Compressive Sensing and Applications

Myungjoo Kang and Myeongmin Kang

1. Compressive Sensing

Consider a one-dimensional, finite-length signal $x \in \mathbb{C}^N$. We will vectorise a two-dimensional image or higher-dimensional data into a long one-dimensional vector. Many real-world signals can be well-approximated by sparse or compressible under a suitable basis. Let $\Psi = [\psi_1|\psi_2|\cdots|\psi_N]$ be an orthonormal basis. Then a signal x can be expressed as $x = \sum_{n=1}^{N} \langle x, \psi_n\rangle \psi_n$. We say that x is k-sparse under Ψ if $\{f_n = \langle x, \psi_n\rangle\}_{n=1,\dots,N}$ has only k-nonzero coefficients, and that x is compressible under Ψ if $\{\langle x, \psi_n\rangle\}_{n=1,\dots,N}$ has a few large coefficients.

Compressive sensing is that a sparse signal can be recovered from what was previously believed to be incomplete information. Consider $\Phi = [\phi_1|\phi_2|\cdots|\phi_N] \in \mathbb{C}^{M\times N}$ for some $M < N$. Then, we can obtains $b = \Phi x = \Phi\Psi f = Af$ where $A = \Phi\Psi$. The measurements Φ is fixed and does not depend on the signal x and then A is selected independent of f. A is referred to as the encoder and obviously encoder is linear. In the encoder, we need to design a good sensing matrix A. The decoder is the attempted recovery of f from its sensing matrix A and b. We define $\|f\|_0 := |\text{supp } f|$ for a signal f. The quantity $\|\cdot\|_0$ is often called ℓ_0-norm although it is actually not a norm. With a sparsity prior, a natural decoder is to search for the sparsest vector f that $b = Af$:

$$\min \|f\|_0 \text{ subject to } b = Af. \tag{1}$$

We need to check that if the problem (1) has a solution, the solution is unique. For given matrix A, spark(A) is the smallest number of columns that are linearly dependent. Using this concept, we get a condition of uniqueness. Let x_0 be a k-sparse N-dimensional vector, let A be a matrix of $M \times N$, and let $y = Ax_0$. If $k < \frac{\text{spark}(A)}{2}$, then x_0 is a unique solution of problem (1). Conversely, if $k \geq \frac{\text{spark}(A)}{2}$, then (1) does not have x_0 as its unique solution. Since the decoder is well-defined for small k, we need an efficient reconstruction algorithm. Unfortunately, the problem (1) is combinatorial problem and NP-hard in general. Essentially two approaches have mainly been pursued: greedy algorithm and convex relaxation. We will introduce greedy algorithms and convex relaxation for solving (1).

2. Greedy Algorithm

A greedy algorithm computes the support of signal iteratively, at each step finding one or more new elements and subtracting their contribution from the measurement vector. Examples include Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), stagewise OMP, regularised OMP, weak OMP. We introduce MP and OMP here.

In 1993, Matching Pursuit is proposed by S Mallat and Z Zhang [17]. Matching Pursuit is an algorithm that decomposes any signal into a linear expansion of atoms that are selected from a redundant dictionary. Let a_i be i-th column of A and x_i be i-th component of x. Assume that $\|a_i\|_2 = 1$ for all i. Equation $Ax = b$ is equivalent to $b = x_1a_1 + \cdots + x_Na_N$. We want to compute a linear expansion of b over a set $\{a_i : i = 1, 2, \dots, N\}$ and their coefficients are sparse. The idea of Matching Pursuit is choosing column of A, in order to best match its inner product structures. For a signal f, f can be decomposed to

$$f = \langle f, a_i\rangle a_i + Rf,$$

where Rf is the residual vector. Clearly, a_i is orthogonal to Rf. So,

$$\|f\|_2^2 = |\langle f, a_i\rangle|^2 + \|Rf\|_2^2$$

by Pythagoras theorem. We have to choose a_i such that $|\langle f, a_i\rangle|$ is maximum in order to minimise $\|Rf\|$. Using this idea, we iteratively choose the column of A that has highest absolute inner product with current residual vector $r_i = b - Ax_i$ and inner product of selected column a_j is added to coefficient x_j.

