Polarimetric Fourier phase retrieval

Julien Flamant, Konstantin Usevich, Marianne Clausel, David Brie

Polarimetric Fourier phase retrieval*

Julien Flamant†, Konstantin Usevich†, Marianne Clausel‡, and David Brie†

Abstract. This work introduces *polarimetric Fourier phase retrieval* (PPR), a physically-inspired model to leverage polarization of light information in Fourier phase retrieval problems. We provide a complete characterization of its uniqueness properties by unraveling equivalencies with two related problems, namely bivariate phase retrieval and a polynomial autocorrelation factorization problem. In particular, we show that the problem admits a unique solution, which can be formulated as a greatest common divisor (GCD) of measurements polynomials. As a result, we propose algebraic solutions for PPR based on approximate GCD computations using the null-space properties Sylvester matrices. Alternatively, existing iterative algorithms for phase retrieval, semidefinite positive relaxation and Wirtinger-Flow, are carefully adapted to solve the PPR problem. Finally, a set of numerical experiments permits a detailed assessment of the numerical behavior and relative performances of each proposed reconstruction strategy. They further demonstrate the fruitful combination of algebraic and iterative approaches towards a scalable, computationally efficient and robust to noise reconstruction strategy for PPR.

Key words. Fourier phase retrieval, polarization, approximate greatest common divisor, semidefinite positive relaxation, Wirtinger Flow

MSC codes. 49N30, 94A12, 12D05

1. Introduction.

The problem of Fourier phase retrieval, i.e., the recovery of a signal given the magnitude of its Fourier transform, has a long and rich history dating back from the 1950s [61]. It has been – and continues to be – of tremendous importance for many applications areas involving optics, such as crystallography [23, 24, 50], astronomy [28, 29], coherent diffraction imaging (also known as lensless imaging) [49, 47], among others. Such problem arises in optics since *phase information* of light cannot be measured directly due to the high oscillating frequency of the electromagnetic field: indeed there is no conventional detector that can sample at a rate of $\sim 10^{12}$ Hz (infrared) up to $\sim 10^{18}$ Hz (hard x-rays). In addition, many imaging applications rely on diffraction measurements in the far-field, where light propagation essentially acts as a Fourier transform operator of the field near the imaged object [33]. Examples include one-dimensional (1D) temporal Fourier transforms performed by spectrometers in ultra-short laser pulse characterization [76] or two-dimensional (2D) spatial Fourier transforms recorded on far-field pixelated detectors in X-ray coherent diffraction imaging [18]. These Fourier-domain detectors, together with the impossibility to measure phase information, yield phaseless Fourier intensity measurements. Therefore, reconstruction of the imaged object requires solving a Fourier phase retrieval problem. See [63] for a comprehensive overview of such problems in optical imaging.

Just like color (wavelength), *polarization* is a fundamental property of light. It encodes

*Corresponding author: Julien Flamant.

Funding: This work was funded by CNRS and GdR ISIS OPENING exploratory research project grant.

†CNRS, Université de Lorraine, CRAN, F-54000 Nancy France (julien.flamant@cnrs.fr, konstantin.usevich@univ-lorraine.fr, david.brie@univ-lorraine.fr).

‡Université de Lorraine, CNRS, IECL, F-54000 Nancy France (marianne.clausel@univ-lorraine.fr)

the geometry of oscillations of the electromagnetic field, which describes an ellipse in the 2D plane perpendicular to the propagation direction for vacuum-like media [20]. As polarized light propagates in media, its polarization can change, thus revealing key properties, such as medium anisotropy or structural properties that are inaccessible to conventional, non-polarized light [30]. As a result, polarized light imaging has found many applications such as material characterization [34], remote sensing [70] or bio-imaging [37]. Despite the important practical interests of polarization, only a few authors have considered leveraging this fundamental attribute of light in phase retrieval problems. The authors in [64, 59] pioneered the use of polarization in Fourier phase retrieval for ultrashort attosecond (10^{-18} s) laser pulse characterization. The motivation for polarimetric measurements arises from a fundamental physical limitation, which prevents the direct use of standard pulse characterization strategies based on nonlinear light-matter interaction such as Frequency-Resolved Optical Gating (FROG) [69] and its variants. Another line of work regards the extension of a scanning coherent diffraction imaging technique, known as *ptychography*, to take into account the polarization of light. This novel imaging modality, called *vectorial ptychography* [26, 27] combines spatially redundant measurements with polarimetric measurements. This allows quantitative imaging of complex anisotropic media, such as biominerals [5, 6]. More generally, recent years have seen a growing interest in the experimental development of computational imaging systems exploiting vectorial and polarization properties of light, such as polarization-sensitive Fourier ptychography [22, 67], polarization-sensitive diffraction tomography [68, 60] or vectorial holography [66], among others.

Related work. Fourier phase retrieval is a long standing problem and therefore has generated a continuous interest from researchers of various horizons, leading to a vast literature ranging from theoretical results to practical imaging algorithms, see [12] for an overview. A recent survey of uniqueness and stability of Fourier phase retrieval can be found in [36]; see also [13] for a discussion of its algebraic properties. A comprehensive tour of existing algorithms is given in [25]; see also [4] for an extensive discussion of related geometric aspects.

One-dimensional Fourier phase retrieval does not admit a unique solution in general [10]. Therefore, many strategies to enforce uniqueness have been devised. These include additional information on the signal, such as knowledge of some entries [11], non-negativity [8], sparsity [42, 56] or minimum phase [38]. Another approach consists in generating additional measurements, e.g., using deterministic masks [39, 15], (randomly) coded diffraction patterns [16] or using redundant, overlapping measurements inspired by ptychography [14, 40].

More closely related to the present work is the use of additional, interference-like measurements in Fourier phase retrieval. The main idea roots in a imaging technique known as *holography*, which involves the coherent interference of the object of interest $\mathbf{x}$ with some reference signal $\mathbf{y}$. Pushing this idea further, authors have developed a strategy ensuring uniqueness in Fourier phase retrieval, called *vectorial phase retrieval* [57] or double-blind holography [45, 58, 54]. More precisely, they show (and exploit) that almost all signals $\mathbf{x}$ and $\mathbf{y}$ can be recovered from four Fourier magnitudes measurements, of $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x} + \jmath\mathbf{y}$ (with $\jmath^2 = -1$) and $\mathbf{x} + \mathbf{y}$, respectively. Similar ideas appear in [41], where the reconstruction problem is formulated using correlations functions instead of Fourier transforms.

While these works share several features with the present paper, they also differ on a number of important points. First, they do not exploit a polarimetric acquisition scheme,

which limits their use in contexts where one in interested in reconstructing the polarized (or bivariate) electromagnetic field (such as in polarized coherent diffraction imaging techniques [65]). In particular, we will show that the proposed polarimetric Fourier phase retrieval model encompasses vectorial phase retrieval as a special case, for a specific choice of *four polarimetric projections*. In addition, while the connection between vectorial phase retrieval and greatest common divisor of polynomials was observed in [41], it was not investigated in detail as the authors focused on a semidefinite programming relaxation. In contrast, algebraic approaches based on greatest common divisor computations are a cornerstone of the proposed methodology for the polarimetric Fourier phase retrieval model.

Contributions. This work introduces a novel Fourier phase retrieval model, called *polarimetric Fourier phase retrieval* (PPR), which takes advantage of the physical measurement of polarization properties in optics. In particular, measurements are readily interpreted in terms of polarimetric Fourier projections of the bivariate electromagnetic field. As such, the proposed model can be implemented using standard optical components, such as polarizers or waveplates. It is flexible: more polarimetric measurements can be performed if desired. We focus on the 1D Fourier case in this paper, as a first step to demonstrate the potential of polarization information in Fourier phase retrieval problems. First, we characterize its uniqueness properties by carefully establishing equivalences with two other problems, namely bivariate Fourier phase retrieval (BPR) and polynomial autocorrelation factorization (PAF). In particular, we show that the PPR problem can be solved through algebraic methods based on approximate greatest common divisor computations. We compare in detail these approaches with tailored adaptations of standard iterative algorithms for Fourier phase retrieval, namely semidefinite positive relaxation and Wirtinger-Flow, to the case of PPR. Finally, numerical experiments demonstrate that combining algebraic and iterative approaches yields a scalable, computationally efficient and robust to noise reconstruction strategy for PPR.

Organization of the paper. A crucial feature of the present paper is the extensive use of equivalences between the polarimetric Fourier phase retrieval (PPR) problem and two other problems, namely bivariate Fourier phase retrieval (BPR) and polynomial autocorrelation factorization (PAF). For reference, these equivalences are stated in Figure 1, with pointers to relevant definitions and equations. Section 2 introduces the PPR model and discusses its physical interpretations in terms of polarimetric measurement. Under some very general conditions, the equivalence with BPR is then established, which permits the study of trivial ambiguities. The relation of PPR with a standard 1D Fourier phase retrieval problem is also discussed. Section 3 starts by reformulating the BPR problem using a polynomial representation, leading to PAF. Then, we leverage uniqueness results on multivariate spectral representations [71] to establish a necessary and sufficient characterization of uniqueness in PAF (Theorem 3.5). Corollary 3.6 states that PAF is almost everywhere unique, and as a result, an algebraic solution can be found using greatest common divisors of measurement polynomials (Proposition 3.7). Section 4 goes back to PPR and exploits uniqueness results to propose a fully algebraic reconstruction method for PPR (Algorithm 1) based on two variations of approximate greatest common divisor computations. Section 5 focus instead on iteratives algorithms for PPR, by tailoring semidefinite relaxation (Algorithm 4) and Wirtinger Flow (Algorithm 5). Section 6 presents several numerical experiments to illustrate and assess the practical performances of the proposed reconstruction strategies. Section 7 collects concluding remarks

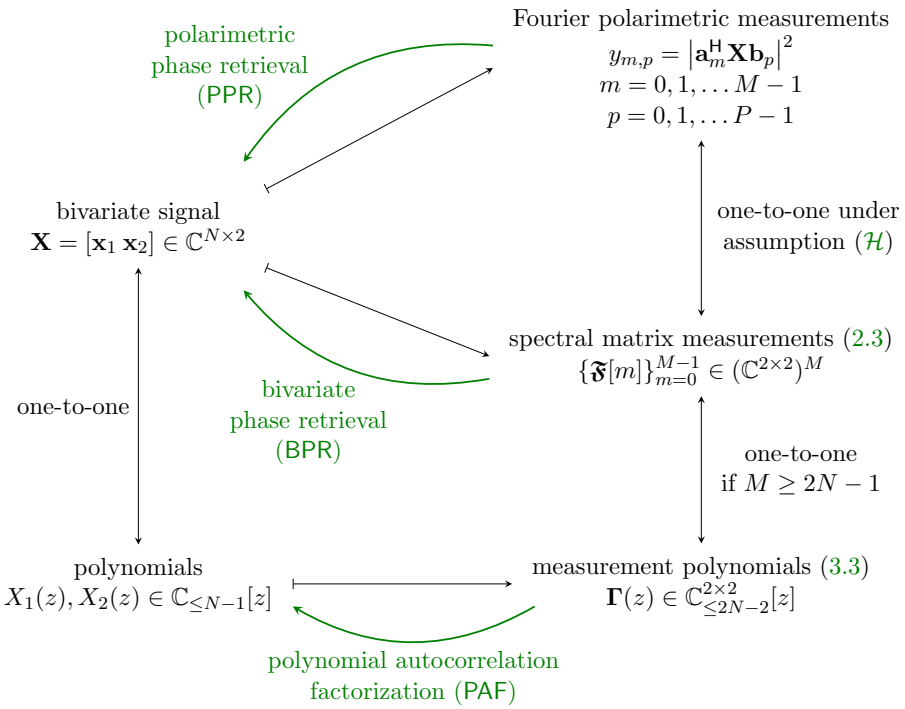

Figure 1. *Equivalences of data and solutions in problems* PPR, BPR *and* PAF.

and Appendices gather technical details and proofs.

Notations. In this paper, we denote by $\mathbb{R}$ the set of real numbers and by $\mathbb{C}$ the set of complex numbers with imaginary unit $\jmath$ such that $\jmath^2 = -1$. Vectors and matrices are denoted in bold lowercase letters and bold capital letters, respectively. Dependence of quantities in terms of a discrete index are indicated by brackets, i.e., $\mathbf{x}[n]$ denotes the n-th entry of the set of vectors $\{\mathbf{x}[n]\}_{n=0}^{N-1}$. Notation $\mathbf{a}^*, \mathbf{A}^*$ indicate the complex conjugate of vector $\mathbf{a}$ and matrix $\mathbf{A}$, respectively. The transpose of a matrix $\mathbf{A}$ is $\mathbf{A}^\top$ and its conjugate transpose is given by $\mathbf{A}^{\mathsf{H}}$. Fourier domain quantities are denoted using capital gothic letters, i.e., the vector $\mathfrak{X}[m] \in \mathbb{C}^2$ denotes the m-th entry of the (one-dimensional) discrete Fourier transform of the vector signal $\{\mathbf{x}[n] \in \mathbb{C}^2\}_{n=0}^{N-1}$, evaluated at a frequency indexed by integer m.

2. Polarimetric Fourier phase retrieval model.

For conciseness, we use from now on the term *phase retrieval* as a synonym for Fourier phase retrieval.

2.1. General formulation.

Consider a discrete bivariate signal $\mathbf{x}[n] = (x_1[n], x_2[n])^\top \in \mathbb{C}^2$ defined for $n = 0, 1, \dots N-1$. Let $\mathbf{X} \in \mathbb{C}^{N\times 2}$ be the matrix representation of $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ obtained by stacking samples row-wise such that

$$\mathbf{X} = \begin{bmatrix} x_1[0] & x_2[0] \\ x_1[1] & x_2[1] \\ \vdots & \vdots \\ x_1[N-1] & x_2[N-1] \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 \end{bmatrix}, \tag{2.1}$$

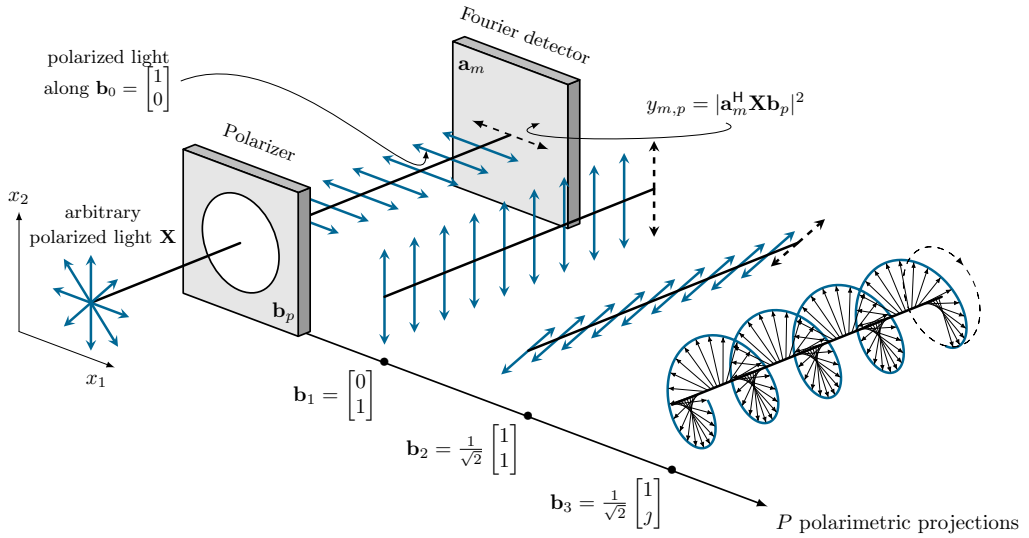

Figure 2. *Physical interpretation of the polarimetric phase retrieval model* (PPR) *in terms of polarization optics. The four polarimetric projections shown correspond to the standard measurement scheme described by* (2.4) *and* (2.5)*, see Example* 1*.*

where $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^N$ collect the two vector components of the signal. We define the *polarimetric (Fourier) phase retrieval* (PPR) problem as the recovery of $\mathbf{X}$ given MP Fourier polarimetric projections. Formally,

$$\text{(PPR)} \qquad \text{find } \mathbf{X} \in \mathbb{C}^{N\times 2} \text{ given measurements } y_{m,p} = \left|\mathbf{a}_m^{\mathsf{H}} \mathbf{X} \mathbf{b}_p\right|^2, \quad m = 0, 1, \dots M-1, \quad p = 0, 1, \dots P-1$$

where $\mathbf{a}_m \in \mathbb{C}^N$ is the discrete Fourier vector corresponding to frequency $f_m = (2\pi m)/M$, such that $a_m[n] = \exp[jnf_m]$ for $n = 0, 1, \dots N-1$. The vector $\mathbf{b}_p \in \mathbb{C}^2$, normalized such that $\|\mathbf{b}_p\|_2^2 = 1$, denotes an arbitrary projection acting on the two vector components of $\mathbf{X}$.

