations around the average

$$\text{var}(x_1, x_2, \ldots, x_n) := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \text{av}(x_1, x_2, \ldots, x_n))^2$$

If $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_n}$ are iid samples from $\mathbf{x}$

$$\mathrm{E}(\text{var}(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})) = \mathrm{Var}(\mathbf{x}) := \mathrm{E}\left((\mathbf{x} - \mathrm{E}(\mathbf{x}))^2\right)$$

The sample standard deviation is

$$\text{std}(x_1, x_2, \ldots, x_n) := \sqrt{\text{var}(x_1, x_2, \ldots, x_n)}$$

Correlation coefficient

Normalized covariance

$$\rho_{(x_1,y_1),\ldots,(x_n,y_n)} := \frac{\operatorname{cov}\left(\left(x_1, y_1\right), \ldots, \left(x_n, y_n\right)\right)}{\operatorname{std}\left(x_1, \ldots, x_n\right) \operatorname{std}\left(y_1, \ldots, y_n\right)}$$

Corollary of Cauchy-Schwarz

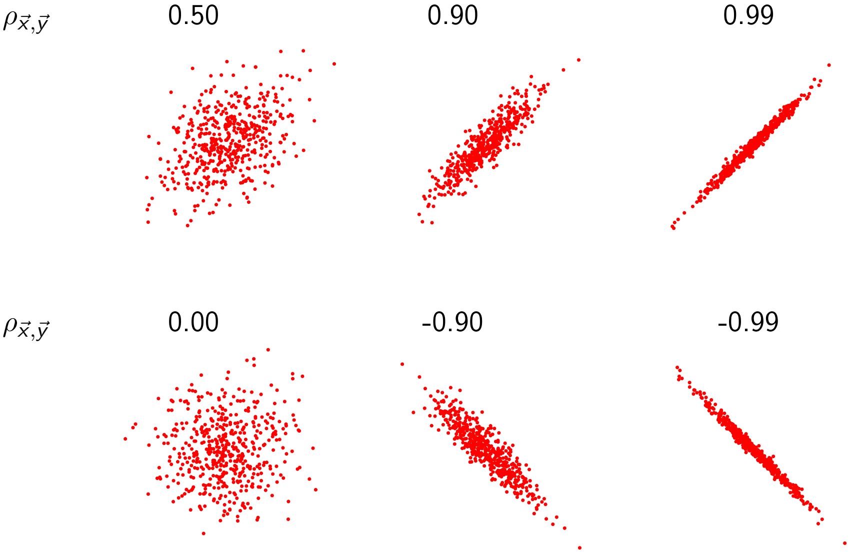
$$-1 \leq \rho_{(x_1,y_1),\ldots,(x_n,y_n)} \leq 1$$

and

$$\rho_{\vec{x},\vec{y}} = -1 \iff y_i = \operatorname{av}\left(y_1, \ldots, y_n\right) - \frac{\operatorname{std}\left(y_1, \ldots, y_n\right)}{\operatorname{std}\left(x_1, \ldots, x_n\right)} \left(x_i - \operatorname{av}\left(x_1, \ldots, x_n\right)\right)$$

$$\rho_{\vec{x},\vec{y}} = 1 \iff y_i = \operatorname{av}\left(y_1, \ldots, y_n\right) + \frac{\operatorname{std}\left(y_1, \ldots, y_n\right)}{\operatorname{std}\left(x_1, \ldots, x_n\right)} \left(x_i - \operatorname{av}\left(x_1, \ldots, x_n\right)\right)$$

