

PHASE RECOVERY FROM A BAYESIAN POINT OF VIEW: THE VARIATIONAL APPROACH

Angélique Drémeau, Florent Krzakala

Sorbonne Universités, UPMC Univ Paris 06, UMR 8550, LPS-ENS F-75005, Paris, France
Laboratoire de Physique Statistique de l'École Normale Supérieure, 29 rue Lhomond, 75005, Paris, France

ABSTRACT

In this paper, we consider the phase recovery problem, where a complex signal vector has to be estimated from the knowledge of the modulus of its linear projections, from a naive variational Bayesian point of view. In particular, we derive an iterative algorithm following the minimization of the Kullback-Leibler divergence under the mean-field assumption, and show on synthetic data with random projections that this approach leads to an efficient and robust procedure, with a reasonable computational cost.

Index Terms— Phase recovery, variational Bayesian approximations, mean-field approximation

1. INTRODUCTION

Reconstructing a complex vector given only the magnitude of measurements, a problem known under the name of *phase recovery problem*, is at the heart of numerous applications, such as crystallography [1], coherent diffractive [2] or optical [3] imaging. Formally, the problem writes as follows: observing $\mathbf{y} \in \mathbb{R}_+^M$ through a $M \times N$ complex-measurement matrix \mathbf{D} , we aim at recovering the signal $\mathbf{x} \in \mathbb{C}^N$, such that

$$\mathbf{y} = |\mathbf{D}\mathbf{x}|. \quad (1)$$

It is a non-convex optimization problem notoriously difficult to solve. Therefore, many algorithms have been devised in the literature to deal with this problem. We can roughly divide them into two main families.

1. *The alternating-projection algorithms* alternate projections on the span of the measurement matrix and on the object domain. Among these approaches, we can mention the works of Gerchberg & Saxton [4], Fienup [5] and Griffin & Lim [6].
2. *The algorithms based on convex relaxations* propose to approximate the phase recovery problem by relaxed problems which can be solved efficiently by standard optimization procedures. Two of the main approaches of this type, namely *PhaseLift* [7, 8] and *PhaseCut* [9], rely in particular on semidefinite programming.

In an attempt to circumvent the non-linearity of the modulus, a Bayesian point of view was recently adopted in [10]. Focusing on the particular context of compressive measurements, the approach exploits sparse priors and resorts to a "loopy-belief-propagation" type algorithm [11, 12] in a continuation of a line of works in compressive sensing [13, 14, 15].

Here instead, we release the compressive constraint and consider the case where the vector of interest is a full, non-sparse vector. We follow the naive variational Bayesian approach based on a mean-field approximation. We show that, as long as our simulation setup is concerned, the proposed algorithm is competitive with state-of-the-art procedures, while emphasizing some desirable properties, namely an higher robustness to noise and a reasonable computational complexity.

2. BAYESIAN POINT OF VIEW

In this section, we recall the Bayesian modeling introduced in [10], which we shall follow in the remaining of the paper, and introduce the estimation problem we propose to solve.

2.1. Model

Reinterpreting problem (1) into a Bayesian framework, we introduce new variables, modeling, on the one hand, the missing phases of the observations, and on the other hand, some acquisition noise. Thus, each absolute-valued measurement y_μ , $\mu \in \{1 \dots M\}$ of \mathbf{y} is expressed as

$$y_\mu = e^{j\theta_\mu} \left(\sum_{i=1}^N x_i d_{\mu i} + n_\mu \right), \quad (2)$$

where $\theta_\mu \in [0, 2\pi)$ stands for its missing conjugate phase, and n_μ is a zero-mean circular Gaussian noise with variance σ_n^2 . We suppose moreover that

$$p(\mathbf{x}) = \prod_{i=1}^N p(x_i) \quad \text{with} \quad p(x_i) = \mathcal{CN}(0, \sigma_x^2), \quad (3)$$

and

$$p(\boldsymbol{\theta}) = \prod_{\mu=1}^M p(\theta_\mu) \quad \text{with} \quad p(\theta_\mu) = \frac{1}{2\pi}. \quad (4)$$

Under these assumptions, the absence of phases in the observations is naturally taken into account in the model since marginalizing on θ_μ leads to a distribution on y_μ which only depends on the moduli of y_μ and $\sum_{i=1}^N x_i d_{\mu i}$.

