MATH 311

Topics in Applied Mathematics

Lecture 18:
Least squares problems.
Norms and inner products.

Overdetermined system of linear equations:

$$\begin{cases} x + 2y = 3 \\ 3x + 2y = 5 \\ x + y = 2.09 \end{cases} \iff \begin{cases} x + 2y = 3 \\ -4y = -4 \\ -y = -0.91 \end{cases}$$

No solution: inconsistent system

Assume that a solution (x_0, y_0) does exist but the system is not quite accurate, namely, there may be some errors in the right-hand sides.

Problem. Find a good approximation of (x_0, y_0).

One approach is the **least squares fit**. Namely, we look for a pair (x, y) that minimizes the sum $(x + 2y - 3)^2 + (3x + 2y - 5)^2 + (x + y - 2.09)^2$.

Least squares solution

System of linear equations:

$$\left\{\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array}\right. \iff A\mathbf{x} = \mathbf{b}$$

For any $\mathbf{x} \in \mathbb{R}^n$ define a **residual** $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$.

The **least squares solution x** to the system is the one that minimizes $\|r(\mathbf{x})\|$ (or, equivalently, $\|r(\mathbf{x})\|^2$).

$$\|r(\mathbf{x})\|^2 = \sum_{i=1}^{m} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i)^2$$

Let A be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$.

Theorem A vector $\hat{\mathbf{x}}$ is a least squares solution of the system $A\mathbf{x} = \mathbf{b}$ if and only if it is a solution of the associated **normal system** $\boxed{A^T A\mathbf{x} = A^T\mathbf{b}.}$

Proof: $A\mathbf{x}$ is an arbitrary vector in $R(A)$, the column space of A. Hence the length of $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is minimal if $A\mathbf{x}$ is the orthogonal projection of $\mathbf{b}$ onto $R(A)$. That is, if $r(\mathbf{x})$ is orthogonal to $R(A)$.

We know that $R(A)^\perp = N(A^T)$, the nullspace of the transpose matrix. Thus $\hat{\mathbf{x}}$ is a least squares solution if and only if

$$A^T r(\hat{\mathbf{x}}) = \mathbf{0} \iff A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \iff A^T A\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Problem. Find the least squares solution to

$$\begin{cases} x + 2y = 3 \\ 3x + 2y = 5 \\ x + y = 2.09 \end{cases}$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 2.09 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 2.09 \end{pmatrix}$$

$$\begin{pmatrix} 11 & 9 \\ 9 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 20.09 \\ 18.09 \end{pmatrix} \iff \begin{cases} x = 1 \\ y = 1.01 \end{cases}$$

Problem. Find the constant function that is the least square fit to the following data

x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c \implies \begin{cases} c = 1 \\ c = 0 \\ c = 1 \\ c = 2 \end{cases} \implies \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$(1, 1, 1, 1) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (c) = (1, 1, 1, 1) \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$c = \frac{1}{4}(1 + 0 + 1 + 2) = 1$ (mean arithmetic value)

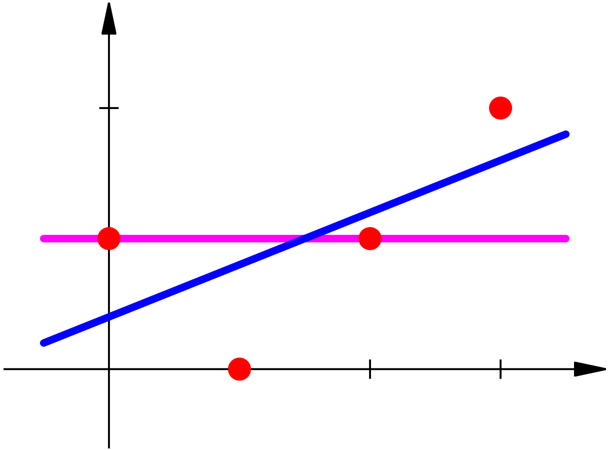
Problem. Find the linear polynomial that is the least square fit to the following data

x	0	1	2	3
$f(x)$	1	0	1	2

$$f(x) = c_1 + c_2 x \implies \begin{cases} c_1 = 1 \\ c_1 + c_2 = 0 \\ c_1 + 2c_2 = 1 \\ c_1 + 3c_2 = 2 \end{cases} \implies \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} 4 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix} \iff \begin{cases} c_1 = 0.4 \\ c_2 = 0.4 \end{cases}$$



Norm

The notion of *norm* generalizes the notion of length of a vector in $\mathbb{R}^n$.