Correlation coefficient

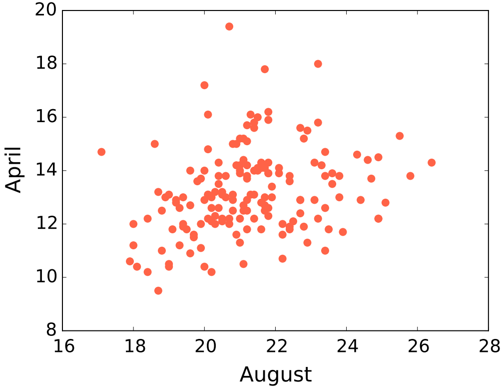


Temperature data

Temperature in Oxford over 150 years

- Feature 1: Temperature in January
- Feature 1: Temperature in August

$$\rho = 0.269$$

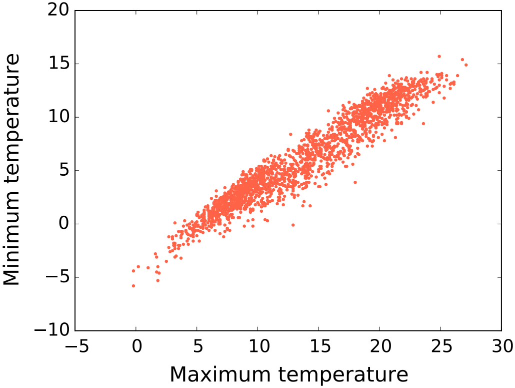


Temperature data

Temperature in Oxford over 150 years (monthly)

- Feature 1: Maximum temperature
- Feature 1: Minimum temperature

$$\rho = 0.962$$



Parallelogram law

A norm $\|\cdot\|$ on a vector space $\mathcal{V}$ is an inner-product norm if and only if

$$2\|\vec{x}\|^2 + 2\|\vec{y}\|^2 = \|\vec{x} - \vec{y}\|^2 + \|\vec{x} + \vec{y}\|^2$$

for any $\vec{x}, \vec{y} \in \mathcal{V}$

ℓ_1 and ℓ_∞ norms

Norms in $\mathbb{R}^n$ or $\mathbb{C}^n$ not induced by an inner product

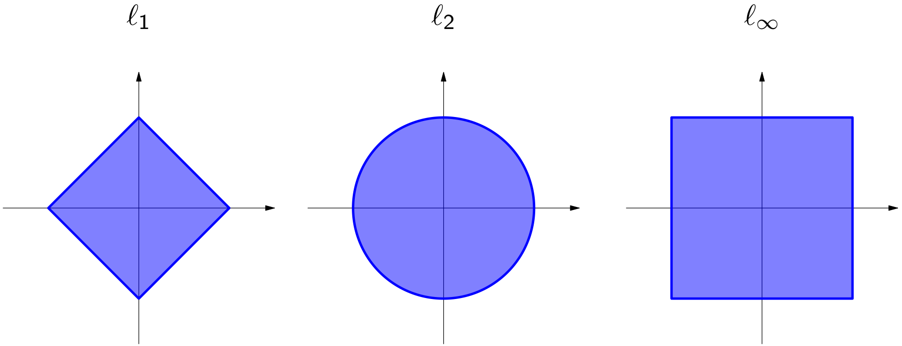
$$||\vec{x}||_1 := \sum_{i=1}^{n} |\vec{x}[i]|$$

$$||\vec{x}||_\infty := \max_i |\vec{x}[i]|$$

Hölder's inequality

$$|\langle \vec{x}, \vec{y} \rangle| \leq ||\vec{x}||_1 \, ||\vec{y}||_\infty$$

Norm balls



Distance

The distance between two vectors $\vec{x}$ and $\vec{y}$ induced by a norm $||\cdot||$ is

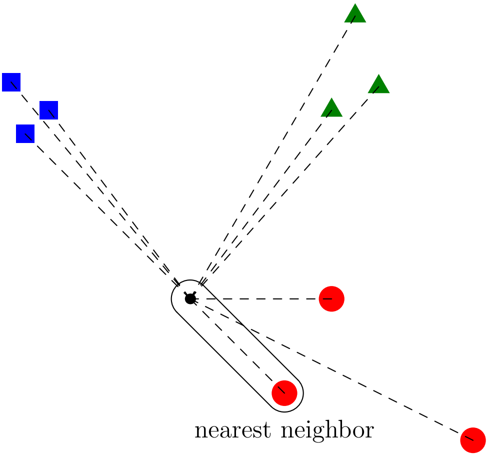
$$d\left(\vec{x}, \vec{y}\right) := ||\vec{x} - \vec{y}||$$

