Sparse Recovery with Fusion Frames

Ulaş Ayaz

ICERM / University of Bonn

09.09.2014

Outline

1. Short bio
2. Introduction to Compressed Sensing
3. Structured Sparsity Models
4. Extensions of Compressed Sensing

About myself

- I finished my PhD under Holger Rauhut at Hausdorff Center for Mathematics, University of Bonn.
- My research lies in mathematical signal processing and compressed sensing.
- I am broadly interested in analysis of sparse and low-dimensional representations of high-dimensional data.

About myself

- I finished my PhD under Holger Rauhut at Hausdorff Center for Mathematics, University of Bonn.
- My research lies in mathematical signal processing and compressed sensing.
- I am broadly interested in analysis of sparse and low-dimensional representations of high-dimensional data.
- Prior to my PhD, I received a Masters degree at the Mathematics department of University of British Columbia under Özgür Yilmaz. My Msc thesis was related to Sigma-Delta quantization.
- My undergraduate degree is from Electrical Engineering and Mathematics at Bogazici University of Istanbul.

About myself

- I finished my PhD under Holger Rauhut at Hausdorff Center for Mathematics, University of Bonn.
- My research lies in mathematical signal processing and compressed sensing.
- I am broadly interested in analysis of sparse and low-dimensional representations of high-dimensional data.
- Prior to my PhD, I received a Masters degree at the Mathematics department of University of British Columbia under Özgür Yilmaz. My Msc thesis was related to Sigma-Delta quantization.
- My undergraduate degree is from Electrical Engineering and Mathematics at Bogazici University of Istanbul.
- For the past 6 months, I worked for a project at SONY in Stuttgart, Germany, where we investigated the applicability of compressed sensing ideas to image sensors.

Introduction

- The conventional way of data acquisition in signal and image processing: Shannon's sampling theorem.

Introduction

- The conventional way of data acquisition in signal and image processing: Shannon's sampling theorem.
- Typically, massive amount of samples are collected through the measurement device, then a compression stage is invoked in which most of this data is discarded.
- This is possible because many man-made and natural images are 'sparse' in the sense that most of its components in an appropriate basis are small or zero.
- Examples: MRI machines, digital cameras, radar..

Cameraman (size= 256×256)

Cameraman (size= 256×256)

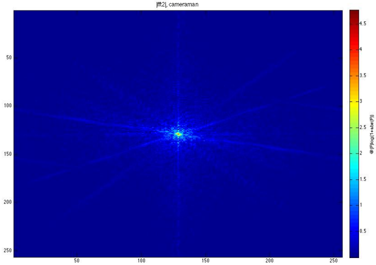

Figure: Fourier coefficients

Cameraman (size= 256×256)

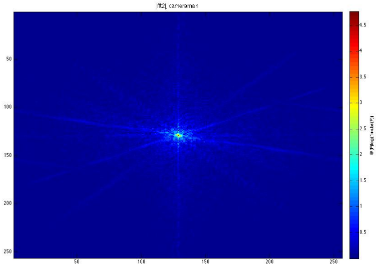

Figure: Fourier coefficients

- **Compressed Sensing** (CS) initiated by Candès, Romberg, Tao and Donoho merges the two steps of acquiring and compressing data into one step by exploiting the sparseness inherent in many signals.

- CS recovers sufficiently sparse vectors from highly incomplete information.
- **Goal:** to solve

$$y = Ax$$

for $x \in \mathbb{R}^N$ when $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times N}$ are given, and when $m \ll N$.

- Many measurement schemes in signal acquisition can be modelled by such a linear system, where x being the signal of interest.

- CS recovers sufficiently sparse vectors from highly incomplete information.
- **Goal:** to solve
$$y = Ax$$
for $x \in \mathbb{R}^N$ when $y \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times N}$ are given, and when $m \ll N$.
- Many measurement schemes in signal acquisition can be modelled by such a linear system, where x being the signal of interest.
- In general, this system is underdetermined, i.e., we have infinitely many solutions.
- We assume that x is s-sparse, that is, $\|x\|_0 := |\text{supp } x| \leq s \ll N$.

- Solving

$$\min_{z \in \mathbb{R}^N} \|z\|_0 \text{ subject to } Az = y,$$

where $Ax = y$, is NP hard.

- Solving

$$\min_{z\in\mathbb{R}^N} \|z\|_0 \text{ subject to } Az = y,$$

where $Ax = y$, is NP hard.