Figure 2 permits to attach precise physical interpretations of PPR measurements in terms of polarization optics. The matrix $\mathbf{X}$ represents the one-dimensional bivariate electromagnetic field, where each row is a vector of $\mathbb{C}^2$ describing an arbitrary polarization state (the so-called Jones vector [30]). This states passes through a polarizer defined by $\mathbf{b}_p \in \mathbb{C}^2$, evaluating the projection of polarization states of $\mathbf{X}$ onto $\mathbf{b}_p$. Finally, light impinges on a Fourier detector described by $\mathbf{a}_m \in \mathbb{C}^N$, leading to squared magnitude PPR measurements $y_{m,p}$.

The measurement model PPR can be easily implemented experimentally. Indeed, Fourier vectors $\{\mathbf{a}_m\}_{m=0}^{M-1}$ correspond to far-field measurements in optics, as encountered in coherent diffraction imaging techniques (for the case of 2D/3D images) or in spectrometry (for the 1D case of ultra-short pulses). On the other hand, the set $\{\mathbf{b}_p\}_{p=0}^{P-1}$ describes the different polarizers (or polarization analysers) required to measure polarization of light. Any arbitrary polarizer (in mathematical terms, any unit-norm vector $\mathbf{b}_p \in \mathbb{C}^2$) can be constructed as as combination of standard optical components, such as linear polarizers or waveplates [30].

Therefore, polarimetric measurements are very flexible: their number, as well as the reference polarization states $\{\mathbf{b}_p\}_{p=0}^{P-1}$ can be tailored at will depending on the context.

2.2. Relation with Fourier matrix measurements. A closely related problem to PPR is the *bivariate phase retrieval* (BPR) problem. Let us introduce the discrete Fourier transform of the bivariate signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ as

$$\mathfrak{X}[m] = \sum_{n=0}^{N-1} \mathbf{x}[n] \exp\left(-2\pi\jmath \frac{mn}{M}\right) = \begin{bmatrix} \mathfrak{X}_1[m] \\ \mathfrak{X}_2[m] \end{bmatrix} = (\mathbf{a}_m^{\mathsf{H}} \mathbf{X})^\top \in \mathbb{C}^2 \tag{2.2}$$

for $m = 0, 1, \dots M-1$. Then let $\mathfrak{F}[m]$ denote the rank-1 complex spectral matrix such that

$$\mathfrak{F}[m] = \mathfrak{X}[m]\mathfrak{X}[m]^{\mathsf{H}} = \begin{bmatrix} |\mathfrak{X}_1[m]|^2 & \mathfrak{X}_1[m]\mathfrak{X}_2[m]^* \\ \mathfrak{X}_2[m]\mathfrak{X}_1[m]^* & |\mathfrak{X}_2[m]|^2 \end{bmatrix} \in \mathbb{C}^{2\times 2}. \tag{2.3}$$

At a given frequency indexed by m, the spectral matrix $\mathfrak{F}[m]$ collects the squared Fourier amplitudes of the two components $\mathbf{x}_1$ and $\mathbf{x}_2$ of the bivariate signal as well as their relative Fourier phase. The recovery of the original bivariate signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ (or equivalently its matrix representation $\mathbf{X}$) from its spectral matrices defines the BPR problem:

$$\text{(BPR)} \qquad \text{find } \mathbf{X} \in \mathbb{C}^{N\times 2} \text{ given spectral matrix measurements } \{\mathfrak{F}[m]\}_{m=0}^{M-1}.$$

The following proposition shows that BPR and PPR are equivalent in the noiseless setting under very general assumptions on the projection vectors $\{\mathbf{b}_p\}_{p=0}^{P-1}$.

Proposition 2.1 (Equivalence between BPR and PPR). *Suppose that the collection of projection vectors $\mathbf{b}_0, \mathbf{b}_1, \dots \mathbf{b}_{P-1} \in \mathbb{C}^2$ satisfies the condition*

$$(\mathcal{H}) \qquad \operatorname{span}_{\mathbb{R}} \left\{ \mathbf{b}_p \mathbf{b}_p^{\mathsf{H}} \right\}_{p=0}^{P-1} = \left\{ \mathbf{M} \in \mathbb{C}^{2\times 2} \mid \mathbf{M}^{\mathsf{H}} = \mathbf{M} \right\},$$

i.e., , the set of P rank-1 matrices $\mathbf{b}_p\mathbf{b}_p^{\mathsf{H}}$ is a generating family (over $\mathbb{R}$) of the space of 2-by-2 Hermitian matrices. Then, under assumption $(\mathcal{H})$, the problem PPR is equivalent to BPR in the sense that $\mathbf{X}$ is a solution of the problem PPR if and only if $\mathbf{X}$ is solution of BPR.

Proof. It is sufficient to show that, under assumption $(\mathcal{H})$, there is a one-to-one correspondence between the data of BPR (spectral matrices $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$) and that of PPR (Fourier polarimetric measurements $\{y_{m,p}\}_{m,p=0}^{M-1,P-1}$). In particular, we prove that for m fixed, the spectral matrix $\mathfrak{F}[m]$ can be obtained from $\{y_{m,p}\}_{p=0}^{P-1}$ and vice-versa. First, remark that

$$y_{m,p} = |\mathbf{a}_m^{\mathsf{H}} \mathbf{X} \mathbf{b}_p|^2 = \mathfrak{X}[m]^\top \mathbf{b}_p \mathbf{b}_p^{\mathsf{H}} \mathfrak{X}^*[m] = \operatorname{Tr} \mathbf{b}_p^* \mathbf{b}_p^\top \mathfrak{F}[m],$$

i.e., measurements $y_{m,p}$ are linear measurements of $\mathfrak{F}[m]$ through sensing matrices $\{\mathbf{b}_p^*\mathbf{b}_p^\top\}_{p=0}^{P-1}$. Conversely, since $\left\{\mathbf{b}_p\mathbf{b}_p^{\mathsf{H}}\right\}_{p=0}^{P-1}$ (and equivalently, $\{\mathbf{b}_p^*\mathbf{b}_p^\top\}_{p=0}^{P-1}$) is a generating family of the space of 2-by-2 Hermitian by matrices by assumption $(\mathcal{H})$, the spectral matrix $\mathfrak{F}[m]$ can be uniquely determined from $\{y_{m,p}\}_{p=0}^{P-1}$ by linear combinations. This concludes the proof. ∎

It is worth noting that the assumption ($\mathcal{H}$) is not restrictive at all. In fact, for $P \geq 4$, the set $\{\mathbf{b}_p\}_{p=0}^{P-1}$ where vectors are i.i.d. Gaussian distributed on $\mathbb{C}^2$ almost surely satisfies ($\mathcal{H}$). The following example gives an explicit choice of projection vectors $\mathbf{b}_p$ for $P = 4$, which has a nice physical interpretation in terms of polarization optics.

Example 1. *Let $P = 4$ and consider the following projection vectors*

$$\mathbf{b}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \mathbf{b}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \ \mathbf{b}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \ \mathbf{b}_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ \jmath \end{bmatrix}. \tag{2.4}$$

The projection vectors $\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2$ and $\mathbf{b}_3$ correspond to Jones vectors of standard polarizers used in optics [20], which are, respectively: horizontal linear polarizer, vertical linear polarizer, 45° linear polarizer and left circular polarizer. See Figure 2 for an illustration. A direct check shows that rank-one matrices $\mathbf{b}_0\mathbf{b}_0^{\mathsf{H}}$, $\mathbf{b}_1\mathbf{b}_1^{\mathsf{H}}$, $\mathbf{b}_2\mathbf{b}_2^{\mathsf{H}}$, $\mathbf{b}_3\mathbf{b}_3^{\mathsf{H}}$ form a basis over the real vector space of 2-by-2 Hermitian matrices, and as a result, they are a generating family of such matrices. PPR *measurements read explicitly*

$$\begin{aligned} y_{m,0} &= |\mathfrak{X}_1[m]|^2, \quad y_{m,1} = |\mathfrak{X}_2[m]|^2, \\ y_{m,2} &= \frac{1}{2}|\mathfrak{X}_1[m] + \mathfrak{X}_2[m]|^2, \quad y_{m,3} = \frac{1}{2}|\mathfrak{X}_1[m] + \jmath\mathfrak{X}_2[m]|^2. \end{aligned} \tag{2.5}$$

These expressions directly give the diagonal terms of $\mathfrak{F}[m]$ as $y_{m,0}$ and $y_{m,1}$. The off-diagonals terms can be recovered easily using polarization identities in the complex case, such that

$$\begin{aligned} \operatorname{real}(\mathfrak{X}_1[m]\mathfrak{X}_2[m]^*) &= \frac{1}{2}\left(|\mathfrak{X}_1[m] + \mathfrak{X}_2[m]|^2 - |\mathfrak{X}_1[m]|^2 - |\mathfrak{X}_2[m]|^2\right) \\ &= y_{m,2} - \frac{1}{2}(y_{m,0} + y_{m,1}), \\ \operatorname{imag}(\mathfrak{X}_1[m]\mathfrak{X}_2[m]^*) &= \frac{1}{2}\left(|\mathfrak{X}_1[m] + \jmath\mathfrak{X}_2[m]|^2 - |\mathfrak{X}_1[m]|^2 - |\mathfrak{X}_2[m]|^2\right) \\ &= y_{m,3} - \frac{1}{2}(y_{m,0} + y_{m,1}). \end{aligned}$$

Remark that the measurement scheme (2.4) yields the same quadratic measurements (2.5) as proposed by several authors [57, 41, 45, 58, 54]. Because of that, BPR is equivalent to the vectorial phase retrieval problem originally introduced in [57]. This shows that PPR encompasses existing measurements strategies as a special case, while bringing extra flexibility in the experimental design of measurements. One of the key benefits of the PPR model is that additional polarimetric measurements can be generated at will using simple off-the-shelf optical components such as linear polarizers or waveplates.

2.3. Trivial ambiguities. Thanks to Proposition 2.1, we can now give a characterization of trivial ambiguities of PPR model by leveraging the equivalent BPR problem. Indeed, one can investigate in a rather simple way the trivial ambiguities that characterize BPR. Formally, these trivial ambiguities correspond to elementary transformations $\{\mathbf{x}[n]\}_{n=0}^{N-1} \to \{\mathbf{x}'[n]\}_{n=0}^{N-1}$ that leave BPR measurements (spectral matrices $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$ defined in (2.3)) unchanged.

Global phase ambiguity. Let $\alpha \in \mathbb{R}$ and consider the bivariate signal $\{\mathbf{x}'[n]\}_{n=0}^{N-1}$ such that $\mathbf{x}'[n] = \exp(\jmath\alpha)\mathbf{x}[n]$ for every n. Then for any m, $\mathfrak{F}'[m] = \mathfrak{X}'[m]\mathfrak{X}'[m]^{\mathsf{H}} = \mathfrak{X}[m]\mathfrak{X}[m]^{\mathsf{H}} = \mathfrak{F}[m]$ since $\mathfrak{X}'[m] = \exp(\jmath\alpha)\mathfrak{X}[m]$ by linearity properties of the Fourier transform.

Shifts. This trivial ambiguity only appears when the bivariate signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ has not full support, i.e., when there exist n_a, n_b with $0 \leq n_a \leq n_b \leq N-1$ such that $\mathbf{x}[n] = \mathbf{0}$ for $n \leq n_a$ and $n \geq n_b$. Assuming this is the case, define the shifted signal $\{\mathbf{x}'[n]\}_{n=0}^{N-1}$ as $\mathbf{x}'[n] = \mathbf{x}[n+n_0]$ where n_0 is a relative integer between $(n_b - N)$ and $(n_a + 1)$ as to ensure proper support. Then, using standard Fourier transform properties one gets that $\mathfrak{X}'[m] = \exp(\jmath 2\pi n_0 m/M)\mathfrak{X}[m]$, so that in turn $\mathfrak{F}'[m] = \mathfrak{F}[m]$ for every m.

Conjugate reflection. Consider now $\{\mathbf{x}'[n]\}_{n=0}^{N-1}$ such that $\mathbf{x}'[n] = \mathbf{x}^*[N-1-n]$. Then for every m, $\mathfrak{X}'[m] = \exp[-\jmath 2\pi(N-1)m/M]\mathfrak{X}^*[m]$. As a result

$$\mathfrak{F}'[m] = \begin{bmatrix} |\mathfrak{X}_1[m]|^2 & \mathfrak{X}_2[m]\mathfrak{X}_1^*[m] \\ \mathfrak{X}_1[m]\mathfrak{X}_2^*[m] & |\mathfrak{X}_2[m]|^2 \end{bmatrix} = \mathfrak{F}[m]^\top. \tag{2.6}$$

This shows that conjugate reflection is not, in general, a trivial ambiguity for BPR. This contrasts with standard univariate Fourier phase retrieval, see [10, 12].

Conjugate reflection can still be a trivial ambiguity provided that the spectral matrix is symmetric for every m, that is $\mathfrak{F}[m] = \mathfrak{F}[m]^\top$. Equivalently, $\mathfrak{F}[m]$ is symmetric if and only if $\mathfrak{X}_1[m]\mathfrak{X}_2^*[m] = \mathfrak{X}_2[m]\mathfrak{X}_1^*[m]$. This means that $\operatorname{imag}(\mathfrak{X}_1[m]\mathfrak{X}_2^*[m]) = 0$, i.e., components $\mathfrak{X}_1[m]$, $\mathfrak{X}_2[m]$ are in phase at every frequency (they have the same complex argument). Interestingly, this condition is interpreted in physical terms as: conjugate reflection is a trivial ambiguity for bivariate phase retrieval if and only if the bivariate signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ is linearly polarized at all frequencies.

2.4. 1D equivalent model for PPR.

Back to the original PPR problem, we see that it defines a new measurement model that performs quadratic scalar projections of the matrix representation $\mathbf{X} \in \mathbb{C}^{N\times 2}$ of the bivariate signal of interest. This *matrix representation* of the underlying signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ can be confusing at first: indeed, the bivariate signal is intrinsically one-dimensional, in the sense that it is a function of a single index n – which can represent time or 1D spatial coordinates, for instance. Thus, a natural question is the following: can PPR be equivalently rewritten as a one-dimensional phase retrieval problem? If so, what is the physical interpretation of such problem?

Let us denote by $\boldsymbol{\xi} = \operatorname{vec}\mathbf{X} \in \mathbb{C}^{2N}$ the long vector obtained by stacking the two columns of $\mathbf{X}$. Using standard vectorization properties of matrix products, one can rewrite PPR measurements as

$$y_{m,p} = |\mathbf{a}_m^{\mathsf{H}}\mathbf{X}\mathbf{b}_p|^2 = |(\mathbf{b}_p^\top \otimes \mathbf{a}_m^{\mathsf{H}})\boldsymbol{\xi}|^2 = |(\mathbf{b}_p^* \otimes \mathbf{a}_m)^{\mathsf{H}}\boldsymbol{\xi}|^2 \tag{2.7}$$

for $m = 0, 1, \ldots M-1$, $p = 0, 1, \ldots P-1$ and where $\mathbf{a} \otimes \mathbf{b}$ stands for the Kronecker product of vectors $\mathbf{a}$ and $\mathbf{b}$. Letting $\mathbf{c}_{m,p} = \mathbf{b}_p^* \otimes \mathbf{a}_m \in \mathbb{C}^{2N}$, the PPR problem is equivalent to

$$\text{(PPR-1D)} \qquad \begin{aligned} &\text{find } \boldsymbol{\xi} \in \mathbb{C}^{2N} \text{ given measurements } y_{m,p} = \left|\mathbf{c}_{m,p}^{\mathsf{H}}\boldsymbol{\xi}\right|^2 \\ &\qquad\qquad m = 0, 1, \ldots M-1, \quad p = 0, 1, \ldots P-1 \end{aligned}.$$

This shows that PPR can be rewritten as a specific instance of 1D phase retrieval with structured measurements vectors $\mathbf{c}_{m,p} \in \mathbb{C}^{2N}$. While being mathematically sound, the equivalent PPR-1D problem brings almost no insights about the bivariate nature of the signal to be

recovered. Moreover, PPR-1D cannot be interpreted as a Fourier phase retrieval problem with masks [3, 39], since measurements vectors $\mathbf{c}_{m,p}$ intertwine Fourier measurements $\mathbf{a}_m$ and polarimetric projections $\mathbf{b}_p$ using Kronecker products. Thus, the study of the theoretical properties of PPR cannot be inferred from standard phase retrieval properties applied to PPR-1D. This requires a dedicated study, which is described in detail in Section 3 and exploited in Section 4 to formulate algebraic solutions to the PPR problem. Nonetheless, as we shall see in Section 5, the equivalent formulation PPR-1D remains particularly useful for designing (iterative) algorithms to solve the original PPR problem.