Classification

Aim: Assign a signal to one of k predefined classes

Training data: n pairs of signals (represented as vectors) and labels: $\{\vec{x}_1, l_1\}$, ..., $\{\vec{x}_n, l_n\}$

Nearest-neighbor classification



Face recognition

Training set: 360 64×64 images from 40 different subjects (9 each)

Test set: 1 new image from each subject

We model each image as a vector in $\mathbb{R}^{4096}$ and use the ℓ_2-norm distance

Face recognition

Training set

Nearest-neighbor classification

Errors: 4 / 40

Test image

Closest image

Orthogonality

Two vectors $\vec{x}$ and $\vec{y}$ are orthogonal if and only if

$$\langle \vec{x}, \vec{y} \rangle = 0$$

A vector $\vec{x}$ is orthogonal to a set $\mathcal{S}$, if

$$\langle \vec{x}, \vec{s} \rangle = 0, \quad \text{for all } \vec{s} \in \mathcal{S}$$

Two sets of $\mathcal{S}_1, \mathcal{S}_2$ are orthogonal if for any $\vec{x} \in \mathcal{S}_1, \vec{y} \in \mathcal{S}_2$

$$\langle \vec{x}, \vec{y} \rangle = 0$$

The orthogonal complement of a subspace $\mathcal{S}$ is

$$\mathcal{S}^{\perp} := \{\vec{x} \mid \langle \vec{x}, \vec{y} \rangle = 0 \quad \text{for all } \vec{y} \in \mathcal{S}\}$$

Pythagorean theorem

If $\vec{x}$ and $\vec{y}$ are orthogonal

$$||\vec{x}+\vec{y}||_{\langle\cdot,\cdot\rangle}^2 = ||\vec{x}||_{\langle\cdot,\cdot\rangle}^2 + ||\vec{y}||_{\langle\cdot,\cdot\rangle}^2$$

Orthonormal basis

Basis of mutually orthogonal vectors with inner-product norm equal to one

If $\{\vec{u}_1, \ldots, \vec{u}_n\}$ is an orthonormal basis of a vector space $\mathcal{V}$, for any $\vec{x} \in \mathcal{V}$

$$\vec{x} = \sum_{i=1}^{n} \langle \vec{u}_i, \vec{x} \rangle \, \vec{u}_i$$

Gram-Schmidt

Builds orthonormal basis from a set of linearly independent vectors $\vec{x}_1, \ldots, \vec{x}_m$ in $\mathbb{R}^n$

1. Set $\vec{u}_1 := \vec{x}_1 / \left|\left|\vec{x}_1\right|\right|_2$

2. For $i = 1, \ldots, m$, compute

$$\vec{v}_i := \vec{x}_i - \sum_{j=1}^{i-1} \langle \vec{u}_j, \vec{x}_i \rangle \, \vec{u}_j$$

   and set $\vec{u}_i := \vec{v}_i / \left|\left|\vec{v}_i\right|\right|_2$

Direct sum

For any subspaces $\mathcal{S}_1, \mathcal{S}_2$ such that

$$\mathcal{S}_1 \cap \mathcal{S}_2 = \{0\}$$

the direct sum is defined as

$$\mathcal{S}_1 \oplus \mathcal{S}_2 := \{\vec{x} \mid \vec{x} = \vec{s}_1 + \vec{s}_2 \quad \vec{s}_1 \in \mathcal{S}_1, \vec{s}_2 \in \mathcal{S}_2\}$$

Any vector $\vec{x} \in \mathcal{S}_1 \oplus \mathcal{S}_2$ has a unique representation

$$\vec{x} = \vec{s}_1 + \vec{s}_2 \quad \vec{s}_1 \in \mathcal{S}_1, \vec{s}_2 \in \mathcal{S}_2$$

