INFONET, GIST

Journal Club (2013. 01. 15)

Probabilistic Reconstruction in Compressed Sensing : Algorithms, Phase Diagrams, and Threshold Achieving Matrices

Authors:	Florent Krzakala, Marc Mezard, Francois Sausset, Yifan Sun and Lenka Zdeborova
Publication:	ArXiv:1206.3953v1 [Cond-mat.stat-mech] 18 Jun 2012
Speaker:	Hyeong-ho, Baek

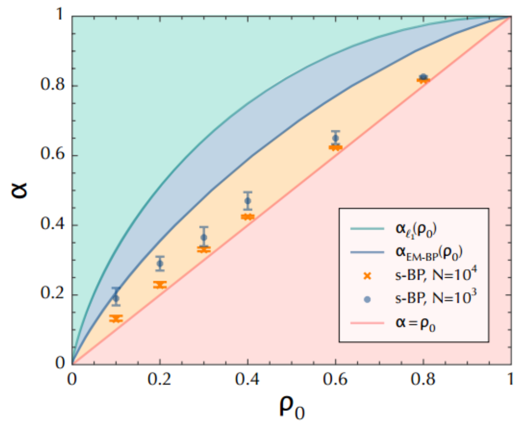
Short summary: In this paper, they present the probabilistic approach to reconstruction and discuss its optimality and robustness. And they detail the derivation of the message passing algorithm for reconstruction. Moreover, they further develop the asymptotic analysis of the corresponding phase diagrams with and without measurement noise, for different distribution of signals.

I. Introduction

The CS problem can be posed as follows: given an $N-$component signal $\mathbf{s}$, one makes M measurements that are grouped into an $M-$component vector $\mathbf{y}$, obtained from $\mathbf{s}$ by a linear transformation using $M \times N$ matrix $\mathbf{F}$, given by $y_\mu = \sum_{i=1}^{N} F_{\mu i} s_i$ with $\mu = 1, 2, ..., M$ .The aim is to reconstruct the signal $\mathbf{s}$ from the knowledge of $\mathbf{F}$ and $\mathbf{y}$. This amounts to inverting the linear system $\mathbf{y} = \mathbf{Fs}$. However, we want to have M as small as possible and when $M < N$ there are fewer equations than unknowns. The system is under-determined and the inverse problem is ill-defined. However, CS deals with sparse signals. In the noiseless case, an exact reconstruction case, an exact reconstruction is possible for such signals as soon as $M > K$.

Candes, Tao, Donoho and collaborators proposed to find the vector satisfying the constraints $\mathbf{y} = \mathbf{Fx}$ which has the smallest l_1 norm. This optimization problem is convex and can be solved using efficient linear programming techniques. For any signal with density $\rho = K / N$, the l_1 reconstruction gives indeed the exact result $\mathbf{x} = \mathbf{s}$ with probability one only if

$\alpha = M / N > \alpha_{l_1}(\rho)$ is, however, larger than ρ. The l_1 reconstruction is thus sub-optimal: it requires more measurements than theoretically necessary.



II. PROBABILISTIC RECONSTRUCTION IN COMPRESSED SENSING

The definition of the compressed sensing problem is as follows

$$y_\mu = \sum_{i=1}^{N} F_{ui} s_i + \xi_u \quad \mu = 1, \ldots, M \,, \tag{1}$$

Where s_i are the signal elements, out of which only K are non-zero. F_{ui} are the elements of a known measurement matrix, y_μ are the known result of measurements, and ξ_u is Gaussian white noise on the measurement with variance Δ_μ. The goal of CS is to find an approach that allows reconstruction with as low values of α as possible.

