Phase Retrieval with a Multivariate Von Mises Prior: From a Bayesian Formulation to a Lifting Solution

Angélique Drémeau, Antoine Deleforge

▶ **To cite this version:**

HAL Id: hal-01653732
https://hal.archives-ouvertes.fr/hal-01653732

Submitted on 5 Dec 2017

PHASE RETRIEVAL WITH A MULTIVARIATE VON MISES PRIOR:
FROM A BAYESIAN FORMULATION TO A LIFTING SOLUTION

Angélique Drémeau[®], Antoine Deleforge[*]*

[®] ENSTA Bretagne and Lab-STICC UMR 6285, Brest, F-29200, France
[*] INRIA Centre Rennes-Bretagne Atlantique, Campus universitaire de Beaulieu, F-35000 Rennes, France

ABSTRACT

In this paper, we investigate a new method for phase recovery when prior information on the missing phases is available. In particular, we propose to take into account this information in a generic fashion by means of a multivariate Von Mises distribution. Building on a Bayesian formulation (a Maximum A Posteriori estimation), we show that the problem can be expressed using a Mahalanobis distance and be solved by a lifting optimization procedure.

Index Terms— Phase retrieval, multivariate Von Mises distribution, Mahalanobis distance, lifting.

1. INTRODUCTION

Since more than twenty years, phase retrieval has been a constantly filled topic. This is because the problem interests numerous application domains, from crystallography [1] to optical imaging [2]. Formally, it can be written as follows: given $\boldsymbol{y} \in \mathbb{R}^M$, recover $\boldsymbol{x} \in \mathbb{C}^K$ such as

$$\boldsymbol{y} = |\mathbf{A}\boldsymbol{x}|, \qquad (1)$$

where $\mathbf{A}$ is a $M \times K$ known complex-measurement matrix. Several answers to this non-convex optimization problem have been proposed, that we can roughly divide into three families: *i)* alternating-projection algorithms, where we can find the works of Gerchberg & Saxton [3], Fienup [4] or Griffin & Lim [5], which alternate projections on the span of the measurement matrix and on the object domain, *ii)* algorithms based on convex relaxations, such as the recent *PhaseLift* [6] and *PhaseCut* [7], which replace the phase recovery problem by relaxed problems that can be efficiently solved by standard optimization procedures, and *iii)* Bayesian approaches, which express the phase recovery problem as the solution of a Bayesian inference problem and apply statistical tools to solve it, such as variational approximations [8, 9].

In the above procedures, the phases are completely missing from the observations: only intensities or amplitudes are acquired. In this paper, we are interested in phase retrieval problems where phases are observed but marred by noise. At the interface between the last two above families, we propose a Bayesian formulation of the problem and resort to a lifting optimization procedure to solve it. A priori knowledge over observed phases through various probabilistic laws have been exploited in previous works [10, 11]. Compared to them, our approach presents two appealing novelties: *i)* it is generic in the sense that it can handle multivariate phase priors and thus arbitrary dependencies; *ii)* the proposed Bayesian optimization problem is cast into a generalization of the recently proposed *PhaseCut* problem [7], for which a number of efficient estimation procedures readily exist, including convex relaxations. The last point is made possible by exploiting a previously unseen connection between a multivariate generalization of the Von Mises distribution and the Mahalanobis distance.

2. BAYESIAN FORMULATION

In this section, we introduce the Bayesian modeling that we propose to exploit in the following and discuss its link to the Mahalanobis distance, particularly interesting for the optimization procedure.

