Sparsity Methods for Systems and Control

What is Sparsity?

Masaaki Nagahara[1,2]

[1]The University of Kitakyushu
nagahara@kitakyu-u.ac.jp
[2]IIT Bombay (Visiting Faculty)

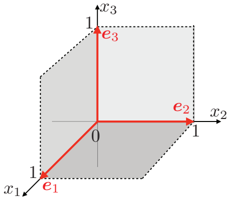
Table of Contents

Table of Contents

Standard basis for $\mathbb{R}^3$

- Standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$:

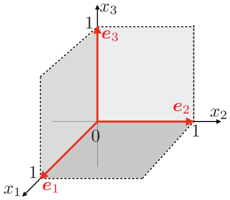
$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$



Standard basis for $\mathbb{R}^3$

- Standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$:

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$



- Any vector $\boldsymbol{y} \in \mathbb{R}^3$ can be represented as

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = y_1\boldsymbol{e}_1 + y_2\boldsymbol{e}_2 + y_3\boldsymbol{e}_3.$$

General basis for $\mathbb{R}^3$

- Any three linearly independent vectors $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, and $\boldsymbol{\phi}_3$ in $\mathbb{R}^3$ form a basis for $\mathbb{R}^3$.
- Any vector $\boldsymbol{y} \in \mathbb{R}^3$ can be represented as

$$\boldsymbol{y} = \beta_1 \boldsymbol{\phi}_1 + \beta_2 \boldsymbol{\phi}_2 + \beta_3 \boldsymbol{\phi}_3$$

- The coefficients $\beta_1, \beta_2, \beta_3$ are uniquely determined.

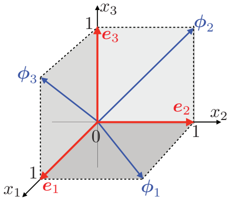
Redundant basis

- Three linearly independent vectors:

$$\phi_1 = e_1 + e_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \phi_2 = e_2 + e_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \phi_3 = e_3 + e_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

- Set of 6 vectors (redundant basis)

$$\{e_1, e_2, e_3, \phi_1, \phi_2, \phi_3\}$$



Redundant basis

- For a vector $\boldsymbol{y} \in \mathbb{R}^3$, we want a signal representation (redundant representation):

$$\boldsymbol{y} = \sum_{i=1}^{3} \alpha_i \boldsymbol{e}_i + \sum_{i=1}^{3} \beta_i \boldsymbol{\phi}_i.$$

- There are infinitely many solutions for α_i and β_i ($i = 1, 2, 3$).

$$(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3) = (y_1, y_2, y_3, 0, 0, 0),$$

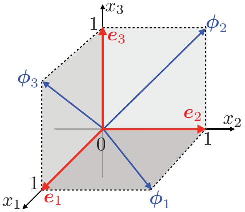
$$(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3) = (-y_3, -y_1, -y_2, y_1, y_2, y_3).$$

Sparse representation

- A vector $\boldsymbol{y} = (1, 1, 1)^\top$ on the plane spanned by $\boldsymbol{e}_1$ and $\boldsymbol{\phi}_2$.
- A coefficient set is obtained as

$$(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3) = (1, 0, 0, 0, 1, 0).$$

This is a sparse representation of $\boldsymbol{y} = (1, 1, 1)^\top$ since it contains many zeros.



Redundant dictionary

- How do you explain this picture by using words in a small dictionary that does not have the word "elephant"?

Redundant dictionary

- A set of vectors $\{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n\}$ in $\mathbb{R}^m$.
- If $m < n$ and m vectors in this set are linearly independent, then this is called a redundant dictionary.
- The elements $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n$ in the dictionary is called atoms (not "words").
- For a vector $\boldsymbol{y} \in \mathbb{R}^m$, we find coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$\boldsymbol{y} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}_i.$$

- If the dictionary is more redundant, then we may obtain a sparser coefficient set.

Sparse representation

- Define a matrix Φ and a vector $\boldsymbol{x}$ as

$$\Phi \triangleq \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \dots & \boldsymbol{\phi}_n \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad \boldsymbol{x} \triangleq \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{R}^n.$$

- Then the relation

$$\boldsymbol{y} = \sum_{i=1}^{n} \alpha_i \boldsymbol{\phi}_i.$$

is compactly rewritten as

$$\boldsymbol{y} = \Phi \boldsymbol{x}.$$

- The matrix Φ is called a dictionary matrix or measurement matrix.

The problem of sparse representation

Problem (Sparse Representation)

Given a vector $\boldsymbol{y} \in \mathbb{R}^m$ and a dictionary matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m > n$. Find the simplest (i.e. sparsest) representation of $\boldsymbol{y}$ that satisfies

$$\boldsymbol{y} = \Phi \boldsymbol{x}.$$

- This problem is also known as compressed sensing, where Φ models an oversampling sensor.