Orthogonal Matching Pursuit [6, 20] is improved Matching Pursuit by orthogonalising the direct projection with a Gram–Schmidt procedure. OMP algorithm iteratively selects the column of A in the same way like MP. The difference is that b (not r_i) is matched by orthogonal projection with columns of A selected in this step and in previous steps. Let Λ_k be columns of A chosen until k steps. Orthogonal projection of b with Λ_k is $A_{\Lambda_k}(A^*_{\Lambda_k}A_{\Lambda_k})^{-1}A^*_{\Lambda_k}b$ and x_k is $(A^*_{\Lambda_k}A_{\Lambda_k})^{-1}A^*_{\Lambda_k}b$, where A_{Λ_k} is the column sub-matrix of A corresponding to Λ_k. Clearly, r_{k+1} is orthogonal to Λ_k. Thus, the resulting Orthogonal Matching Pursuit converges with a finite number of iterations less than rank A.

In 2003, it was proved that assuming that $\|x\|_0 < 0.5(1 + \frac{1}{\mathrm{spark}(A)})$, OMP (and MP) are guaranteed to find the sparsest solution in [8]. In 2007, J Tropp and A Gilbert proved in [9] that assuming that A is Gaussian, for $\delta \in (0, 0.36)$ and $M \geq Ck\ln(N/\delta)$, OMP can reconstruct the sparse signal with probability exceeding $1 - 2\delta$. Similar result holds when A is Bernoulli and $M \geq Ck^2\ln(N/\delta)$.

MP and OMP are fast and easy to implement. But they do not work when there are noisy measurements. They work for Gaussian and Bernoulli measurement matrices but it is not known whether they succeed in the important class of partial Fourier measurement matrices.

3. ℓ_1 Relaxation

The ℓ_1 minimisation approach considers the solution of

$$\min \|f\|_1 \text{ subject to } b = Af. \tag{2}$$

This is a convex optimisation problem and can be seen as a convex relaxation of (1). In the real-valued case, (2) is casted by a linear program and in the complex-valued case, it is casted by a second order cone program. Of course, we hope that the solution of (2) coincides with the solution of (1). Here, we provide an intuitive explanation to expect that the use of (2) will indeed promote sparsity. Suppose dimension of signal f is 2 and dimension of measurement vector b is 1. Except for situations where $\ker A$ is parallel to one of faces of the poly type $\{x : \|x\|_1 = 1\}$, there is a unique solution of (2), which is sparse solution. Of course, for $p < 1$, when the regulariser of (1) is changed by $\|f\|_p$, there is also a unique sparse solution. Since $\|\cdot\|_p$ is neither norm nor convex for $p < 1$, that problem is hard to solve.

The use of ℓ_1 minimisation appears already in the PhD thesis of B Logan in connection with sparse frequency estimation, where he observed that ℓ_1 minimisation may recover exactly a frequency sparse signal from undersampled data provided the sparsity is small. Donoho and B Logan provide the earliest theoretical work on sparse recovery using ℓ_1 minimisation. It is found in 1990 that the idea to recover sparse Fourier spectra from undersampled non-equispace samples. In statics, use of ℓ_1 minimisation and related methods was popularised with the work (LASSO) of Tibshirani. In image processing, the use of total variation minimisation, which is connected to ℓ_1 minimisation, appears in the work of Rudin, Osher and Fatemi.

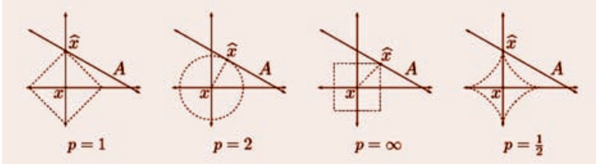

Fig. 1. The solution of ℓ_p (quasi-)norm minimisation by one dimensional subspace for $p = 1, 2, \infty$ and $\frac{1}{2}$.

Many people provided the condition to recover sparse solution by ℓ_1 minimisation adopting various contents. D Donoho and X Hou provided that condition using the content mutual coherence. Mutual coherence of A assuming that the columns of A are normalised is given by

$$\mu(A) = \max_{1\leq i<j\leq n} |\langle a_i, a_j\rangle|,$$

where a_i is i-th column of A. They proved in [1] that assuming that $f \in \mathbb{R}^n$ is k-sparse vector such that $Af = b$ and $k < \frac{1}{2}(1 + \frac{1}{\mu(A)})$, then f is the unique solution of (2).