3. Uniqueness and polynomial formulation.

This section studies the uniqueness properties of noiseless PPR under the set of assumptions ($\mathcal{H}$) defined in Section 2.2. Thanks to Proposition 2.1, we see that any solution of the problem PPR is a solution of the problem BPR, and vice-versa. This formal equivalence permits to study uniqueness properties of the original PPR through BPR. Following standard practice in Fourier phase retrieval problems, Section 3.1 reformulates BPR using a polynomial formalism. Theorem 3.2 shows that under the usual oversampling condition $M \geq 2N-1$, BPR is equivalent to a polynomial autocorrelation factorization (PAF) problem. Section 3.2 then provides general uniqueness results for PAF and demonstrates that it can be solved using simple greatest common divisor computations.

3.1. Bivariate phase retrieval as a polynomial factorization problem.

This section follows standard practice in Fourier phase retrieval problems [10, 12, 8, 11, 9] and adopts the polynomial representation of Fourier transforms to study the uniqueness properties of the BPR problem. Formally, let $\mathbb{C}_{\leq N-1}[z]$ be the space of polynomials of degree at most $N-1$. First, let us define the polynomials $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ as generating polynomials of the components of the bivariate signal $\mathbf{x}[n] = (x_1[n], x_2[n])^\top \in \mathbb{C}^2$, $n = 0, 1, \dots N-1$

$$X_1(z) = \sum_{n=0}^{N-1} x_1[n] z^n, \quad X_2(z) = \sum_{n=0}^{N-1} x_2[n] z^n. \tag{3.1}$$

Similarly, define their conjugate reflections $\widetilde{X}_1, \widetilde{X}_2 \in \mathbb{C}_{\leq N-1}[z]$, obtained by reversing the order and conjugating the coefficients of $X_1(z)$ and $X_2(z)$:

$$\widetilde{X}_1(z) = \sum_{n=0}^{N-1} x_1^*[N-n-1] z^n, \quad \widetilde{X}_1(z) = \sum_{n=0}^{N-1} x_2^*[N-n-1] z^n. \tag{3.2}$$

Then we define the following matrix polynomial $\mathbf{\Gamma} \in \mathbb{C}^{2\times 2}_{\leq 2N-2}[z]$

$$\mathbf{\Gamma}(z) = \begin{bmatrix} \Gamma_{11}(z) & \Gamma_{12}(z) \\ \Gamma_{21}(z) & \Gamma_{22}(z) \end{bmatrix} = \begin{bmatrix} X_1(z)\widetilde{X}_1(z) & X_1(z)\widetilde{X}_2(z) \\ X_2(z)\widetilde{X}_1(z) & X_2(z)\widetilde{X}_2(z) \end{bmatrix} = \begin{bmatrix} X_1(z) \\ X_2(z) \end{bmatrix} \begin{bmatrix} \widetilde{X}_1(z) & \widetilde{X}_2(z) \end{bmatrix}, \tag{3.3}$$

where each element of the matrix is a polynomial $\Gamma_{ij} \in \mathbb{C}_{\leq 2N-2}[z]$. The coefficients of these polynomials are simply the covariance functions (auto-covariances and cross-covariances) of the vector components $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^N$ that define the bivariate signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$. Moreover, the spectral matrices $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$ of BPR are linked to the evaluations of the polynomial $\mathbf{\Gamma}(z)$.

Lemma 3.1. *The coefficients Γ_{ij} of the matrix polynomial $\mathbf{\Gamma} \in \mathbb{C}^{2\times 2}_{\leq 2N-2}[z]$ are given by*

$$\Gamma_{ij}(z) = \sum_{n=0}^{2N-2} \gamma_{ij}[n-N+1]z^n \text{ with } \gamma_{ij}[n] = \sum_{k\in\mathbb{Z}} x_i[k+n]x_j^*[k], \tag{3.4}$$

where $x_i[n] = 0$ for $n < 0$ and $n \geq N$ by convention, and the covariance functions $\gamma_{ij}[n]$ are defined for $n = -N+1, \ldots, N-1$. Moreover, the spectral matrices $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$ of BPR can be expressed for $m = 0, 1, \ldots M-1$ as

$$\mathfrak{F}[m] = e^{\jmath 2\pi \frac{m(N-1)}{M}} \mathbf{\Gamma}(e^{-\jmath 2\pi \frac{m}{M}}). \tag{3.5}$$

Lemma 3.1 extends to the bivariate case the well-known correspondence between autocovariance polynomials and Fourier amplitude in univariate Fourier phase retrieval (see for instance [10, 12]). For completeness, we give a formal proof in Appendix A.

We will refer to $\mathbf{\Gamma}(z)$ and its entries $\Gamma_{ij}(z)$ as *measurement polynomials*. Eq. (3.5) shows that the coefficients of $\Gamma_{ij} \in \mathbb{C}_{\leq 2N-2}[z]$ can be uniquely identified from the spectral matrix measurements $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$ of BPR provided that the number of Fourier measurements M exceeds the degree of these polynomials by at least one, i.e.,

$$M \geq 2N-1. \tag{3.6}$$

This is the well-known oversampling condition in standard univariate Fourier phase retrieval, see e.g. [12]. As a result, one can establish the equivalence between BPR and a polynomial recovery problem called Polynomial Autocorrelation Factorization (PAF).

Theorem 3.2. *For $M \geq 2N-1$, BPR is equivalent to the following problem*

(PAF) find $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ given measurement polynomial $\mathbf{\Gamma}(z)$ *defined as* (3.3) .

In other terms, there is a one-to-one correspondence between the data ($\mathbf{\Gamma}(z)$ and $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$) as well as the sets of solutions of the problems (polynomials $X_1(z), X_2(z)$ and bivariate signal components $\mathbf{x}_1, \mathbf{x}_2$).

Appendix A provides a proof of this result for completeness. Figure 1 summarizes this equivalence between BPR and PAF problems, and recall how data and solutions of respective problems connect to the initial PPR problem.

3.2. General uniqueness result. The PAF formulation is very helpful for establishing the uniqueness conditions of BPR and, in turn, that of PPR under the nonrestrictive assumption ($\mathcal{H}$). Notably, PAF enables a complete characterization of uniqueness properties in terms of algebraic properties of complex polynomials. To simplify the presentation in the following, uniqueness properties refer jointly to PPR, BPR and PAF problems. In this section, we reproduce several important results from [71] regarding the uniqueness of polynomial autocorrelation factorizations problems. The notion of greatest common divisor (GCD) of complex polynomials plays a pivotal role in establishing and interpreting these statements. In what follows, we use the following definition of the GCD of polynomials taken from [71, 72].

Definition 3.3. *Consider two polynomials $A_1, A_2 \in \mathbb{C}_{\leq D}[z]$, where at least one of them is nonzero. The GCD of $A_1(z)$ and $A_2(z)$ is a polynomial $Q(z) \in \mathbb{C}_{\leq K}[z]$, with highest possible K, such that there exists two polynomials $R_1, R_2 \in \mathbb{C}_{\leq D-K}[z]$ satisfying*

$$A_1(z) = Q(z)R_1(z) \text{ and } A_2(z) = Q(z)R_2(z), \tag{3.7}$$

and $Q(z)$ and has exactly $\min(L_1, L_2)$ zero leading coefficients (where L_1 and L_2 is the number of zero leading coefficients in A_1 and A_2 respectively).

Remark 3.4. The GCD exists and is unique up to a multiplication by a scalar in $\mathbb{C}\backslash\{0\}$. Therefore by writing $Q(z) = \gcd(A_1, A_2)$ we mean that $Q(z)$ is a GCD up to this ambiguity. All the usual properties of the GCD apply, despite the special treatment of zero leading coefficients, see [71]. In particular, the polynomials $R_1(z)$ and $R_2(z)$, called quotient polynomials, satisfy $\gcd(R_1, R_2) = 1$ and are said to be *co-prime*. Extension of the notion of GCD to multiple polynomials is straightforward.

Theorem 3.5 ([71]). *The following equivalences are true:*

1. PAF *admits a unique solution (up to trival ambiguities);*
2. *$H(z) = \gcd(\Gamma_{11}, \Gamma_{12}, \Gamma_{21}, \Gamma_{22})$ has no roots outside the unit circle;*
3. *$X_1(z)$ and $X_2(z)$ have no common roots outside the unit circle and the leading coefficient of $\gcd(X_1, X_2)$ is nonzero.*

The proof of this result can be found in [71], where the generalization of PAF to the case of R polynomials is considered. Note that the uniqueness condition given in Theorem 3.5 clarifies previous statements made in the literature [57, 41]. In particular, in [57, Theorem 1] it was claimed that coprimeness of the polynomials $X_1(z)$ and $X_2(z)$ was a necessary and sufficient for uniqueness of the solution. Theorem 3.5 shows that it was just a sufficient condition, because unimodular roots do not affect uniqueness. This agrees with a similar behavior observed for univariate one-dimensional Fourier phase retrieval [10], where unimodular roots do not contribute to the number of non-trivial solutions. However, unlike univariate one-dimensional Fourier phase retrieval, the bivariate case is almost everywhere unique, as shown in the following corollary.

Corollary 3.6 ([71]). *The PAF problem admits a unique solution for almost every polynomials $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$.*

The proof essentially comes down to observing that the set of polynomials $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ with at least one common root is an algebraic variety of dimension at most $2N-1$; hence it is of measure zero. Put it differently, this shows that PAF has the appealing property that almost all polynomials $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ can be uniquely recovered from measurement polynomials $\Gamma_{11}(z), \Gamma_{12}(z), \Gamma_{21}(z)$ and $\Gamma_{22}(z)$.

In practice, if one picks polynomials $X_1(z)$ and $X_2(z)$ at random from some continuous probability distribution, then they can be almost surely uniquely recovered through PAF. Moreover, they are almost surely co-prime, i.e., $\gcd(X_1, X_2) = 1$. In this very general case, the following proposition shows that recovery is possible through simple GCD computations.

Proposition 3.7 (GCD-based recovery). *Let $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ such that $\gcd(X_1, X_2) = 1$. Then $X_1(z)$ and $X_2(z)$ can be uniquely recovered as*

$$X_1(z) = c_1 \gcd(\Gamma_{11}, \Gamma_{12}) \text{ and } X_2(z) = c_2 \gcd(\Gamma_{21}, \Gamma_{22}). \tag{3.8}$$

where $c_1, c_2 \in \mathbb{C}$ can be determined explicitely (up to one global phase) from measurement polynomials.

Proof. Suppose that $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ such that $\gcd(X_1, X_2) = 1$. This implies that $\gcd(\widetilde{X}_1, \widetilde{X}_2) = 1$. Therefore, $\gcd(\Gamma_{11}, \Gamma_{12}) = \gcd(X_1\widetilde{X}_1, X_1\widetilde{X}_2) = c_1 X_1(z)$ since $\widetilde{X}_1(z)$ and $\widetilde{X}_2(z)$ are co-prime. The same argument yields $\gcd(\Gamma_{21}, \Gamma_{22}) = c_2 X_2(z)$. Constants c_1 and c_2 can be determined from the knowledge of correlation functions $\gamma_{ij}[n]$, or equivalently, measurement polynomials $\Gamma_{ij}(z)$; see Section 4.3 for explicit expressions. ∎

Proposition 3.7 is a central result. It indicates that the PAF problem, and by extension, BPR and PPR can be solved using polynomial algebraic techniques. This distinctive feature arises as a direct consequence of accounting for polarization in Fourier phase retrieval problems. This original direction is further explored in Section 4, where we devise algebraic approaches to solve the noisy PPR problem using approximate GCD computations.

4. Solving PPR with algebraic methods. A central result of the previous section is Proposition 3.7, which states that polynomials $X_1(z)$ and $X_2(z)$ can be uniquely recovered (up to trivial ambiguities) as GCDs of measurements polynomials $\Gamma_{11}(z), \Gamma_{12}(z), \Gamma_{21}(z)$ and $\Gamma_{22}(z)$. The set of equivalencies summarized in Figure 1 further demonstrates that, in absence of noise, such *algebraic approaches* can be readily used to solve the initial PPR problem. In the context of noisy PPR measurements, this section shows how to leverage the notion of *approximate* GCD [72] for solving the polarimetric phase retrieval problem thanks to computational linear algebra methods. In the sequel, we assume that PPR measurements are corrupted by additive i.i.d. Gaussian noise such that for $m = 0, 1, \dots M-1$ and $p = 0, 1, \dots P-1$,

$$y_{m,p} = |\mathbf{a}_m^{\mathsf{H}} \mathbf{X} \mathbf{b}_p|^2 + n_{m,p}, \quad n_{m,p} \sim \mathcal{N}(0, \sigma^2), \tag{4.1}$$

where σ^2 is the Gaussian noise variance. The signal-to-noise ratio (SNR) is then defined as

$$\mathrm{SNR} = \frac{\sum_{m=0}^{M-1} \sum_{p=0}^{P-1} |\mathbf{a}_m^{\mathsf{H}} \mathbf{X} \mathbf{b}_p|^4}{MP\sigma^2}. \tag{4.2}$$

Algorithm 1 summarizes the use of algebraic approaches to solve noisy PPR. They operate in two steps. First, one first needs to obtain an estimate $\hat{\mathbf{\Gamma}}(z)$ of the measurement polynomial matrix $\mathbf{\Gamma}(z)$ given noisy scalar PPR measurements $y_{m,p}$, $m = 0, 1, \dots, M-1$, $p = 0, 1, \dots, P-1$. Section 4.1 addresses this question. The second step exploits approximate GCDs computations of measurement polynomials to recover estimates $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ of the coefficients of polynomials $X_1(z)$ and $X_2(z)$ (or equivalently, the two components of the bivariate signal $\{\mathbf{x}[n]\}_{n=0}^{N-1}$). Section 4.2 introduces the main theoretical tools for this task, namely the notion of Sylvester matrices and their (left or right) kernel properties, in a general context. Section 4.3 then gives two practical algebraic algorithms to recover estimates of the bivariate signal of interest.

4.1. Reconstruction of measurement polynomials. Recall that by Lemma 3.1 measurement polynomials $\Gamma_{ij}(z)$ can be readily expressed in terms of auto-covariance functions $\{\gamma_{11}[n]\}, \{\gamma_{22}[n]\}$ and cross-covariance functions $\{\gamma_{12}[n]\}, \{\gamma_{21}[n]\}$. Thus, recovery of polynomials $\Gamma_{ij}(z)$ is identical to the recovery of $\{\gamma_{ij}[n]\}_{n\in\mathbb{Z}}$ for $i, j = 1, 2$. Equivalently, by discrete Fourier transformation, one must retrieve the spectral matrix $\mathfrak{F}[m]$ for $m = 0, 1, \dots, M-1$ from PPR measurements.

Algorithm 1: Algebraic approaches for noisy PPR

Input: polarimetric measurements $y_{m,p}$, $m = 0, 1, \dots M-1$, $p = 0, 1, \dots P-1$

Step 1: reconstruction of measurements polynomials (Section 4.1);

for *m = 0,. . . , M-1* **do**

  use P polarimetric measurements to obtain an estimate $\hat{\mathfrak{F}}[m]$ as (4.7);

end

Obtain estimates $\{\hat{\gamma}_{ij}[n]\}_{n=1-N}^{N-1}$ of covariance functions for $i, j = 1, 2$ by inverse FFT of entries of $\{\hat{\mathfrak{F}}[m]\}_{m=0}^{M-1}$ (possibly resampled to $2N-1$ points if $M > 2N-1$);

Define measurement polynomials $\hat{\Gamma}_{ij}(z)$ with coefficients $\{\hat{\gamma}_{ij}[n-N+1]\}_{n=0}^{2N-2}$, see (3.4);

Step 2: approximate GCD computations (Section 4.2 and Section 4.3);

Construct the estimated matrix polynomial $\hat{\mathbf{\Gamma}}(z)$ using step 1;

Obtain $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ as outputs of one the following methods: right-kernel Sylvester (Algorithm 2) or left-kernel Sylvester (Algorithm 3);

Result: estimates $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$

Consider noisy measurements given by (4.1). Since $|\mathbf{a}_m^{\mathsf{H}} \mathbf{X} \mathbf{b}_p|^2 = \operatorname{Tr} \mathbf{b}_p^* \mathbf{b}_p^\top \mathfrak{F}[m]$, an estimate $\hat{\mathfrak{F}}[m]$ of $\mathfrak{F}[m]$ is found for every m by minimizing the following quadratic-loss

$$\hat{\mathfrak{F}}[m] = \underset{\substack{\tilde{\mathfrak{F}}[m]=\tilde{\mathfrak{F}}[m]^{\mathsf{H}} \\ \operatorname{rank} \tilde{\mathfrak{F}}[m]=1}}{\arg\min} \sum_{p=0}^{P-1} \left(y_{m,p} - \operatorname{Tr} \mathbf{b}_p^* \mathbf{b}_p^\top \tilde{\mathfrak{F}}[m] \right)^2 , \tag{4.3}$$

where the Hermitian and rank-one constraint ensures the estimated spectral matrix $\hat{\mathfrak{F}}[m]$ has the right structure for future polynomial GCD computations.