- **Method:** ℓ_1-minimization (Basis Pursuit).

$$\min_{z\in\mathbb{R}^N} \|z\|_1 \text{ subject to } Az = y,$$

where $Ax = y$ and $\|z\|_1 = \sum_i |z_i|$.

- The ℓ_1-minimization reconstructs the original x under appropriate conditions on A and on the sparsity of x.

- Solving

$$\min_{z \in \mathbb{R}^N} \|z\|_0 \text{ subject to } Az = y,$$

  where $Ax = y$, is NP hard.

- **Method:** ℓ_1-minimization (Basis Pursuit).

$$\min_{z \in \mathbb{R}^N} \|z\|_1 \text{ subject to } Az = y,$$

  where $Ax = y$ and $\|z\|_1 = \sum_i |z_i|$.

- The ℓ_1-minimization reconstructs the original x under appropriate conditions on A and on the sparsity of x.

- One goal is to obtain bounds on required m for sparse recovery. Hard to come up with deterministic constructions of A.

- Random constructions, e.g. Gaussian and Bernoulli, prove to be useful.
- We call a matrix **Gaussian** if all entries are independently chosen from standard normal distribution and **Bernoulli** if all entries are ± 1 with equal probability. These are examples of **subgaussian** random variables, which satisfy

$$\mathbb{P}(|X| \geq t) \leq \beta e^{-\theta t^2} \quad \text{for all } t > 0$$

for some constants $\beta, \theta > 0$.

- Random constructions, e.g. Gaussian and Bernoulli, prove to be useful.
- We call a matrix **Gaussian** if all entries are independently chosen from standard normal distribution and **Bernoulli** if all entries are ± 1 with equal probability. These are examples of **subgaussian** random variables, which satisfy

$$\mathbb{P}(|X| \geq t) \leq \beta e^{-\theta t^2} \quad \text{for all } t > 0$$

  for some constants $\beta, \theta > 0$.

- Such matrices with $m \geq Cs \ln(N/s)$ guarantee recovery of s-sparse vectors via ℓ_1-minimization with high probability.

- Random constructions, e.g. Gaussian and Bernoulli, prove to be useful.
- We call a matrix **Gaussian** if all entries are independently chosen from standard normal distribution and **Bernoulli** if all entries are ± 1 with equal probability. These are examples of **subgaussian** random variables, which satisfy

$$\mathbb{P}(|X| \geq t) \leq \beta e^{-\theta t^2} \text{ for all } t > 0$$

  for some constants $\beta, \theta > 0$.
- Such matrices with $m \geq Cs \ln(N/s)$ guarantee recovery of s-sparse vectors via ℓ_1-minimization with high probability.
- This bound is also optimal.

- However, there is a big gap between theoretical guarantees and practical performance of CS techniques in terms of the constant C.
- Obtaining good constants is important for practical purposes.

Theorem

(A., Rauhut'11) Let $A \in \mathbb{R}^{m \times N}$ be a Gaussian matrix. Then an s-sparse vector x can be recovered via ℓ_1-minimization w.h.p. if

$$m > 2s \ln(4N).$$

- However, there is a big gap between theoretical guarantees and practical performance of CS techniques in terms of the constant C.
- Obtaining good constants is important for practical purposes.

Theorem

(A., Rauhut'11) Let $A \in \mathbb{R}^{m \times N}$ be a Gaussian matrix. Then an s-sparse vector x can be recovered via ℓ_1-minimization w.h.p. if

$$m > 2s \ln(4N).$$

- The constant $C = 2$ is optimal due to Donoho and Tanner's asymptotic analysis via random polytopes. The log-term falls slightly short of the optimal one $\log(N/s)$.

- However, there is a big gap between theoretical guarantees and practical performance of CS techniques in terms of the constant C.
- Obtaining good constants is important for practical purposes.

Theorem

(A., Rauhut'11) Let $A \in \mathbb{R}^{m \times N}$ be a Gaussian matrix. Then an s-sparse vector x can be recovered via ℓ_1-minimization w.h.p. if

$$m > 2s \ln(4N).$$

- The constant $C = 2$ is optimal due to Donoho and Tanner's asymptotic analysis via random polytopes. The log-term falls slightly short of the optimal one $\log(N/s)$.
- A similar result holds for subgaussian matrices, e.g., Bernoulli matrices.

Structured Sparsity

- Often signals possess more structure than just plain sparsity.