2.1. Observation model

Let M sensors record K complex signals through linear instantaneous mixing, in the presence of both additive noise and multiplicative phase noise. The noisy observation $\boldsymbol{y} \in \mathbb{C}^M$ is then expressed as

$$\boldsymbol{y} = \mathbf{Diag}\{\boldsymbol{\phi}\}^H \mathbf{A}\boldsymbol{x} + \boldsymbol{n} \qquad (2)$$

where $\mathbf{A} \in \mathbb{C}^{M \times K}$ is the mixing matrix, $\boldsymbol{x} \in \mathbb{C}^K$ is the source signal, $\boldsymbol{n} \in \mathbb{C}^M$ is the noise vector, $\boldsymbol{\phi} = [e^{j\theta_1}, \ldots, e^{j\theta_M}]^\top$ is the phase vector with $\boldsymbol{\theta} \triangleq [\theta_1, \ldots, \theta_M]^\top \in] -\pi, \pi]^M$, the operator $\mathbf{Diag}\{.\}$ transforms row- or column-vectors into diagonal matrices and $\cdot^H$ denotes the complex conjugate transpose. For simplicity, we assume that

the additive noise is zero-mean *i.i.d.* circular complex Gaussian with variance σ_n^2. Note that generalizing subsequent derivations to an arbitrary noise covariance matrix Γ_n instead is straightforward with appropriate changes of variable.

2.2. Von Mises prior

In the literature, model (2) has been already considered in phase retrieval problems with a uniform prior on the phases $\boldsymbol{\theta}$ (see *e.g.* [8, 9]). Here, we look for a more informative model enforcing uncertain structures on and between phases.

Considering phases naturally leads to directional statistics. Among them, the most familiar one is probably the Von-Mises distribution, defined independently for each variable $\theta_m, m \in \{1, \dots, M\}$ as

$$p(\theta_m) = \frac{1}{2\pi I_0(\kappa_m)} \exp\left(\kappa_m \cos(\theta_m - \mu_m)\right), \quad (3)$$

where $\kappa_m \in \mathbb{R}$ and $\mu_m \in]-\pi, \pi]$ are parameters of the distribution, and $I_0(.)$ is the modified Bessel function of the first kind of order 0. This distribution has been considered in the literature, *e.g.* in [10]. In practice, it is well-adapted to situations where we want to take into account prior information (such as the mean through parameter μ_m or the variance through κ_m) on the phases independently of one another. Its extension to the multivariate case is not straightforward and can take different forms [12]. In this paper, we assume $\boldsymbol{\theta}$ to be distributed according to

$$p(\boldsymbol{\theta}) = \frac{1}{C(\boldsymbol{\kappa}, \boldsymbol{\Delta})} \exp\left(\boldsymbol{\kappa}^\top \boldsymbol{c}(\boldsymbol{\theta}, \boldsymbol{\mu}) \right.$$
$$\left. - \boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{\mu})^\top \boldsymbol{\Delta} \boldsymbol{s}(\boldsymbol{\theta}, \boldsymbol{\mu}) - \boldsymbol{c}(\boldsymbol{\theta}, \boldsymbol{\mu})^\top \boldsymbol{\Delta} \boldsymbol{c}(\boldsymbol{\theta}, \boldsymbol{\mu})\right), \quad (4)$$

where $C(\boldsymbol{\kappa}, \boldsymbol{\Delta})$ is a normalizing constant and functions $\boldsymbol{c}$ and $\boldsymbol{s}$ are respectively defined by, $\forall m \in \{1, \dots, M\}$,

$$c_m(\boldsymbol{\theta}, \boldsymbol{\mu}) = \cos(\theta_m - \mu_m), \quad s_m(\boldsymbol{\theta}, \boldsymbol{\mu}) = \sin(\theta_m - \mu_m). \quad (5)$$

The matrix $\boldsymbol{\Delta}$ is real-symmetric with zeros on its diagonal and captures dependencies between phases. Without loss of generality[1], we will assume in the sequel that $\boldsymbol{\mu} = [0, \dots, 0]^\top \triangleq \mathbf{0}_M$. This multivariate extension of the Von-Mises distribution was suggested at the end of [12], but does not seem to have been extensively studied or used. We prefer it here over other alternatives due to the following result (proof in appendix A):