To solve (4.3), we adopt a heuristic but simple strategy similar to practical polarimetric reconstruction techniques used in optics [62, 31]. First, we exploit the *Stokes parameters* representation of 2-by-2 Hermitian matrices, which read for an arbitrary Hermitian matrix $\mathbf{M} \in \mathbb{C}^{2\times 2}$

$$\mathbf{M} = \frac{1}{2} \begin{bmatrix} S_0 + S_1 & S_2 + \jmath S_3 \\ S_2 - \jmath S_3 & S_0 - S_1 \end{bmatrix} \qquad S_0, S_1, S_2, S_3 \in \mathbb{R}. \tag{4.4}$$

This set of four real-valued parameters are widely used in optics to describe the different polarization states of light. Formally, Stokes parameters define a bijective map $\mathcal{S} : \{\mathbf{M} \in \mathbb{C}^{2\times 2} | \mathbf{M} = \mathbf{M}^{\mathsf{H}}\} \to \mathbb{R}^4$ such that $\mathcal{S}(\mathbf{M}) = (S_0, S_1, S_2, S_3)^\top$. This allows to express the noiseless measurements as a simple scalar product between Stokes vectors, i.e.,

$$\operatorname{Tr} \mathbf{b}_p^* \mathbf{b}_p^\top \tilde{\mathfrak{F}}[m] = \frac{1}{2} \left[\mathcal{S} \left(\mathbf{b}_p^* \mathbf{b}_p^\top \right) \right]^\top \mathcal{S} \left(\tilde{\mathfrak{F}}[m] \right). \tag{4.5}$$

Therefore, for m fixed, we can set $\mathbf{y}_{m,:} = (y_{m,0}, y_{m,1}, \dots, y_{m,P-1})^\top \in \mathbb{R}_+^P$ as the vector collecting the P polarimetric measurements. Then one defines the polarization measurement matrix

$\mathbf{D} \in \mathbb{R}^{P\times 4}$ such that its p-th row reads $\mathbf{D}_p = \frac{1}{2}\left[\mathcal{S}\left(\mathbf{b}_p^*\mathbf{b}_p^\top\right)\right]^\top$. Note that the matrix $\mathbf{D}$ does not depend on Fourier frequency index m. This leads to rewriting problem (4.3) as

$$\hat{\tilde{\mathfrak{F}}}[m] = \underset{\substack{\tilde{\mathfrak{F}}[m]=\tilde{\mathfrak{F}}[m]^{\mathsf{H}} \\ \operatorname{rank}\tilde{\mathfrak{F}}[m]=1}}{\arg\min} \left\| \mathbf{y}_{m,:} - \mathbf{D}\mathcal{S}\left(\tilde{\mathfrak{F}}[m]\right)\right\|_2^2 . \tag{4.6}$$

A possibly sub-optimal yet very simple solution to (4.6) consists in finding the best rank-one approximation of the classical least square estimator of Stokes parameters, i.e.,

$$\hat{\tilde{\mathfrak{F}}}[m] = \operatorname{rank1}\left\{\mathcal{S}^{-1}\left(\mathbf{D}^\dagger \mathbf{y}_{m,:}\right)\right\}, \tag{4.7}$$

where $\mathbf{D}^\dagger$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{D}$ and $\mathcal{S}^{-1}$ is the inverse Stokes mapping defined by (4.4). The operator $\operatorname{rank1}\{\mathbf{M}\}$ finds the best rank-one approximation of a given matrix $\mathbf{M}$ with respect to the Frobenius norm. For the present 2-by-2 Hermitian matrix case, the solution is given by keeping the first singular vector of $\mathbf{M}$, that is $\operatorname{rank1}(\mathbf{M}) = \sigma_0 \mathbf{u}_0 \mathbf{u}_0^{\mathsf{H}}$, where σ_0 and $\mathbf{u}_0$ are respectively the largest singular value and its corresponding singular vector. Then, estimates $\{\hat{\gamma}_{ij}[n]\}_{n=1-N}^{N-1}$ of covariance functions for $i,j = 1,2$ are directly obtained by inverse discrete Fourier transformation of entries of the spectral matrices $\{\hat{\tilde{\mathfrak{F}}}[m]\}_{m=0}^{M-1}$ (possibly resampled to $2N-1$ points if $M > 2N-1$). Finally, Eq. (3.4) permits to define estimated polynomials $\hat{\Gamma}_{ij}(z)$ as polynomials in $\mathbb{C}_{\leq 2N-2}[z]$ with vector of coefficients $\left[\hat{\gamma}_{ij}[1-N] \quad \hat{\gamma}_{ij}[2-N] \quad \dots \quad \hat{\gamma}_{ij}[N-1]\right]$.

4.2. Sylvester matrices and GCD. Proposition 3.7 shows that, in the noiseless case, polynomials $X_1(z)$ and $X_2(z)$ can be uniquely recovered as GCDs of the measurement polynomial matrix $\mathbf{\Gamma}(z)$. In the noisy PPR measurement case, it further suggests that polynomials $X_1(z)$ and $X_2(z)$ can be estimated, or *approximately recovered* from the estimated matrix polynomial $\hat{\mathbf{\Gamma}}(z)$ computed in Section 4.1. Importantly, the term *approximate* refers here to the practical infeasibility of computing exact GCDs due to numerical instabilities related to machine precision, so that approximate GCD computations must be used instead. There are many possible approaches for approximate GCD, see e.g., [52] for a recent review. In this work, we follow [72] and carry approximate GCD computations by taking advantage of the kernel (or null-space) properties of Sylvester matrices.

The following section reviews the relevant theory. Practical use of these results in the context of PPR is given in Section 4.3.

For simplicity, we assume polynomials $A, B \in \mathbb{C}_{\leq L}[z]$ of same degree L. Then we define the Sylvester-like matrices, parameterized by an integer $D \leq L$ (possibly negative) as

$$\mathcal{S}_D(A,B) = \left[\begin{array}{ccc|ccc} a_0 & & & b_0 & & \\ \vdots & \ddots & & \vdots & \ddots & \\ a_L & & a_0 & b_L & & b_0 \\ & \ddots & \vdots & & \ddots & \vdots \\ & & a_L & & & b_L \end{array}\right] \in \mathbb{C}^{(2L-D+1)\times 2(L-D+1)}. \tag{4.8}$$

The Sylvester-like matrices are tightly linked with multiplication matrices of polynomials, see Appendix B for more details. When $D = 1$ (i.e., the matrix is square $2L \times 2L$), the matrix is

the well-known Sylvester matrix. There are, however, two important extensions of the classic case:

- When $1 \leq D \leq L$, the matrix is tall (the number of columns does not exceed the number of rows), and it is called the *Sylvester subresultant* matrix.
- If $D \leq 1$ (in general, chosen to be negative), the matrix is fat (the number of rows does not exceed the number of columns), and such a matrix is called *extended Sylvester* matrix.

For an overview of such matrices and the corresponding literature, we refer to [72] (note that unlike [72] we use the same notation for subresultant and extended Sylvester matrices). The following theorem is classic.

Theorem 4.1 (Sylvester). *Two polynomials $A, B \in \mathbb{C}_{\leq L}[z]$ have a non-trivial common divisor if and only if $\mathcal{S}_1(A,B)$ is rank deficient. Moreover the degree K of $\gcd(A,B)$ is equal to the rank defect of $\mathcal{S}_1(A,B)$, i.e.,*

$$K = 2L - \operatorname{rank} \mathcal{S}_1(A,B)$$

and $\gcd(A,B) \in \mathbb{C}_{\leq K}[z]$.

Unfortunately, Theorem 4.1 does not give an explicit way to compute $\gcd(A,B)$. In fact, explicit determination of the GCD requires the use of Sylvester matrices $\mathcal{S}_D(A,B)$ in the general case $D \neq 1$. More precisely, Proposition 4.2 and Proposition 4.4 below show that the GCD can be retrieved from the left or right kernel of carefully constructed Sylvester matrices. In what follows, we assume that the GCD has degree K and note $Q(z) = \gcd(A,B) \in \mathbb{C}_{\leq K}[z]$. Moreover, we define

$$F(z) = \frac{A(z)}{Q(z)}, \quad G(z) = \frac{B(z)}{Q(z)}$$

the corresponding quotient polynomials. We begin with the result on the right kernel of Sylvester subresultant matrices.

Proposition 4.2 (Right kernel, see e.g. [72, Lemma 4.6]). *The rank of the Sylvester subresultant matrix $\mathcal{S}_K(A,B)$ is equal to $2(L-K+1)-1$ (i.e., it has rank defect equal to 1). Moreover, for the (unique up to scalar factor) nonzero vector in the right kernel*

$$\mathcal{S}_K(A,B) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = 0; \tag{4.9}$$

with $\mathbf{u}, \mathbf{v} \in \mathbb{C}^{L-K+1}$, the corresponding polynomials are multiples of the quotient polynomials:

$$U(z) = -cG(z), \quad V(z) = cF(z), \tag{4.10}$$

where $c \in \mathbb{C}$ is some constant.

Remark 4.3. In view of the connection between Sylvester and multiplication matrices (see Appendix B), Proposition 4.2 implies that the polynomials $U, V \in \mathbb{C}_{\leq L-K}[z]$ defined in (4.10) are the only ones (up to multiplication by a nonzero scalar) that satisfy $A(z)U(z)+B(z)V(z) = 0$ (see (4.9)).

For the case of extended Sylvester matrices ($D \leq 1$), the result on the left kernel matrices is less known in the form that we are using here. This is the reason why we also provide a short proof of the following proposition in Appendix B.

Proposition 4.4 (Left kernel). *Let $D \leq 1$ (i.e., $\mathcal{S}_D(A,B)$ is fat with $2L-D+1$ rows). Then the rank of $\mathcal{S}_D(A,B)$ is equal to*

$$\operatorname{rank} \mathcal{S}_D(A,B) = 2L - D + 1 - K; \tag{4.11}$$

therefore the dimension of the left kernel (i.e., the rank defect) is equal to K (the degree of the GCD). Moreover, let $\mathbf{u}_{-1}, \ldots, \mathbf{u}_{-K} \in \mathbb{C}^{2L-D+1}$ be a basis of the left kernel of $\mathcal{S}_D(A,B)$ (for example, its last K left singular vectors) and define the following matrix

$$\mathbf{H} = \begin{bmatrix} \mathcal{H}_{K+1}(\mathbf{u}_{-1}) & \cdots & \mathcal{H}_{K+1}(\mathbf{u}_{-K}) \end{bmatrix} \in \mathbb{C}^{(K+1)\times K(2L-D+1-K)}, \tag{4.12}$$

where each block is a Hankel matrix, i.e.,$\mathcal{H}_{K+1}(\mathbf{u})$ denotes the Hankel matrix with $K+1$ rows built from a vector $\mathbf{u} \in \mathbb{C}^{2L-D+1}$ such that

$$\mathcal{H}_{K+1}(\mathbf{u}) = \begin{bmatrix} u[0] & u[1] & \cdots & u[2L-D-K] \\ u[1] & u[2] & \cdots & u[2L-D-K+1] \\ \vdots & \vdots & & \vdots \\ u[K] & u[K+1] & \cdots & u[2L-D] \end{bmatrix} \in \mathbb{C}^{(K+1)\times(2L-D+1-K)}. \tag{4.13}$$

Then we have $\operatorname{rank} \mathbf{H} = K$ and the left kernel of $\mathbf{H}$ is spanned by the the vector of coefficients $\mathbf{q} \in \mathbb{C}^{K+1}$ of the GCD $Q(z)$.

The next section exploits these properties of the kernel of Sylvester matrices to formulate algebraic algorithms for the PPR problem.

4.3. Algebraic algorithms.

In this section, we propose two algorithms for estimating coefficients of polynomials $X_1(z)$ and $X_2(z)$ from the estimated matrix polynomial $\hat{\mathbf{\Gamma}}(z)$ computed in Section 4.1. Both algorithms rely on the use of the singular value decomposition (SVD) to find the left or right kernels of Sylvester matrices constructed from $\hat{\mathbf{\Gamma}}(z)$. Thus the proposed reconstuction methods may appear as suboptimal since the Sylvester structure is not taken into account when computing the (low-rank) kernels. This limitation could be overcome with structured low-rank approximations [48], to be specifically tailored for the PPR problem. Such a study would fall outside the scope of the present work. Still, as demonstrated by the numerical experiments presented in Section 6, the SVD already provides excellent reconstruction performance in many scenarios, while maintaining a reasonable computational burden.

4.3.1. Right kernel Sylvester.

The first algorithm is based on the properties of the right kernel of Sylvester matrices described in Proposition 4.2. It uses the fact that $X_1(z)$ and $X_2(z)$ are (without noise) quotient polynomials of

$$\Gamma_{11}(z) = X_1(z)\widetilde{X}_1(z) \text{ and } \Gamma_{21}(z) = X_2(z)\widetilde{X}_1(z).$$

One can remark that $X_1(z)$ and $X_2(z)$ are also quotient polynomials of $\Gamma_{12}(z) = X_1(z)\widetilde{X}_2(z)$ and $\Gamma_{22}(z) = X_1(z)\widetilde{X}_2(z)$, which adds some freedom in the choice of measurement polynomials. For the sake of simplicity, we will work with estimated polynomials $\hat{\Gamma}_{11}(z)$ and $\hat{\Gamma}_{21}(z)$ in the following.

Algorithm 2: Right kernel Sylvester

Input: estimated matrix polynomial $\hat{\mathbf{\Gamma}}(z) \in \mathbb{C}^{2\times 2}_{\leq 2N-2}$.
Build the matrix $\mathbf{S} = \mathcal{S}_{N-1}(\hat{\Gamma}_{11}, \hat{\Gamma}_{21}) \in \mathbb{C}^{(3N-2)\times 2N}$;
Take $\mathbf{v} = \mathbf{v}_{2N} \in \mathbb{C}^{2N}$ to be the $2N$-th right singular vector of $\mathbf{S}$ (corresponding to the last nontrivial singular value);
Partition $\mathbf{v}$ as $\mathbf{v} = (-\mathbf{v}_2, \mathbf{v}_1)$, where $\mathbf{v}_1 = c\widehat{\mathbf{x}}_1$ and $\mathbf{v}_2 = c\widehat{\mathbf{x}}_2$ with $c \in \mathbb{C}$;
Determine $|c|$ by proper norm scaling as

$$|c| = \left(\frac{\|\mathbf{v}_1\|_2^2 + \|\mathbf{v}_2\|_2^2}{\hat{\gamma}_{11}[0] + \hat{\gamma}_{22}[0]} \right)^{\frac{1}{2}}$$

Set $\widehat{\mathbf{x}}_1 = \mathbf{v}_1/|c|$ and $\widehat{\mathbf{x}}_2 = \mathbf{v}_2/|c|$;
Result: estimates $\widehat{\mathbf{x}}_1$ and $\widehat{\mathbf{x}}_2$

Algorithm 3: Left kernel Sylvester

Input: estimated matrix polynomial $\hat{\mathbf{\Gamma}}(z) \in \mathbb{C}^{2\times 2}_{\leq 2N-2}$.
for *j = 1,2* **do**
  Build the matrix $\mathbf{S} = \mathcal{S}_1(\hat{\Gamma}_{j1}, \hat{\Gamma}_{j2}) \in \mathbb{C}^{(4N-4)\times(4N-4)}$;
  Take the last $N-1$ left singular vectors of $\mathbf{S}$, i.e., $\mathbf{u}_{3N-2}, \ldots, \mathbf{u}_{4N-4}$;
  Construct the Hankel matrix $\mathbf{H}$ as (4.12);
  Retrieve $\mathbf{w}_j = c_j\widehat{\mathbf{x}}_j$, $c_j \in \mathbb{C}$ as the last left singular vector of $\mathbf{H}$.
end
Determine constants c_1, c_2 as

$$c_1 = \frac{\|\mathbf{w}_1\|_2}{\sqrt{\hat{\gamma}_{11}[0]}} \text{ and } c_2 = \frac{\|\mathbf{w}_2\|_2}{\sqrt{\hat{\gamma}_{22}[0]}} \exp\left[\jmath(\arg\hat{\gamma}_{12}[0] - \arg\mathbf{w}_2^{\mathsf{H}}\mathbf{w}_1)\right]$$

Set $\widehat{\mathbf{x}}_1 = \mathbf{w}_1/c_1$ and $\widehat{\mathbf{x}}_2 = \mathbf{w}_2/c_2$;
Result: estimates $\widehat{\mathbf{x}}_1$ and $\widehat{\mathbf{x}}_2$

The complete right kernel Sylvester approach is summarized in Algorithm 2. It exploits the fact that $\gcd(\Gamma_{11}, \Gamma_{21}) = X_1(z)$, i.e., it is a polynomial of degree $N-1$. Therefore according to Proposition 4.2, the last nontrivial singular vector of the Sylvester matrix $\mathcal{S}_{N-1}(\hat{\Gamma}_{11}, \hat{\Gamma}_{21})$ provides an estimate of the (one-dimensional) right kernel, which in turn gives, up to one complex multiplicative constant, an estimation $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ of the vectors of coefficients defining polynomials $X_1(z)$ and $X_2(z)$. This constant is then computed (up to one unit-modulus factor due to the trivial rotation ambiguity) by scaling the 2-norm of $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ thanks to the value at the origin ($n = 0$) of estimated auto-covariance functions $\hat{\gamma}_{11}[n]$ and $\hat{\gamma}_{22}[n]$.