We shall adopt a probabilistic inference approach to reconstruct the signal.

$$\widehat{P}(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^{N} [(1-\rho)\delta(x_i) + \rho\phi(x_i)] \prod_{\mu=1}^{M} \frac{1}{\sqrt{2\pi\Delta_\mu}} e^{-\frac{1}{2\Delta_\mu}(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2} \,, \tag{2}$$

Where Z, the partition function, is a normalization constant. Here we model the signal as stochastic with iid entries, the fraction of non-zero entries being $\rho > 0$ and their distribution being ϕ.

We stress that in general the signal properties are not known and hence we do not assume that the signal model matches the empirical signal distribution, $\rho = \rho_0, \Delta = \Delta_0, \phi = \phi_0$. One crucial point in our approach is using $\rho < 1$ which includes the fact that on searches a sparse signal in the model of the signal.

A. The Bayesian optimality and the Nishimori condition

The probabilistic approach can also be recovered from a Bayesian point of view. Indeed, given $\mathbf{F}$ and $\mathbf{y}$, from Bayes theorem, we have

$$P(\mathbf{x} \,|\, \mathbf{F}, \mathbf{y}) = \frac{P(\mathbf{x} \,|\, \mathbf{F}) P(\mathbf{y} \,|\, \mathbf{F}, \mathbf{x})}{P(\mathbf{y} \,|\, \mathbf{F})} \tag{3}$$

The value of measurements $\mathbf{y}$ given the knowledge of the matrix $\mathbf{F}$ and the signal $\mathbf{x}$ is, by definition of the problem, given by $P(\mathbf{y} \,|\, \mathbf{F}, \mathbf{x}) = \prod_{\mu=1}^{M} \delta(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)$ in the noiseless case, and by

$$P(y \,|\, F, x) = \prod_{\mu=1}^{M} \frac{1}{\sqrt{2\pi\Delta_\mu}} e^{-\frac{1}{2\Delta_\mu}(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2} \tag{4}$$

With random Gaussian measurement noise of variance Δ_μ, for measurement μ. To express the probability $P(\mathbf{x} \,|\, \mathbf{F})$ we consider that the signal dose not depend on the measurement matrix. And we model the signal as an iid:

$$P(x \,|\, F) = \prod_{i=1}^{N} \left[\left(1 - \rho\right) \delta(x_i) + \rho \phi(x_i) \right] \tag{5}$$

Thus the posterior probability of $\mathbf{x}$ after the measurement of $\mathbf{y}$ is given by

$$P(\mathbf{x}\,|\,\mathbf{F},\mathbf{y}) = \frac{1}{Z(\mathbf{y},\mathbf{F})}\prod_{i=1}^{N}\left[(1-\rho)\delta(x_i)+\rho\phi(x_i)\right]\prod_{\mu=1}^{M}\frac{1}{\sqrt{2\pi\Delta_\mu}}e^{-\frac{1}{2\Delta_\mu}(y_\mu-\sum_{i=1}^{N}F_{\mu i}x_i)^2}, \qquad (6)$$

Where $Z(\mathbf{y},\mathbf{F}) = P(\mathbf{y}\,|\,\mathbf{F})$ is again the normalization constant.

An estimator $\mathbf{x}^*$ that minimizes mean-squared error with respect to the original signal $\mathbf{s}$, defined as $E=\sum_{i=1}^{N}(x_i-s_i)^2/N$, is then obtained from averages of x_i with respect to the probability measure $P(\mathbf{x}\,|\,\mathbf{F},\mathbf{y})$, i.e.,

$$x_i^* = \int dx_i x_i v_i(x_i), \qquad (7)$$

Where $v_i(x_i)$ is the marginal probability distribution of the variable i

$$v_i(x_i) \equiv \int_{\{x_j\}_{j\neq i}} P(\mathbf{x}\,|\,\mathbf{F},\mathbf{y}). \qquad (8)$$

III. THE BELIEF PROPAGATION RECONSTRUCTION ALGORITHM FOR COMPRESSED SENSING

Exact computation of the averages $x_i^* = \int dx_i x_i v_i(x_i)$ requires exponential time and is thus intractable. To approximate the expectations we will use a variant of the belief propagation(BP) algorithm. Indeed, message passing has been shown very efficient in terms of both precision and speed for the CS problem.

