

Mathematical Foundations of Data Sciences



Gabriel Peyré
CNRS & DMA
École Normale Supérieure
gabriel.peyre@ens.fr
www.gpeyre.com
www.numerical-tours.com

November 9, 2017

Chapter 15

Compressed Sensing

This chapter details an important class of inverse problems, which corresponds to using “random” forward operators Φ . This is interesting from an applicative point of view since it allows to model a novel class of imaging devices which can potentially have improved resolution with respect to traditional operators (e.g. low-pass filters for usual cameras) when using in conjunction with sparse regularization technics. This is also interesting from a theoretical point of view, since the mathematical analysis becomes much simpler than with deterministic operators, and one can have good recovery and stability performances. Let us however stress that the “physical” creation of hardware that fulfils the theoretical hypothesis, in particular in medical imaging, is still largely open (put aside some restricted areas), although the theory gives many insightful design guides to improve imaging devices.

The main references for this chapter are [29, 23, 38].

15.1 Motivation and Potential Applications

15.1.1 Single Pixel Camera

In order to illustrate the exposition, we will discuss the “single pixel camera” prototype developed at Rice University [?], and which is illustrated by the figure 15.1 (left). It is an important research problem of developing a new class of cameras allowing to obtain both the sampling and the compression of the image. Instead of first sampling very finely (ie with very large Q) the analog signal \tilde{f} to obtain a $f \in \mathbb{R}^Q$ image then compressing enormously (ie with M small) using (??), we would like to dispose directly of an economic representation $y \in \mathbb{R}^P$ of the image, with a budget P as close to M and such that one is able to “decompress” y to obtain a good approximation of the image f_0 .

The “single-pixel” hardware performs the compressed sampling of an observed scene \tilde{f}_0 (the letter “R” in Figure 15.1), which is a continuous function indicating the amount of light $\tilde{f}_0(s)$ reaching each point $s \in \mathbb{R}^2$ of the focal plane of the camera. To do this, the light is focused against a set of Q micro-mirrors aligned on the focal plane. These micro-mirrors are not sensors. Unlike conventional sampling (described in Section ??), they do not record any information, but they can each be positioned to reflect or absorb light. To obtain the complete sampling/compression process, one very quickly changes P times the configurations of the micro-mirrors. For $p = 1, \dots, P$, one sets $\Phi_{p,q} \in \{0, 1\}$, depending on whether the micromirror at position q has been placed in the absorbing (0) or reflective (value 1) position at step p of the acquisition. The total light reflected at step p is then accumulated into a single sensor (hence the name “single pixel”, in fact it is rather a “single sensor”), which achieves a linear sum of the reflected intensities to obtain the recorded $y_p \in \mathbb{R}$ value. In the end, if the light intensity arriving on the surface c_q of the mirror indexed by $f_q = \int_{c_q} \tilde{f}_0(s)ds$ is denoted (as in the ?? section) as q , the equation that links the discrete image $f \in \mathbb{R}^Q$

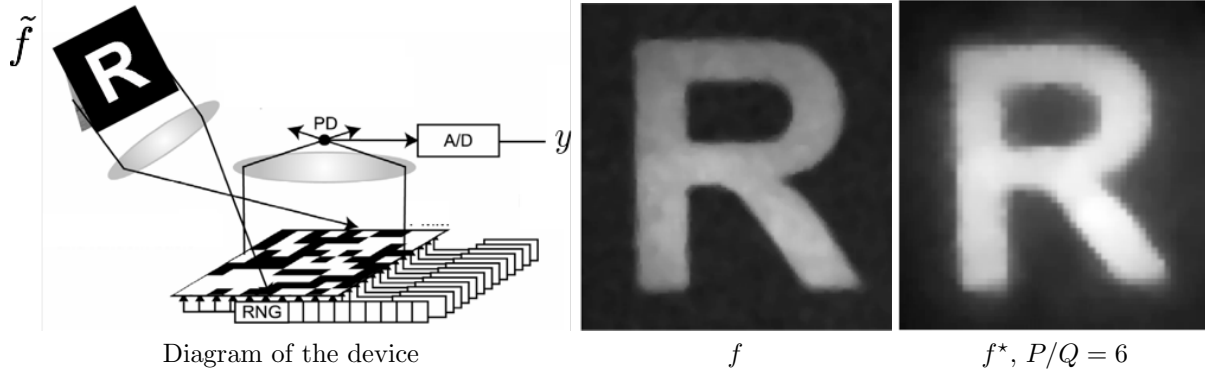


Figure 15.1: Left: diagram of the single-pixel acquisition method. Center: image $f_0 \in \mathbb{R}^Q$ “ideal” observed in the focal plane of the micro-mirrors. Right: image $f_0^* = \Psi x^*$ reconstructed from observation $y \in \mathbb{R}^P$ with a compression factor $P/Q = 6$ using ℓ^1 -type regularization.