We present analysis of ℓ_1 minimisation adopting the concept null space property. A matrix A is said to satisfy the Null Space Property (NSP) of order k with $\gamma \in (0, 1)$ if $\|\eta_T\|_1 \leq \gamma\|\eta_{T^c}\|_1$ for all set $T \subset \{1, 2, \ldots, N\}$ with cardinality of $T \leq k$, and for all $\eta \in \ker A$. The following sparse recovery result of ℓ_1 minimisation is based on this concept.

Let $A \in \mathbb{C}^{M\times N}$ be a matrix and $f \in \mathbb{C}^N$ and $b = Af$, f^* be a k-sparse solution of (1). A satisfies the null space property of order k if and only if f^* is the unique solution of (2).

The NSP is actually equivalent to sparse recovery using ℓ_1. This fact seems to have first appeared explicitly in [14]. The term null space property was coined by A Cohen, W Dahmen, and R DeVore. But, the NSP is somewhat difficult to handle directly. In 2005, E Candés and T Tao proposed the concept restricted isometry property. The restricted isometry constant δ_k of a matrix A is the smallest number satisfying

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

for all k-sparse vector z. A matrix A is said to satisfy the Restricted Isometry Property (RIP) of order k with constant δ_k if $\delta_k \in (0,1)$. In contrast to the NSP, the RIP is not necessary condition for sparse recovery by ℓ_1. Many people proposed a sufficient condition of exact sparse ℓ_1-recovery using the concept RIP. E Candés, M Rudelson, T Tao and R Vershynin first note the following fact in [3].

Assume A satisfies the RIP of order $3k$ and order $4k$ with $\delta_{3k}+3\delta_{4k} < 2$. Let $f \in \mathbb{C}^N$ and $b = Af$, and f^* be a k-sparse solution of (1). Then, f^* is the unique solution of (2).

E Candés provided in [2] that sparse recovery using ℓ_1 is guaranteed as $\delta_{2k} < \sqrt{2}-1$. The sufficient condition was improved to $\delta_{2k} < \frac{2}{\sqrt{2}+3}$ in [16]. In 2010, S Foucart proved in [15] that every sparse vector can be recovered by ℓ_1 if $\delta_{2k} < \frac{3}{4+\sqrt{6}}$.

E Candés and T Tao also proposed another sufficient condition on the RIP adopting the concept restricted orthogonality constants in [4]. The k, k'-restricted orthogonality constants $\theta_{k,k'}$ of a matrix A defines the smallest number such that

$$|\langle Ax, Ay\rangle| \le \theta_{k,k'}\|x\|_2\|y\|_2$$

holds for all k-sparse vector x and k'-sparse vector y with disjoint supports. They gave the sufficient condition $\delta_k + \theta_{k,k} + \theta_{k,2k} < 1$ on the RIP. This condition was later improved to $\delta_{1.5k} + \theta_{k,1.5k} < 1$ in [18].

Many people deal with RIP of Gaussian matrix, Bernoulli matrix and partial Fourier matrix. Gaussian matrix is that the entries of it are chosen as i.i.d. (independent and identically distributed) Gaussian random variables with expectation 0 and variance $\frac{1}{M}$. Similarly, Bernoulli matrix is that the entry of it takes the value $\frac{1}{\sqrt{M}}$ or $-\frac{1}{\sqrt{M}}$ with equal probability $\frac{1}{2}$. Partial Fourier matrix is submatrix of discrete Fourier transform matrix consisting of random rows. R Baraniuk, M Davenport, R DeVore, and M Wakin proved the following statement.

Let $A \in \mathbb{R}^{M\times N}$ be a Gaussian or Bernoulli matrix. For given $0 < \delta < 1$, there exist constants C, C_1 depending only on δ such that RIP δ_k of A less than δ with probability exceeding $1 - e^{-C_1 m}$ provided $M \ge Ck\ln(\frac{N}{k})$.

Therefore, k-sparse vector can be recovered using ℓ_1 minimisation for Gaussian or Bernoulli matrix with overwhelming probability if $M \ge Ck\ln(\frac{N}{k})$ for some universal constant C.