A. Belief Propagation recursion

The canonical BP equation for the probability measure $P(\mathbf{x}\,|\,\mathbf{F},\mathbf{y})$ are expressed in terms of $2MN$ "messages", $m_{\mu\to j}(x_j)$ and $m_{j\to\mu}(x_j)$, which are probability distribution functions.

$$m_{\mu\to i}(x_i)=\frac{1}{Z^{\mu\to i}}\int\prod_{j\neq i}dx_j e^{-\frac{1}{2\Delta_\mu}\left(\sum_{j\neq i}F_{\mu j}x_j+F_{\mu i}x_i-y_\mu\right)^2}\prod_{j\neq i}m_{j\to\mu}(x_j), \tag{9}$$

$$m_{i\to\mu}(x_i)=\frac{1}{Z^{i\to\mu}}\left[(1-\rho\delta(x_i)+\rho\phi(x_i)\right]\prod_{\gamma\neq\mu}m_{\gamma\to i}(x_i)\,, \tag{10}$$

Where $Z^{\mu\to i}$ and $Z^{i\to\mu}$ are normalization factors ensuring that $\int dx_i m_{\mu\to i}(x_i)=1$, $\int dx_i m_{i\to\mu}(x_i)=1$.

Using the Hubbard-Stratonovich transformation

$$e^{-\frac{w^2}{2\Delta}}=\frac{1}{\sqrt{2\pi\Delta}}\int d\lambda e^{-\frac{\lambda^2}{2\Delta}+\frac{i\lambda w}{\Delta}} \tag{11}$$

For $w=\left(\sum_{j\neq i}F_{\mu j}x_j\right)$ we can simplify Eq.(9) as

$$m_{\mu\to i}(x_i)=\frac{1}{Z^{\mu\to i}\sqrt{2\pi\Delta}}e^{-\frac{1}{2\Delta_\mu}(F_{\mu i}x_i-y_\mu)^2}\int d\lambda e^{-\frac{1}{2\Delta_\mu}}\prod_{j\neq i}\left[\int dx_j m_{j\to\mu}(x_j)e^{\frac{F_{\mu j}x_j}{\Delta_\mu}(y_\mu-F_{\mu i}x_i+i\lambda)}\right] \tag{12}$$

The integration over scalar x_j takes the form of the moment generating function. Therefore,

$$m_{\mu\to i}(x_i)=\frac{1}{Z^{\mu\to i}\sqrt{2\pi\Delta}}e^{-\frac{1}{2\Delta_\mu}(F_{\mu i}x_i-y_\mu)^2}\int d\lambda e^{-\frac{\lambda^2}{2\Delta_\mu}}\prod_{j\neq i}E_{X_j}\left[e^{\frac{F_{\mu j}x_j}{\Delta_\mu}(y_\mu-F_{\mu i}x_i+i\lambda)}\right] \tag{13}$$

By assuming that each scalar X_j is Gaussian distributed, the moment generating function is expressed using means and variance. Thus, introducing means and variances as "messages"

$$a_{i\to\mu}\equiv\int dx_i x_i m_{i\to\mu}(x_i)\,, \tag{14}$$

$$v_{i\to\mu}\equiv\int dx_i x_i^2 m_{i\to\mu}(x_i)-a_{i\to\mu}^2 \tag{15}$$

We obtain

$$m_{\mu\to i}(x_i)=\frac{1}{Z^{\mu\to i}\sqrt{2\pi\Delta}}e^{-\frac{1}{2\Delta_\mu}(F_{\mu i}x_i-y_\mu)^2}\int d\lambda e^{-\frac{\lambda^2}{2\Delta_\mu}}\prod_{j\neq i}\left[e^{\frac{F_{\mu j}a_{j\to\mu}}{\Delta_\mu}(y_\mu-F_{\mu i}x_i+i\lambda)+\frac{F_{\mu j}^2 v_{j\to\mu}}{2\Delta_\mu^2}(y_\mu-F_{\mu i}x_i+i\lambda)}\right] \tag{16}$$