“seen through the mirrors” to the P measures $y \in \mathbb{R}^P$ is

$$\forall p = 1, \dots, P, \quad y_p \approx \sum_q \Phi_{p,n} \int_{c_n} \tilde{f}_0(s) ds = (\Phi f_0)_p,$$

(here \approx accounts for some noise), which corresponds to the usual forward model of inverse problems

$$y = \Phi f_0 + w \in \mathbb{R}^P$$

where w is the noise vector. It is important to note that the mirrors do not record anything, so in particular the f_0 discrete image is never calculated or recorded, since the device directly calculates the compressed representation y from the analog signal \tilde{f}_0 . The term w models here the acquisition imperfections (measurement noise). The compressed sampling therefore corresponds to the transition from the observed scene \tilde{f}_0 to the compressed vector y . The “decompression” corresponds to the resolution of an inverse problem, whose goal is to find a good approximation of f_0 (the discrete image “ideal” as seen by the micro-mirrors) from y .

15.1.2 Sparse Recovery

In order to reconstruct an approximation of the (unknown) image f_0 , following Section 11.2, we assume it is sparse in some dictionary Ψ . Denoting $A \stackrel{\text{def.}}{=} \Psi \Phi \in \mathbb{R}^{P \times N}$, this leads us to consider the usual ℓ^1 regularized problem (11.10)

$$x_\lambda \in \operatorname{argmin}_{x \in \mathbb{R}^Q} \frac{1}{2\lambda} \|y - Ax\|^2 + \|x\|_1, \quad (\mathcal{P}_\lambda(y))$$

so that the reconstructed image is $f_\lambda = \Psi x_\lambda$. We also sometimes consider the constraint problem

$$x_\varepsilon \in \operatorname{argmin}_{\|Ax - y\| \leq \varepsilon} \|x\|_1, \quad (\mathcal{P}^\varepsilon(y))$$

where, for the sake of simplicity, we set $\varepsilon = \|w\|$ (which we assume is known). From a mathematical point of view, these problem are equivalent in the sense that there exists a bijection between λ and ε which links it solution. But in practice, this bijection is not explicitly known and depends on y .

Here, it is important to remember that A is drawn from a random matrix ensemble. For an arbitrary Ψ , it is hard to analyze this random distribution. If Ψ is orthogonal, and the distribution of the columns of Φ are invariant by rotation (which is the case if the entries are i.i.d. Gaussian), then A has the same distribution as Φ . In the following, we thus directly models the distribution of A and assumes it has some nice property (typically it is close to being Gaussian i.i.d.).

15.2 Dual Certificate Theory and Non-Uniform Guarantees

15.2.1 Random Projection of Polytopes

When there is no noise, $w = 0$ a way to tackle the problem is to use the characterization of solutions of $(\mathcal{P}^0(Ax_0)) = (\mathcal{P}_0(Ax_0))$ given in Section 14.1.2. According to Proposition 43, identifiable vectors with sparsity $\|x_0\|_0 = s$ corresponds to s -dimensional faces of the ℓ^1 balls B_1 which are mapped to face of the projected polytope AB_1 . This leads to a combinatorial problems to count the number of face of random polytope. Donoho and Tanner were able to perform a sharp analysis of this problem. They showed the existence of two regimes, using two functions C_A, C_M so that, with high probability (i.e. a probability converging exponentially fast to 1 with (n, p)) on the matrix A

- All x_0 so that $\|x_0\|_0 \leq C_A(P/N)P$ are identifiable.
- Most x_0 so that $\|x_0\|_0 \leq C_M(P/N)P$ are identifiable.