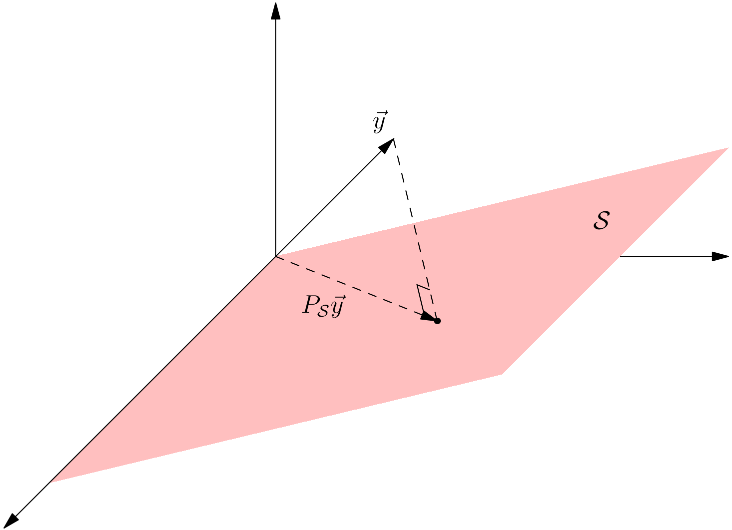
Orthogonal projection

The orthogonal projection of $\vec{x}$ onto a subspace $\mathcal{S}$ is a vector denoted by $\mathcal{P}_{\mathcal{S}}\,\vec{x}$ such that

$$\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x} \in \mathcal{S}^{\perp}$$

The orthogonal projection is unique

Orthogonal projection



Orthogonal projection

Any vector $\vec{x}$ can be decomposed into

$$\vec{x} = \mathcal{P}_{\mathcal{S}}\,\vec{x} + \mathcal{P}_{\mathcal{S}^\perp}\,\vec{x}.$$

For any orthonormal basis $\vec{b}_1, \ldots, \vec{b}_m$ of $\mathcal{S}$,

$$\mathcal{P}_{\mathcal{S}}\,\vec{x} = \sum_{i=1}^{m} \left\langle \vec{x}, \vec{b}_i \right\rangle \vec{b}_i$$

The orthogonal projection is a linear operation. For $\vec{x}$ and $\vec{y}$

$$\mathcal{P}_{\mathcal{S}}\,(\vec{x} + \vec{y}) = \mathcal{P}_{\mathcal{S}}\,\vec{x} + \mathcal{P}_{\mathcal{S}}\,\vec{y}$$

Dimension of orthogonal complement

Let $\mathcal{V}$ be a finite-dimensional vector space, for any subspace $\mathcal{S} \subseteq \mathcal{V}$

$$\dim(\mathcal{S}) + \dim\left(\mathcal{S}^\perp\right) = \dim(\mathcal{V})$$

Orthogonal projection is closest

The orthogonal projection $\mathcal{P}_{\mathcal{S}}\,\vec{x}$ of a vector $\vec{x}$ onto a subspace $\mathcal{S}$ is the solution to the optimization problem

$$\begin{aligned} \underset{\vec{u}}{\text{minimize}} \qquad & ||\vec{x} - \vec{u}||_{\langle\cdot,\cdot\rangle} \\ \text{subject to} \qquad & \vec{u} \in \mathcal{S} \end{aligned}$$

Proof

Take any point $\vec{s} \in \mathcal{S}$ such that $\vec{s} \neq \mathcal{P}_{\mathcal{S}}\,\vec{x}$

$$||\vec{x} - \vec{s}||^2_{\langle\cdot,\cdot\rangle}$$

Proof

Take any point $\vec{s} \in \mathcal{S}$ such that $\vec{s} \neq \mathcal{P}_{\mathcal{S}}\,\vec{x}$

$$||\vec{x} - \vec{s}||^2_{\langle\cdot,\cdot\rangle} = ||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x} + \mathcal{P}_{\mathcal{S}}\,\vec{x} - \vec{s}||^2_{\langle\cdot,\cdot\rangle}$$

