Convex Relaxation and Linear Programming with Applications to Sparse Recovery

Optimal Control: From Calculus of Variations Theory to Numerical Optimization Methods and Tools, with Application to Motion Planning and Control

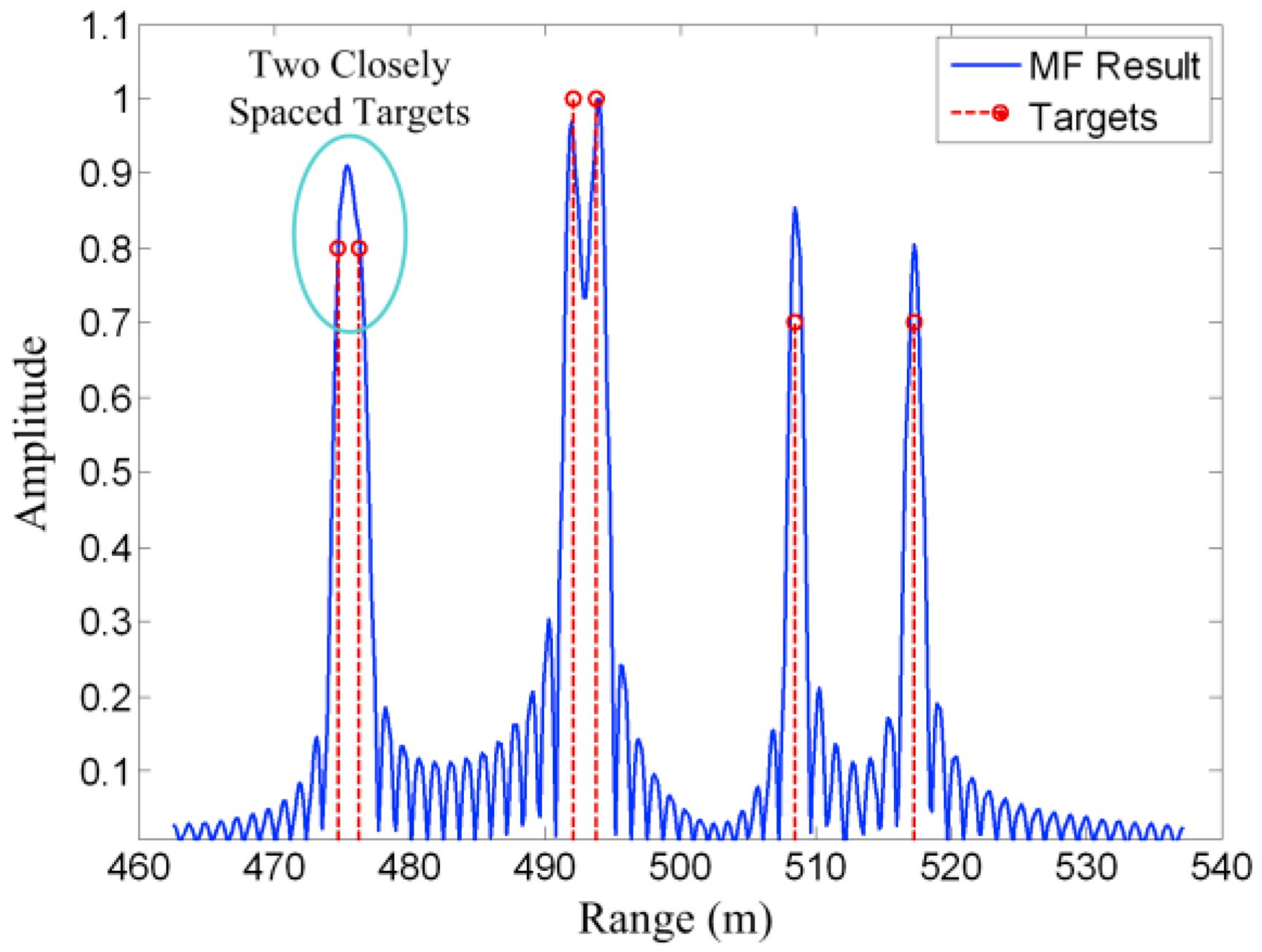
Francesco Mancuso - *francesco.mancuso@phd.unipi.it*
Department of Information Engineering (DII) - University of Pisa
RaSS National Lab - CNIT

