Department of Mathematics
Technical University of Munich

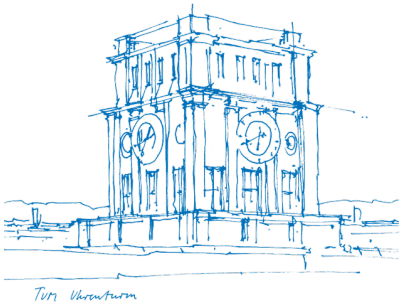
Compressed Sensing

Seminar of Mathematics in Data Science
(Prof. Dr. Massimo Fornasier, PD Dr. Peter Massopust)

Oleksii Davydenko

Department of Mathematics
Technical University of Munich

June 30th, 2021



Outline

Problem Statement

Goal: Recover a (sparse) signal from its reduced representation (measure).

Fundamental theorem of linear algebra: in order to effectively reconstruct a signal we need a sequence of measurements that is at least as long as the original signal.

Assume sparsity -> recover signal from "incomplete" measures!

This applies primarily in situations when repeated measures are costly or harmful (like in the case of magnetic resonance imaging) or when measures are naturally sparse (like radar scans).

Applications
on naturally sparse data

- Magnetic Resonance Imaging (MRI)
- Radar data
- Wireless communication
- Astronomical signal and image processing
- Camera design and imaging
- Collaborative filtering
- Matrix completion

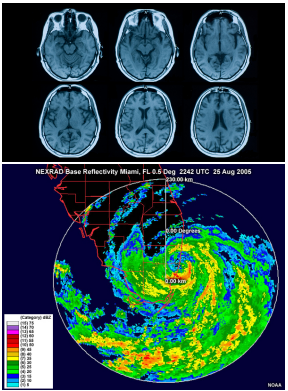


Image sources: [1] and [2]

Outline

Why l_1 optimization?

Our signal transformation process can be described as $y = Ax$, where $x \in \mathbb{C}^N$, $A \in \mathbb{C}^{m \times N}$ - the *measurement matrix*, $y \in \mathbb{C}^m$ - *measurement vector*. We are especially interested in the case when $m << N$. With the additional assumption that x is k-sparse, restoring it from y becomes realistic. The l_1-minimization approach considers the solution of

$$min\|z\|_1 \text{ subject to } Az = y$$

which is a convex problem and can be seen as a convex relaxation of

$$min\|z\|_0 \text{ subject to } Az = y$$

l_1-minimization tends to promote sparse solutions.

Sparse Vectors

We denote the support of the vector x as $supp(x) = \{j : x_j \neq 0\}$, and

$$\|x\|_0 := |supp(x)|$$

denotes the "l_0-norm" of the vector, even though formally it is not a norm.
Then a vector x is called *k-sparse* if $\|x\|_0 \leq k$. We can now define the best k-term approximation error of a vector $x \in \mathbb{C}^N$ as

$$\sigma_k(x)_p = \inf_{z \in \Sigma_k} \|x - z\|_p,$$

where Σ_k is a set of all k-sparse vectors

$$\Sigma_k := \{x \in \mathbb{C}^N : \|x\|_0 \leq k\}.$$

The Null Space Property

The null space property provides the necessary and sufficient conditions on the reconstruction of signals using the techniques of l_1-relaxation.

Definition 1. A matrix $A \in \mathbb{C}^{m \times N}$ is said to satisfy the null space property (NSP) of order k with constant $\gamma \in (0, 1)$ if

$$\|\eta_T\|_1 \leq \gamma \|\eta_{T^c}\|_1$$

for all sets $T \subset \{1, ..., N\}, \#T \leq k$ and for all $\eta \in kerA$.

Theorem 1. Let $A \in \mathbb{C}^{m \times N}$ be a matrix that satisfies the NSP of order k with constant $\gamma \in (0, 1)$. Let $x \in \mathbb{C}^N$ and $y = Ax$ and let x^* be a solution of the l_1-minimization problem. Then

$$\|x - x^*\|_1 \leq \frac{2(1+\gamma)}{1-\gamma} \sigma_k(x)_1$$

In particular, if x is k-sparse then $x^* = x$.

The Null Space Property

The null space property provides the necessary and sufficient conditions on the reconstruction of signals using the techniques of l_1-relaxation.

Definition 1. A matrix $A \in \mathbb{C}^{m \times N}$ is said to satisfy the null space property (NSP) of order k with constant $\gamma \in (0, 1)$ if

$$\|\eta_T\|_1 \leq \gamma \|\eta_{T^c}\|_1$$

for all sets $T \subset \{1, ..., N\}, \#T \leq k$ and for all $\eta \in kerA$.