2.2. Problem formulation

Within model (2)-(4), the recovery of the complex signal of interest \mathbf{x} can be expressed as the solution of the following marginalized Maximum A Posteriori (MAP) estimation problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{y}), \quad (5)$$

$$\text{with } p(\mathbf{x}|\mathbf{y}) = \int_{\boldsymbol{\theta}} p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}). \quad (6)$$

Because of the marginalization on the hidden variables $\boldsymbol{\theta}$, the direct computation of $p(\mathbf{x}|\mathbf{y})$ is intractable. We can however envisage different sub-optimal techniques to solve (5). In this paper, we focus on the solutions brought by variational approaches and derive the procedure related to a particular one: the mean-field approximation.

3. VARIATIONAL APPROXIMATIONS

3.1. General formulation

Keeping in mind our previous notations, the variational approximations aim at approximating the posterior joint distribution $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})$ by the distribution $\hat{q}(\mathbf{x}, \boldsymbol{\theta})$ leading to the minimum of the Kullback-Leibler (KL) divergence conditionally to a set of given constraints \mathcal{F} :

$$\hat{q}(\mathbf{x}, \boldsymbol{\theta}) = \underset{q \in \mathcal{F}}{\operatorname{argmin}} \int_{\mathbf{x}} \int_{\boldsymbol{\theta}} q(\mathbf{x}, \boldsymbol{\theta}) \log \left(\frac{q(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})} \right) d\mathbf{x} d\boldsymbol{\theta}. \quad (7)$$

Depending on \mathcal{F} , the minimization (7) gives raise to different approximations [16]. We mention here two of them.

- With $\mathcal{F} = \{q | q = \prod_{i=1}^N q_i(x_i) \prod_{\mu=1}^M q_\mu(\theta_\mu)\}$, we define a Mean-Field (MF) approximation and problem (7) can be efficiently solved using the ‘‘Variational Bayes Expectation-Maximization’’ (VB-EM) algorithm [17].
- With $\mathcal{F} = \{q | q = \frac{\prod_{a=1}^A q_a(\mathbf{x}_a) \prod_{b=1}^B q_b(\boldsymbol{\theta}_b)}{\prod_{i=1}^N q_i(x_i)^{\alpha_i-1} \prod_{\mu=1}^M q_\mu(\theta_\mu)^{\beta_\mu-1}}\}$ where $[\mathbf{x}_1 \dots \mathbf{x}_A]$ (resp. $[\boldsymbol{\theta}_1 \dots \boldsymbol{\theta}_B]$) partitions the variables \mathbf{x} (resp. $\boldsymbol{\theta}$) and α_i (resp. β_μ) is the degree of variable node x_i (resp. θ_μ), problem (7) refers to the minimization of the Bethe free energy, which can be solved by generalized approximate message passing (GAMP) algorithms [14, 18]. This is the approach followed in [10].

To the best of our knowledge, the MF approximation has never been considered in a phase retrieval context, within

model (2)-(4). We detail the structure of the resulting algorithm in the next subsection. With respect to the more involved Bethe approach, the variational MF one has some clear advantages: it is conceptually much simpler and leads immediately to a fast iterative algorithm, without further approximation. It is also more mathematically grounded as it leads to a *provably convergent algorithm*, and to a guaranteed *provable bound* of the KL divergence (this is not the case of the Bethe approach [19]), all that, we shall see, with very similar performances (at least in the non-sparse cases we are considering here).