For instance, they show that $C_A(1/4) = 0.065$ and $C_M(1/4) = 0.25$. This analysis can be shown to be sharp in high dimension, i.e. when $\|x_0\|_0 > C_M(P/N)P$, then x_0 is not identifiable with high probability (this corresponds to a phase transition phenomena). For large dimensions (N, P) , the scaling given by C_M describe very well what one observe in practice. For $P = N/4$ (compression of a factor 4), one retrieve in practice all vector with sparsity smaller than P/N . The fonction C_M can be computed numerically, and it can be shown to have a logarithmic grows $C_M(r) \sim \log(r)$ for small r . This suggests that for high compression regime, one recovers with ℓ^1 minimization almost all vector with a sparsity $\|x_0\|_0$ proportional (up to log factor) to the number of measurements P .

15.2.2 Gaussian Width

For Gaussian matrices (or at least rotation invariant distributions).

15.2.3 Dual Certificates

In order to analyze recovery performance, one can looks not only for ℓ^2 stability ($\|x_\lambda - x_0\| \sim \|w\|$) but also that x_λ has the same support as x_0 when $\|w\|$ is small. As detailed in Section 14.2.3, this requires to ensure that the pre-certificate η_F defined in (14.24) is non-degenerated, i.e.

$$\|\eta_F\|_\infty \leq 1 \quad \text{where} \quad \eta_F = A^* A_I (A_I^* A_I)^{-1} \text{sign}(x_{0,I}). \quad (15.1)$$

Figure 14.4 suggests that this should be the case if P is large enough with respect to $\|x_0\|$. This theorem backup this observation.

Coherence-based analysis. We first perform a crude analysis using the so-called coherence of the matrix $A = (a_j)_{j=1}^N$ where the $a_j \in \mathbb{R}^P$ are the columns of A , which we assume to be normalized $\|a_j\| = 1$

$$\mu \stackrel{\text{def.}}{=} \max_{0 \leq i \neq j} |\langle a_i, a_j \rangle| = \|A^* A - \text{Id}_N\|_\infty \quad (15.2)$$

where $\|C\|_\infty = \max_{i,j} |C_{i,j}|$. The coherence is 0 for an orthogonal matrix, and is always larger than 1, $\mu \in [0, 1]$. The smaller the coherence, the better conditioned the inverse problem $Ax = y$ is, and the more likely is the certificate η_F to be non-degenerate, as shown by the following proposition.

Proposition 52. *One has, denoting $s = \|x_0\|_0 = |I|$ where $I = \text{supp}(x_0)$, for $\mu < \frac{1}{s-1}$,*

$$\|\eta_{F,I^c}\|_\infty \leq \frac{s\mu}{1 - (s-1)\mu}. \quad (15.3)$$

In particular, if $s < \frac{1}{2} \left(1 + \frac{1}{\mu}\right)$, $\|\eta_{F,I^c}\|_\infty < 1$ and one can thus apply the recovery Theorem 43.

Proof. We recall that the ℓ^∞ operator norm (see Remark 8) is

$$\|B\|_\infty = \max_i \sum_j |B_{i,j}|.$$

We denote $C = A^*A$. One has

$$\|A_{I^c}^* A_I\|_\infty = \max_{j \in I^c} \sum_{i \in I} C_{i,j} \leq s\mu \quad \text{and} \quad \|\text{Id}_s - A_I^* A_I\|_\infty = \max_{j \in I} \sum_{i \in I, i \neq j} C_{i,j} \leq (s-1)\mu$$

One also has

$$\begin{aligned} \|(A_I^* A_I)^{-1}\|_\infty &= \|((\text{Id}_s - A_I^* A_I) - \text{Id}_s)^{-1}\|_\infty = \left\| \sum_{k \geq 0} (\text{Id}_s - A_I^* A_I)^k \right\|_\infty \\ &\leq \sum_{k \geq 0} \|\text{Id}_s - A_I^* A_I\|_\infty^k \leq \sum_{k \geq 0} ((s-1)\mu)^k = \frac{1}{1 - (s-1)\mu} \end{aligned}$$

which is legit because the matrix series indeed converge since $(s-1)\mu < 1$. Using these two bounds, one has

$$\|\eta_{F,I^c}\|_\infty = \|A_{I^c}^* A_I (A_I^* A_I)^{-1} \text{sign}(x_{0,I})\|_\infty \leq \|A_{I^c}^* A_I\|_\infty \|(A_I^* A_I)^{-1}\|_\infty \|\text{sign}(x_{0,I})\|_\infty \leq (s\mu) \times \frac{1}{1 - (s-1)\mu} \times 1.$$