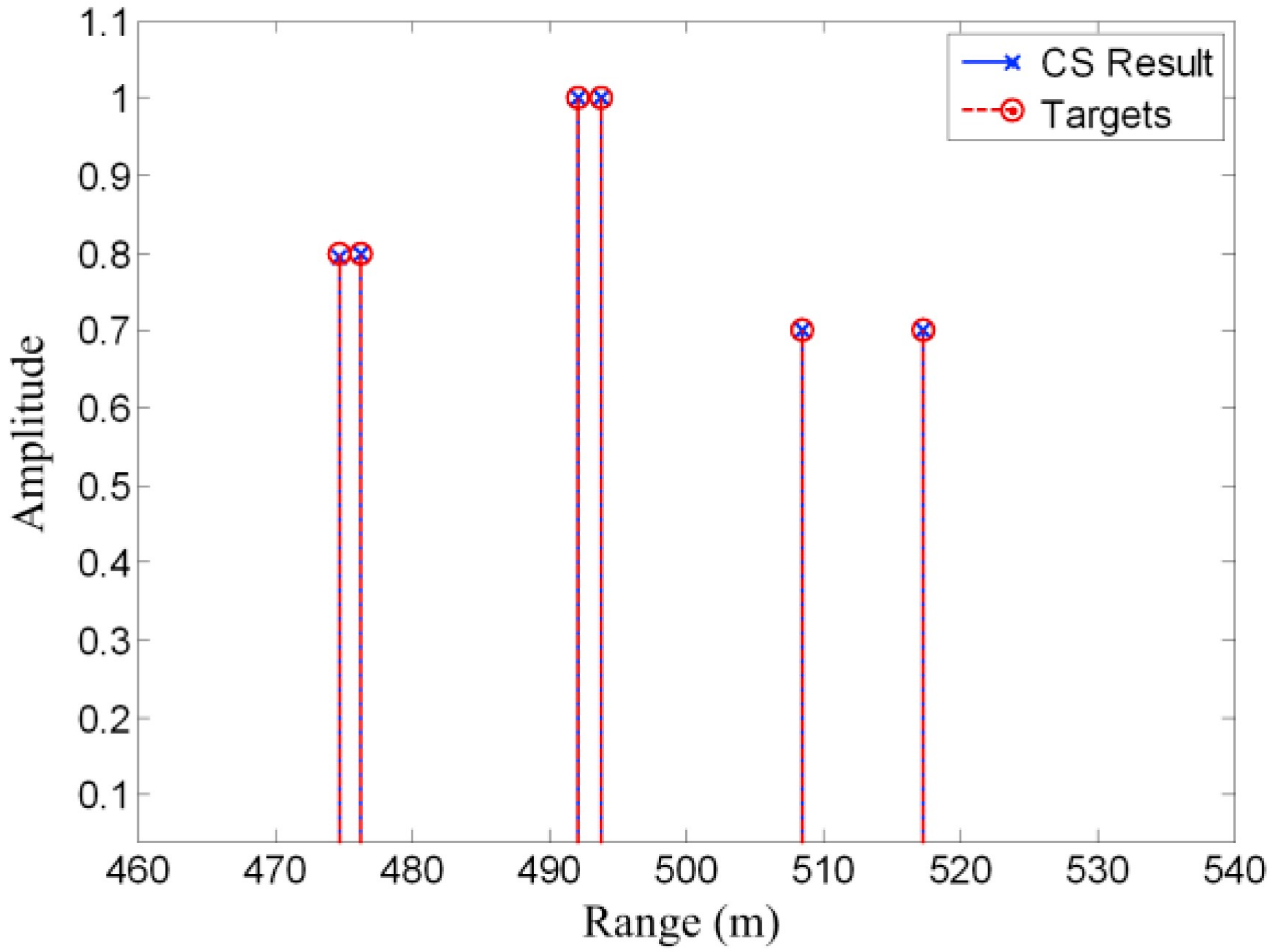
Presentation Outline

- **Compressed Sensing**
- **Linear Programming and ℓ_1-Minimisation**
- **ℓ_1-Minimisation in pop culture: the Netflix Challenge**