Definition. Let V be a vector space. A function $\alpha : V \to \mathbb{R}$ is called a **norm** on V if it has the following properties:

(i) $\alpha(\mathbf{x}) \geq 0$, $\alpha(\mathbf{x}) = 0$ only for $\mathbf{x} = \mathbf{0}$ (positivity)
(ii) $\alpha(r\mathbf{x}) = |r|\,\alpha(\mathbf{x})$ for all $r \in \mathbb{R}$ (homogeneity)
(iii) $\alpha(\mathbf{x} + \mathbf{y}) \leq \alpha(\mathbf{x}) + \alpha(\mathbf{y})$ (triangle inequality)

Notation. The norm of a vector $\mathbf{x} \in V$ is usually denoted $\|\mathbf{x}\|$. Different norms on V are distinguished by subscripts, e.g., $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2$.

Examples. $V = \mathbb{R}^n$, $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

- $\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \dots, |x_n|)$.

Positivity and homogeneity are obvious.
The triangle inequality:

$$|x_i + y_i| \le |x_i| + |y_i| \le \max_j |x_j| + \max_j |y_j|$$

$$\implies \max_j |x_j + y_j| \le \max_j |x_j| + \max_j |y_j|$$

- $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$.

Positivity and homogeneity are obvious.
The triangle inequality: $|x_i + y_i| \le |x_i| + |y_i|$

$$\implies \sum_j |x_j + y_j| \le \sum_j |x_j| + \sum_j |y_j|$$

Examples. $V = \mathbb{R}^n$, $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

- $\|\mathbf{x}\|_p = \big(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\big)^{1/p}$, $p > 0$.

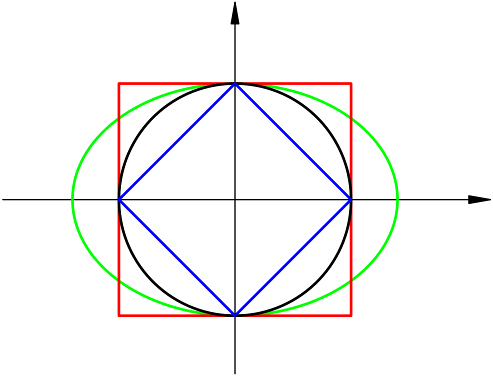
Theorem $\|\mathbf{x}\|_p$ is a norm on $\mathbb{R}^n$ for any $p \geq 1$.

Remark. $\|\mathbf{x}\|_2 =$ Euclidean length of $\mathbf{x}$.

Definition. A **normed vector space** is a vector space endowed with a norm.

The norm defines a distance function on the normed vector space: $\mathrm{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.
Then we say that a sequence $\mathbf{x}_1, \mathbf{x}_2, \ldots$ converges to a vector $\mathbf{x}$ if $\mathrm{dist}(\mathbf{x}, \mathbf{x}_n) \to 0$ as $n \to \infty$.

Unit circle: $\|\mathbf{x}\| = 1$



$\|\mathbf{x}\| = (x_1^2 + x_2^2)^{1/2}$	black
$\|\mathbf{x}\| = \left(\frac{1}{2}x_1^2 + x_2^2\right)^{1/2}$	green
$\|\mathbf{x}\| = \lvert x_1\rvert + \lvert x_2\rvert$	blue
$\|\mathbf{x}\| = \max(\lvert x_1\rvert, \lvert x_2\rvert)$	red

Examples. $V = C[a, b]$, $f : [a, b] \to \mathbb{R}$.

- $\|f\|_\infty = \max\limits_{a \le x \le b} |f(x)|$.
- $\|f\|_1 = \displaystyle\int_a^b |f(x)|\, dx$.
- $\|f\|_p = \left(\displaystyle\int_a^b |f(x)|^p\, dx \right)^{1/p}$, $p > 0$.