3.2. Variational Bayes EM algorithm

The VB-EM algorithm is an iterative procedure which successively updates the factors of the considered MF approximation. Particularized to model (2)-(4) and the MF approximation defined above, it leads to the following update equations¹:

$$q(\theta_\mu) = \frac{1}{2\pi I_0(\frac{2}{\sigma_n^2} |y_\mu^* \langle z_\mu \rangle|)} \exp \left(\frac{2}{\sigma_n^2} \Re(y_\mu^* \langle z_\mu \rangle e^{j\theta_\mu}) \right), \quad (8)$$

$$q(x_i) = \mathcal{CN}(m_i, \Sigma_i), \quad (9)$$

where

$$\Sigma_i = \frac{\sigma_n^2 \sigma_x^2}{\sigma_n^2 + \sigma_x^2 \mathbf{d}_i^H \mathbf{d}_i}, \quad (10)$$

$$m_i = \frac{\sigma_x^2}{\sigma_n^2 + \sigma_x^2 \mathbf{d}_i^H \mathbf{d}_i} \langle \mathbf{r}_i \rangle^H \mathbf{d}_i, \quad (11)$$

$$\langle \mathbf{r}_i \rangle = \bar{\mathbf{y}} - \sum_{k \neq i} m_k \mathbf{d}_k, \quad (12)$$

with

$$\bar{\mathbf{y}} = \left[y_\mu e^{(j \arg(y_\mu^* \langle z_\mu \rangle))} \frac{I_1(\frac{2}{\sigma_n^2} |y_\mu^* \langle z_\mu \rangle|)}{I_0(\frac{2}{\sigma_n^2} |y_\mu^* \langle z_\mu \rangle|)} \right]_{\mu=\{1 \dots N\}} \quad (13)$$

$$\langle z_\mu \rangle = \sum_{i=1}^N m_i d_{\mu i}. \quad (14)$$

In the equations above, I_0 (resp. I_1) stands for the modified Bessel function of the first kind for order 0 (resp. 1), H denotes the conjugate transpose and * the scalar conjugate.

At this point of discussion, it can be interesting to compare the proposed procedure and the sparse-decomposition algorithm proposed in [20]. Although solving different problems (phase recovery for the one, sparse decomposition for the other), the algorithms rely both on MF approximations, and share indeed notable structural similarities.

¹For a sake of clarity, we drop here the iteration indices.

i) In particular, (9) - (12) are identical in both algorithms. The only difference lies in the definition of the vector $\langle \mathbf{r}_i \rangle$. More than a simple current residual, $\langle \mathbf{r}_i \rangle$ takes here into account complex-valued observations $\bar{\mathbf{y}}$, in which the “reconstructed” phase $\arg(y_\mu^* \langle z_\mu \rangle)$, reminding of the MAP estimate of θ_μ , is weighted by the ratio of the two Bessel functions.

ii) Interestingly, model (2)-(4) can be straightforwardly extended to the compressive framework, by replacing the Gaussian prior model by a Bernoulli-Gaussian one. Doing so, a compressive version of the procedure can be simply obtained by adding an update equation of the distribution on the sparse-representation support as proposed in [20].

Coming back to problem (5), an approximation of $p(\mathbf{x}|\mathbf{y})$ then simply follows from

$$p(\mathbf{x}|\mathbf{y}) \simeq q(\mathbf{x}) = \prod_i q(x_i), \quad (15)$$

resulting in the estimation $\hat{\mathbf{x}}$ such that, $\forall i \in \{1, \dots, N\}$,

$$\hat{x}_i = \operatorname{argmax}_{x_i} q(x_i) = m_i. \quad (16)$$

In this way, the MF approximation leads finally to a solution which tends to maximize the posterior probability of *each* x_i (rather than of the whole vector \mathbf{x}). In fact (see [20, 18] for a similar discussion) at each iterations, the KL divergence decays, and since it is bounded, the algorithm must converge.