One has

$$\frac{s\mu}{1 - (s-1)\mu} \iff 2s\mu < 1 + \mu$$

which gives the last statement. \square

One can show that one always has

$$\mu \geq \sqrt{\frac{N-P}{P(N-1)}}. \quad (15.4)$$

For Gaussian matrix $A \in \mathbb{R}^{P \times N}$, one has for large $(N, P) \rightarrow +\infty$

$$\mu \sim \sqrt{\log(PN)/P}$$

which shows that Gaussian matrix are close to being optimal for the bound (15.4) if $N \gg P$. Applying Proposition 52 thus shows that ℓ^1 regularization is able to recover with a stable support vector with less than $s \sim O(\sqrt{P})$ (ignoring log terms) non-zero coefficients. In face, we will show now that it does much better and recover a proportional number $s \sim O(P)$. This is because the coherence bound (15.3) is very pessimistic.

Randomized analysis of the Fuchs certificate. We consider here a class of sufficiently “random” matrices.

Definition 22 (sub-Gaussian random matrix). *A random matrix $\sqrt{P} A$ is said to be sub-Gaussian if its entries are independent such that $\mathbb{E}(A_{i,j}) = 0$ (zero mean) $\mathbb{E}(A_{i,j}^2) = 1/P$ and*

$$\mathbb{P}(|\sqrt{P} A_{i,j}| \geq t) \leq \beta e^{-\kappa t^2}.$$

Note that its entries does not needs to be identically distributed, but the sub-Gaussiannity parameter (β, κ) should not depend on (i, j) . Note also the presence of the normalization factor \sqrt{P} , which is here to ensure $\mathbb{E}(\|a_j\|^2) = 1$ where a_j are the columns of A .

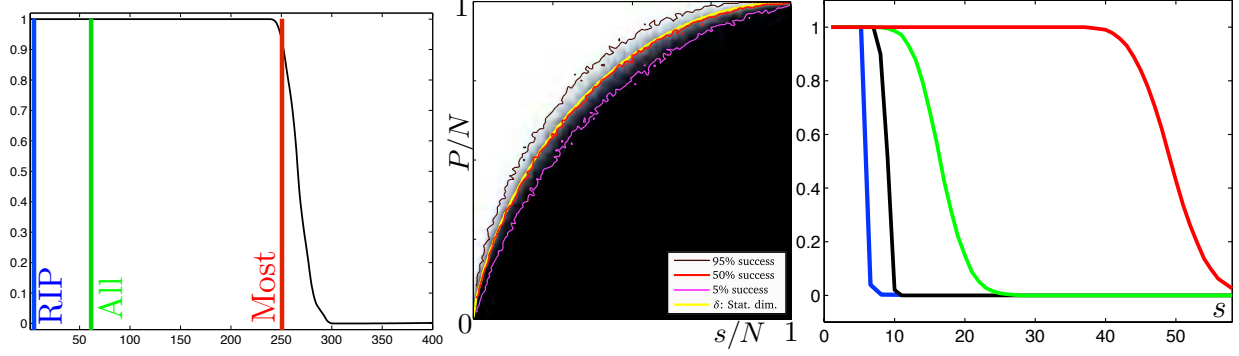


Figure 15.2: Phase transition. For the figure on the right shows probability as function of sparsity that certain criteria hold true, blue: w-ERC, black: ERC, green $|\eta_F| \leq 1$, red: identifiability.

Theorem 44. For a given $x_0 \in \mathbb{R}^N$, denoting $s = \|x_0\|_0$, and assuming A is sub-Gaussian, then provided that

$$P \geq 2s \log(N)$$

condition (15.1) holds with probability

$$1 - ??$$

so that x_λ has the same support and sign as x_0 when $(\|w\|, \|w\|/\lambda)$ is small enough.

Note that the constant c depends only on the sub-Gaussianity parameter (β, κ) .

This is a non-uniform recovery guarantee, in the sense that one first choose a vector x_0 , then draws the matrix A , and the recovery results holds with high probability. In contrast blabla

15.3 RIP Theory for Uniform Guarantees

15.3.1 RIP Constants

The RIP constant δ_s of a matrix $\Phi \in \mathbb{R}^{P \times N}$ is defined as

$$\forall z \in \mathbb{R}^N, \quad \|z\|_0 \leq s \implies (1 - \delta_s)\|z\|^2 \leq \|\Phi z\|^2 \leq (1 + \delta_s)\|z\|^2 \quad (15.5)$$

The following theorem states that for a sub-Gaussian random matrix, these RIP constant grows slowly with s .