Compressed Sensing

Are we violating the Nyquist-Shannon sampling theorem?





images are from **[1]**, examples of CS application in radar imaging

Compressed Sensing

aka compressive sensing, sparse sampling or whatever

Compressed sensing is a signal processing technique for acquiring and reconstructing a signal, by finding solutions to **underdetermined linear systems**. Through optimization, the **sparsity** of a signal can be exploited to recover it from far fewer samples than required by the Nyquist–Shannon (and Whittaker actually) sampling theorem[1].

SPOILER ALERT: we are violating nothing. This is a misconception, because the sampling theorem guarantees perfect reconstruction **given sufficient, not necessary, conditions**. A sampling method fundamentally different from classical fixed-rate sampling cannot "violate" the sampling theorem.

[1] https://en.wikipedia.org/wiki/Compressed_sensing

Compressed Sensing

Sparse and incoherent

CS relies on two principles: ***sparsity***, which pertains to the signals of interest, and ***incoherence***, which pertains to the sensing modality [2]:

- **Sparsity** expresses the idea that the “information rate” of a continuous time signal may be much smaller than suggested by its bandwidth, or that a discrete-time signal depends on a number of degrees of freedom which is comparably much smaller than its (finite) length.

- **Incoherence** extends the duality between time and frequency and expresses the idea that objects having a sparse representation in a certain domain must be spread out in another domain, just as a Dirac or a spike in the time domain is spread out in the frequency domain.

Compressed Sensing

Recovery of sparse signal

Consider the problem of recovering an unknown sparse signal $x_0(t) \in \mathbb{R}^m$. We have n linear measurement of the form $y = Ax_0$, where the $a_k \in \mathbb{R}^m$ are known test signals. **There are many more unknowns than observations** (i.e., $n \ll m$, the linear system is vastly undetermined) [4]. We want to look for the ***sparsest*** solution of the linear system:

$$\min |\text{supp}(x)|$$

$$\text{s.t. } Ax = y$$

where $\text{supp}(x)$ denotes the support of x, meaning where $x \neq 0$. The quantity that we want to minimise is the ℓ_0 norm of x.

Compressed Sensing

ℓ_0-Minimisation is unfeasible

Unfortunately, ℓ_0-minimisation is a **NP-hard problem** [5], thus we'd like to replace the ℓ_0 norm with a more tractable norm. The ℓ_1 has two useful properties:

- Minimising the ℓ_1 **promotes sparsity**
- The ℓ_1 norm can be minimised by using **computationally efficient** algorithms