Lemma 1 *Let $\boldsymbol{\mu} = \mathbf{0}_M$, and let $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \dots, \hat{\theta}_M]^\top$ maximize the multivariate Von Mises distribution (4). We have:*

$$\hat{\boldsymbol{\phi}} \triangleq [e^{j\hat{\theta}_1}, \dots, e^{j\hat{\theta}_M}]^\top = \underset{\substack{\boldsymbol{\phi} \\ |\phi_i|^2=1\ \forall i}}{\operatorname{argmin}} \ ||\boldsymbol{\phi} - \mathbf{1}_M||_{\Gamma_\phi}^2$$

---

[1]Assuming $\boldsymbol{\mu} \neq [0, \dots, 0]^\top$ amounts to considering the observation model $\boldsymbol{y} = \mathbf{Diag}\{\tilde{\boldsymbol{\phi}}\}^H \tilde{\mathbf{A}} \boldsymbol{x}$ with $\tilde{\boldsymbol{\phi}} = \mathbf{Diag}\{\boldsymbol{u}\}^H \boldsymbol{\phi}$, $\tilde{\mathbf{A}} = \mathbf{Diag}\{\boldsymbol{u}\} \mathbf{A}$ and $\boldsymbol{u} = [e^{j\mu_1}, \dots, e^{j\mu_M}]$.

where $||\cdot||_{\Gamma_\phi}$ *denotes the Mahalanobis distance with covariance* Γ_ϕ, $\mathbf{1}_M \triangleq [1, \dots, 1]^\top$, *and* $\forall(i, k) \in \{1, \dots, M\}^2$

$$(\Gamma_\phi^{-1})_{ik} = \left\{ \begin{array}{ll} \Delta_{ik} & \text{if } k \neq i, \\ \frac{1}{2}\kappa_i - \sum_{l \neq i} \Delta_{il} & \text{if } k = i. \end{array} \right. \quad (6)$$

In other words, maximizing the density (4) can be cast as a quadratically-constrained norm-minimization problem. This type of problem is central in the classical phase retrieval literature (see [7]), but does not seem to appear when using other multivariate generalizations of the Von Mises distribution as phase priors, *e.g.*, the one studied in [12].

3. PHASE AND SIGNAL ESTIMATION

3.1. Maximum a posteriori

Using Lemma 1, it follows that the Maximum A Posteriori (MAP) estimate of $\boldsymbol{\phi}$ within model (2) and (4) writes:

$$\hat{\boldsymbol{\phi}}_{\text{MAP}} = \underset{\substack{\boldsymbol{\phi} \\ |\phi_i|^2=1\ \forall i}}{\operatorname{argmax}} \ \log p(\boldsymbol{\phi}|\boldsymbol{y}), \quad (7)$$

$$= \underset{\substack{\boldsymbol{\phi} \\ |\phi_i|^2=1\ \forall i}}{\operatorname{argmin}} \ \frac{1}{\sigma_n^2} ||\boldsymbol{y} - \mathbf{Diag}\{\boldsymbol{\phi}\}^H \mathbf{A}\boldsymbol{x}||_2^2$$

$$+ ||\boldsymbol{\phi} - \mathbf{1}_M||_{\Gamma_\phi}^2. \quad (8)$$

Following a similar idea as in [7], we couple this MAP estimation of the phase vector $\boldsymbol{\phi}$ with a Maximum Likelihood estimation of the source signal $\boldsymbol{x}$:

$$\hat{\boldsymbol{x}}_{\text{ML}} = \underset{\boldsymbol{x}}{\operatorname{argmax}} \ \log p(\boldsymbol{y}; \boldsymbol{x}), \quad (9)$$

$$= \underset{\boldsymbol{x}}{\operatorname{argmin}} \ ||\boldsymbol{y} - \mathbf{Diag}\{\boldsymbol{\phi}\}^H \mathbf{A}\boldsymbol{x}||_2^2, \quad (10)$$

$$= \mathbf{A}^+ \mathbf{Diag}\{\boldsymbol{\phi}\} \boldsymbol{y}, \quad (11)$$

where $\mathbf{A}^+$ stands for the Moore-Penrose pseudo-inversion of matrix $\mathbf{A}$. Reinjecting this estimate in the MAP problem (8) leads to