One of the key advantages of this algorithm lies in its simplicity. Indeed, it only requires a single SVD of a $(3N-2) \times 2N$ matrix and thus has overall computational complexity $\mathcal{O}(N^3)$.

4.3.2. Left kernel Sylvester. The second algorithm exploits the properties of the left kernel of extended (fat) Sylvester matrices (i.e., $\mathcal{S}_D$ for $D \leq 1$) detailed in Proposition 4.4. For simplicity and to reduce the size of the involved matrices we set $D = 1$ in what follows. Nonetheless, the proposed approach can be adapted to any value of $D \leq 1$ if needed.

Algorithm 3 summarizes the complete procedure. In essence, it follows the theoretical result of Proposition 3.7. In particular, compared to the right kernel Sylvester approach, estimated coefficients $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ are obtained by two separate GCD computations: the vector of coefficients $\hat{\mathbf{x}}_1$ is obtained by computing the GCD of estimated measurement polynomials $\hat{\Gamma}_{11}(z)$ and $\hat{\Gamma}_{12}(z)$, whereas $\hat{\mathbf{x}}_2$ is obtained by computing the GCD of $\hat{\Gamma}_{21}(z)$ and $\hat{\Gamma}_{22}(z)$. Importantly, the two GCDs are determined up to a multiplicative complex constant, say c_1 and c_2, which can be determined jointly using PPR measurements.

The computation of each GCD requires three steps [72]: a first SVD to determine the $N-1$ last left singular vectors of the Sylvester matrix $\mathcal{S}_1$; the construction of a fat, horizontally stacked Hankel matrix $\mathbf{H}$ with N rows from these $N-1$ singular vectors; a second SVD to obtain the N coefficients of the GCD as the last left singular vector of $\mathbf{H}$. Once GCDs have been obtained, determination of constants c_1 and c_2 (up to a common global phase factor) is carried out by properly scaling the norms of estimated coefficients $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ (using $\hat{\gamma}_{11}[0]$ and $\hat{\gamma}_{22}[0]$) and adjusting the phase factor $\arg c_1 c_2^*$ thanks to the value at $n = 0$ of the estimated cross-covariance function $\hat{\gamma}_{12}[n]$.

The complexity of the left kernel Sylvester method described in Algorithm 3 is higher for two main reasons. First, as explained above, it requires the computations of two SVDs for each one of the two GCDs determinations. Moreover, while the first SVD has a cost of $\mathcal{O}(N^3)$, the second SVD is performed on a large fat stacked Hankel matrix $\mathbf{H}$, with complexity $\mathcal{O}(N^4)$, which dominates the overall computational burden of Algorithm 3.

5. Solving PPR with iterative algorithms.

We now address the design of iterative algorithms to solve the noisy PPR problem. Section 5.1 and Section 5.2 exploit the PPR-1D representation of the original problem to provide a semidefinite programming (SDP) relaxation and Wirtinger flow algorithm, respectively.

5.1. SDP relaxation.

Semidefinite programming (SDP) approaches for phase retrieval have been increasingly popular for over a decade [15, 16, 75]. In the classical 1D phase retrieval case, SDP approaches exploit that even though measurements are quadratic in the unknown signal $\mathbf{x} \in \mathbb{C}^N$, they are linear in the rank-one matrix $\mathbf{x}\mathbf{x}^{\mathsf{H}}$. For PPR, the 1D equivalent representation PPR-1D enables to formulate a SDP relaxation of the original problem, by observing that

$$|\mathbf{c}_{m,p}^{\mathsf{H}}\boldsymbol{\xi}|^2 = \operatorname{Tr} \mathbf{c}_{m,p}\mathbf{c}_{m,p}^{\mathsf{H}}\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{H}} = \operatorname{Tr} \mathbf{C}_{m,p}\boldsymbol{\Xi}, \tag{5.1}$$

i.e., noiseless measurements can be rewritten as a linear function of the lifted positive semidefinite rank-one matrix $\boldsymbol{\Xi} = \boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{H}} \in \mathbb{C}^{2N\times 2N}$. Following the classical PhaseLift methodology [16, 15], the original nonconvex PPR problem can be relaxed into a SDP convex program as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\sum_{m=0}^{M-1}\sum_{p=0}^{P-1}\left(y_{m,p} - \operatorname{Tr} \mathbf{C}_{m,p}\boldsymbol{\Xi}\right)^2 + \lambda\|\boldsymbol{\Xi}\|_\star, \\ \text{subject to} \quad & \boldsymbol{\Xi} \succeq 0 \end{aligned} \tag{5.2}$$

Algorithm 4: SDP relaxation for PPR

Input: measurements $\mathbf{y} \in \mathbb{R}^{MP}$, lifted measurement matrices $\mathbf{C}_{m,p} \in \mathbb{C}^{2N\times 2N}$, regularization parameter $\lambda \geq 0$.

set arbitrary $\boldsymbol{\Xi}^{(0)}$;

$\boldsymbol{\Psi}^{(0)} \leftarrow \boldsymbol{\Xi}^{(0)}$;

$k \leftarrow 0$;

while *stopping criterion is not satisfied* **do**

$\boldsymbol{\Xi}^{(k+1)} = \mathbf{prox}_{t_k g}\left(\boldsymbol{\Psi}^{(k)} - t_k \nabla f(\boldsymbol{\Psi}^{(k)})\right)$ where the proximal operator is given by (5.8);

$\eta_{k+1} = \frac{1+\sqrt{1+4\eta_k^2}}{2}$;

$\boldsymbol{\Psi}^{(k+1)} = \boldsymbol{\Xi}^{(k+1)} + \left(\frac{\eta_k - 1}{\eta_{k+1}}\right)\left(\boldsymbol{\Xi}^{(k+1)} - \boldsymbol{\Xi}^{(k)}\right)$;

$k \leftarrow k+1$;

end

$\hat{\boldsymbol{\xi}} \leftarrow \mathrm{rank1}\left(\boldsymbol{\Xi}^{(k)}\right)$;

Result: estimate $\hat{\boldsymbol{\xi}}$

where $\lambda \geq 0$ is an hyperparameter that allows to control the trade-off between the likelihood of observations and the nuclear norm regularization $\|\cdot\|_\star$. Note that since $\boldsymbol{\Xi}$ is constrained to be positive semidefinite, the nuclear norm regularization is equivalent to the trace-norm regularization used in [15] since $\|\boldsymbol{\Xi}\|_\star = \operatorname{Tr}\boldsymbol{\Xi}$ in this case. The SDP program (5.2) takes a standard form: therefore it can be solved in many ways, including interior point methods [74], first-order methods [51] or using disciplined convex programming solvers such as `CVXPY`[1]. For completeness, we provide below an explicit algorithm to solve (5.2) using a proximal gradient approach [7, Chapter 10]. It closely follows the approach described in [15, 32].

The objective function in (5.2) can be rewritten as the sum $f(\boldsymbol{\Xi}) + g(\boldsymbol{\Xi})$ with

$$f(\boldsymbol{\Xi}) = \frac{1}{2}\sum_{m=0}^{M-1}\sum_{p=0}^{P-1}\left(y_{m,p} - \operatorname{Tr}\mathbf{C}_{m,p}\boldsymbol{\Xi}\right)^2, \quad g(\boldsymbol{\Xi}) = \lambda\|\boldsymbol{\Xi}\|_\star + \iota_{\succeq 0}(\boldsymbol{\Xi}), \tag{5.3}$$

where $\iota_{\succeq 0}(\cdot)$ denotes the indicator function on the positive semidefinite cone. This ensures the formal equivalence between (5.2) and the unconstrained minimization problem

$$\min_{\boldsymbol{\Xi}\in\mathbb{C}^{2N\times 2N}} f(\boldsymbol{\Xi}) + g(\boldsymbol{\Xi})\,. \tag{5.4}$$

The convex optimization problem (5.4) can be efficiently solved by proximal gradient methods, which take advantage of the splitting between f and g of the objective function. More precisely,

[1] https://www.cvxpy.org/

we use the fast proximal gradient method which consist, at iteration k:

$$\boldsymbol{\Xi}^{(k+1)} = \underset{t_k g}{\mathbf{prox}}\left(\boldsymbol{\Psi}^{(k)} - t_k \nabla f(\boldsymbol{\Psi}^{(k)})\right), \tag{5.5}$$

$$\eta_{k+1} = \frac{1+\sqrt{1+4\eta_k^2}}{2}, \tag{5.6}$$

$$\boldsymbol{\Psi}^{(k+1)} = \boldsymbol{\Xi}^{(k+1)} + \left(\frac{\eta_k - 1}{\eta_{k+1}}\right)\left(\boldsymbol{\Xi}^{(k+1)} - \boldsymbol{\Xi}^{(k)}\right), \tag{5.7}$$

where t_k is a step-size which is chosen such that the proximal gradient step (5.5) obey some sufficient decrease condition; see e.g. [7, p. 271] for details. Our choice for the function g in (5.4) enables a simple expression for the associated proximal operator (see [32]):

$$\begin{aligned}\underset{\tau g}{\mathbf{prox}}(\mathbf{X}) &= \min_{\mathbf{Z}\succeq 0} \tau\lambda\|\mathbf{Z}\|_\star + \|\mathbf{Z}-\mathbf{X}\|_2^2 \\ &= \mathbf{Ushrink}(\Sigma, \tau\lambda)\mathbf{U}^{\mathsf{H}},\end{aligned} \tag{5.8}$$

where in the last equation, $\mathbf{U}\Sigma\mathbf{U}^{\mathsf{H}}$ is the eigenvalue decomposition of $\mathbf{X}$ and the shrink operator is defined entry-wise by $\mathrm{shrink}(\sigma_i, \tau\lambda) = \max\{\mathrm{real}(\sigma_i) - \tau\lambda, 0\}$

Choice of regularization parameter λ. In this work, we fix the value of the regularization parameter to $\lambda = 1/\mathrm{SNR}$: we found empirically that this choice provides good results in most scenarios, as it provides a reasonable tradeoff between likelihood of observations and the nuclear norm regularization in the objective function of (5.2).

Convergence. Obviously, as (5.2) is a convex program, the precision towards the optimal cost value can become arbitrarily good as one increases the number of iterations. In practice, one needs to stop the algorithm when a prescribed tolerance ε is reached. To this aim we implemented stopping criteria that carefully monitor a normalized residual, see [32] for details. Moreover, it may happen that the estimated lifted matrix $\hat{\boldsymbol{\Xi}}$ generated by the sequence of $\boldsymbol{\Xi}^{(k)}$ is not rank one: in this case, one first computes the rank-one approximation of $\hat{\boldsymbol{\Xi}}$ (e.g. using SVD) to obtain the estimated signal $\hat{\boldsymbol{\xi}}$.

Complexity. The computational cost of the proposed algorithm concentrates on the proximal gradient step (5.5), where the evaluation of the proximal operator and the computation of the gradient ∇f share the computational burden. More precisely, the eigenvalue decomposition of a $2N \times 2N$ matrix together with the shrink operator leads to $\mathcal{O}(N^3)$ calculations. The computation of the gradient leads to MP trace evaluations of order $\mathcal{O}(N^2)$ flops, meaning that the number of flops per iteration is of order $\mathcal{O}(MPN^2 + N^3)$.

The full procedure is summarized in Algorithm 4.

5.2. Wirtinger flow for PPR. Exploiting further the 1D equivalent representation PPR-1D of the PPR problem, another approach consists in minimizing directly the following non-convex quadratic objective

$$\min_{\boldsymbol{\xi}\in\mathbb{C}^{2N}} F(\boldsymbol{\xi}) = \frac{1}{2}\||\mathbf{y} - |\mathbf{C}\boldsymbol{\xi}|^2\|_2^2, \tag{5.9}$$

Algorithm 5: Wirtinger Flow for PPR: PPR-WF

Input: measurements $\mathbf{y} \in \mathbb{R}^{MP}$, measurment matrix $\mathbf{C} \in \mathbb{C}^{MP\times 2N}$, tolerance ε
set $\boldsymbol{\xi}^{(0)}$ using the desired initialization method;
$\boldsymbol{\xi}^{(1)} \leftarrow \boldsymbol{\xi}^{(0)}$;
$k \leftarrow 1$;
while $\|\boldsymbol{\xi}^{(i+1)} - \boldsymbol{\xi}^{(i)}\|_2 > \varepsilon \|\boldsymbol{\xi}^{(i)}\|_2$ **do**
  $\beta_k \leftarrow \frac{k+1}{k+3}$;
  $\boldsymbol{\psi}^{(k)} \leftarrow \boldsymbol{\xi}^{(k)} + \beta_k \left(\boldsymbol{\xi}^{(k)} - \boldsymbol{\xi}^{(k-1)}\right)$;
  compute optimal step-size μ_k (5.14);
  $\boldsymbol{\xi}^{(k+1)} \leftarrow \boldsymbol{\psi}^{(k)} - \mu_k \nabla F\left(\boldsymbol{\psi}^{(k)}\right)$;
  $k \leftarrow k + 1$;
end
$\hat{\boldsymbol{\xi}} \leftarrow \boldsymbol{\xi}^{(k)}$;
Result: estimate $\hat{\boldsymbol{\xi}}$

where $\mathbf{y} \in \mathbb{R}^{MP}$ gathers PPR measurements and where the rows of $\mathbf{C} \in \mathbb{C}^{MP\times 2N}$ are given by $\mathbf{c}_{m,p}^{\mathsf{H}}$, see Section 2.4. Provided that one can find a initial point $\boldsymbol{\xi}^{(0)}$ close enough from the global minimizer of (5.9), a simple strategy based on gradient descent can be used to solve PPR. However, such an approach requires special care since the optimization variable $\boldsymbol{\xi}$ is complex-valued. In fact, the objective function in (5.9) is real-valued, and thus it is not differentiable with respect to complex analysis. Instead, one needs to resort to the so-called $\mathbb{CR}$ or *Wirtinger*-calculus [44] to provide a meaningful extension of gradient-descent-type algorithms to the complex case. This is precisely the approach proposed in [17] to solve standard phase retrieval, where the complex gradient descent is called *Wirtinger flow* (WF).

Leveraging the original WF approach, we propose below a complex-gradient descent algorithm which solves the nonconvex problem (5.9). Compared to the original paper [17], we incorporate optimal step size selection [43] together with a proposed acceleration scheme [77]. We further propose an efficient strategy for initialization based on the algebraic methods for PPR described in Section 4. The superiority of these initializations over standard ones (e.g. spectral initialization as proposed in [17]) will be demonstrated in Section 6.2.

The proposed PPR-WF algorithm is as follows. Starting from two initial points $\boldsymbol{\xi}^{(0)}$, $\boldsymbol{\xi}^{(1)}$, the k-th iteration reads

$$\beta_k = \frac{k+1}{k+3}, \tag{5.10}$$

$$\boldsymbol{\psi}^{(k)} = \boldsymbol{\xi}^{(k)} + \beta_k \left(\boldsymbol{\xi}^{(k)} - \boldsymbol{\xi}^{(k-1)}\right), \tag{5.11}$$

$$\boldsymbol{\xi}^{(k+1)} = \boldsymbol{\psi}^{(k)} - \mu_k \nabla F\left(\boldsymbol{\psi}^{(k)}\right), \tag{5.12}$$

where β_k is a sequence of accelerated parameters and μ_k is a carefully chosen stepsize, see further below. Compared to the standard WF algorithm, PPR-WF takes advantage of the

acceleration procedure first proposed in [77] in the context of ptychographic phase retrieval (but using a magnitude loss function instead of a square magnitude loss function as used here). Note that the complex gradient of F can be computed explicitly as

$$\nabla F(\boldsymbol{\psi}) = \mathbf{C}^{\mathsf{H}} \left[\left(|\mathbf{C}\boldsymbol{\psi}|^2 - \mathbf{y} \right) \odot \mathbf{C}\boldsymbol{\psi} \right], \tag{5.13}$$

where the symbol $\odot$ denotes entry-wise product between vectors.