Typical example of sub-Gaussian random matrix are Gaussian or Bernoulli matrices. Let us stress that, although this is an abuse of notation, here we assume that A is a random matrix, and not a deterministic one as previously considered.

Theorem 45. If A is a sub-Gaussian random matrix, then it satisfies $\delta_s \leq 1$ provided

$$P \geq C\delta^{-2}s \ln(eN/s) \text{ with probability } 1 - 2e^{-\delta^2 \frac{m}{2C}}. \quad (15.6)$$

15.3.2 RIP implies stable recovery

RIP implies dual certificate guarantees.

In the following, we fix a vector $x \in \mathbb{R}^N$ and denote $y = \Phi x + w$ the measurement, with $\|w\| \leq \varepsilon$. We denote $x_s \in \mathbb{R}^N$ the best s -term approximation of x , obtained by only keeping the s largest coefficients in magnitude from x and setting the others to 0.

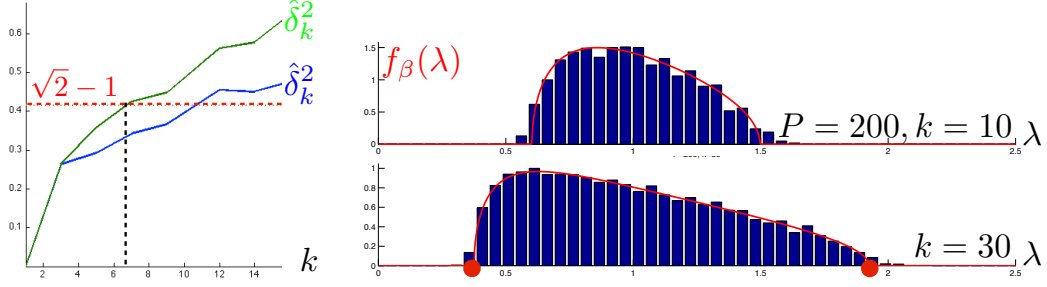


Figure 15.3: Left: evolution of lower bounds $\hat{\delta}_k$ on the RIP constant. Right: empirical distribution of singular values of Gaussian matrices.

We consider a solution x^* of

$$\min_{\|\Phi \tilde{x} - y\| \leq \varepsilon} \|\tilde{x}\|_1.$$

This note recall the proof from [7] of the following theorem

Theorem 46 ([7]). *If $\delta_{2s} \leq \sqrt{2} - 1$ then there exists C_0, C_1 such that*

$$\|x^* - x\|_1 \leq \frac{C_0}{\sqrt{s}} \|x_s - x\| + C_1 \varepsilon.$$

The remaining of this section is devoted to proving this theorem.

Notations. We denote in the following $h = x^* - x$ and denote T_0 the largest s coefficients of x in magnitude (so that $x_s = x_{T_0}$), T_1 the s largest coefficients of $h_{T_0^c}$, T_2 the following s largest coefficients of $h_{T_0^c}$ and so on. We denote $T = T_0 \cup T_1$ which is an index set of size $2s$.

Lemma 5. *One has*

$$\sum_{j \geq 2} \|h_{T_j}\| \leq \frac{1}{\sqrt{s}} \|h_{T_0^c}\|_1$$

Proof. By the definition of T_j for $j \geq 2$, one has, for all $j \geq 2$

$$\forall i \in T_{j-1}, \quad \|h_{T_j}\|_\infty \leq h_i,$$

and hence

$$\|h_{T_j}\|_\infty \leq \frac{1}{s} \|h_{T_{j-1}}\|_1.$$

This proves that

$$\|h_{T_j}\| \leq \sqrt{s} \|h_{T_j}\|_\infty \leq \frac{1}{\sqrt{s}} \|h_{T_{j-1}}\|_1$$

and thus

$$\sum_{j \geq 2} \|h_{T_j}\| \leq \frac{1}{\sqrt{s}} \sum_{j \geq 1} \|h_{T_j}\|_1 = \frac{1}{\sqrt{s}} \|h_{T_0^c}\|_1.$$