Structured Sparsity

- Often signals possess more structure than just plain sparsity.
- In the *block sparsity* model one assumes that a signal consists of blocks, of which only a few are nonzero.
- Consider signals $\mathbf{x} = (x_j)_{j=1}^N \in \mathbb{R}^{dN}$, where the components $x_j \in \mathbb{R}^d$ are vectors themselves. We say that $\mathbf{x}$ is *s-sparse* if it is s-sparse in the block sense, i.e., $\|\mathbf{x}\|_0 \leq s$, where

$$\|\mathbf{x}\|_0 = \mathrm{card}\{j \in [N] : x_j \neq 0\}.$$

Structured Sparsity

- Often signals possess more structure than just plain sparsity.
- In the *block sparsity* model one assumes that a signal consists of blocks, of which only a few are nonzero.
- Consider signals $\mathbf{x} = (x_j)_{j=1}^N \in \mathbb{R}^{dN}$, where the components $x_j \in \mathbb{R}^d$ are vectors themselves. We say that $\mathbf{x}$ is *s-sparse* if it is s-sparse in the block sense, i.e., $\|\mathbf{x}\|_0 \leq s$, where

$$\|\mathbf{x}\|_0 = \mathrm{card}\{j \in [N] : x_j \neq 0\}.$$

- This model is strongly related to the *joint sparsity* model, where one assumes that nonzeros coefficients appear at the same location within each of the channels, e.g., three color channels of an RGB image.

Compressed sensing of structured signals

- In my research, I worked on a refinement of the block sparsity model was introduced which is related to the concept of fusion frames.
- In the fusion frame sparsity model we assume that the signal is block sparse and, in addition, lies in

$$\mathcal{H} = \{\mathbf{x} = (x_j)_{j=1}^{N} : x_j \in W_j, \ \forall j \in [N]\} \subset \mathbb{R}^{dN},$$

i.e., every nonzero block x_j of the signal is assumed to lie in a certain k-dimensional subspace W_j of $\mathbb{R}^d$.

Compressed sensing of structured signals

- In my research, I worked on a refinement of the block sparsity model was introduced which is related to the concept of fusion frames.
- In the fusion frame sparsity model we assume that the signal is block sparse and, in addition, lies in

$$\mathcal{H} = \{\mathbf{x} = (x_j)_{j=1}^N : x_j \in W_j, \ \forall j \in [N]\} \subset \mathbb{R}^{dN},$$

  i.e., every nonzero block x_j of the signal is assumed to lie in a certain k-dimensional subspace W_j of $\mathbb{R}^d$.
- We measure the coherence with the parameter

$$\lambda = \max_{i \neq j} \|P_i P_j\|_{2 \to 2},$$

  where P_i is the orthogonal projection onto the subspace W_i.

Setup and results

- I am particularly interested in determining whether one can reduce the number of measurements needed for sparse recovery if one knows that the subspaces are incoherent.

Setup and results

- I am particularly interested in determining whether one can reduce the number of measurements needed for sparse recovery if one knows that the subspaces are incoherent.

- We collect m linear measurements of a vector $\mathbf{x} \equiv (x_j)_{j=1}^N \in \mathcal{H}$

$$\mathbf{y} = (y_i)_{i=1}^m = \left(\sum_{j=1}^N a_{ij} x_j \right)_{i=1}^m \tag{1}$$

  and wish to recover it by solving a *mixed* ℓ_1/ℓ_2 *minimization* which is the natural extension of ℓ_1 minimization to the block vectors.

Setup and results

- I am particularly interested in determining whether one can reduce the number of measurements needed for sparse recovery if one knows that the subspaces are incoherent.

- We collect m linear measurements of a vector $\mathbf{x} \equiv (x_j)_{j=1}^N \in \mathcal{H}$

$$\mathbf{y} = (y_i)_{i=1}^m = \left(\sum_{j=1}^N a_{ij} x_j \right)_{i=1}^m \tag{1}$$

and wish to recover it by solving a *mixed* ℓ_1/ℓ_2 *minimization* which is the natural extension of ℓ_1 minimization to the block vectors.

- The ℓ_1/ℓ_2 norm of a block vector $\mathbf{x}$ is

$$\|\mathbf{x}\|_{2,1} = \sum_i \|x_i\|_2.$$

- Proposed algorithm then is

$$\hat{x} = \text{argmin}_{\mathbf{z}\in\mathcal{H}} \|\mathbf{z}\|_{2,1} \quad s.t. \quad \left(\sum_{j=1}^{N} a_{ij} z_j \right)_{i=1}^{m} = (y_i)_{i=1}^{m}.$$

- Earlier results by Boufounos et. al. show that it is sufficient for recovery to take $m \gtrsim s \ln(N)$ random (Gaussian or Bernoulli) measurements.