Optimal step-size selection. We combine acceleration for WF with the optimal step-size selection proposed in [43] for the standard WF algorithm. For completeness, we reproduce here the main ingredients underpinning optimal step size selection in (5.12) and refer the reader to [43] for further details. At iteration k, the optimal stepsize μ_k is defined by line search, i.e.,

$$\mu_k = \arg\min_{\mu} F\left(\boldsymbol{\xi}^{(k+1)}\right) = F\left(\boldsymbol{\psi}^{(k)} - \mu \nabla F\left(\boldsymbol{\psi}^{(k)}\right)\right). \tag{5.14}$$

The authors in [43] showed that the 1D optimization problem (5.14) boils down to finding the roots of a univariate cubic polynomial with real coefficients, the latter being completely determined by the knowledge of $\boldsymbol{\psi}^{(k)}$, $\nabla F\left(\boldsymbol{\psi}^{(k)}\right)$ and $\mathbf{y}$, see [43, Eq. (17)]. Roots can be determined in closed-form, and two cases can occur: *(a)* there is only one real root, and thus it gives the optimal step-size μ_k; *(b)* there are three real roots, and in this case μ_k is set to the real root associated to the minimum objective value. Note that optimal selection for WF is somewhat inexpensive, with computational cost dominated by the calculation of the cubic polynomial coefficients scaling as $\mathcal{O}(MP)$.

Initialization. Since PPR-WF attempts at minimizing a nonconvex quadratic objective (5.9), the choice of initial points $\boldsymbol{\xi}^{(0)}$, $\boldsymbol{\xi}^{(1)}$ is crucial to hope that PPR-WF will be able to recover a global minimizer of the objective function. For simplicity, we set $\boldsymbol{\xi}^{(1)} = \boldsymbol{\xi}^{(0)}$, so that we only discuss the selection of $\boldsymbol{\xi}^{(0)}$. Five different initialization strategies for PPR-WF are considered:

- *spectral initialization* [17]: this standard approach consists in computing the eigenvector $\mathbf{v}$ corresponding to the largest eigenvalue of the matrix

$$\mathbf{Y} = \frac{1}{MP} \sum_{r=0}^{MP-1} y_r \mathbf{c}_r \mathbf{c}_r^{\mathsf{H}} \tag{5.15}$$

  and to rescale it properly to set

$$\boldsymbol{\xi}^{(0)} = \frac{\mathbf{v}}{\lambda}, \quad \lambda = \left(N \frac{\sum_{r=0}^{MP-1} y_r}{\sum_{r=0}^{MP-1} \|\mathbf{c}_r\|^2} \right)^{1/2}. \tag{5.16}$$

- *random phase initialization*: we first generate a random measurement phase vector $\boldsymbol{\phi} \in \mathbb{R}^{MP}$ with i.i.d. entries $\phi_r \sim \mathcal{U}([0, 2\pi])$. Then, we set

$$\boldsymbol{\xi}^{(0)} = \mathbf{C}^{\dagger} \tilde{\mathbf{y}}, \quad \tilde{\mathbf{y}} = \mathbf{y} \odot \exp(\jmath \boldsymbol{\phi}), \tag{5.17}$$

  where $\mathbf{C}^{\dagger}$ is the pseudo-inverse of $\mathbf{C}$.

- *right kernel Sylvester initialization*: we set $\boldsymbol{\xi}^{(0)}$ as the result of Algorithm 1 where approximate GCDs computations are performed using the right kernel Sylvester method (Algorithm 2).
- *left kernel Sylvester initialization*: we set $\boldsymbol{\xi}^{(0)}$ as the result of Algorithm 1 where approximate GCDs computations are performed using the left kernel Sylvester method (Algorithm 3).
- *SDP initialization*: we set $\boldsymbol{\xi}^{(0)}$ as the output the SDP approach described in Section 5.1 and summarized in Algorithm 4.

Convergence monitoring. We monitor convergence of PPR-WF by computing at each iteration k, the normed residual $\|\boldsymbol{\xi}^{(k+1)} - \boldsymbol{\xi}^{(k)}\|_2 / \|\boldsymbol{\xi}^{(k)}\|_2$ and stop the algorithm when it goes below a prescribed tolerance $\varepsilon \ll 1$.

Complexity. The computational cost per iteration of PPR-WF is dominated by the evaluation of the complex gradient (5.13), which scales as $\mathcal{O}(MPN)$. Note that the optimal step-size selection procedure scales as $\mathcal{O}(MP)$, meaning that the whole cost of PPR-WF remains $\mathcal{O}(MPN)$ per iteration. Algorithm 5 summarizes the proposed PPR-WF algorithm.

6. Numerical experiments.

We provide in this section several numerical experiments that address how PPR can be solved in practice using both algebraic and algorithmic approaches described in Section 4 and Section 5, respectively. Importantly, we demonstrate that the use of the Wirtinger Flow algorithm together with a right-Sylvester initialization provides an excellent trade-off between mean squared error (MSE) and computational burden. This combination of algorithmic and algebraic reconstruction methods provides a scalable, asymptotically MSE optimal, and parameter free inversion procedure for PPR.

Just like in standard phase retrieval, the global phase ambiguity in PPR requires to properly realign any estimated signal $\hat{\mathbf{X}}'$ with the ground truth $\mathbf{X}$ in order to provide a meaningful squared reconstruction error value. We define the realigned estimated signal $\hat{\mathbf{X}}$ as

$$\hat{\mathbf{X}} = e^{\jmath\Phi_0}\hat{\mathbf{X}}' \text{ with } \Phi_0 = \underset{\phi\in[0,2\pi)}{\arg\min} \|e^{\jmath\phi}\hat{\mathbf{X}}' - \mathbf{X}\|_F^2 . \tag{6.1}$$

The squared reconstruction error is then defined in terms of the Frobenius norm as $\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2$. Note that in practice, the minimization involved in the realignment procedure can simply be performed by evaluating the complex phase of the standard inner product between the vectors $\hat{\boldsymbol{\xi}}'$ and $\boldsymbol{\xi}$ obtained from matrices $\hat{\mathbf{X}}'$ and $\mathbf{X}$, respectively.

This section is organized as follows. Section 6.1 presents the reconstruction of a realistic bivariate pulse from noiseless PPR measurements using the different approaches presented in the paper. Section 6.2 then discusses the choice of initialization in PPR-WF. Section 6.3 benchmarks the robustness to noise of the proposed reconstructions methods. Finally, Section 6.4 provides a first study of the impact of the number of PPR measurements on reconstruction performances.

6.1. Reconstruction of bivariate pulse.

As a first experiment, we consider the reconstruction of a bivariate pulse from noiseless PPR measurements. The signal to be recovered defines a typical complex-valued bivariate analytic signal associated to the bivariate electromagnetic field to be estimated in ultra-short electromagnetic pulses experiments, see e.g. [64, 76]. It is defined for $N = 64$ points and we consider the simple noise-free measurement scheme (2.5)

with $M = 2N-1$ and $K = 4$. The bivariate pulse exhibits slow variations of the instantaneous polarization state, ensuring uniqueness of the PPR solution. We investigate the capacity of the methods introduced in Section 4 and Section 5 to properly recover the bivariate signal of interest. Note that for Wirtinger Flow, we only consider spectral initialization—remaining strategies will be extensively benchmarked in Section 6.2 below.

Figure 3 depicts the different reconstructed bivariate signals obtained by each method along with the associated squared error $(\hat{\mathbf{x}}[n] - \mathbf{x}[n])^2$ for every time index n, where the estimated signal $\hat{\mathbf{x}}$ is realigned with the ground truth $\mathbf{x}$ using (6.1). Except Wirtinger Flow with spectral initialization, all methods successfully recover the original bivariate signal, where successful recovery in the noiseless context is decided whenever $\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2 < 10^{-20}$. Left and right kernel Sylvester provide similar reconstruction quality, with a slight advantage to left kernel Sylvester. The SDP approach performs also well, yet three or four order of magnitude of squared error above the previous approaches. Due to the very low error levels involved here, this has little consequence; however, compared to the aforementioned methods SDP exhibits both larger memory usage and overall computational cost, which makes it a less attractive option to solve this PPR problem in the noiseless scenario. Strikingly, one can observe that the Wirtinger Flow approach relying on spectral initialization is not able to recover the ground truth signal. Intuitively, it may be explained by the fact that spectral initialization provides an initial point too far from the global optimum, resulting in Wirtinger Flow to get stuck in a local minima instead. This first experiment suggests that the performance of WF-based methods for PPR is tightly related to the quality of initial points, which we will investigate in detail in the next section.

6.2. Comparison of initialization strategies for PPR-WF. Choice of initial points in nonconvex problems is usually a difficult but crucial task, as it directly impacts whether or not the considered algorithm will be able to recover the global optimum of the problem. The proposed PPR-WF algorithm does not avoid this key bottleneck, as already illustrated by the bivariate pulse recovery experiment depicted in Figure 3. To assess the role played by initial points in PPR-WF, we carefully benchmark the five initialization methods described in Section 5.2, that is spectral initialization, random phase initialization, SDP, left and right kernel Sylvester. We generated a random Gaussian complex-valued signal $\mathbf{X} \in \mathbb{C}^{N\times 2}$ with i.i.d. entries of length $N = 32$ such that $\|\mathbf{X}\|_F = 1$ which was fixed for all experiments. PPR noisy measurements (4.1) were considered for the simple measurement scheme (2.5) with $M = 2N - 1$, $P = 4$. We investigated three values of SNR, of $10, 40$ and 60 dB respectively. For each SNR value, we generated 100 independent noisy measurements and run the proposed PPR-WF algorithm using the five aforementioned initialization procedures.

Figure 4 depicts obtained reconstruction results for the three SNR scenarios, where we compare initialization methods in terms of cost function evolution $F(\boldsymbol{\xi}^{(k)})$ and normed residual $\|\boldsymbol{\xi}^{(k+1)} - \boldsymbol{\xi}^{(k)}\|_2 / \|\boldsymbol{\xi}^{(k)}\|_2$ decrease. Note that we imposed a identical number of 2500 iterations of PPR-WF for each approach to ensure fair comparisons. We also plot the empirical distribution of squared error values for each initialization for further comparison of the quality of the reconstructed signal (recall that squared error values are calculated after proper realignment of the estimated signal with the ground truth). For SNR = 10 dB (which is a very challenging scenario for PPR), there are no noticeable difference between initialization strategies: they

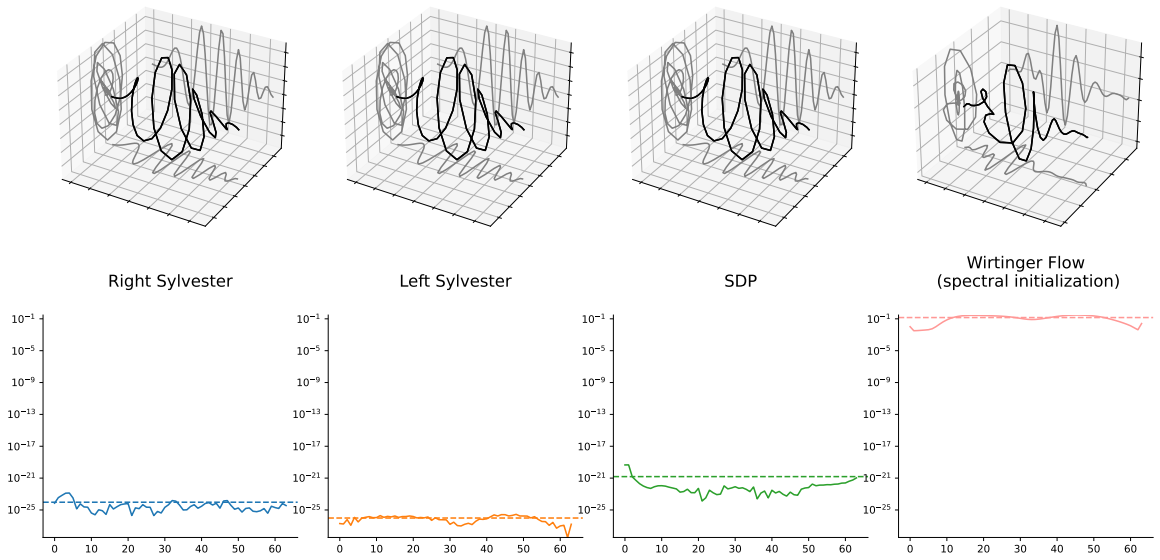

Figure 3. *Reconstruction of a bivariate pulse ($N = 64$) from noiseless PPR measurements ($M = 2N - 1, P = 4$) using the different methods described in this paper. The reconstructed signal trace and squared error per index n are shown for each approach. Dashed lines show the corresponding MSE value.*

provide similar results in terms of cost value decrease, residual evolution and error distribution. For SNR = 40 dB, two categories of initializations with distinctive behaviors start to appear: SDP and Sylvester-based approaches on one side, and on the other side, spectral and random phase initializations. On average, SDP and Sylvester-based initial points provides smaller optimal values, faster decrease of the residual and better reconstruction results in terms of squared error. This behavior is accentuated for SNR = 60 dB, where spectral and random phase initialization are unable to ensure convergence of PPR-WF to the global optimum. This agrees with the observations made in Figure 3 in the noiseless case for spectral initialization. Note that the poor performance of spectral initialization is not that surprising since it was originally designed for random phase retrieval measurements [17], and not for the case of deterministic Fourier measurements as in the present PPR setting.

These results demonstrate the importance of the choice of the initial point in PPR-WF towards good convergence properties and recovery performance. Overall, spectral and random phase strategies are systematically outperformed by SDP and left/right kernel Sylvester initializations. The latter initialization strategies provide similar performances in terms of the three figures-of-merit used here; however they exhibit very different computational costs (see Figure 6 and next section). We shall see that the very limited cost of the right kernel Sylvester approach (only requiring $\mathcal{O}(N^3)$ operations, i.e., that of a single iteration of the SDP Algorithm 4) makes it an excellent initialization for PPR-WF in many scenarios.