It is also possible to show that if all k-sparse x can be recovered from $y = Ax$ using l_1-minimization the necessarialy A satisfies the NSP of order k with some constant $\gamma \in (0, 1)$. This means that NSP is actually equivalent to sparse l_1-recovery.

Restricted Isometry Property

Since the NSP is difficult to show directly, the restricted isometry property (RIP) is used, which is easier to handle and it also implies stability under noise.

Definition 2. The restricted isometry constant δ_k of a matrix $A \in \mathbb{C}^{m \times N}$ is the smallest number such that

$$(1 - \delta_k)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta_k)\|z\|_2^2,$$

for all $z \in \Sigma_k$.
A matrix A is said to satisfy the restricted isometry property of order k with constant σ_k if $\sigma_k \in (0, 1)$. σ_k can also be equivalently defined as

$$\delta_k = \max_{T \subset \{1,\dots,N\}, \#T \leq k} \|A_T^* A_T - I\|_{2 \to 2},$$

which means that all column submatrices of A with at most k columns are required to be well-conditioned.

Connection between RIP and NSP

The RIP implies the NSP as shown in the following lemma.

Lemma 1. Assume that $A \in \mathbb{C}^{m \times N}$ satisfies the RIP of order $K = k + h$ with constant $\sigma_k \in (0, 1)$. Then A has the NSP of order k with constant $\gamma = \sqrt{\frac{k}{h} \frac{1+\sigma_K}{1-\sigma_K}}$

Outline

Many ways to do l_1-optimization

- Chambolle and Pock's Primal Dual algorithm
- Iteratively Reweighted Least Squares (IRLS)
- LARS method
- Orthogonal Matching Pursuit
- CoSaMP
- Iterative hard thresholding

Proximal mapping

For some convex function G the proximal mapping is defined as

$$P_G(\tau, z) := arg \min_{x \in \mathbb{C}^N} \{\tau G(x) + \frac{1}{2}\|x - z\|_2^2\}$$

and

$$F^*(\xi) = sup\{Re\langle z, \xi\rangle - F(z) : z \in \mathbb{C}^m\}$$

denotes the convex (Fenchel) conjugate of F.

Chambolle and Pock's algorithm

A goal of the primal dual approach is to solve the optimization problem in the form

$$\min_{x \in \mathbb{C}^N} F(Ax) + G(x). \tag{1}$$

It performs an iteration on the dual variable, the primal variable and an auxiliary primal variable. The initial points are defined as $x^0, \bar{x}^0 = x \in \mathbb{C}^N, \xi^0 \in \mathbb{C}^m$ and parameters $\tau, \sigma > 0, \theta \in [0, 1]$, we iteratively compute

$$\xi^{n+1} := P_{F^*}(\sigma; \xi^n + \sigma A \bar{x}^n), \tag{2}$$

$$x^{n+1} := P_G(\tau; x^n - \tau A^* \xi^{n+1}), \tag{3}$$

$$\bar{x}^{n+1} := x^{n+1} + \theta(x^{n+1} - x^n), \tag{4}$$

Chambolle and Pock's algorithm

This algorithm converges to the saddle point of the primal dual problem, and can be interpreted as a gradient descent method for solving the primal problem, combined with a gradient ascent method to simultaneously solve the dual problem. Its convergence has been proven (see the seminar paper).

Iteratively reweighted least squares (IRLS)

This iterative algorithm assumes that A satisties the NSP and is guaranteed to reconstruct vectors with the same error estimate as l_1-minimization. The algorithm has a guaranteed linear rate of convergence which can even be improved to superlinear rate with a small modification.
Denote $\mathcal{F}(y) = \{x : Ax = y\}$ and $\mathcal{N} = kerA$. Observe that $|t| = \frac{t^2}{|t|}$ for $t \neq 0$. Hence, an l_1-minimization can be recasted into a weighted l_2-minimization, with the hope that

$$arg \min_{x \in \mathcal{F}(y)} \sum_{j=1}^{N} |x_j| \approx arg \min_{x \in \mathcal{F}(y)} \sum_{j=1}^{N} x_j^2 |x_j^*|^{-1},$$

as soon as x^* is the desired l_1-norm minimizer. The advantage of the reformulation is that minimizing the smooth function t^2 is much easier than minimizing the nonsmooth function $|t|$. However, neither one disposes of x^* a priori (this is the vector we are interested in computing), and we cannot expect that $x_j^* \neq 0$ for all $j = 1, ..., N$, since we hope for k-sparse solutions.