- Proposed algorithm then is

$$\hat{x} = \mathrm{argmin}_{\mathbf{z}\in\mathcal{H}} \|\mathbf{z}\|_{2,1} \quad s.t. \quad \left(\sum_{j=1}^{N} a_{ij} z_j\right)_{i=1}^{m} = (y_i)_{i=1}^{m}.$$

- Earlier results by Boufounos et. al. show that it is sufficient for recovery to take $m \gtrsim s \ln(N)$ random (Gaussian or Bernoulli) measurements.

- In my PhD thesis, I showed that $m \gtrsim (1 + \lambda s) \ln(N)$ many measurements suffice with subgaussian matrices.

- The number of measurements decreases with increasing incoherence (smaller λ).

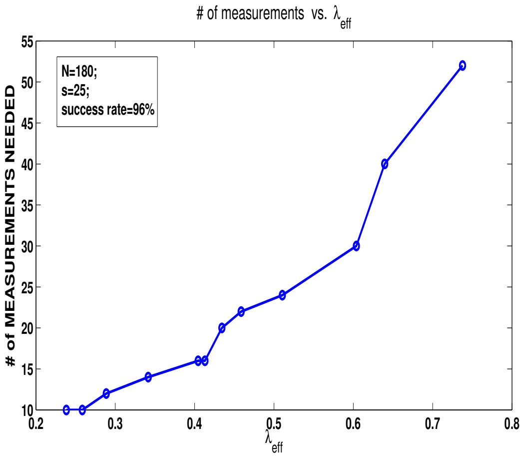

Figure: m vs. λ_{eff}

- For a fixed sparsity s, the relation between λ_{eff} and number of measurements 'm' needed for recovery

Necessary conditions for sparse recovery

- I also gave lower bounds for the necessary number of measurements for any type of linear measurement scheme (not necessarily random) and showed that sufficient conditions are close to being optimal under special conditions.

Necessary conditions for sparse recovery

- I also gave lower bounds for the necessary number of measurements for any type of linear measurement scheme (not necessarily random) and showed that sufficient conditions are close to being optimal under special conditions.

Theorem

(A., Rauhut'13) In order to recover any s-sparse vector in a fusion frame $(W_j)_{j=1}^N$ *in* $\mathbb{R}^d$ *via* $(L1)$ *and any measurement matrix* $A \in \mathbb{R}^{m\times N}$*, one needs at least*

$$m \geq c_1 \frac{s}{d} \ln\left(\frac{N}{c_2 s}\right) + c_3 \frac{ks}{d}, \tag{2}$$

where $\dim(W_j) = k$*.*

Necessary conditions for sparse recovery

- I also gave lower bounds for the necessary number of measurements for any type of linear measurement scheme (not necessarily random) and showed that sufficient conditions are close to being optimal under special conditions.

Theorem

(A., Rauhut'13) In order to recover any s-sparse vector in a fusion frame $(W_j)_{j=1}^N$ *in* $\mathbb{R}^d$ *via* $(L1)$ *and any measurement matrix* $A \in \mathbb{R}^{m\times N}$*, one needs at least*

$$m \geq c_1 \frac{s}{d} \ln\left(\frac{N}{c_2 s}\right) + c_3 \frac{ks}{d}, \tag{2}$$

where $\dim(W_j) = k$.

- Next, we will see how close Condition (2) is to our previous result.
- Observe that λ does not appear in (2) unlike the sufficient condition

$$m \gtrsim (1 + \lambda s) \ln(N).$$

- However, in a fusion frame where $N \geq d/k$, the parameter λ cannot get arbitrarily small.
- The bounds on the packing diameter of equi-dimensional subspaces (of dimension k) in $\mathbb{R}^d$ give that

$$\lambda^2 \geq \frac{kN - d}{dN - d} \sim \frac{k}{d}.$$

- However, in a fusion frame where $N \geq d/k$, the parameter λ cannot get arbitrarily small.
- The bounds on the packing diameter of equi-dimensional subspaces (of dimension k) in $\mathbb{R}^d$ give that

$$\lambda^2 \geq \frac{kN - d}{dN - d} \sim \frac{k}{d}.$$

- When the lower bound holds, the sufficient condition becomes

$$m \gtrsim \sqrt{\frac{k}{d}}\, s \ln(N). \tag{3}$$

- However, in a fusion frame where $N \geq d/k$, the parameter λ cannot get arbitrarily small.