Performing the Gaussian integral over λ, we obtain

$$m_{\mu\to i}(x_i)=\frac{1}{\tilde{Z}^{\mu\to i}}e^{-\frac{x_i^2}{2}A_{\mu\to i}+B_{\mu\to i}x_i}\ ,\quad \tilde{Z}^{\mu\to i}=\sqrt{\frac{2\pi}{A_{\mu\to i}}}e^{\frac{B_{\mu\to i}^2}{2A_{\mu\to i}}} \tag{17}$$

Where

$$A_{\mu\to i}=\frac{F_{\mu i}^2}{\Delta_\mu+\sum_{j\neq i}F_{\mu j}^2 v_{j\to\mu}} \tag{18}$$

$$B_{\mu\to i}=\frac{F_{\mu i}(y_\mu-\sum_{j\neq i}F_{\mu j}a_{j\to\mu})}{\Delta_\mu+\sum_{j\neq i}F_{\mu j}^2 v_{j\to\mu}} \tag{19}$$

To close the equation on messages $a_{i\to\mu}$ and $v_{i\to\mu}$ we notice that

$$m_{i\to\mu}(x_i)=\frac{1}{\tilde{Z}^{i\to\mu}}[(1-\rho)\delta(x_i)+\rho\phi(x_i)]e^{-\frac{x_i^2}{2}\sum_{\gamma\neq\mu}A_{\gamma\to i}+x_i\sum_{\gamma\neq\mu}B_{\gamma\to i}} \tag{20}$$

Message $a_{i\to\mu}$ and $v_{i\to\mu}$ are respectively the mean and variance of the probability distribution $m_{i\to\mu}(x_i)$. It also useful to define the local beliefs a_i and v_i as

$$a_i\equiv\int dx_i x_i m_i(x_i) \tag{21}$$

$$v_i\equiv\int dx_i x_i^2 m_i(x_i)-a_i^2, \tag{22}$$

Where

$$m_i\left(x_i\right)=\frac{1}{\tilde{Z}^i}[(1-\rho)\delta(x_i)+\rho\phi(x_i)]e^{-\frac{x_i^2}{2}\sum_\gamma A_{\gamma\to i}+x_i\sum_\gamma B_{\gamma\to i}} \tag{23}$$

Let us define the probability distribution

$$\mathrm{M}_\phi\left(\sum{}^2,R,x\right)=\frac{1}{\hat{Z}(\sum{}^2,R)}[(1-\rho)\delta(x)+\rho\phi(x)]\frac{1}{\sqrt{2\pi}\sum}e^{-\frac{(x-R)^2}{2\sum^2}}, \tag{24}$$

$$\sum{}^2:=\frac{1}{\sum_{\gamma\neq\mu}A_{\gamma\to i}},\quad R:=\frac{\sum_{\gamma\neq\mu}B_{\gamma\to i}}{\sum_{\gamma\neq\mu}A_{\gamma\to i}}$$

Where $\hat{Z}(\sum{}^2,R,x)$ is normalization. We define the average and variance of M_ϕ as

$$f_a(\sum{}^2,R)\equiv\int dxx\mathrm{M}(\sum{}^2,R,x) \tag{25}$$

$$f_c(\sum{}^2,R)\equiv\int dxx^2\mathrm{M}(\sum{}^2,R,x)-f_a^2(\sum{}^2,R) \tag{26}$$