Iteration

If we had a good approximation w_j^n of $|(x_j^*)^2 + \epsilon_n^2|^{-1/2} \approx |x_j^*|^{-1}$, for some $\epsilon_n > 0$, we could compute

$$x^{n+1} = arg \min_{x \in \mathcal{F}(y)} \sum_{j=1}^{N} x_j^2 w_j^n, \tag{5}$$

and then update $\epsilon_{n+1} \leq \epsilon_n$ by a certain rule specified later. Furthermore, we set

$$w_j^{n+1} = |(x_j^{n+1})^2 + \epsilon_{n+1}^2|^{-1/2}, \tag{6}$$

and iterate the process. Ideally, a good chouce of $\epsilon_n \to 0$ will allow the iterative computation of the l_1-minimizer.

Weighted l_2-minimization

If we assume that the weight w is strictly positive ($w_j > 0$ for all $j \in \{1, ..., N\}$), then $l_2(w)$ is a Hilbert space with the inner product

$$\langle u, v \rangle_w := \sum_{j=1}^{N} w_j u_j v_j.$$

We define

$$x^w := arg \min_{z \in \mathcal{F}(y)} \|z\|_{2,w},$$

where $\|z\|_{2,w} = \langle z, z \rangle_w^{1/2}$. Because the $\|\cdot\|_{2,w}$-norm is strictly convex, the minimizer x^w is necesarily unique; it is characterized by the orthogonality conditions

$$\langle x^w, \eta \rangle_w = 0, \text{ for all } \eta \in \mathcal{N}$$

IRLS obeys a linear convergence rate.

Alternating Optimization

We analyze the algorithm (5 and 6) by observing that

$$|t| = \min_{w>0} \frac{1}{2}(wt^2 + w^{-1}),$$

there the minimum is attained for $w = \frac{1}{|t|}$. Based on this simple relationship, given a real number $\epsilon > 0$ and a weight factor $w \in \mathbb{R}^N$, with $w_j > 0, j = 1, ..., N$, we introduce the functional

$$\mathcal{J}(z, w, \epsilon) := \frac{1}{2} \sum_{j=1}^{N} (z_j^2 w_j + \epsilon^2 + w_j^{-1}), z \in \mathbb{R}^N$$

The algorithm described by 5 and 6 can be recast as an alternating method for choosing optimizers and weights based on the functional $\mathcal{J}$. To provide more detail, recall that $r(z)$ denotes the nonincreasing rearrangement of a vector $z \in \mathbb{R}^N$.

IRLS Algorithm

Initialize $w^0 := (1, ..., 1)$. Set $\epsilon := 1$. For $n = 0, 1, ...$, recursively define

$$\epsilon_{n+1} := \min\{\epsilon_n, \frac{r_{K+1}(x^{n+1})}{N}\}, \tag{7}$$

and

$$x^{n+1} := arg \min_{z \in \mathcal{F}(y)} \mathcal{J}(z, w^n, \epsilon_n) = arg \min_{z \in \mathcal{F}(y)} \|z\|_{2,w^n} \tag{8}$$

where K is a fixed integer that will be specified later. Finally, set

$$w^{n+1} := arg \min_{w>0} \mathcal{J}(x^{n+1}, w, \epsilon_{n+1}). \tag{9}$$

IRLS Algorithm

The algorithm stops if $\epsilon_n = 0$; in this case, define $x^j := x^n$ for $j > n$. In general, this algorithm generates an infinite sequence $(x^n)_{n\in\mathbb{N}}$ of vectors.

At every step the algorithm requires the solution of a weighted least squares problem. In matrix form

$$x^{n+1} = D_n^{-1} A^* (A D_n^{-1} A^*)^{-1} y,$$

where D_n is the $N \times N$ diagonal matrix, the j-th diagonal entry of which is w_j^n. Once x^{n+1} is found, the weight w^{n+1} is given by

$$w_j^{n+1} = [(x_j^{n+1})^2 + \epsilon_{n+1}^2]^{-1/2}, j = 1, ..., N.$$

Outline

Extensions of Compressed Sensing

Matrix input. We could consider restoring not only vector input, but extend it to matrices of minimal rank consistent with a given underdetermined linear system of equations.

Matrix completion. In the matrix completion setup the measurements are the pointwise observations of entries of the matrix. The RIP fails completely in this setting, and 'localized' low-rank matrices in the null space of S cannot be recovered by any method whatsoever. However, if certain conditions on the left and right singular vectors if the underlying low-rank matrix are imposed, essentially requiring that such vectors are uncorrelated with the canonical basis, then it was shown that such incoherent matrices of rank at most k can be recovered from m randomly chosen entries with high probability provided

$$m \geq Ck \max\{n, p\} log^2(\max\{n, p\}).$$

Thank you for your attention!