□

Lemma 6. *One has*

$$\|h_{T_0^c}\|_1 \leq \|h_{T_0}\|_1 + 2\|x_{T_0^c}\|_1$$

Proof. One has

$$\begin{aligned}
\|x\|_1 &\geq \|x + h\| && \text{because } x^* \text{ is a minimizer} \\
&= \|(x + h)_{T_0}\|_1 + \|(x + h)_{T_0^c}\|_1 \\
&\geq \|x_{T_0}\|_1 - \|h_{T_0}\|_1 + \|h_{T_0^c}\|_1 - \|x_{T_0^c}\|_1 && \text{using the triangular inequality.}
\end{aligned}$$

Decomposing the left hand size $\|x\|_1 = \|x_{T_0}\|_1 + \|x_{T_0^c}\|_1$, one obtains the result. \square

Lemma 7. *If z and z' have disjoint supports and $\|z\| \leq s$ and $\|z'\|_0 \leq s$,*

$$|\langle \Phi z, \Phi z' \rangle| \leq \delta_{2s} \|z\| \|z'\|.$$

Proof. Using the RIP (15.5) since $z \pm z'$ has support of size $2s$ and the fact that $\|z \pm z'\|^2 = \|z\|^2 + \|z'\|^2$, one has

$$(1 - \delta_{2s}) (\|z\|^2 + \|z'\|^2) \leq \|\Phi z \pm \Phi z'\|^2 \leq (1 + \delta_{2s}) (\|z\|^2 + \|z'\|^2).$$

One thus has using the parallelogram equality

$$|\langle \Phi z, \Phi z' \rangle| = \frac{1}{4} \|\Phi z + \Phi z'\|^2 - \|\Phi z - \Phi z'\|^2 \leq \delta_{2s} \|z\| \|z'\|.$$

\square

Theorem 46 requires bounding $\|h\|$. We bound separately $\|h_T\|$ and $\|h_{T^c}\|$.

Part 1: bounding $\|h_T\|$. One has

$$\begin{aligned}
\|h_{T^c}\| &= \left\| \sum_{j \geq 2} h_{T_j} \right\| \leq \sum_{j \geq 2} \|h_{T_j}\| && \text{using the triangular inequality} \\
&\leq \frac{1}{\sqrt{s}} \|h_{T_0^c}\|_1 && \text{using Lemma 5} \\
&\leq \frac{1}{\sqrt{s}} \|h_{T_0}\|_1 + \frac{2}{\sqrt{s}} \|x_{T_0^c}\|_1 && \text{using Lemma 6} \\
&\leq \frac{1}{\sqrt{s}} \|h_{T_0}\|_1 + 2e_0 && \text{denoting } e_0 = \frac{1}{\sqrt{s}} \|x_{T_0^c}\|_1 \\
&\leq \|h_{T_0}\| + 2e_0 && \text{using Cauchy-Schwartz} \\
&\leq \|h_T\| + 2e_0 && \text{because } T_0 \subset T.
\end{aligned}$$

The final bound reads

$$\|h_{T^c}\| \leq \|h_T\| + 2e_0. \quad (15.7)$$

Part 2: bounding $\|h_{T^c}\|$. One has

$$\begin{aligned}
\|h_T\|^2 &\leq \frac{1}{1 - \delta_{2s}} \|\Phi h_T\|^2 && \text{using the RIP (15.5)} \\
&= \frac{A - B}{1 - \delta_{2s}} && \text{using } \Phi h_T = \Phi h - \sum_{j \geq 2} \Phi h_{T_j},
\end{aligned}$$

where we have introduced

$$A = \langle \Phi h_T, \Phi h \rangle \quad \text{and} \quad B = \langle \Phi h_T, \sum_{j \geq 2} \Phi h_{T_j} \rangle.$$

One has

$$\begin{aligned}
|A| &\leq \|\Phi h_T\| \|\Phi h\| && \text{using Cauchy-Schwartz} \\
&\leq \sqrt{1 + \delta_{2s}} \|h_T\| \|\Phi h\| && \text{using the RIP (15.5)} \\
&\leq \sqrt{1 + \delta_{2s}} \|h_T\| 2\varepsilon && \text{using } \|\Phi h\| \leq \|\Phi x - y\| + \|\Phi x^* - y\| \leq 2\varepsilon
\end{aligned}$$