3.3. Noise estimation

As emphasized in [20], the estimation of model parameters, and in particular the noise variance, can easily be embedded within the VB-EM procedure (8)-(12). Particularized to model (2)-(4), this leads to

$$\hat{\sigma}_n^2 = \operatorname{argmax}_{\sigma_n^2} \int_{\boldsymbol{\theta}} \int_{\mathbf{x}} q(\boldsymbol{\theta}) q(\mathbf{x}) \log p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y} | \sigma_n^2) d\mathbf{x} d\boldsymbol{\theta}, \quad (17)$$

$$= \frac{1}{M} \left(\mathbf{y}^H \mathbf{y} - 2 \Re(\bar{\mathbf{y}}^H \sum_i m_i \mathbf{d}_i) + \sum_i \Sigma_i \mathbf{d}_i^H \mathbf{d}_i + \sum_i \sum_j m_i^* m_j \mathbf{d}_i^H \mathbf{d}_j \right). \quad (18)$$

Although it may not seem necessary if the noise is known, the iterative estimation of the noise variance is in fact fundamental and improves the behavior of the algorithm that would be, otherwise, trapped close to poor solutions. This point, that was discussed in detail in [18] in the case of compressed sensing, can be intuitively understood by noticing that $\hat{\sigma}_n^2$ is a measure of the (mean) discrepancies between the observations and the assumed model.

4. EXPERIMENTS

In this section, we study the performance of the proposed algorithm by extensive computer simulations. We evaluate and compare the performance of 4 different algorithms: Gerchberg-Saxton [4], PhaseCut [9] and the variational Bayesian approaches: prGAMP introduced in [10] and considered here in its non-compressive version, and the proposed MF-based one, denoted by prVBEM. The algorithms present different complexities. The implementation of PhaseCut relies on interior-point methods, with a complexity growing as $\mathcal{O}(M^{3.5} \log(1/\epsilon))$ where ϵ is the target precision [9]. Gerchberg-Saxton, prGAMP and prVBEM share similar complexities, of order $\mathcal{O}(M^2)$ *per iteration*. We use for PhaseCut and prGAMP the implementations proposed by their authors on their webpages². Within these implementations, the stopping criterium is set to its default value for prGAMP; after preliminary testing and considering the best tradeoff between recovery performance and computational time, Gerchberg-Saxton stops after 3000 iterations and PhaseCut is run until the target precision drops below 10^{-2} . Finally prVBEM iterates as long as the KL divergence (7) decreases more than a difference of 10^{-8} between two consecutive iterations.

In the sequel, we emphasize two desirable properties of the proposed algorithm. To that end, we consider the following general experiment setup. Observations are generated according to model (2)-(4) with the following parameters: $N = 64$, $M = \alpha N$ with $\alpha \in \{1, \dots, 6\}$ and $\sigma_x^2 = 1$. The elements of the dictionary \mathbf{D} are *i.i.d.* realizations of a zero-mean circular Gaussian distribution with variance M^{-1} . For a fair comparison of the algorithms, prGAMP and prVBEM are both not aware of the values of σ_n^2 and σ_x^2 used to generate the observations.

We assess the performance in terms of the reconstruction of the signal \mathbf{x} . In particular, we consider the correlation between the estimated signal and the one used to generate the data, as a function of the number of measurements M . This figure of merit is evaluated from 100 trials for each simulation points.

4.1. Robustness to noise

We first compare the performance of the algorithms in three different contexts: the noiseless case, *i.e.*, $\sigma_n^2 = 0$ (Fig. 1(a)), and two noisy cases where $\sigma_n^2 = 0.3$ (Fig. 1(b)) and $\sigma_n^2 = 0.7$ (Fig. 1(c)).