Proof

Take any point $\vec{s} \in \mathcal{S}$ such that $\vec{s} \neq \mathcal{P}_{\mathcal{S}}\,\vec{x}$

$$\begin{aligned}
||\vec{x} - \vec{s}||^2_{\langle\cdot,\cdot\rangle} &= ||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x} + \mathcal{P}_{\mathcal{S}}\,\vec{x} - \vec{s}||^2_{\langle\cdot,\cdot\rangle} \\
&= ||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x}||^2_{\langle\cdot,\cdot\rangle} + ||\mathcal{P}_{\mathcal{S}}\,\vec{x} - \vec{s}||^2_{\langle\cdot,\cdot\rangle}
\end{aligned}$$

Proof

Take any point $\vec{s} \in \mathcal{S}$ such that $\vec{s} \neq \mathcal{P}_{\mathcal{S}}\,\vec{x}$

$$\begin{aligned} ||\vec{x} - \vec{s}||^2_{\langle \cdot,\cdot \rangle} &= ||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x} + \mathcal{P}_{\mathcal{S}}\,\vec{x} - \vec{s}||^2_{\langle \cdot,\cdot \rangle} \\ &= ||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x}||^2_{\langle \cdot,\cdot \rangle} + ||\mathcal{P}_{\mathcal{S}}\,\vec{x} - \vec{s}||^2_{\langle \cdot,\cdot \rangle} \\ &> ||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x}||^2_{\langle \cdot,\cdot \rangle} \quad \text{if } \vec{s} \neq \mathcal{P}_{\mathcal{S}}\,\vec{x} \end{aligned}$$

Denoising

Aim: Estimating a signal from perturbed measurements

If the noise is additive, the data are modeled as the sum of the signal $\vec{x}$ and a perturbation $\vec{z}$

$$\vec{y} := \vec{x} + \vec{z}$$

The goal is to estimate $\vec{x}$ from $\vec{y}$

Assumptions about the signal and noise structure are necessary

Denoising via orthogonal projection

Assumption: Signal is well approximated as belonging to a predefined subspace $\mathcal{S}$

Estimate: $\mathcal{P}_{\mathcal{S}}\,\vec{y}$, orthogonal projection of the noisy data onto $\mathcal{S}$

Error:

$$||\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{y}||_2^2 = ||\mathcal{P}_{\mathcal{S}^\perp}\,\vec{x}||_2^2 + ||\mathcal{P}_{\mathcal{S}}\,\vec{z}||_2^2$$

Proof

$$\vec{x} - \mathcal{P}_{\mathcal{S}}\, \vec{y}$$

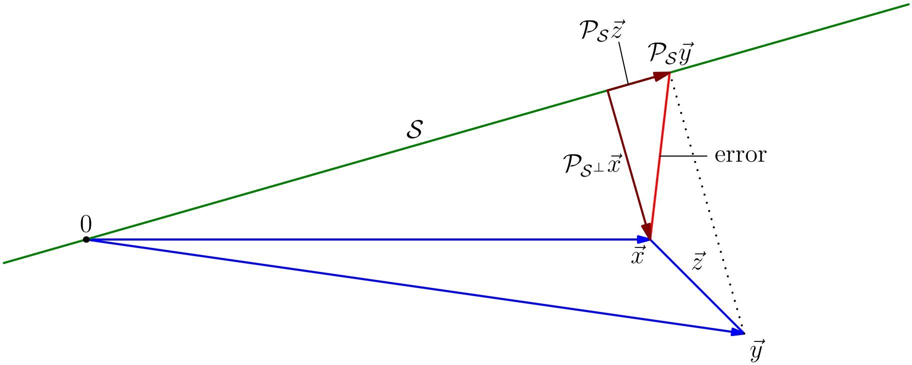
Proof

$$\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{y} = \vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{z}$$