E Candés and T Tao proposed that partial Fourier matrix satisfies the RIP of order $3k$ and order $4k$ with $\delta_{3k}+3\delta_{4k} < 2$ with probability at least $1-N^{-ct}$ if $M \ge Ctk\ln^6 N$ for some $t > 1$. M Rudelson and R Vershynin improved that condition about M is $M \ge Ctk\ln N \ln(Ctk\ln N)(\ln k)^2$ for some $N, t > 1, k > 2$. Thus, if A is a partial Fourier matrix and M satisfies the preceeding condition, the problem (1) is equivalent to its convex relaxation (2) for all k-sparse signal with high probability.

4. Application

Compressive sensing can be used in all applications where the task is the reconstruction of a signal or an image from linear measurement. There should be reason to believe that the signal is sparse in a suitable basis. At first, we consider image restoration and image inpainting. We consider $y = Hu + \epsilon$ where y is the observed image, u is the original image, ϵ is the noise, H is the degrading operator (e.g. convolution with some kernel). The image restoration is the process to recover original image u using y and H. Image inpainting is the process of recover missing pixels of given image. For given image x, let Λ be the index set of all available data. Since the data for the indices in Λ^c is not believed, we can only use data $P_\Lambda x$ for the indices in Λ, where P_Λ, is called "row selctor", is a matrix which comprises a subset of the rows for the indices in Λ of an identity matrix. Mixing image restoration problem and image inpainting problem, we want to seek original image u using the data $P_\Lambda y = P_\Lambda(Hu+\epsilon)$. Actually, P_Λ is $M\times N$ matrix for some $M < N$. The sparsity prior of images in tight frame has been used in many image restoration and inpainting problem. Assume that ϵ is 0. We set the problem following

$$\min_u \|Wu\|_0 \text{ subject to } P_\Lambda Hu = P_\Lambda y, \tag{3}$$

where W is an tight frame. Problem (3) is casted by unconstrained problem

$$\min_u \|Wu\|_0 + \frac{1}{2}\|P_\Lambda Hu - P_\Lambda y\|_2^2.$$

H Ji, Z Shen and Y Xu get good results solving convex relaxation of this problem. Figure 2 is a result by H Ji, Z Shen and Y Xu.

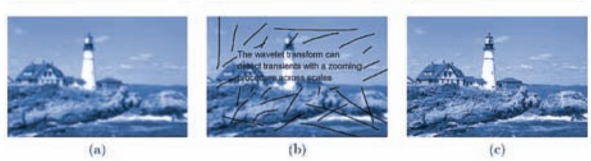

Fig. 2. (a) Blurred image by out of focus kernel, (b) Blurred and scratched image, (c) Reconstructed image.

Second application is Magnetic Resonance Imaging (MRI). MRI is a medical imaging technique used in radiology to visualise detailed internal structures. In MRI, samples are collected directly in Fourier frequency domain (k-space) of object. The scan time in MRI is proportional to the number of Fourier coefficients. Using compressive sensing technique, we can reduce the number of samples and scan time. Real MR images are known to be sparse in discrete cosine transform (DCT) and wavelet transform. We write this problem in the form,

$$\min_f \|f\|_0 \text{ subject to } R\mathcal{F}Wf = y,$$

where $\mathcal{F}$ is Fourier transform matrix, R is random row selector, W is a DCT matrix or wavelet transform matrix, $u = Wf$ is reconstruction image. Several people have also observed that it is often useful to include Total Variation $|\nabla \cdot| = \sum \sqrt{|\nabla_{x_1} \cdot|^2 + |\nabla_{x_2} \cdot|^2}$. Using these facts, T Goldstein and S Osher solve the problem,

$$\min_u \|Wu\|_1 + |\nabla u| \text{ subject to } R\mathcal{F}f = y, \tag{4}$$

where W is a haar wavelet transform matrix. Figure 3 is a result solving the problem (4).

Further applications include analogue to digital conversion, single-pixel imaging, data compression, astronomical signal, geophysical data analysis and compressive radar imaging. The point of compressive sensing is that even though the amount of data is very small, we can have most of the information contained in the object. Thus, compressive sensing has many potential applications in various fields.