Table of Contents

Linear equations

- linear equations with unknowns x_1, x_2, and x_3:

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_3 &= 0 \end{aligned}$$

- There are infinitely many solutions
- All solutions

$$x_1 = t, \quad x_2 = -2t + 3, \quad x_3 = t,$$

  where $t \in \mathbb{R}$ is a parameter.
- Such a system of equations is called an underdetermined system.
- How can we specify one unique vector among these?

Minimum solution

- Let us consider a detective, like Edogawa Conan[1], who solve this problem.
- The two proofs (equations) are insufficient and he should seek one more independent proof.
- If he gets one more proof saying *the criminal is the smallest one among the suspects*.
- The ℓ^2-norm

$$\begin{aligned}\|\boldsymbol{x}\|_2^2 &= x_1^2 + x_2^2 + x_3^2 \\ &= t^2 + (-2t+3)^2 + t^2 \\ &= 6(t-1)^2 + 3.\end{aligned}$$

  is minimized by $t = 1$.
- The unique solution is $(x_1, x_2, x_3) = (1, 1, 1)$.

[1]See: `https://en.wikipedia.org/wiki/Case_Closed`

Linear equations in matrix form

- Linear equations in a matrix form:

$$\Phi \boldsymbol{x} = \boldsymbol{y}.$$

- Φ is an $m \times n$ matrix where $m < n$ (we call this a fat matrix).
- Assume Φ has full row rank, that is,

$$\operatorname{rank}(\Phi) = m.$$

- For any vector $\boldsymbol{y} \in \mathbb{R}^m$, there exists at least one solution $\boldsymbol{x}$ that satisfies $\Phi \boldsymbol{x} = \boldsymbol{y}$.
- In fact, there are infinitely many solutions.

Table of Contents

Norms in $\mathbb{R}^n$

Norm

A norm in $\mathbb{R}^n$ should satisfy

1. For any vector $\boldsymbol{x} \in \mathbb{R}^n$ and any number $\alpha \in \mathbb{R}$, $\|\alpha \boldsymbol{x}\| = |\alpha| \|\boldsymbol{x}\|$.
2. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.
3. $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \boldsymbol{0}$.

- The ℓ^2 norm (or Euclidean norm)

$$\|\boldsymbol{x}\|_2 \triangleq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

- The ℓ^1 norm

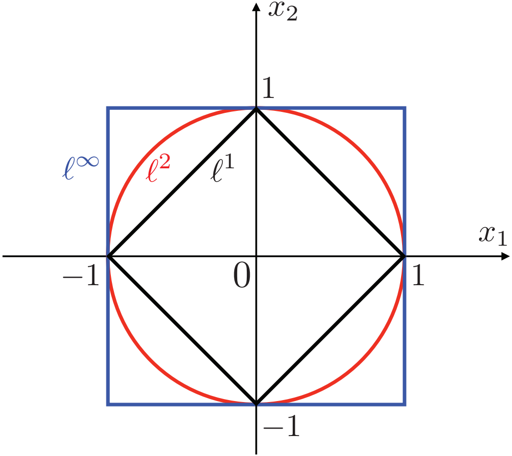
$$\|\boldsymbol{x}\|_1 \triangleq |x_1| + |x_2| + \ldots + |x_n|.$$

- The ℓ^∞ norm (or maximum norm)

$$\|\boldsymbol{x}\|_\infty \triangleq \max\{|x_1|, |x_2|, \ldots, |x_n|\}$$

Norms in $\mathbb{R}^n$

Contour curves ($\|\boldsymbol{x}\|_p = 1$) of ℓ^1, ℓ^2, ℓ^∞ norms.



ℓ^0 norm

- Consider a vector $\boldsymbol{x} = [x_1, x_2, \dots, x_n]^\top \in \mathbb{R}^n$.
- The ℓ^0 norm of $\boldsymbol{x}$ is defined by

$$\|\boldsymbol{x}\|_0 \triangleq |\mathrm{supp}(\boldsymbol{x})|,$$

  where supp is the support of $\boldsymbol{x}$, namely,

$$\mathrm{supp}(\boldsymbol{x}) \triangleq \left\{ i \in \{1, 2, \dots, n\} : x_i \neq 0 \right\},$$

  and $|\cdot|$ denotes the number of elements.
- The ℓ^0 norm counts the number of non-zero elements in $\boldsymbol{x}$.
- The ℓ^0 norm *does not* satisfy the first property in the definition of norm, and it is sometimes called the ℓ^0 pseudo-norm.

$$\|2\boldsymbol{x}\|_0 = \|\boldsymbol{x}\|_0 \neq 2\|\boldsymbol{x}\|_0.$$

ℓ^0 optimization problem

Now the problem of sparse representation is formulated as follows:

ℓ^0 optimization problem

Given a vector $\boldsymbol{y} \in \mathbb{R}^m$ and a full-row-rank matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m < n$. Find the optimizer $\boldsymbol{x}^*$ of the optimization problem:

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \ \|\boldsymbol{x}\|_0 \ \text{ subject to } \ \boldsymbol{y} = \Phi \boldsymbol{x}.$$

This problem is called the ℓ^0 optimization.