Theorem $\|f\|_p$ is a norm on $C[a, b]$ for any $p \ge 1$.

Inner product

The notion of *inner product* generalizes the notion of dot product of vectors in $\mathbb{R}^n$.

Definition. Let V be a vector space. A function $\beta : V \times V \to \mathbb{R}$, usually denoted $\beta(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$, is called an **inner product** on V if it is positive, symmetric, and bilinear. That is, if

(i) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ only for $\mathbf{x} = \mathbf{0}$ (positivity)

(ii) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ (symmetry)

(iii) $\langle r\mathbf{x}, \mathbf{y} \rangle = r\langle \mathbf{x}, \mathbf{y} \rangle$ (homogeneity)

(iv) $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$ (distributive law)

An **inner product space** is a vector space endowed with an inner product.

Examples. $V = \mathbb{R}^n$.

- $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$.

- $\langle \mathbf{x}, \mathbf{y} \rangle = d_1 x_1 y_1 + d_2 x_2 y_2 + \cdots + d_n x_n y_n$,

where $d_1, d_2, \ldots, d_n > 0$.

- $\langle \mathbf{x}, \mathbf{y} \rangle = (D\mathbf{x}) \cdot (D\mathbf{y})$,

where D is an invertible $n \times n$ matrix.

Remarks. (a) Invertibility of D is necessary to show that $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \implies \mathbf{x} = \mathbf{0}$.

(b) The second example is a particular case of the third one when $D = \mathrm{diag}(d_1^{1/2}, d_2^{1/2}, \ldots, d_n^{1/2})$.

Problem. Find an inner product on $\mathbb{R}^2$ such that $\langle \mathbf{e}_1, \mathbf{e}_1 \rangle = 2$, $\langle \mathbf{e}_2, \mathbf{e}_2 \rangle = 3$, and $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle = -1$, where $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$.

Let $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$.
Then $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$, $\mathbf{y} = y_1\mathbf{e}_1 + y_2\mathbf{e}_2$.
Using bilinearity, we obtain

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle x_1\mathbf{e}_1 + x_2\mathbf{e}_2, y_1\mathbf{e}_1 + y_2\mathbf{e}_2 \rangle$$

$$= x_1\langle \mathbf{e}_1, y_1\mathbf{e}_1 + y_2\mathbf{e}_2 \rangle + x_2\langle \mathbf{e}_2, y_1\mathbf{e}_1 + y_2\mathbf{e}_2 \rangle$$

$$= x_1y_1\langle \mathbf{e}_1, \mathbf{e}_1 \rangle + x_1y_2\langle \mathbf{e}_1, \mathbf{e}_2 \rangle + x_2y_1\langle \mathbf{e}_2, \mathbf{e}_1 \rangle + x_2y_2\langle \mathbf{e}_2, \mathbf{e}_2 \rangle$$

$$= 2x_1y_1 - x_1y_2 - x_2y_1 + 3x_2y_2.$$

Examples. $V = C[a, b]$.

- $\langle f, g \rangle = \int_a^b f(x)g(x)\,dx$.

- $\langle f, g \rangle = \int_a^b f(x)g(x)w(x)\,dx$,

where w is bounded, piecewise continuous, and $w > 0$ everywhere on $[a, b]$.

w is called the **weight** function.

Theorem Suppose $\langle \mathbf{x}, \mathbf{y} \rangle$ is an inner product on a vector space V. Then

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle \quad \text{for all } \mathbf{x}, \mathbf{y} \in V.$$

Proof: For any $t \in \mathbb{R}$ let $\mathbf{v}_t = \mathbf{x} + t\mathbf{y}$. Then

$$\langle \mathbf{v}_t, \mathbf{v}_t \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + 2t\langle \mathbf{x}, \mathbf{y} \rangle + t^2\langle \mathbf{y}, \mathbf{y} \rangle.$$

The right-hand side is a quadratic polynomial in t (provided that $\mathbf{y} \neq \mathbf{0}$). Since $\langle \mathbf{v}_t, \mathbf{v}_t \rangle \geq 0$ for all t, the discriminant D is nonpositive. But $D = 4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle$.