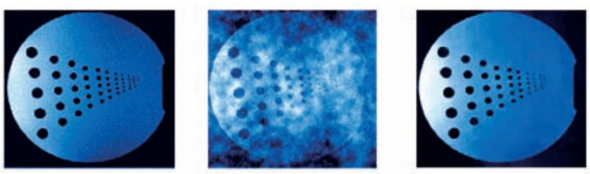

Fig. 3. Left: original image, middle: linear reconstruction using 30% of the k-space data, right: compressive sensing reconstruction using same data of middle.

References

[1] D. Donoho and X. Huo, Uncertainty principles and ideal atomic decomposition, *IEEE Trans. Inform. Theory* **47** (2001).

[2] E. Candés, The restricted isometry property and its implications for compressive sensing, *C. R. Acad. Sci. Paris S'er. I Math.* **346** (2008).

[3] E. Candés, M. Rudelson, T. Tao and R. Vershynin, Error correction via linear programming, *Proc. 46th Annual IEEE Symposium on Foundations of Computer Science, IEEE* (2005).

[4] E. Candés and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory* **51** (2005).

[5] E. Candés and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Trans. Inform. Theory* **52** (2006).

[6] G. Davis, S. Mallat and Z. Zhang, Adaptive time-frequency decompositions, *SPIE Journal of Optical Engineering* **33** (1994).

[7] H. Ji, Z. Shen and Y. Xu, Wavelet frame based image restoration with missing/damaged pixels, *East Asia Journal on Applied Mathematics* **1** (2011).

[8] J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory* **50** (2004).

[9] J. Tropp and A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory* **53** (2007).

[10] M. Rudelson and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, *Comm. Pure Appl. Math.* **61** (2008).

[11] M. Lustig, D. Donoho and J. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magn. Reson. Med.* **58** (2007).

[12] R. Baraniuk, Compressive sensing, *IEEE Signal Processing Magazine* **24** (2007).

[13] R. Baraniuk, M. Davenport, R. DeVore and M. Wakin, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* **28** (2008).

[14] R. Gribonval and M. Nielsen, Sparse representations in unions of bases, *IEEE Trans. Inform. Theory* **49** (2003).

[15] S. Foucart A note on guaranteed sparse recovery via ℓ_1-minimization, *Appl. Comput. Harmon. Anal.* **29** (2010).

[16] S. Foucart and M. Lai, Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$, *Appl. Comput. Harmon. Anal.* **26** (2009).

[17] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process* (1993).

[18] T. Cai, G. Xu and J. Zhang, On recovery of sparse signals via ℓ_1 minimization, *IEEE Trans. Inform. Theory* **55** (2009).

[19] T. Goldstein and S. Osher, The split Bregman method for L1-Regularized problems, *SIAM J. Imaging. Sci.* **2** (2009).

[20] Y. Pati, R. Rezaiifar and P. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in *Proc. 27th Asilomar Conference on Signals, Systems and Computers*, ed. A. Singh, (IEEE Computer Society Press, Los Alamitos, CA, USA, 1993).

[21] M. Davenport, M. Duarte, Y. Eldar and G. Kutyniok, Introduction to compressed sensing, Chapter in *Compressed Sensing: Theory and Applications* (Cambridge University Press, 2011).

[22] M. Fornasier and H. Rauhut, Compressive sensing, Chapter in Part 2 of the *Handbook of Mathematical Methods in Imaging* (Springer, 2011).

*Translated from the Newsletter of Korean Mathematical Society **138** (2011) 2–7*

Myungjoo Kang

Seoul National University, Korea

Myungjoo Kang received the BS degree in mathematics from Seoul National University, Seoul, Korea, and the PhD degree in mathematics from the University of California, Los Angeles, in 1996. He was with the Electrical and Computer Engineering Department from the University of California, San Diego, as a Postdoctoral Researcher, from 1996–2000. He has been an Assistant Professor at the Department of Mathematical Sciences, Seoul National University, from 2003–2007, and an Associate Professor, from 2008–present. His research interests are in mathematical image processing as well as numerical schemes and computational fluid dynamics.

Myeongmin Kang

Seoul National University, Korea

Myeongmin Kang received the BS degree in mathematics and computer science from Seoul National University, Seoul, Korea, in 2008, and she has been an integrated M/PhD student at the Department of Mathematical Sciences, Seoul National University, from 2008–present. Her research interests are in mathematical image processing such as compressive sensing.