As we can see in Fig. 1(a), the noiseless case is an ideal case: within the considered setup, as soon as $M > 3N$, almost all algorithms obtain a correlation higher than 0.9. Exception is made by the Gerchberg-Saxton algorithm, which

²resp. <http://www.di.ens.fr/data/software/> (released June 18th, 2012) and http://gampmatlab.wikia.com/wiki/Generalized_Approximate_Message_Passing (released May 21st, 2014)

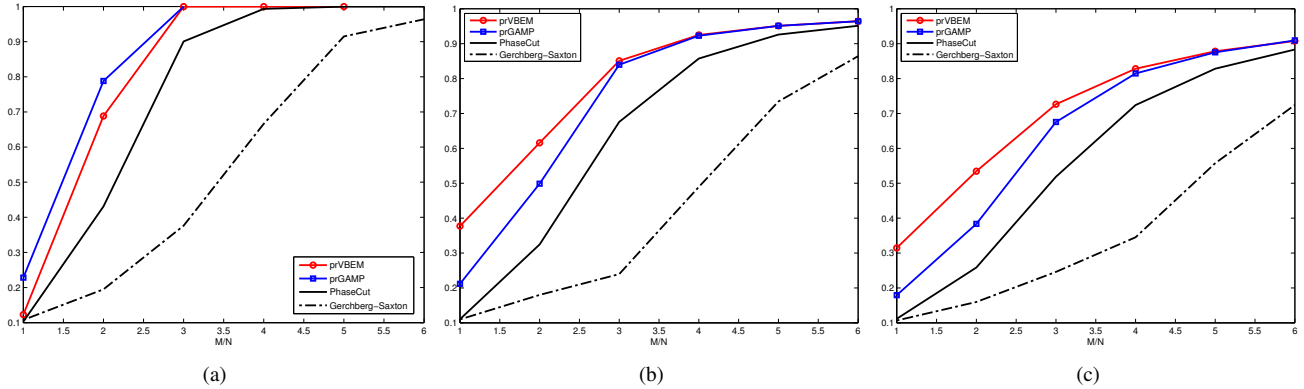


Fig. 1. Correlation between estimated and original signals, as a function of the number of measurements M (x-axis is M/N with $N = 64$): noiseless case (a), $\sigma_n^2 = 0.3$ (b) and $\sigma_n^2 = 0.7$ (c).

requires in this setup at least $5N$ measurements to perform an acceptable reconstruction.

Most of practical experimental schemes however lead to very noisy measurements. In such scenarios, the phase recovery may become difficult. This is illustrated in Fig. 1(b) and (c) where we can observe a general degradation of the performance. Beyond this rough tendency, we note interestingly that a gap seems to open up between the performance obtained by prVBEM and the other algorithms. Although the proposed algorithm does not reach a correlation equal to 1, it appears relatively more robust and outperforms all other algorithms.

4.2. Computational time

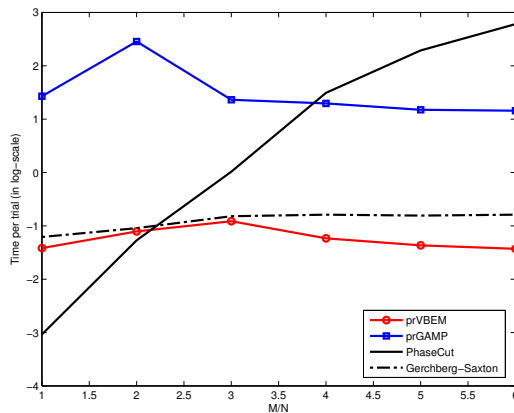


Fig. 2. Average running time (log-scale) as a function of the number of measurements M (x-axis is M/N with $N = 64$).

Considering the previous noiseless setup, we focus then on the computational costs of the different algorithms. In Fig. 2, we see that prVBEM presents the lowest running time, in

particular lower than prGAMP while sharing a similar complexity order per update step. It is of course difficult to compare the running times of algorithms which do not have the same stopping criteria. However, this comparison highlights the general good convergence properties of the VB-EM algorithm, which, in that phase-recovery case, requires less than 300 iterations to significantly decrease the KL divergence (7). When dealing with high-dimensional data, this behavior may be highly desirable.