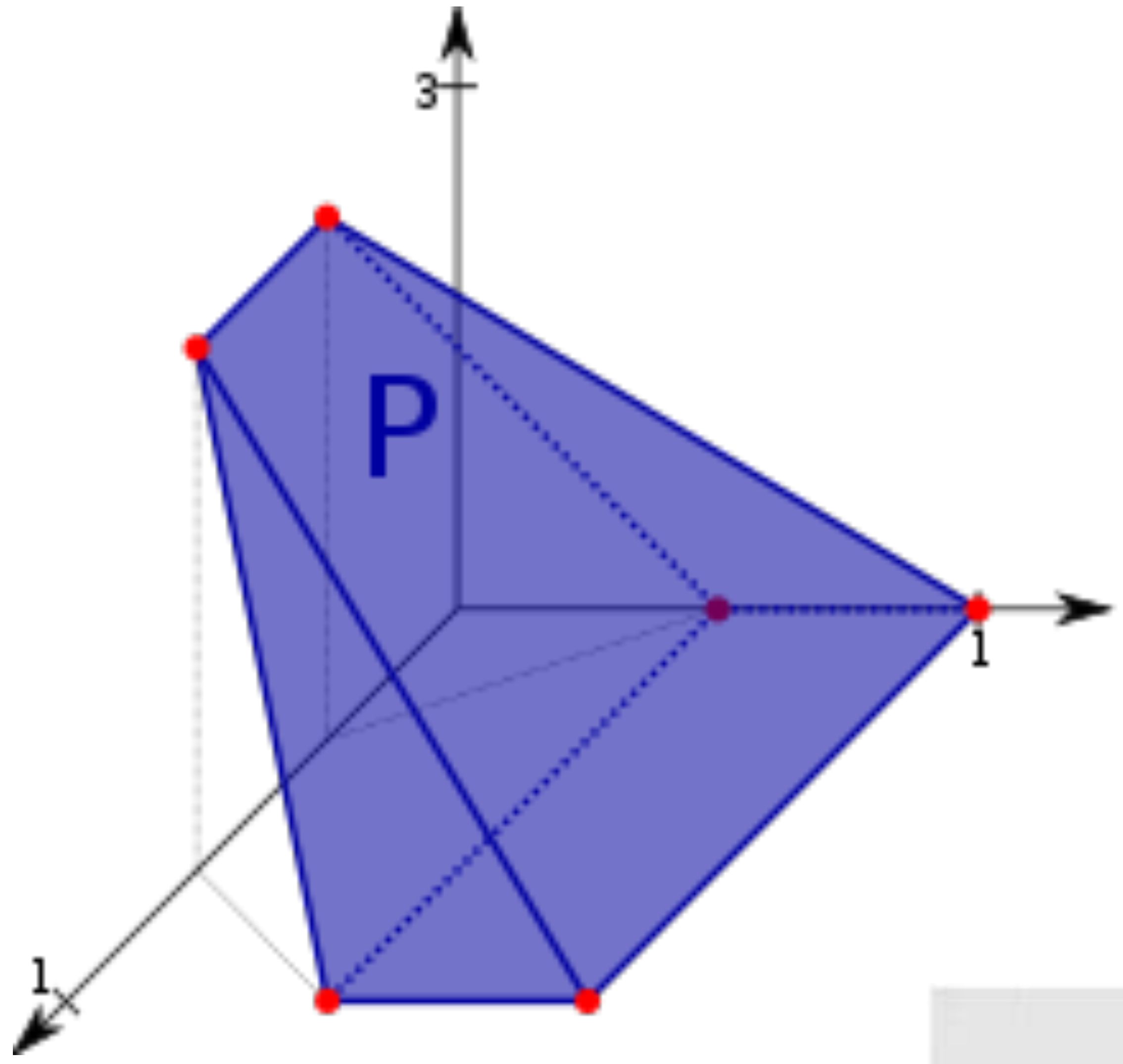
$$\min \|x\|_1 = \min \sum_{j=1}^{n} |x_j|^{*}$$

$$\text{s.t. } Ax = y$$

*provided that $A \in \mathbb{R}^{n \times m}$ obeys a **uniform uncertainty principle** [3].

Linear Programming* and ℓ_1-Minimisation

***Or how to determine the most effective utilisation of military resources during WW2[1].**



[1] https://en.wikipedia.org/wiki/George_Dantzig

Linear Programming and ℓ_1-Minimisation

Linear Programming ingredients

- **Decision variables:**

$$x_1, \ldots, x_n \in \mathbb{R}$$

- **Constraints:**

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \qquad \text{or} \qquad \sum_{j=1}^{n} a_{ij} x_j = b_i$$

- **Objective function:**

$$\min \sum_{j=1}^{n} c_j x_j$$

Linear Programming and ℓ_1-Minimisation

LP ℓ_1-Minimisation formulation

$$\cancel{\min \|x\|_1 = \sum_{j=1}^{n} |x_j|} \longrightarrow$$

IT IS NOT LINEAR

$$\min \sum_{j=1}^{n} z_j \qquad \begin{aligned} z_j - x_j &\geq 0 \\ z_j + x_j &\geq 0 \end{aligned} \qquad Ax = y \longrightarrow$$

IT IS LINEAR

ℓ_1-Minimisation in pop culture: the Netflix Challenge

Would you recommend this movie?[1].