$$\hat{\boldsymbol{\phi}}_{\text{MAP}} = \underset{\substack{\boldsymbol{\phi} \\ |\phi_i|^2=1\ \forall i}}{\operatorname{argmin}} \ \frac{1}{\sigma_n^2} ||(\mathbf{I}_M - \mathbf{A}\mathbf{A}^+)\mathbf{Diag}\{\boldsymbol{y}\}\boldsymbol{\phi}||_2^2$$

$$+ \left\| (\mathbf{I}_M \quad -\mathbf{1}_M) \cdot \left(\begin{array}{c} \boldsymbol{\phi} \\ 1 \end{array} \right) \right\|_{\Gamma_\phi}^2, \quad (12)$$

where $\mathbf{I}_M$ stands for the identity matrix and the Mahalanobis distance term has been re-written to be homogeneous in $\boldsymbol{u} = [\boldsymbol{\phi} \quad 1]^\top \in \mathbb{C}^{M+1}$. Using this trick, it follows that solving (12) is equivalent to solving the $M+1$-dimensional problem

$$\hat{\boldsymbol{u}} = \underset{\substack{\boldsymbol{u} \\ |u_i|^2=1\ \forall i}}{\operatorname{argmin}} \ \boldsymbol{u}^H \mathbf{Q}\boldsymbol{u}, \text{ where} \quad (13)$$

$$\mathbf{Q} = \begin{pmatrix} \mathbf{M} + \sigma_n^2 \Gamma_\phi^{-1} & -\sigma_n^2 \Gamma_\phi^{-1} \mathbf{1}_M \\ -\sigma_n^2 \mathbf{1}_M^\top \Gamma_\phi^{-1} & \sigma_n^2 \sum_{i,k} (\Gamma_\phi^{-1})_{ik} \end{pmatrix} \in \mathbb{C}^{(M+1)^2}, \quad (14)$$

with $\mathbf{M} = \mathbf{Diag}\{\boldsymbol{y}^H\}(\mathbf{I}_M - \mathbf{AA}^+)\mathbf{Diag}\{\boldsymbol{y}\}$. It is easily verified that if $\hat{\boldsymbol{u}}$ is solution of (13), then $\hat{\boldsymbol{\phi}}_{\text{MAP}} = \hat{\boldsymbol{u}}_{1:M}/\hat{u}_{M+1}$ is solution of (12).

Interestingly, when $\Gamma_\phi^{-1} = \mathbf{0}$ (uninformative prior on phases), (13) is equivalent to the program proposed by Waldspurger et al. [7] for classical phase retrieval. They refer to this complex quadratically-constrained quadratic program as *PhaseCut*, in reference to its real counterpart which is known to be equivalent to the classical graph-partition problem *MaxCut* [13]. These non-convex problems are NP-hard in general, difficult to solve in practice, and have been extensively studied, yielding a number of efficient optimization schemes for particular instances. The most straightforward approach consists in iteratively minimizing (13) with respect to each u_i alternatively, which can be done in closed-form [7]. Since the problem is non-convex, this method is bound to converge to a local minimum which depends on the initialization.

3.2. Lifting solution

A particularly popular alternative to solve (13) is referred to as *Lifting*, and consists in solving the following convex *semidefinite program* (SDP) instead :

$$\begin{array}{cl} \underset{\mathbf{U} \succeq \mathbf{0}}{\operatorname{argmin}} & \operatorname{trace}\{\mathbf{QU}\} \\ \operatorname{diag}\{\mathbf{U}\} = \mathbf{1}_M \end{array} \quad (15)$$

where $\succeq \mathbf{0}$ denotes positive semi-definiteness. Note that (15) is a *relaxation* of (13), in the sense that if $\hat{\mathbf{U}} = \hat{\boldsymbol{u}}\hat{\boldsymbol{u}}^H$ is a rank-1 solution of (15), then $\hat{\boldsymbol{u}}$ is a solution of (13). However, $\hat{\mathbf{U}}$ may not always be rank-1 in practice. In the classical prior-less phase retrieval case where $\Gamma_\phi^{-1} = \mathbf{0}$, the combined extensive research efforts in [6] and [7] lay theoretical grounds providing conditions on $\mathbf{A}$ for which solving (15) enables stable recovery of the phase vector $\boldsymbol{\phi}$ and signal vector $\boldsymbol{x}$ with high probability. Extending these theories to the proposed Bayesian generalization necessitates a deep research investigation, which cannot be tackled within this short paper. Rather, an experimental validation of the lifting approach in the multivariate Von-Mises phase retrieval setting is conducted in Section 4.

