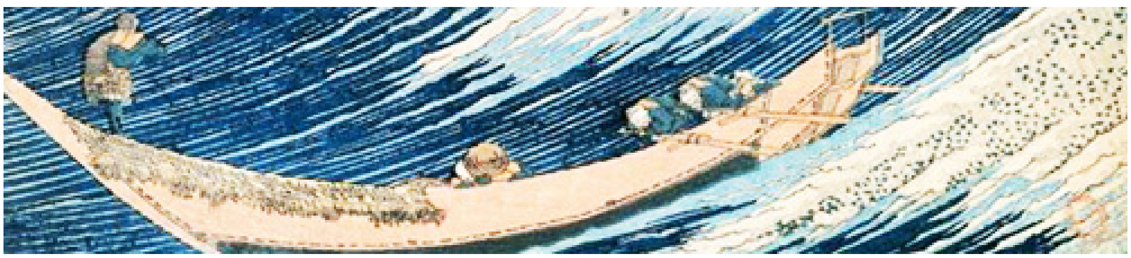
Mathematical Foundations of Data Sciences



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

October 7, 2017

Chapter 3

Compressed Sensing

Main ref: [10, 9, 12]
TODO.

3.1 Motivation and Potential Applications

3.2 Dual Certificate Theory and Non-Uniform Guarantees

Polytopes, descent cone and Gaussian width.

3.3 RIP Theory for Uniform Guarantees

3.3.1 RIP Constants

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

$$\forall\, z \in \mathbb{R}^N, \quad \|z\|_0 \leqslant s \quad \Longrightarrow \quad (1-\delta_s)\|z\|^2 \leqslant \|\Phi z\|^2 \leqslant (1+\delta_s)\|z\|^2 \tag{3.1}$$

3.3.2 RIP implies stable recovery

RIP implies dual certificate guarantees.

In the following, we fixe a vector $x \in \mathbb{R}^N$ and denote $y = \Phi x + w$ the measurement, with $\|w\| \leqslant \varepsilon$. 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.

We consider a solution $x^\star$ of

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

This note recall the proof from [4] of the following theorem

Theorem 1 ([4])**.** *If $\delta_{2s} \leqslant \sqrt{2} - 1$ then there exists C_0, C_1 such that*

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

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

Notations. We denote in the following $h = x^\star - x$ 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 1. *One has*

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

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

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

and hence

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

This proves that

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

and thus

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

□

Lemma 2. *One has*

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

Proof. One has

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

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

Lemma 3. *If z and z' have disjoints supports and $\|z\| \leqslant s$ and $\|z'\|_0 \leqslant s$,*

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

Proof. Using the RIP (3.1) since $z \pm z'$ has support of size $2s$ and the fact that $\|z \pm z'\|^2 = \|z\|^2 + \|z'\|^2$, one has

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

One thus has using the parallelogram equality

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

□

Theorem 1 requires bounding $\|h\|$. We bound separately $\|h_T\|$ and $\|h_{T^c}\|$.

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

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

The final bound reads

$$\|h_{T^c}\| \leqslant \|h_T\| + 2e_0. \tag{3.2}$$

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

$$\begin{aligned}
\|h_T\|^2 \leqslant \frac{1}{1-\delta_{2s}}\|\Phi h_T\|^2 && \text{using the RIP (3.1)}\\
= \frac{A-B}{1-\delta_{2s}} && \text{using } \Phi h_T = \Phi h - \sum_{j\geqslant 2}\Phi h_{T_j},
\end{aligned}$$

where we have introduced

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

One has

$$\begin{aligned}
|A| \leqslant \|\Phi h_T\|\|\Phi h\| && \text{using Cauchy-Schwartz}\\
\leqslant \sqrt{1+\delta_{2s}}\|h_T\|\|\Phi h\| && \text{using the RIP (3.1)}\\
\leqslant \sqrt{1+\delta_{2s}}\|h_T\| 2\varepsilon && \text{using } \|\Phi h\| \leqslant \|\Phi x - y\| + \|\Phi x^\star - y\| \leqslant 2\varepsilon
\end{aligned}$$

The final bound reads

$$|A| \leqslant 2\varepsilon\sqrt{1+\delta_{2s}}\|h_T\| \tag{3.3}$$

One has

$$\begin{aligned}
|B| \leqslant |\langle \Phi h_{T_0},\, \sum_{j\geqslant 2}\Phi h_{T_j}\rangle| + |\langle \Phi h_{T_1},\, \sum_{j\geqslant 2}\Phi h_{T_j}\rangle| && \text{using the triangular inequality}\\
\leqslant \sum_{j\geqslant 2} |\langle \Phi h_{T_0},\, \Phi h_{T_j}\rangle| + |\langle \Phi h_{T_0},\, \Phi h_{T_j}\rangle| && \text{using the triangular inequality}\\
\leqslant \sum_{j\geqslant 2} \delta_{2s}\|h_{T_0}\|\|h_{T_j}\| + \delta_{2s}\|h_{T_1}\|\|h_{T_j}\| && \text{using Lemma 3}\\
= \delta_{2s}(\|h_{T_0}\| + \|h_{T_1}\|)\sum_{j\geqslant 2}\|h_{T_j}\| && \\
\leqslant \delta_{2s}\sqrt{2}\|h_T\|\sum_{j\geqslant 2}\|h_{T_j}\| && T_0 \text{ and } T_1 \text{ are disjoints}\\
\leqslant \frac{\sqrt{2}\ \delta_{2s}}{\sqrt{s}}\|h_T\|\|h_{T_0^c}\|_1 && \text{using Lemma 1}
\end{aligned}$$

The final bound reads

$$|B| \leqslant \frac{\sqrt{2}\ \delta_{2s}}{\sqrt{s}} \|h_T\| \|h_{T_0^c}\|_1. \tag{3.4}$$

Putting together (3.3) and (3.4) one obtains

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

thus

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

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

$$\|h_T\| \leqslant \frac{\alpha}{1-\rho}\varepsilon + \frac{2\rho}{1-\rho} e_0. \tag{3.5}$$

Conclusion. One has

$$\begin{aligned}
\|h\| &\leqslant \|h_T\| + \|h_{T^c}\| && \text{using the triangular inequality} \\
&\leqslant 2\|h_T\| + 2e_0 && \text{using (3.2)} \\
&\leqslant \frac{2\alpha}{1-\rho}\varepsilon + 2\frac{1+\rho}{1-\rho} e_0 && \text{using (3.5)}
\end{aligned}$$

which proves the theorem.

3.3.3 Gaussian Matrices RIP

3.3.4 Fourier sampling RIP

Bibliography

[1] Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. SIAM, 2014.

[2] 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.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[4] 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.

[5] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.*, 20:89–97, 2004.

[6] 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.

[7] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319, 2016.

[8] Philippe G Ciarlet. Introduction à l'analyse numérique matricielle et à l'optimisation. 1982.

[9] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Birkhäuser Basel, 2013.

[10] Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[11] Neal Parikh, Stephen Boyd, et al. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239, 2014.

[12] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, Frank Lenzen, and L Sirovich. *Variational methods in imaging*. Springer, 2009.