Proof

$$\begin{aligned}\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{y} &= \vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{z} \\ &= \mathcal{P}_{\mathcal{S}^\perp}\,\vec{x} - \mathcal{P}_{\mathcal{S}}\,\vec{z}\end{aligned}$$

Error


$\mathcal{P}_{\mathcal{S}}\vec{z}$
$\mathcal{P}_{\mathcal{S}}\vec{y}$
$\mathcal{S}$
$\mathcal{P}_{\mathcal{S}^\perp}\vec{x}$
error
0
$\vec{x}$
$\vec{z}$
$\vec{y}$

Face denoising

Training set: 360 64×64 images from 40 different subjects (9 each)

Noise: iid Gaussian noise

$$\text{SNR} := \frac{||\vec{x}||_2}{||\vec{z}||_2} = 6.67$$

We model each image as a vector in $\mathbb{R}^{4096}$

Face denoising

We denoise by projecting onto:

- $\mathcal{S}_1$: the span of the 9 images from the same subject
- $\mathcal{S}_2$: the span of the 360 images in the training set

Test error:

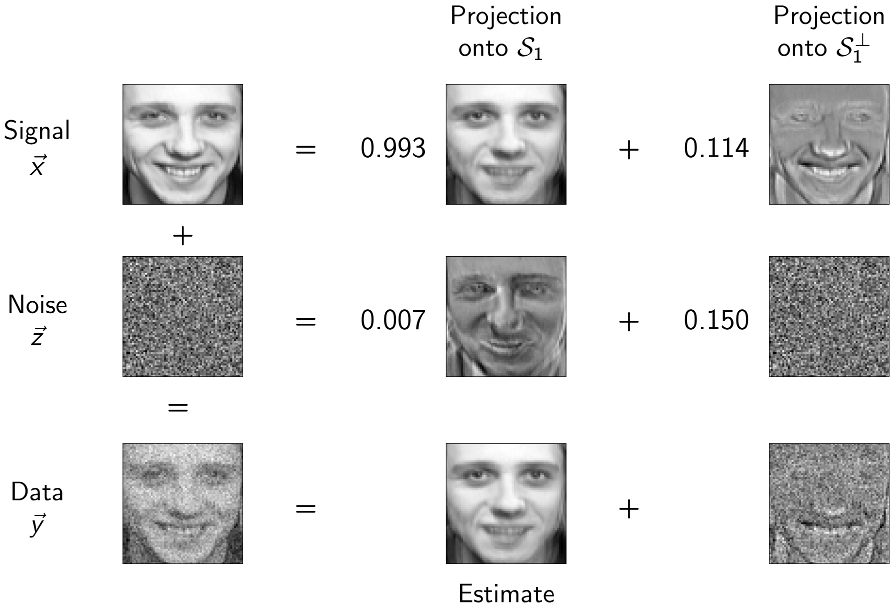
$$\frac{||\vec{x} - \mathcal{P}_{\mathcal{S}_1}\,\vec{y}||_2}{||\vec{x}||_2} = 0.114$$

$$\frac{||\vec{x} - \mathcal{P}_{\mathcal{S}_2}\,\vec{y}||_2}{||\vec{x}||_2} = 0.078$$

$\mathcal{S}_1$

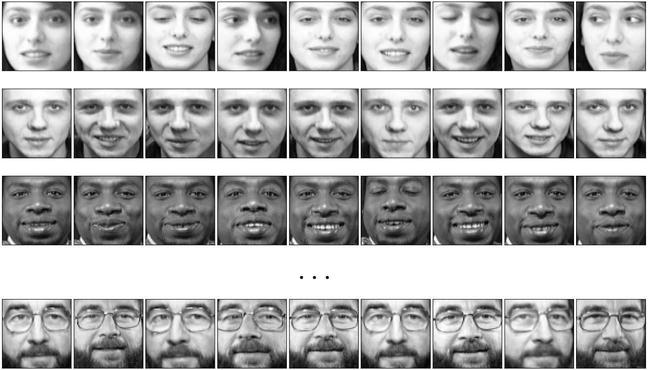
$\mathcal{S}_1 :=$ span $\Bigg($  $\Bigg)$