3.3. Algorithms

A large number of efficient generic SDP solvers are available, including interior-point methods [14] or augmented Lagrangian methods [15]. As mentioned in [7], the *block-coordinate descent* (BCD) method proposed in [16] is particularly simple and efficient for problems of the form (15), and
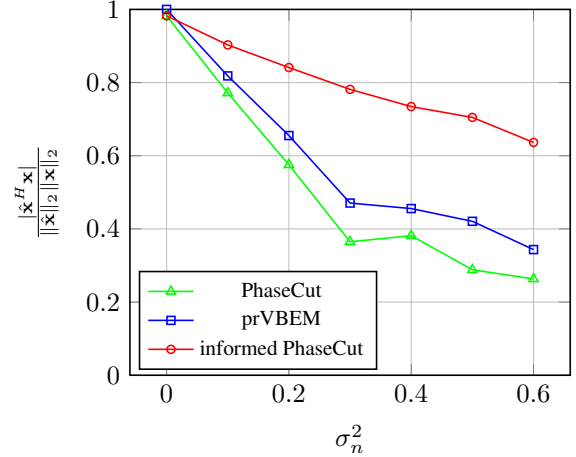
Fig. 1. (Averaged) normalized correlation as a function of the variance σ_n^2 for the *i.i.d.* 1D Von Mises prior.

is therefore used here. In practice, when the obtained solution $\hat{\mathbf{U}}$ is not rank-1, a natural approach consists in selecting the leading eigenvector of $\mathbf{U}$.
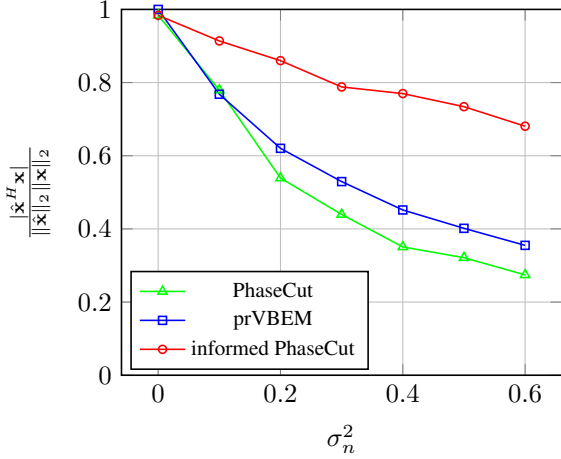
4. EXPERIMENTS

In this section, we propose two different experimental setups to assess the relevance of the above procedure. More precisely, we consider two particular cases of the multivariate Von-Mises prior (4) : the 1D Von-Mises distribution and the Markov chain.

For both setups, we confront it to two state-of-the-art phase retrieval algorithms, namely *PhaseCut* [7] and *prVBEM* [9]. The first one relies on the same optimization procedure as the one proposed here, but does not exploit any information on the phases to recover. The second one shares the same Bayesian formulation (2) as the algorithm proposed here but considers a non-informative, uniform distribution on the phases. In the sequel, we will refer to our approach as "informed PhaseCut".

We consider the following general experimental setup. Observations are generated according to model (2) with $M = 256$ and $K = 64$. The elements of the dictionary $\mathbf{A}$ (resp. vector $\boldsymbol{x}$) are *i.i.d.* realizations of a zero-mean circular Gaussian distribution with variance M^{-1} (resp. 1). We assess the performance in terms of the reconstruction of the signal $\boldsymbol{x}$. In particular, we consider the correlation between the estimated signal and the one used to generate the data,

$$\frac{|\hat{\mathbf{x}}^H \mathbf{x}|}{\|\hat{\mathbf{x}}\|_2 \|\mathbf{x}\|_2},$$

as a function of the noise variance σ_n^2. This figure of merit is evaluated from 50 trials for each simulation points.