Table of Contents

How to solve it?

ℓ^0 optimization problem

Given a vector $\boldsymbol{y} \in \mathbb{R}^m$ and a full-row-rank matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m < n$. Find the optimizer $\boldsymbol{x}^*$ of the optimization problem:

$$\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \ \|\boldsymbol{x}\|_0 \ \text{subject to} \ \boldsymbol{y} = \Phi \boldsymbol{x}.$$

- We can try an exhaustive search for this optimization.

Exhaustive search: example

- Find the minimum-ℓ^0 solution (x_1, x_2, x_3) that satisfies

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_3 &= 0 \end{aligned}$$

- First, try $(x_1, x_2, x_3) = (0, 0, 0)$. This is not a solution.
- Second, try $(x_1, 0, 0)$, $(0, x_2, 0)$, and $(0, 0, x_3)$.
  1. If $(x_1, 0, 0)$ is a solution, then $x_1 = 3$ and $x_1 = 0$. This is not a solution.
  2. If $(0, x_2, 0)$ is a solution, then $x_2 = 3$. This is a solution.
  3. If $(0, 0, x_3)$ is a solution, then $x_3 = 3$ and $x_3 = 0$. This is not a solution.
- The solution to the ℓ^0 optimization is $(0, 3, 0)$.

Exhaustive Search (step 1)

- If $\boldsymbol{y} = \boldsymbol{0}$, then output $\boldsymbol{x}^* = \boldsymbol{0}$ as the optimal solution and quit.
- Otherwise, proceed to the next step.

Exhaustive Search (step 2)

- Find a vector $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_0 = 1$ that satisfies the equation $\boldsymbol{y} = \Phi\boldsymbol{x}$. That is, set

$$\boldsymbol{x}_1 \triangleq \begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{x}_2 \triangleq \begin{bmatrix} 0 \\ x_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \quad \boldsymbol{x}_n \triangleq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ x_n \end{bmatrix}$$

  and search $x_i \in \mathbb{R}$ $(i = 1, 2, \ldots, n)$ that satisfies

$$\boldsymbol{y} = \Phi\boldsymbol{x}_i = x_i\boldsymbol{\phi}_i.$$

- If a solution exists for some i, output $\boldsymbol{x}^* = \boldsymbol{x}_i$ as the solution and quit.
- Otherwise, proceed the next step.

Exhaustive Search (step 3)

- Find a vector $\boldsymbol{x}$ with $\|\boldsymbol{x}\|_0 = 2$ that satisfies the equation $\boldsymbol{y} = \Phi\boldsymbol{x}$. That is, set

$$\boldsymbol{x}_{1,2} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{x}_{1,3} \triangleq \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \quad \boldsymbol{x}_{n-1,n} \triangleq \begin{bmatrix} 0 \\ \vdots \\ 0 \\ x_{n-1} \\ x_n \end{bmatrix}$$

  and search $x_i, x_j \in \mathbb{R}$ $(i, j = 1, 2, \ldots, n)$ that satisfies

$$\boldsymbol{y} = \Phi\boldsymbol{x}_{i,j} = x_i\boldsymbol{\phi}_i + x_j\boldsymbol{\phi}_j.$$

- If a solution exists for some i, j, then output $\boldsymbol{x}^* = \boldsymbol{x}_{i,j}$ and quit.
- Otherwise, proceed the next step.

Exhaustive Search (step k)

- Do similar procedures for $\|\boldsymbol{x}\|_0 = k$, $k = 3, 4, \ldots, m$.

Is exhaustive search useful?

- It is easily implemented.
- The computation time to find a solution grows exponentially with problem size m.
- Suppose $m = 100$. Then it roughly takes $2^{100} \approx 1.3 \times 10^{30}$ iterations (at the worst). If we can do one iteration in 10^{-15} seconds (by a super computer), then we obtain the solution after 1.3×10^{15} seconds, or 30 million years.
- The study of sparse representation is to solve such a big problem in a reasonable time.

Conclusion

- Sparsity of a vector is measured by its ℓ^0 norm.
- In sparse representation, a redundant dictionary of vectors is used.
- In sparse representation, the smallest number of vectors are automatically chosen from a redundant dictionary that represent a given vector (ℓ^0 optimization).
- The exhaustive search to solve ℓ^0 optimization requires computational time that exponentially increases as the problem size increases.