The closed form of the BP update is

$$a_{i\to\mu}=f_a\left(\frac{1}{\sum_{\gamma\neq\mu}A_{\gamma\to i}},\frac{\sum_{\gamma\neq\mu}B_{\gamma\to i}}{\sum_{\gamma\neq\mu}A_{\gamma\to i}}\right),\quad a_i=f_a\left(\frac{1}{\sum_\gamma A_{\gamma\to i}},\frac{\sum_\gamma B_{\gamma\to i}}{\sum_\gamma A_{\gamma\to i}}\right), \tag{27}$$

$$v_{i\to\mu}=f_c\left(\frac{1}{\sum_{\gamma\neq\mu}A_{\gamma\to i}},\frac{\sum_{\gamma\neq\mu}B_{\gamma\to i}}{\sum_{\gamma\neq\mu}A_{\gamma\to i}}\right),\quad v_i=f_c\left(\frac{1}{\sum_\gamma A_{\gamma\to i}},\frac{\sum_\gamma B_{\gamma\to i}}{\sum_\gamma A_{\gamma\to i}}\right) \tag{28}$$

For a general signal model $\phi(x_i)$ the functions f_a and f_c can be computed using a numerical integration over x_i. Eqs. (14-15) together with (18-19) and (20) lead to closed iterative message passing equation, which can be solved by iterations. There equation can be used for any signal $\mathbf{s}$, and any matrix $\mathbf{F}$. When a fixed point of the BP equations is reached, the elements of the original signal are estimated as $x_i^* = a_i$, and the corresponding variance v_i can be used to quantify the correctness of the estimate. Perfect reconstruction is found when the message converge to a fixed point such that $a_i = s_i$ and $v_i = 0$.

IV. DISCUSSION

After meeting, please write discussion in the meeting and update your presentation file.

Reference

[1] Krzakala F., Mezard M., Sausset F., Sun Y. & Zdeborova L. Statistical physics-based reconstruction in compressed sensing. Phys. Rev. X 021005 (2012).

[2] Candes E. J. & Wakin M. B. An Introduction To Compressive Sampling. IEEE Signal Processing Magazine 25, 21-30 (2008).

[3] Donoho D., Maleki A. & Montanari A. Message passing algorithms for compressed sensing: I. motivation and construction. In Information Theory Workshop (ITW), 2010 IEEE, 1 {5 (2010).

[4] Donoho D. L., Javanmard A. & Montanari A. Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing (2011). ArXiv:1112.0708v1 [cs.IT]

[5] Montanari A. & Bayati M. Message-passing algorithms for compressed sensingThe dynamics of message passing on dense graphs, with applications to compressed sensing (2010). ArXiv:1001.3448.

[6] Rangan S. Estimation with random linear mixing, belief propagation and compressed sensing. In Information Sciences and Systems (CISS), 2010 44th Annual Conference on, 1-6 (2010).

[7] Montanari A. & Bayati M. Message-passing algorithms for compressed sensingThe dynamics of message passing on dense graphs, with applications to compressed sensing (2010). ArXiv:1001.3448

[8] Bayati M., Lelarge M. & Montanari A. Universality in message passing algorithms (2012). In preparation.

[9] Wu Y. & Verdu S. Renyi information dimension: fundamental limits of almost lossless analog compression. IEEE Transactions on Information Theory 56, 37213747 (2010).

[10] Wu Y. & Verdu S. Optimal Phase Transitions in Compressed Sensing (2011). ArXiv:1111.6822v1 [cs.IT].

[11] Wu Y. & Verdu S. MMSE Dimension. IEEE Transactions on Information Theory 57, 4857 - 4879 (2011).

[12] Baron D., Sarvotham S. & Baraniuk R. Bayesian Compressive Sensing Via Belief Propagation. IEEE Transactions on Signal Processing 58, 269 - 280 (2010).

[13] Rangan S., Fletcher A. & Goyal V. Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing. arXiv:0906.3234v2 (2009).

[14] Donoho D. L., Johnstone I. & Montanari A. Accurate Prediction of Phase Transitions in Compressed Sensing via a Connection to Minimax Denoising (2011). ArXiv:1111.1041v1 [cs.IT].