Fig. 2. (Averaged) normalized correlation as a function of the variance σ_n^2 for the Markov chain prior.

4.1. 1D Von Mises prior

As a first experimental setup, we consider the case where the phase noise is distributed on each sensor independently of one another according to the Von Mises law (3) with parameter $\mu_i = 0$ and $\kappa_i = 1$, $\forall i \in \{1, \ldots, M\}$.

Figure 1 presents the performance of the three algorithms with this particular prior distribution. As expected, *informed PhaseCut* outperforms the other algorithms, proving a good inclusion of the prior additional information. More particularly, the gap between them increases with the noise variance: for $\sigma_n^2 = 0.6$, *informed PhaseCut* achieves a correlation around 0.7 against 0.3 for *PhaseCut* and *prVBEM*.

4.2. Markov chain

In a second experimental setup, we consider the particular case where only the first two subdiagonals of $\boldsymbol{\Delta}$ are non-zero. Considering a small variance of the phases $\boldsymbol{\theta}$, straightforward calculus leads to

$$p(\boldsymbol{\theta}) \simeq \frac{1}{C(\boldsymbol{\kappa}, \boldsymbol{\Delta})} \tag{16}$$

$$\exp\left(-\sum_i \left((\Gamma_\phi^{-1})_{ii}\theta_i^2 - 2(\Gamma_\phi^{-1})_{i(i-1)}\theta_i\theta_{i-1} + o(\theta_i^2)\right)\right),$$

where Γ_ϕ^{-1} is linked to the parameters $\boldsymbol{\Delta}$, $\boldsymbol{\kappa}$ through (6). This expression can be directly identified to a Markov chain such as $\forall i \in \{2, \ldots, M\}$, $\theta_i = a\,\theta_{i-1} + \omega_i$, where $\omega_i \sim \mathcal{N}(0, \sigma_\theta^2)$, $\forall i$ and $\theta_1 \sim \mathcal{N}(0, \sigma_\theta^2)$, providing that

$$(\Gamma_\phi^{-1})_{ik} = \begin{cases} -\frac{a}{2\sigma_\theta^2} & \text{if } k = i+1 \text{ or } k = i-1, \\ \frac{1+a^2}{2\sigma_\theta^2} & \text{if } k = i \neq M, \\ \frac{1}{2\sigma_\theta^2} & \text{if } k = i = M, \\ 0 & \text{elsewhere.} \end{cases} \tag{17}$$

We suppose here that $a = 0.8$, $\sigma_\theta^2 = 0.1$.

Figure 2 confirms the good behavior of *informed Phase-Cut* observed in the first experiment setup: taking into account the structure of the missing phases, it allows a better estimation (in the sense of the correlation) of the signal of interest $\boldsymbol{x}$. The advantage brought by such prior inclusion increases with the noise variance: *informed PhaseCut* reveals here again more robustness.

5. CONCLUSION

In this paper, we have presented a novel algorithm able to solve the phase recovery problem with a multivariate Von Mises prior distribution. To that end, we have showed that this particular prior information can be efficiently integrated into a Maximum A Posteriori estimation by means of a Mahalanobis distance. The proposed solution relies on a lifting procedure and, to the extent of our experiments, reveals a coherent behavior with regard to non-informed state-of-the-art algorithms.