The final bound reads

$$|A| \leq 2\varepsilon \sqrt{1 + \delta_{2s}} \|h_T\| \quad (15.8)$$

One has

$$\begin{aligned}
|B| &\leq |\langle \Phi h_{T_0}, \sum_{j \geq 2} \Phi h_{T_j} \rangle| + |\langle \Phi h_{T_1}, \sum_{j \geq 2} \Phi h_{T_j} \rangle| && \text{using the triangular inequality} \\
&\leq \sum_{j \geq 2} |\langle \Phi h_{T_0}, \Phi h_{T_j} \rangle| + |\langle \Phi h_{T_1}, \Phi h_{T_j} \rangle| && \text{using the triangular inequality} \\
&\leq \sum_{j \geq 2} \delta_{2s} \|h_{T_0}\| \|h_{T_j}\| + \delta_{2s} \|h_{T_1}\| \|h_{T_j}\| && \text{using Lemma 7} \\
&= \delta_{2s} (\|h_{T_0}\| + \|h_{T_1}\|) \sum_{j \geq 2} \|h_{T_j}\| \\
&\leq \delta_{2s} \sqrt{2} \|h_T\| \sum_{j \geq 2} \|h_{T_j}\| && T_0 \text{ and } T_1 \text{ are disjoint} \\
&\leq \frac{\sqrt{2} \delta_{2s}}{\sqrt{s}} \|h_T\| \|h_{T_0^c}\|_1 && \text{using Lemma 5}
\end{aligned}$$

The final bound reads

$$|B| \leq \frac{\sqrt{2} \delta_{2s}}{\sqrt{s}} \|h_T\| \|h_{T_0^c}\|_1. \quad (15.9)$$

Putting together (15.8) and (15.9) one obtains

$$\|h_T\|^2 \leq \frac{\|h_T\|}{1 - \delta_{2s}} \left(\sqrt{1 + \delta_{2s}} 2\varepsilon + \frac{2}{\sqrt{s}} \delta_{2s} \|h_{T_0^c}\|_1 \right)$$

thus

$$\begin{aligned}
\|h_T\| &\leq \alpha\varepsilon + \frac{\rho}{\sqrt{s}} \|h_{T_0^c}\|_1 && \text{denoting } \begin{cases} \alpha = 2 \frac{\sqrt{1 + \delta_{2s}}}{1 - \delta_{2s}}, \\ \rho = \frac{\sqrt{2} \delta_{2s}}{1 - \delta_{2s}}. \end{cases} \\
&\leq \alpha\varepsilon + \frac{\rho}{\sqrt{s}} \|h_{T_0}\|_1 + \frac{2\rho}{\sqrt{s}} \|x_{T_0^c}\|_1 && \text{using Lemma 6} \\
&\leq \alpha\varepsilon + \rho \|h_{T_0}\| + 2\rho e_0 && \text{using Cauchy-Schwartz} \\
&\leq \alpha\varepsilon + \rho \|h_T\| + 2\rho e_0 && \text{because } T_0 \subset T.
\end{aligned}$$

Note that since $\delta_{2s} < \sqrt{2} - 1$, one has $\rho < 1$. This implies

$$\|h_T\| \leq \frac{\alpha}{1 - \rho} \varepsilon + \frac{2\rho}{1 - \rho} e_0. \quad (15.10)$$

Conclusion. One has

$$\begin{aligned}
\|h\| &\leq \|h_T\| + \|h_{T^c}\| && \text{using the triangular inequality} \\
&\leq 2\|h_T\| + 2e_0 && \text{using (15.7)} \\
&\leq \frac{2\alpha}{1 - \rho} \varepsilon + 2 \frac{1 + \rho}{1 - \rho} e_0 && \text{using (15.10)}
\end{aligned}$$

which proves the theorem.