6.3. Recovery performance with noisy measurements. We now investigate the recovery performances (in terms of MSE and computation time) of the different proposed algorithms

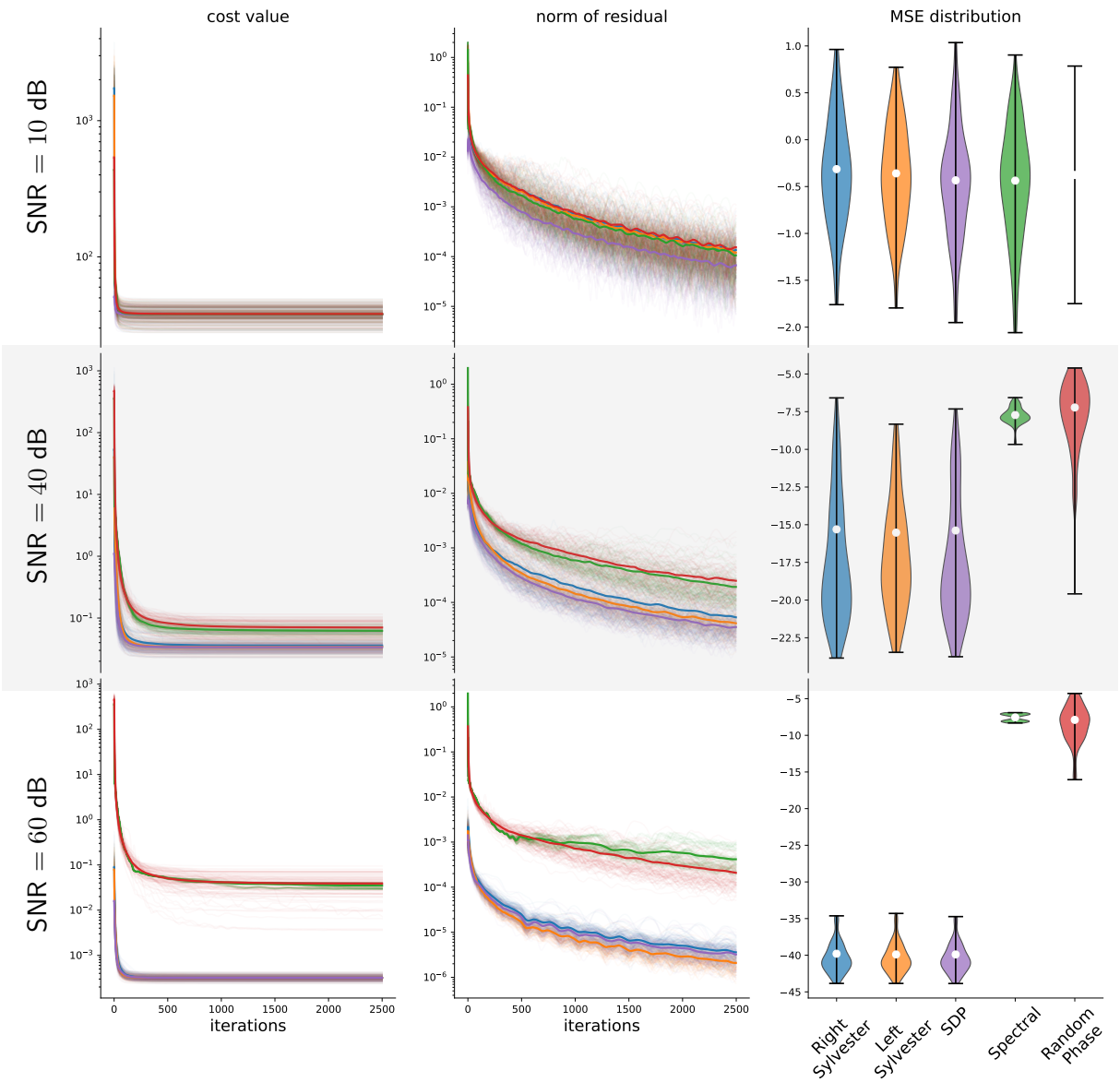

Figure 4. *Comparison of initialization strategies for PPR-WF for the recovery of an arbitrary random bivariate signal of length $N = 32$ with $M = 2N - 1$ and $P = 4$ noisy measurements. We benchmark spectral initialization, random phase initialization, left and right-kernel Sylvester initialization, and SDP initialization strategies in terms of cost function evolution, normed residual decrease and squared error distribution. Rows corresponds to values of SNR of $10, 40$ and 60 dB, respectively. For each SNR value, left and middle panels present the evolution of the cost function and residual value with iterations, respectively. For each initialization method, thin colored lines indicate trajectories for each one of the 100 independent trials, and thick colored lines display their average respective average. The right panel provides violin plots representing a kernel density estimate of squared error distribution associated to each initialization strategy. White dots indicate MSE values and horizontal bars give extreme values for each squared error distribution.*

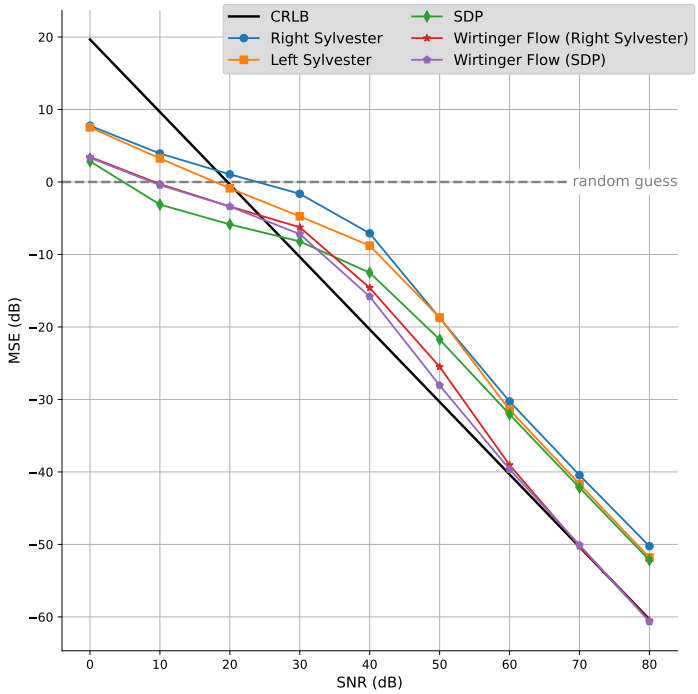

Figure 5. *Evolution of the MSE with the SNR for the different PPR reconstruction methods proposed in this paper. For Wirtinger Flow (PPR-WF), two initialization strategies (SDP and right kernel Sylvester) are considered. The ground truth signal is a randomly generated bivariate signal with $N = 32$. The simple measurement scheme for $M = 2N - 1$ and $P = 4$ was used. The thick black line indicates the corresponding Cramèr-Rao lower bound analytically derived in Appendix C.*

for PPR when dealing with noisy measurements. We consider an additive white Gaussian noise model (4.1) for which the SNR is defined in (4.2). We generated a ground truth signal $\mathbf{X} \in \mathbb{C}^{N\times 2}$ with i.i.d. Gaussian entries of length $N = 32$ such that $\|\mathbf{X}\|_F = 1$ and selected the simple, $M = 2N - 1, P = 4$ measurement scheme (2.5). For a given SNR value, the MSE associated with each one of the proposed methods to solve PPR was obtained by averaging of 100 independent reconstructions. Following our analysis of initialization strategies in Section 6.2, we considered two initializations for PPR-WF (Algorithm 5): an algebraic one using right kernel Sylvester, and a second one exploiting the output of the SDP approach (Algorithm 4).

Figure 5 displays the evolution of MSE for values of SNR ranging from 0 dB to 80 dB. As expected, the MSE decreases as the SNR increases, independently from the considered method. Overall, algorithmic methods (SDP and PPR-WF, independently from the initialization method) outperform algebraic ones (left and right kernel Sylvester) in terms of MSE values. More precisely, algebraic methods are not informative in the "low-SNR" regime (SNR $\leq$ 30 dB) as they provide (relative) MSE values above 0 dB, meaning that they do not provide a better reconstruction than a simple i.i.d. random guess scaled to the ground truth norm. Furthermore we observe that SDP is more robust to noise than PPR-WF, even when PPR-WF is initialized with the SDP output. This agrees with the fact that SDP methods are known to be robust to noise in general. In terms of initialization strategies of PPR-WF,

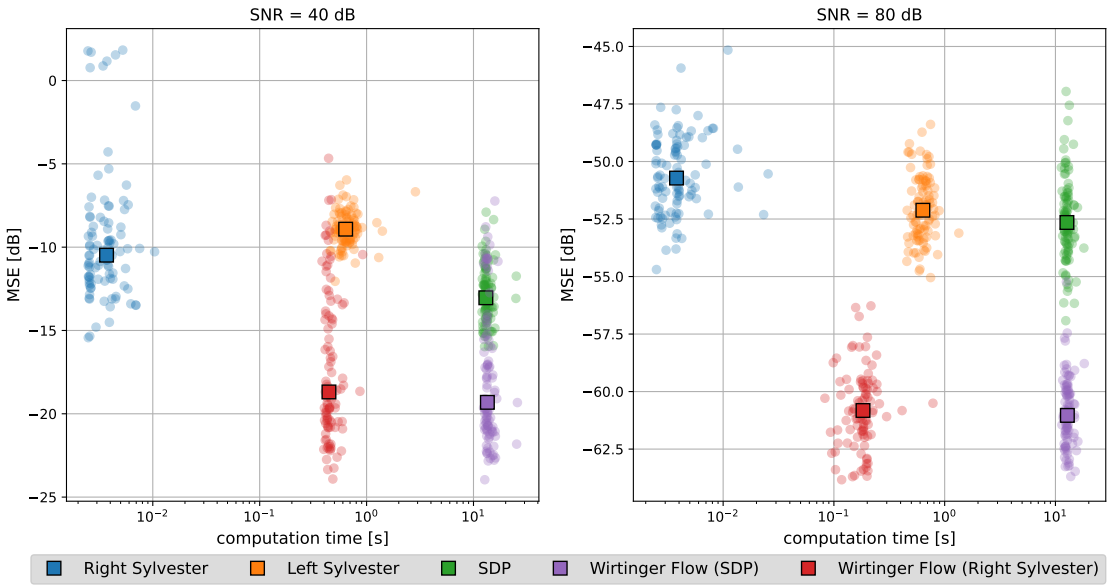

Figure 6. *Comparison of performances of the different reconstruction strategies in the MSE - computation time plane. Results corresponds to the same data plotted in Figure 5, for SNR = 40 dB (left panel) and SNR = 80 dB (right panel). For each reconstruction strategy, points represent one of the 100 independent trials; the boxed markers depicts the median error - median computation time of each method.*

SDP provides a small advantage over the right kernel Sylvester initialization for SNR values ranging from 30 to 60dB; in the low-SNR or high-SNR regime, both initialization strategies are equivalent in terms of MSE performance. The high-SNR regime ($\geq$ 60 dB) highlights several other distinctive behaviors. First, we observe that beyond SNR = 40 dB, PPR-WF outperforms all other methods, including SDP, by a few dB up to about 10 dB of relative MSE in the asymptotic regime. Second, left-kernel Sylvester and SDP provide similar performance, only improving the right-kernel Sylvester method by a small margin.

For completeness, we also provide the Cramèr-Rao lower bound (CRLB) for the noisy PPR measurement model (4.1) to characterize a lower bound on the MSE of any unbiased estimator of the ground truth signal. An analytical derivation of the resulting CRLB is given in Appendix C. Figure 5 displays the CRLB on top of MSE values obtained for each reconstruction method. We observe that the CRLB is not informative below SNR $\leq$ 20 dB as all methods provide smaller MSE values—it simply means that the CRLB is particularly pessimistic in this regime. On the contrary, the CRLB provides a meaningful lower bound in the high-SNR regime. Importantly, it demonstrates that PPR-WF is an asymptotically optimal reconstruction method for PPR—independently from the initialization strategy—since it attains the CRLB for SNR $\geq$ 60 dB.

Figure 6 depicts the performances of the different reconstruction strategies in the MSE - computation time plane. Two SNR values of SNR = 40 dB ("mid" SNR regime) and SNR = 80 dB (high-SNR regime) are considered. For each reconstruction method, a single point correspond to one of the 100 independent trials. Execution runtimes have been obtained on a 2021

Apple M1 Max MacBookPro with 32Gb RAM, using `Python 3.11.1` and `NumPy 1.24.1`. PPR-WF runtimes include the computation of the initialization (right kernel Sylvester or SDP). In terms of computation burden, the right kernel Sylvester approach outperforms all other methods by at least 2 orders of magnitude. SDP is the slowest method tested, with a computational cost greater than 4 orders of magnitude for right kernel Sylvester and 2 orders of magnitude for left kernel Sylvester and PPR-WF with the right kernel Sylvester initialization. In addition, when using the SDP initialization, the cost of PRR-WF is almost entirely dominated by the cost of SDP, which prevents one from benefiting of the low-complexity of the PRR-WF algorithm. Finally, Figure 6 shows that the best trade-off between MSE performance and computation time is attained by combining PPR-WF with the inexpensive right kernel Sylvester initialization.

6.4. Influence of number of measurements. One of the key advantages of the polarimetric measurement model in PPR is that one can easily increase the number of measurements MP by performing more polarimetric projections, i.e., by increasing P. In fact, in practical experiments it may be oftentimes easier to set up a new polarizer state $\mathbf{b}_p$ than changing the actual detector, which would be required if one desires to increase the number of Fourier measurements M. Therefore, a natural question is the following: if one seeks to increase the total number of measurements MP, is it better—in terms of MSE—to increase the number of Fourier measurements M or to increase the number of polarimetric projections P? Alternatively, can performance be simply improved by averaging over repeated independent measurements? This is a vast topic related to the question of experimental design, which requires a specific treatment which is outside the scope of the present paper. Nonetheless, we provide in the sequel a first study of the influence of the number of measurements in PPR for completeness.

Following the MSE performance analysis in Section 6.3, we use the same randomly generated ground truth signal $N = 32$ and investigate the performances for three cases leading to the same total number of measurements MP:

- $M = 2N - 1, P = 12$ case: we use the correspondence between the 2-sphere and unit vectors of $\mathbb{C}^2$ to take advantage of optimal spherical tesselations such as HEALPix [35]. In physical terms, it can be interpreted as finding one of the many possible Jones vector $\mathbf{b}_p$ corresponding to the Stokes parameters defining the rank-one matrix $\mathbf{b}_p\mathbf{b}_p^{\mathsf{H}}$. Formally, given Cartesian coordinates $(s_p^x, s_p^y, s_p^z) \in \mathbb{R}^3$ of a point on the unit 2-sphere, we define the projection vector $\mathbf{b}_p$ as:

$$\mathbf{b}_p = \frac{1}{\sqrt{2}\sqrt{1+s_p^z}} \begin{bmatrix} \jmath s_p^x \\ s_p^y + (1+s_p^z)\jmath \end{bmatrix} \quad \text{if } s_p^z \neq -1, \quad \mathbf{b}_p = \begin{bmatrix} \jmath \\ 0 \end{bmatrix} \quad \text{if } s_p^z = -1\,. \tag{6.2}$$

  Note that our choice of $P = 12$ corresponds to the first level of HEALPix sphere discretization.
- $M = 3(2N - 1), P = 4$ case: we keep the simple polarimetric measurement scheme (2.4) and increase the number M of Fourier domain measurements.
- $M = 2N - 1, P = 4$ [×3] case: we keep the simple minimal measurement scheme for PPR—which was used in Section 6.3—and repeat independently each measurement sample three times, therefore leading to the same total number of measurements as

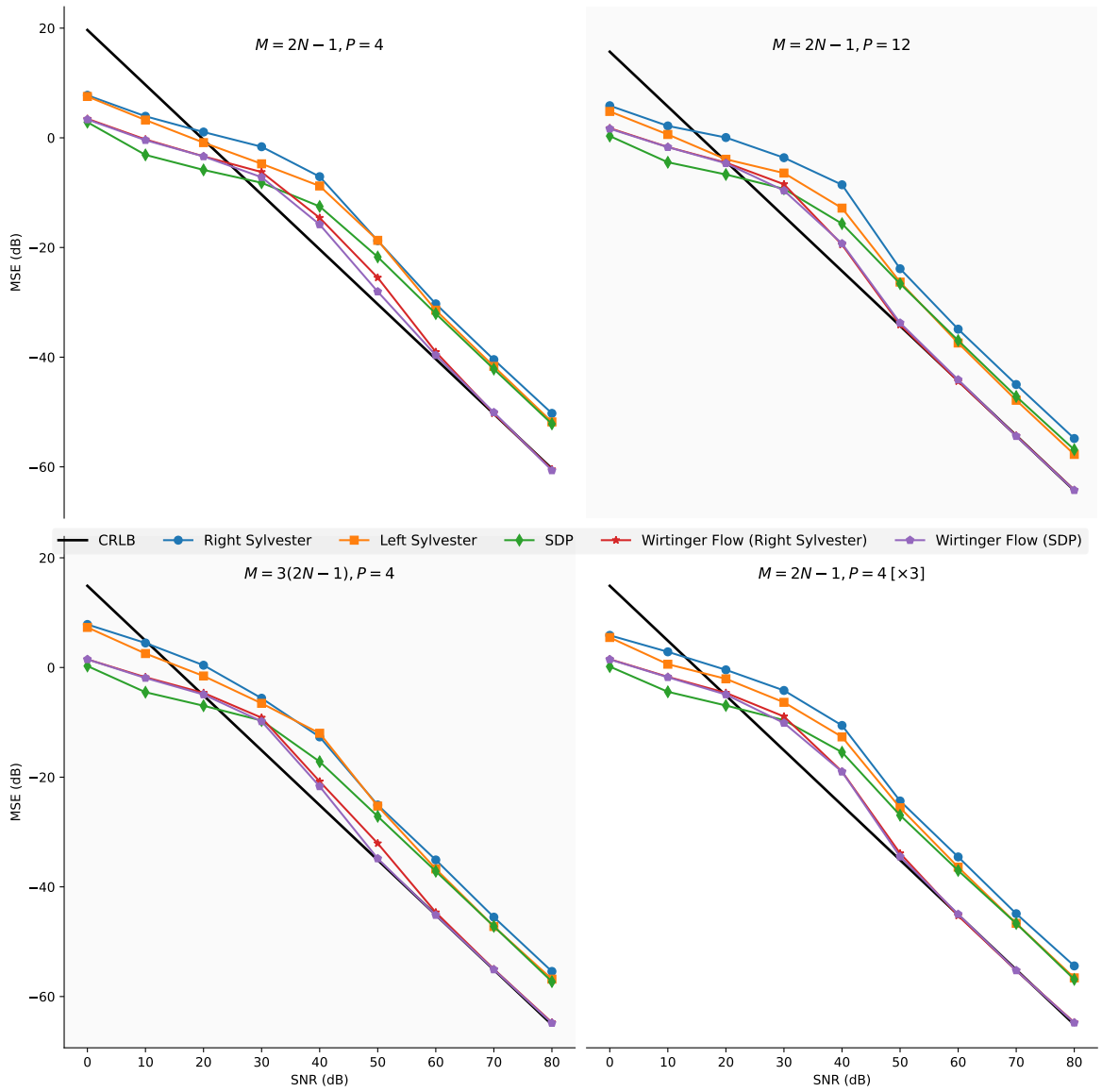

Figure 7. *Comparison of the evolution of the MSE with respect to SNR for the four measurements schemes* $M = 2N - 1, P = 4$ *(top left),* $M = 2N - 1, P = 12$ *(top right),* $M = 3(2N - 1), P = 4$ *(bottom left) and* $M = 2N - 1, P = 4\,[\times 3]$ *(i.e., the first scheme where each measurement sample is repeated independently three times). Experiments all use* $N = 32$ *and follow the same protocol as described in Section 6.3.*

the two strategies above.