A. PROOF OF LEMMA 1

We have:

$$||\boldsymbol{\phi} - \mathbf{1}_M||^2_{\Gamma_\phi} \tag{18}$$

$$= (\boldsymbol{\phi} - \mathbf{1}_M)^H \Gamma_\phi^{-1}(\boldsymbol{\phi} - \mathbf{1}_M)$$

$$= \text{trace}\{\Gamma_\phi^{-1}\} + \sum_{i,k}(\Gamma_\phi^{-1})_{ik} - 2\sum_i(\Gamma_\phi^{-1})_{ii}\cos(\theta_i)$$

$$- \sum_i\sum_{k\neq i}(\Gamma_\phi^{-1})_{ik}e^{-j(\theta_i-\theta_k)} - \sum_i\sum_{k\neq i}(\Gamma_\phi^{-1})_{ik}e^{-j\theta_i}$$

$$- \sum_i\sum_{k\neq i}(\Gamma_\phi^{-1})_{ki}e^{j\theta_i},$$

$$= \text{trace}\{\Gamma_\phi^{-1}\} + \sum_{i,k}(\Gamma_\phi^{-1})_{ik} - 2\sum_i\left(\sum_k(\Gamma_\phi^{-1})_{ik}\right)\cos(\theta_i)$$

$$+ \sum_i\sum_{k\neq i}(\Gamma_\phi^{-1})_{ik}\cos(\theta_i - \theta_k),$$

where we have assumed that $(\Gamma_\phi^{-1})_{ki} = (\Gamma_\phi^{-1})_{ik}$, $\forall(i, k) \in \{1, \ldots, M\}^2$, or, in other words, $(\Gamma_\phi^{-1})_{ik} \in \mathbb{R}$. Identifying then parameters $\boldsymbol{\kappa}$ and $\boldsymbol{\Delta}$ of the multivariate distribution (4) with (18), it comes straightforwardly that, under the condition (6),

$$||\boldsymbol{\phi} - \mathbf{1}_M||^2_{\Gamma_\phi} \propto -\log p(\boldsymbol{\theta})$$

where $\propto$ denotes here equality up to a constant. This means that we can use indifferently the multivariate Von-Mises distribution (4) or the Mahalanobis distance as a cost function if we add to the latter the constraint $|\phi_i| = 1$, $\forall i \in \{1, \ldots, M\}$. $\square$

B. 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] A. Drémeau, A. Liutkus, D. Martina, O. Katz, C. Schuelke, F. Krzakala, S. Gigan, and L. Daudet, "Reference-less measurement of the transmission matrix of a highly scattering material using a dmd and phase retrieval techniques," *Optics Express*, vol. 23, pp. 11898–11911, 2015.

[3] 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.

[4] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, pp. 2758–2769, 1982.

[5] 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.

[6] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1241–1274, 2013.

[7] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, maxcut and complex semidefinite programming," *Mathematical Programming*, vol. 149, no. 1-2, pp. 47–81, 2015.

[8] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," in *Communication, Control, and Computing (Allerton)*, October 2012.

[9] A. Drémeau and F. Krzakala, "Phase recovery from a bayesian point of view: the variational approach," in *Proc. IEEE Int'l Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015, pp. 3661–3665.

[10] T. Gerkmann, "Bayesian estimation of clean speech spectral coefficients given a priori knowledge of the phase," *IEEE Transactions On Signal Processing*, vol. 62, no. 16, pp. 4199 – 4208, 2014.

[11] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE Journal on selected areas in communications*, vol. 23, no. 9, pp. 1748 – 1757, 2005.

[12] K. V Mardia, G. Hughes, C. C Taylor, and H. Singh, "A multivariate von mises distribution with applications to bioinformatics," *Canadian Journal of Statistics*, vol. 36, no. 1, pp. 99–109, 2008.

[13] M. X. Goemans and D. P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *J. ACM*, vol. 42, pp. 1115 – 1145, 1995.

[14] C. Helmberg, F. Rendl, R. J. Vanderbei, and H. Wolkowicz, "An interior-point method for semidefinite programming," *SIAM Journal on Optimization*, vol. 6, pp. 342 – 361, 1996.

[15] Y. Nesterov, "Smoothing technique and its applications in semidefinite optimization," *Mathematical Programming*, vol. 110, pp. 245 – 259, 2007.

[16] Z. Wen, D. Goldfarb, and K. Scheinberg, *Block coordinate descent methods for semidefinite programming*, Springer, 2012.