Cauchy-Schwarz Inequality:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}\, \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}.$$

Cauchy-Schwarz Inequality:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}\,\sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}.$$

Corollary 1 $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}|\,|\mathbf{y}|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Equivalently, for all $x_i, y_i \in \mathbb{R}$,

$$(x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

Corollary 2 For any $f, g \in C[a, b]$,

$$\left(\int_a^b f(x) g(x)\, dx \right)^2 \leq \int_a^b |f(x)|^2\, dx \cdot \int_a^b |g(x)|^2\, dx.$$

Norms induced by inner products

Theorem Suppose $\langle \mathbf{x}, \mathbf{y} \rangle$ is an inner product on a vector space V. Then $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ is a norm.

Proof: Positivity is obvious. Homogeneity:

$$\|r\mathbf{x}\| = \sqrt{\langle r\mathbf{x}, r\mathbf{x} \rangle} = \sqrt{r^2 \langle \mathbf{x}, \mathbf{x} \rangle} = |r| \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

Triangle inequality (follows from Cauchy-Schwarz's):

$$\|\mathbf{x} + \mathbf{y}\|^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle$$

$$= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$$

$$\leq \langle \mathbf{x}, \mathbf{x} \rangle + |\langle \mathbf{x}, \mathbf{y} \rangle| + |\langle \mathbf{y}, \mathbf{x} \rangle| + \langle \mathbf{y}, \mathbf{y} \rangle$$

$$\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \, \|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.$$

Examples. • The length of a vector in $\mathbb{R}^n$,

$$|\mathbf{x}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

is the norm induced by the dot product

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

• The norm $\|f\|_2 = \left(\int_a^b |f(x)|^2 \, dx \right)^{1/2}$ on the vector space $C[a,b]$ is induced by the inner product

$$\langle f, g \rangle = \int_a^b f(x) g(x) \, dx.$$

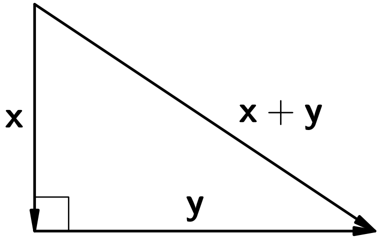
Angle

Since $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$, we can define the *angle* between nonzero vectors in any vector space with an inner product (and induced norm):

$$\angle(\mathbf{x}, \mathbf{y}) = \arccos \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \, \|\mathbf{y}\|}.$$

Then $\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \, \|\mathbf{y}\| \cos \angle(\mathbf{x}, \mathbf{y})$.

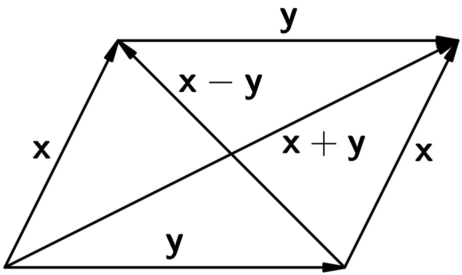
In particular, vectors $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** (denoted $\mathbf{x} \perp \mathbf{y}$) if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.



Pythagorean Theorem:

$$\mathbf{x} \perp \mathbf{y} \implies \|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Proof:
$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y}\rangle \\ &= \langle \mathbf{x},\mathbf{x}\rangle + \langle \mathbf{x},\mathbf{y}\rangle + \langle \mathbf{y},\mathbf{x}\rangle + \langle \mathbf{y},\mathbf{y}\rangle \\ &= \langle \mathbf{x},\mathbf{x}\rangle + \langle \mathbf{y},\mathbf{y}\rangle = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.\end{aligned}$$



Parallelogram Identity:

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

Proof: $\|\mathbf{x}+\mathbf{y}\|^2 = \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$.

Similarly, $\|\mathbf{x}-\mathbf{y}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle - \langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$.

Then $\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\langle \mathbf{x}, \mathbf{x} \rangle + 2\langle \mathbf{y}, \mathbf{y} \rangle = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$.