We refer the interested reader to the author's webpage³ for the implementation of prVBEM.

5. CONCLUSION

Using the classical variational mean-field approach to Bayesian inference, we have presented a new iterative algorithm to estimate complex vectors from the moduli of their linear projections (the so-called phase retrieval problem). This leads to a principled, convergent, and efficient procedure. As far as our simulations are concerned, we have shown in particular that the algorithm performs very well, both in terms of running time and stability with respect to the state-of-art. A natural continuation would be to consider the sparse compressive case, and to apply our approach to relevant experimental situations, such as, for instance, optical ones [21].

6. ACKNOWLEDGEMENTS

This work has been supported in part by the ERC under the European Union's 7th Framework Programme Grant Agreement 307087-SPARCS. AD is currently working at ENSTA Bretagne, Lab-STICC (UMR 6285), 2 rue Francois Verny, F-29200 Brest, France.

³<http://angelique.dreameau.free.fr>

7. REFERENCES

- [1] R. W. Harrison, "Phase problem in crystallography," *Journal of the Optical Society of America A*, vol. 10, no. 5, pp. 1046–1055, 1993.
- [2] H. M. Quiney, "Coherent diffractive imaging using short wavelength light sources: a tutorial review," *Journal of Modern Optics*, vol. 57, no. 13, July 2010.
- [3] O. Katz, P. Heidmann, M. Fink, and S. Gigan, "Non-invasive single-shot imaging through scattering layers and around corners via speckle correlations," *Nature Photonics*, vol. 8, pp. 784 – 790, 2014.
- [4] R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.
- [5] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.
- [6] D. Griffin and J. Lim, "Signal estimation from modified short-time fourier transform," *IEEE Transactions On Acoustics, Speech and Signal Processing*, vol. 32, no. 2, pp. 236–243, 1984.
- [7] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," *Inverse Problems*, vol. 27, no. 1, 2011.
- [8] E. Candès, T. Strohmer, and V. Voroninski, "Phaselift: exact and stable signal recovery from magnitude measurements via convex programming," *Communications in Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, August 2013.
- [9] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming Series A- Springer*, December 2013.
- [10] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," in *Communication, Control, and Computing (Allerton)*, October 2012.
- [11] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.
- [12] M. Mezard and A. Montanari, *Information, physics, and computation*, Oxford University Press, 2009.
- [13] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *Proc. IEEE Information Theory Workshop (ITW)*, Cairo, Egypt, January 2010, pp. 1–5.
- [14] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE Int'l Symposium on Information Theory (ISIT)*, July 2011.
- [15] F. Krzakala, M. Mezard, F. Sausset, Y. F. Sun, and L. Zdeborova, "Statistical physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, no. 021005, May 2012.
- [16] J. S. Yedidia, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions On Information Theory*, vol. 57, no. 7, July 2005.
- [17] M. J. Beal and Z. Ghahramani, "The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures," *Bayesian Statistics*, vol. 7, pp. 453–463, 2003.
- [18] F. Krzakala, A. Manoel, E.W. Tramel, and L. Zdeborova, "Variational free energies for compressed sensing," in *Information Theory (ISIT), 2014 IEEE International Symposium on*, June 2014, pp. 1499–1503.
- [19] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1-2, pp. 1–305, 2008.
- [20] A. Drémeau, C. Herzet, and L. Daudet, "Boltzmann machine and mean-field approximation for structured sparse decompositions," *IEEE Trans. On Signal Processing*, vol. 60, no. 7, pp. 3425–3438, July 2012.
- [21] A. Liutkus, D. Martina, S. Popoff, G. Chardon, O. Katz, G. Lerosey, S. Gigan, L. Daudet, and I. Carron, "Imaging with nature: Compressive imaging using a multiply scattering medium," *Scientific reports*, vol. 4, 2014.