- The bounds on the packing diameter of equi-dimensional subspaces (of dimension k) in $\mathbb{R}^d$ give that

$$\lambda^2 \geq \frac{kN - d}{dN - d} \sim \frac{k}{d}.$$

- When the lower bound holds, the sufficient condition becomes

$$m \gtrsim \sqrt{\frac{k}{d}}\, s \ln(N). \tag{3}$$

- The gap between (3) and the necessary condition

$$m \geq c_1 \frac{s}{d} \ln\left(\frac{N}{c_2 s}\right) + c_3 \frac{ks}{d}$$

is small.

Extensions of CS

- CS falls into the area of estimating low dimensional signals from underdetermined information.
- *Sparse* or *group sparse* signals are examples of such low dimensionality.

Extensions of CS

- CS falls into the area of estimating low dimensional signals from underdetermined information.
- *Sparse* or *group sparse* signals are examples of such low dimensionality.
- In recent years there have been further extensions of CS ideas to *low rank matrix recovery*, *matrix completion*, *1-bit compressed sensing* and *phase retrieval* problems.

Extensions of CS

- CS falls into the area of estimating low dimensional signals from underdetermined information.
- *Sparse* or *group sparse* signals are examples of such low dimensionality.
- In recent years there have been further extensions of CS ideas to *low rank matrix recovery*, *matrix completion*, *1-bit compressed sensing* and *phase retrieval* problems.
- In low rank matrix recovery, rather than recovering a sparse vector $x \in \mathbb{R}^N$, we aim at recovering a matrix $X \in \mathbb{R}^{N_1 \times N_2}$ from incomplete information.
- We take some linear measurements $y = \mathcal{A}(X) \in \mathbb{R}^m$.
- The sparsity assumption is replaced by low rank, i.e., $\mathrm{rank}(X) = r \ll \min\{N_1, N_2\}$.

- In order to recovery X from y, solving

$$\min \operatorname{rank}(Z) \;\; s.t. \;\; \mathcal{A}(Z) = y$$

is not plausible. We replace $\operatorname{rank}(X)$ by the nuclear norm $\|X\|_*$, that is ℓ_1-norm of the singular values of X and solve

$$\min \|Z\|_* \;\; s.t. \;\; \mathcal{A}(Z) = y.$$

- For recovery, random maps $\mathcal{A}$ provide optimal bounds

$$m \geq Cr \min\{N_1, N_2\}.$$

- In order to recovery X from y, solving

$$\min \mathrm{rank}(Z) \;\; s.t. \;\; \mathcal{A}(Z) = y$$

is not plausible. We replace $\mathrm{rank}(X)$ by the nuclear norm $\|X\|_*$, that is ℓ_1-norm of the singular values of X and solve

$$\min \|Z\|_* \;\; s.t. \;\; \mathcal{A}(Z) = y.$$

- For recovery, random maps $\mathcal{A}$ provide optimal bounds

$$m \geq Cr \min\{N_1, N_2\}.$$

- A special case of low rank recovery is *matrix completion* if the measurement map $\mathcal{A}$ samples the entries $\mathcal{A}(X)_\ell = X_{j,k}$ for some indices j, k.

- A popular example is the online movie rating systems and the related Netflix problem.

- In 1-bit CS, one aims to recover a sparse signal $x \in \mathbb{R}^N$ from its sign measurements

$$y = \mathrm{sign}(Ax)$$

  where $A \in \mathbb{R}^{m \times N}$ with $m < N$.

- One can restrict the attention on the signals x on the unit sphere, since the scale information is lost during quantization.

- In 1-bit CS, one aims to recover a sparse signal $x \in \mathbb{R}^N$ from its sign measurements
$$y = \operatorname{sign}(Ax)$$
where $A \in \mathbb{R}^{m \times N}$ with $m < N$.

- One can restrict the attention on the signals x on the unit sphere, since the scale information is lost during quantization.

- Recent works of Boufounos et. al, Plan, Vershynin..

- There are convex programs and random matrices A to recover x with $m \gtrsim s \ln(N)$ measurements.

- *The phase retrieval problem* is about recovering a general signal, an image for example, from the magnitude of its Fourier transform.
- Many applications in X-ray crystallography, diffraction imaging, where devices may not be able to capture the phase of the signal.

- *The phase retrieval problem* is about recovering a general signal, an image for example, from the magnitude of its Fourier transform.
- Many applications in X-ray crystallography, diffraction imaging, where devices may not be able to capture the phase of the signal.
- Recently Candes et. al. have linked this problem to low rank matrix recovery by so called *phase lifting* method and extended ideas from therein.
- In all of these areas and in particularly my research, many ideas from random matrices, convex optimization, geometric functional analysis are commonly used and developed.

Thank you!