Denoising via projection onto $\mathcal{S}_1$



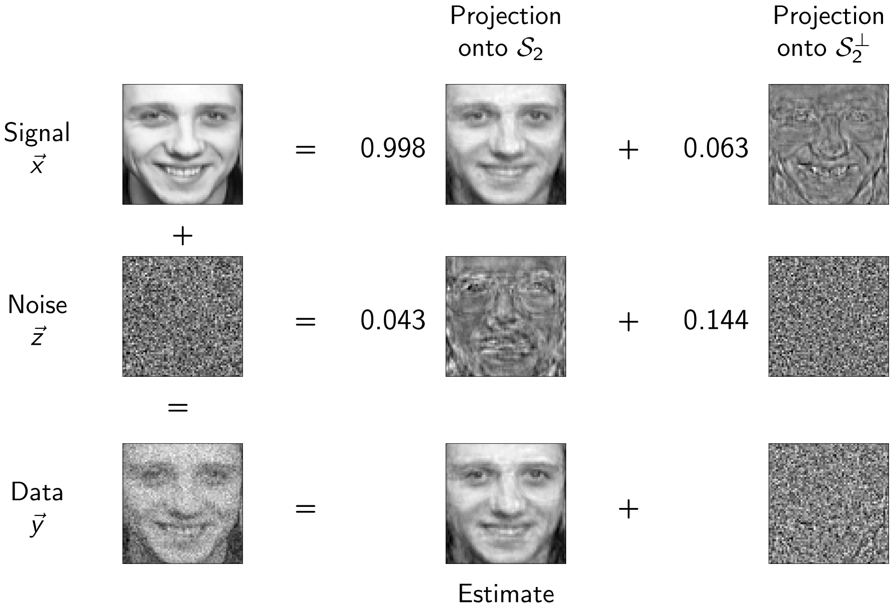
$\mathcal{S}_2$

$\mathcal{S}_2 :=$ span (



)

Denoising via projection onto $\mathcal{S}_2$



$\mathcal{P}_{\mathcal{S}_1}\,\vec{z}$ and $\mathcal{P}_{\mathcal{S}_2}\,\vec{z}$

$\mathcal{P}_{\mathcal{S}_1}\,\vec{z}$ $\mathcal{P}_{\mathcal{S}_2}\,\vec{z}$

$$0.007 = \frac{\|\mathcal{P}_{\mathcal{S}_1}\,\vec{z}\|_2}{\|\vec{x}\|_2} < \frac{\|\mathcal{P}_{\mathcal{S}_2}\,\vec{z}\|_2}{\|\vec{x}\|_2} = 0.043$$

$$\frac{0.043}{0.007} = 6.14 \approx \sqrt{\frac{\dim(\mathcal{S}_2)}{\dim(\mathcal{S}_1)}} \qquad \text{(not a coincidence)}$$

$\mathcal{P}_{\mathcal{S}_1^\perp}\, \vec{x}$ and $\mathcal{P}_{\mathcal{S}_2^\perp}\, \vec{x}$

$\mathcal{P}_{\mathcal{S}_1^\perp}\, \vec{x}$

$\mathcal{P}_{\mathcal{S}_2^\perp}\, \vec{x}$

$$0.063 = \frac{\left|\left|\mathcal{P}_{\mathcal{S}_2^\perp}\, \vec{x}\right|\right|_2}{||\vec{x}||_2} < \frac{\left|\left|\mathcal{P}_{\mathcal{S}_1^\perp}\, \vec{x}\right|\right|_2}{||\vec{x}||_2} = 0.190$$

$\mathcal{P}_{\mathcal{S}_1}\,\vec{y}$ and $\mathcal{P}_{\mathcal{S}_2}\,\vec{y}$

$\vec{x}$

$\mathcal{P}_{\mathcal{S}_1}\,\vec{y}$

$\mathcal{P}_{\mathcal{S}_2}\,\vec{y}$