15.3.3 Gaussian Matrices RIP

15.3.4 Fourier sampling RIP

Bibliography

- [1] P. Alliez and C. Gotsman. Recent advances in compression of 3d meshes. In N. A. Dodgson, M. S. Floater, and M. A. Sabin, editors, *Advances in multiresolution for geometric modelling*, pages 3–26. Springer Verlag, 2005.
- [2] P. Alliez, G. Ucelli, C. Gotsman, and M. Attene. Recent advances in remeshing of surfaces. In *AIM@SHAPE report*. 2005.
- [3] Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. SIAM, 2014.
- [4] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
- [5] Stephen Boyd and Lieven Vandenbergh. *Convex optimization*. Cambridge university press, 2004.
- [6] E. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Commun. on Pure and Appl. Math.*, 57(2):219–266, 2004.
- [7] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l’Académie des Sciences, Serie I*(346):589–592, 2006.
- [8] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *SIAM Multiscale Modeling and Simulation*, 5:861–899, 2005.
- [9] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97, 2004.
- [10] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.
- [11] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.
- [12] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.
- [13] F. R. K. Chung. Spectral graph theory. *Regional Conference Series in Mathematics, American Mathematical Society*, 92:1–212, 1997.
- [14] Philippe G Ciarlet. Introduction à l’analyse numérique matricielle et à l’optimisation. 1982.
- [15] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM Multiscale Modeling and Simulation*, 4(4), 2005.

- [16] P. Schroeder et al. D. Zorin. Subdivision surfaces in character animation. In *Course notes at SIGGRAPH 2000*, July 2000.
- [17] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. on Pure and Appl. Math.*, 57:1413–1541, 2004.
- [18] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.*, 4(3):245–267, 1998.
- [19] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, Dec 1994.
- [20] Heinz Werner Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.
- [21] M. Figueiredo and R. Nowak. An EM Algorithm for Wavelet-Based Image Restoration. *IEEE Trans. Image Proc.*, 12(8):906–916, 2003.
- [22] M. S. Floater and K. Hormann. Surface parameterization: a tutorial and survey. In N. A. Dodgson, M. S. Floater, and M. A. Sabin, editors, *Advances in multiresolution for geometric modelling*, pages 157–186. Springer Verlag, 2005.
- [23] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.
- [24] I. Guskov, W. Sweldens, and P. Schröder. Multiresolution signal processing for meshes. In Alyn Rockwood, editor, *Proceedings of the Conference on Computer Graphics (Siggraph99)*, pages 325–334. ACM Press, August8–13 1999.
- [25] A. Khodakovsky, P. Schröder, and W. Sweldens. Progressive geometry compression. In *Proceedings of the Computer Graphics Conference 2000 (SIGGRAPH-00)*, pages 271–278, New York, July 23–28 2000. ACM Press.
- [26] L. Kobbelt. $\sqrt{3}$ subdivision. In Sheila Hoffmeyer, editor, *Proc. of SIGGRAPH’00*, pages 103–112, New York, July 23–28 2000. ACM Press.
- [27] M. Lounsbery, T. D. DeRose, and J. Warren. Multiresolution analysis for surfaces of arbitrary topological type. *ACM Trans. Graph.*, 16(1):34–73, 1997.
- [28] S. Mallat. *A Wavelet Tour of Signal Processing, 3rd edition*. Academic Press, San Diego, 2009.
- [29] Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [30] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. on Pure and Appl. Math.*, 42:577–685, 1989.
- [31] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.
- [32] Gabriel Peyré. *L’algèbre discrète de la transformée de Fourier*. Ellipses, 2004.
- [33] Gabriel Peyré and Marco Cuturi. Computational optimal transport. 2017.
- [34] J. Portilla, V. Strela, M.J. Wainwright, and Simoncelli E.P. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Proc.*, 12(11):1338–1351, November 2003.
- [35] E. Praun and H. Hoppe. Spherical parametrization and remeshing. *ACM Transactions on Graphics*, 22(3):340–349, July 2003.

- [36] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
- [37] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkhäuser, NY*, 2015.
- [38] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, Frank Lenzen, and L Sirovich. *Variational methods in imaging*. Springer, 2009.
- [39] P. Schröder and W. Sweldens. Spherical Wavelets: Efficiently Representing Functions on the Sphere. In *Proc. of SIGGRAPH 95*, pages 161–172, 1995.
- [40] P. Schröder and W. Sweldens. Spherical wavelets: Texture processing. In P. Hanrahan and W. Purgathofer, editors, *Rendering Techniques '95*. Springer Verlag, Wien, New York, August 1995.
- [41] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- [42] A. Sheffer, E. Praun, and K. Rose. Mesh parameterization methods and their applications. *Found. Trends. Comput. Graph. Vis.*, 2(2):105–171, 2006.
- [43] Jean-Luc Starck, Fionn Murtagh, and Jalal Fadili. *Sparse image and signal processing: Wavelets and related geometric multiscale analysis*. Cambridge university press, 2015.
- [44] W. Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computation Harmonic Analysis*, 3(2):186–200, 1996.
- [45] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546, 1997.