[1] https://en.wikipedia.org/wiki/Netflix_Prize

ℓ_1-Minimisation in pop culture: the Netflix Challenge

Matrix completion[1] and movie recommendation system

There is an unknown **ground truth** matrix M, analogous to the unknown sparse signal in compressive sensing. The input is a matrix $\hat{M}$, derived from M by erasing some of its entries — the erased values are unknown, and the remaining values are known. The goal is to recover the matrix M from $\hat{M}$ [6].

Netflix was interested in the matrix M where rows are customers, columns are movies, and an entry of the matrix describes how much a customer would like a given movie. If a customer has rated a movie, then that entry is known; otherwise, it is unknown. Thus, most of the entries of M are missing in $\hat{M}$.

[1] https://en.wikipedia.org/wiki/Matrix_completion

ℓ_1-Minimisation in pop culture: the Netflix Challenge

Rank Minimisation

The key assumption is that M has **low rank**. We might try to recover M from $\hat{M}$ by solving the following optimization problem:

$$\min \operatorname{rank}(M)$$

$$\text{s.t. } M \text{ agrees with } \hat{M} \text{ on its known entries}$$

Again, this rank-minimisation problem is **NP-hard**.

Can we view the rank-minimisation as an ℓ_0-minimisation problem, and then switch to the ℓ_1 norm instead?

ℓ_1-Minimisation in pop culture: the Netflix Challenge

Rank Minimisation

The answer is: **YES!**

Let's write the Singular Value Decomposition (SVD) of M:

$$M = USV^T$$

The rank of M is equal to the number of non-zero entries entries of S, so we can rewrite the optimization problem as:

$$\min \|\Sigma(M)\|_0$$

s.t. M agrees with $\hat{M}$ on its known entries

where $\Sigma(M)$ is the set of singular values of M.

ℓ_1-Minimisation in pop culture: the Netflix Challenge

Nuclear Norm Minimisation

And now we consider the ℓ_1 relaxation[1] of the previous problem:

$$\min \|\Sigma(M)\|_1$$

s.t. M agrees with $\hat{M}$ on its known entries

This problem is called **Nuclear Norm Minimisation**, and it minimises the sum of the singular values (i.e., the Nuclear Norm) maintaining the consistency with the known information.

[1] https://en.wikipedia.org/wiki/Matrix_completion#Algorithms_for_Low-Rank_Matrix_Completion

References

[1] Yang, J.; Jin, T.; Xiao, C.; Huang, X. *Compressed Sensing Radar Imaging: Fundamentals, Challenges, and Advances*. *Sensors* 2019, *19*, 3100. https://doi.org/10.3390/s19143100

[2] E.J. Candes, M. Wakin. (2008). Wakin, M.B.: *An introduction to compressive sampling*. IEEE Signal Process. Mag. 25(2), 21-30. Signal Processing Magazine, IEEE. 25. 21 - 30. https://doi.org/10.1109/MSP.2007.914731 .

[3] E.J. Candes, J. Romberg and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, in IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489-509, Feb. 2006, https://doi.org/10.1109/TIT.2005.862083

[4] E.J. Candes, J. Romberg and T. Tao (2006), *Stable signal recovery from incomplete and inaccurate measurements*. Comm. Pure Appl. Math., 59: 1207-1223. https://doi.org/10.1002/cpa.20124

[5] T. Roughgarden, G. Valiant, CS168: *The Modern Algorithmic Toolbox Lecture #17: Compressive Sensing,* 2022, https://web.stanford.edu/class/cs168/l/l17.pdf

[6] T. Roughgarden, G. Valiant, CS168: *The Modern Algorithmic Toolbox Lecture #18: Linear and Convex Programming, with Applications to Sparse Recovery,* 2022, https://web.stanford.edu/class/cs168/l/l18.pdf