Figure 7 depicts the MSE as a function of SNR for the three measurement setups described above, where results from the experiment in Section 6.3 have been reproduced for better comparison. As expected, increasing the total number of measurements MP improves overall

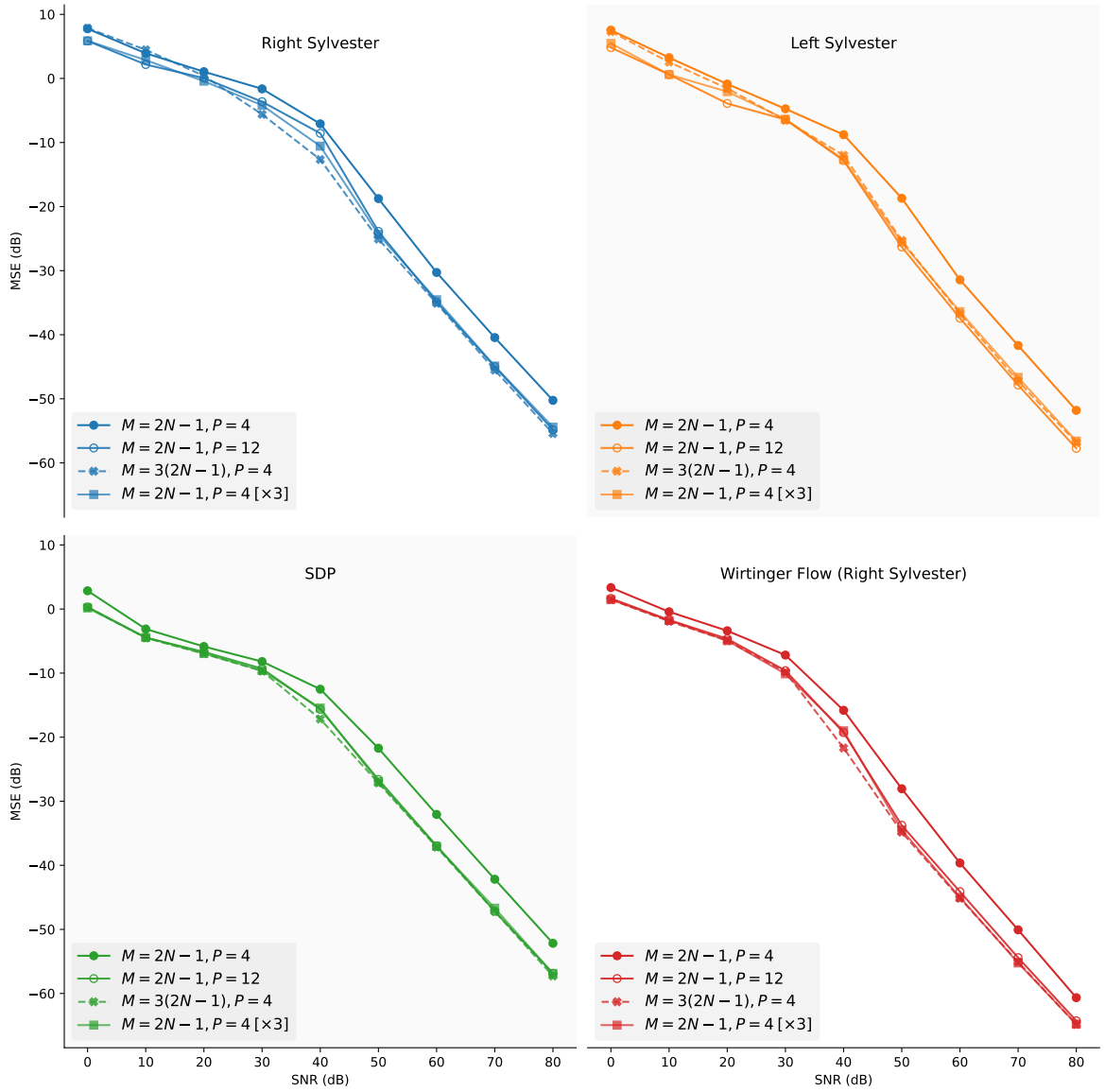

Figure 8. *Side-by-side comparison of each proposed reconstruction method for the four measurements schemes* $M = 2N-1, P = 4$, $M = 2N-1, P = 12$, $M = 3(2N-1), P = 4$ *and* $M = 2N-1, P = 4[\times 3]$.

performance: this can be directly checked by remarking that the CRLB corresponding to $M = 2N-1, P = 12$, $M = 3(2N-1), P = 4$ and $M = 2N-1, P = 4\,[\times 3]$ cases is lower that of the $M = 2N-1, P = 4$ setup presented in Figure 5. Moreover, the different proposed reconstructions methods for PPR behave similarly with one another as in our description made in Section 6.3. In particular, we note that PPR-WF—for both initialization strategies—also attains the CRLB in these three new setups, proving again that it establishes a convenient

879 approach to solve PPR.
880 Figure 8 provides a side-by-side comparison of these three measurement schemes for each
881 reconstruction method along with the minimal $M = 2N - 1, P = 4$ scheme for comparisons.
882 For PPR-WF, only the right kernel Sylvester initialization is depicted to simplify the pre-
883 sentation, without affecting conclusions. First, on the PPR-WF panel, we can remark that
884 all three measurement schemes lead to similar asymptotic performance (i.e., similar CRLB),
885 with a slight disadvantage to the $M = 2N - 1, P = 12$ case. Second, we note that for algo-
886 rithmic approaches (SDP and PPR-WF), the difference concentrates in the mid-SNR regime,
887 i.e., between 30 dB and 50 dB, where oversampling in the Fourier domain offers slight MSE
888 improvement over increasing the number of polarimetric projections or repeating measure-
889 ments. On the other hand, for algebraic approaches we observe two slight different behaviors
890 for SNR $\geq$ 20 dB: while oversampling in the Fourier domain gives the best MSE results for
891 the left kernel Sylvester approach, the strategy with increased polarimetric measurements pro-
892 vides the best asymptotic MSE results for the right kernel Sylvester approach. Nonetheless,
893 differences between measurement strategies remain tenuous, so that—in first approximation—
894 they can be considered all equivalent for this signal example. The choice of one strategy over
895 the other ones may be decided from experimental constraints or by exploiting some known
896 properties of the solution, e.g., narrow-band or average polarization properties.

897 **7. Conclusion.** This paper introduces a new model for Fourier phase retrieval called po-
898 larimetric phase retrieval (PPR), which takes advantage of polarization measurements in ap-
899 plications involving polarized light. The theoretical study of PPR relies on drawing careful
900 equivalences with two other problems, namely bivariate phase retrieval (BPR) and polynomial
901 autocorrelation factorization (PAF). In the noiseless case, these problems are found to be
902 equivalent under very general conditions, which are summarized in Figure 1. A crucial result
903 is Theorem 3.5: it shows that PAF admits a unique solution under very general conditions.
904 Therefore, the original PPR problem admits a unique solution for almost every signals. More-
905 over, the PAF representation enables the use of algebraic reconstruction strategies for PPR
906 based on GCD computations (Proposition 3.7). This original research direction is explored
907 in detail in Section 4, where we propose two fully algebraic (i.e., non-iterative) algorithms for
908 PPR relying on SVDs of Sylvester-like matrices. For completeness, Section 5 carefully adapts
909 classical phase retrieval algorithms (SDP relaxation and Wirtinger-Flow) to solve the PPR
910 problem. Section 6 provides extensive numerical experiments to benchmark the performances
911 of each approach. These results demonstrate that, if one is interested in a scalable, computa-
912 tionally efficient and robust to noise reconstruction strategy, then both algebraic and iterative
913 approaches should be combined. In practice, the best trade-off between reconstruction per-
914 formance and computational burden for PPR combines Wirtinger Flow (PPR-WF, Algorithm
915 5) with a carefully designed initialization based on right kernel Sylvester (Algorithm 1 with
916 GCDs computations performed using Algorithm 2).
917 We believe that PPR opens promising new avenues for the exploitation of light polarization
918 in Fourier phase retrieval problems. It enables the use of algebraic methods based on GCDs
919 computations to solve the Fourier phase retrieval problem. While this research direction is
920 particularly exciting, it also raises important challenges. For instance, an important issue to
921 be addressed lies in improving the performance of algebraic methods at low SNR, e.g. with

more robust estimation of the measurement polynomials or adding some prior information about the signal to be recovered (e.g., smoothness). A second challenge lies in extending the presented approaches to the case of polarized images, which is not straightforward at all since properties of polynomials with multiple variables (and their GCDs) differ considerably from their single variable counterpart. Combined with recent advances in non-convex gradient descent optimization (e.g., using partial normalizations of the gradient [19]), algebraic approaches pave the way for computationally efficient reconstruction strategies in polarimetric Fourier phase retrieval problems.

Appendix A. Relation between Fourier measurements and measurements polynomials.

Proof of Lemma 3.1. Recall that the discrete Fourier transform of $\{\mathbf{x}[n]\}_{n=0}^{N-1}$ is denoted by $\boldsymbol{\mathfrak{X}}[m] = [\mathfrak{X}_1[m], \mathfrak{X}_2[m]]^\top \in \mathbb{C}^2$ for $m = 0, 1, \dots, M-1$, see (2.2). Then the Fourier entries can be related to polynomials $X_1(z)$ and $X_2(z)$ by comparing (2.2) with (3.1):

$$\mathfrak{X}_1[m] = X_1\left(e^{-\jmath 2\pi \frac{m}{M}}\right), \quad \mathfrak{X}_2[m] = X_2\left(e^{-\jmath 2\pi \frac{m}{M}}\right),$$

for any $m = 0, 1, \dots, M-1$. Similarly, comparing (2.2) with (3.2), their conjugates can be expressed through the conjugate reflection polynomials $\widetilde{X}_1(z)$ and $\widetilde{X}_2(z)$

$$\mathfrak{X}_1^*[m] = X_1^*\left(e^{-\jmath 2\pi \frac{m}{M}}\right) = \sum_{n=0}^{N-1} x_1[n]^* e^{2\pi\jmath \frac{nm}{M}} = e^{\jmath 2\pi \frac{m(N-1)}{M}} \widetilde{X}_1\left(e^{-\jmath 2\pi \frac{m}{M}}\right),$$

$$\mathfrak{X}_2^*[m] = X_2^*\left(e^{-\jmath 2\pi \frac{m}{M}}\right) = \sum_{n=0}^{N-1} x_2[n]^* e^{2\pi\jmath \frac{nm}{M}} = e^{\jmath 2\pi \frac{m(N-1)}{M}} \widetilde{X}_2\left(e^{-\jmath 2\pi \frac{m}{M}}\right).$$

As a result, thanks to (2.3), BPR measurements can be expressed in terms of measurement polynomials $\Gamma_{ij}(z)$ as follows:

$$\begin{aligned}
\boldsymbol{\mathfrak{F}}[m] &= \begin{bmatrix} |\mathfrak{X}_1[m]|^2 & \mathfrak{X}_1[m]\mathfrak{X}_2[m]^* \\ \mathfrak{X}_2[m]\mathfrak{X}_1[m]^* & |\mathfrak{X}_2[m]|^2 \end{bmatrix} \\
&= e^{\jmath 2\pi \frac{m(N-1)}{M}} \begin{bmatrix} X_1\left(e^{-\jmath 2\pi \frac{m}{M}}\right) \widetilde{X}_1\left(e^{-\jmath 2\pi \frac{m}{M}}\right) & X_1\left(e^{-\jmath 2\pi \frac{m}{M}}\right) \widetilde{X}_2\left(e^{-\jmath 2\pi \frac{m}{M}}\right) \\ X_2\left(e^{-\jmath 2\pi \frac{m}{M}}\right) \widetilde{X}_1\left(e^{-\jmath 2\pi \frac{m}{M}}\right) & X_2\left(e^{-\jmath 2\pi \frac{m}{M}}\right) \widetilde{X}_2\left(e^{-\jmath 2\pi \frac{m}{M}}\right) \end{bmatrix} \\
&= e^{\jmath 2\pi \frac{m(N-1)}{M}} \boldsymbol{\Gamma}(e^{-\jmath 2\pi \frac{m}{M}}),
\end{aligned}$$

which completes the proof. ∎

Proof of Theorem 3.2. The proof essentially comes down to showing the one-to-one correspondences summarized in Figure 1. More precisely, we show the one-to-one correspondence between the data (measurement matrix polynomial $\boldsymbol{\Gamma}(z)$ in PAF, spectral matrices $\{\boldsymbol{\mathfrak{F}}[m]\}_{m=0}^{M-1}$ in BPR) as well as the one-to-one correspondence between sets of solutions (polynomials $X_1(z)$ and $X_2(z)$ in PAF, vectors components $\mathbf{x}_1$ and $\mathbf{x}_2$ in BPR). First note that the mapping between $\mathbb{C}^N$ and $\mathbb{C}_{\leq N-1}[z]$ is a linear one-to-one map (and is an isomorphism):

$$\mathbf{a} = \begin{bmatrix} a[0] & a[1] & \cdots & a[N-1] \end{bmatrix}^\top \mapsto A(z) = a[0] + za[1] + \cdots + z^{N-1}a[N-1].$$

Hence, the signals $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{C}^N$ can be uniquely recovered from the polynomials $X_1, X_2 \in \mathbb{C}_{\leq N-1}[z]$ and vice versa. Similarly, thanks to (3.5), the Fourier covariance measurements $\{\mathfrak{F}[m]\}_{m=0}^{M-1}$ are a linear transformation of the sequence

$$\left\{\mathbf{\Gamma}\left(e^{-\jmath 2\pi \frac{m}{M}}\right)\right\}_{m=0}^{M-1}$$

of evaluations of the matrix polynomial $\mathbf{\Gamma}(z)$ at a set of M distinct points $\{e^{-\jmath 2\pi \frac{m}{M}}\}_{m=0}^{M-1}$ on the complex plane. If $M \geq 2N-1$ (the degree of the polynomials plus one), then it is known that the coefficients of the polynomials can be uniquely recovered from the evaluations at M distinct points, and therefore the following map is an injection

$$\begin{aligned} \mathbb{C}^{2\times 2}_{\leq 2N-2} &\to (\mathbb{C}^{2\times 2})^M \\ \mathbf{\Gamma}(z) &\mapsto \{\mathfrak{F}[m]\}_{m=0}^{M-1}, \end{aligned}$$

which completes the proof. ∎

Appendix B. Multiplication matrices, Sylvester matrices and greatest common divisors.

For $\mathbf{q} \in \mathbb{C}^{D+1}$ and an integer T, we define the following $(D+T+1) \times (T+1)$ matrix

$$\mathbf{M}_T(\mathbf{q}) = \underbrace{\begin{bmatrix} q_0 & & \\ \vdots & \ddots & \\ q_K & & q_0 \\ & \ddots & \vdots \\ & & q_K \end{bmatrix}}_{T+1 \text{ columns}}.$$

This matrix is called the *multiplication matrix*, as it is represents multiplication by the polynomial $Q \in \mathbb{C}_{\leq D}[z]$, whose coefficients are in $\mathbf{q}$. Indeed, for any $F \in \mathbb{C}_{\leq T}[z]$ and $A(z) = F(z)Q(z)$, the coefficient vectors $\mathbf{f} \in \mathbb{C}^{T+1}$ and $\mathbf{a} \in \mathbb{C}^{T+D+1}$ are linked with

$$\mathbf{a} = \mathbf{M}_T(\mathbf{q})\mathbf{f},$$

see also [21] for more details on multiplication matrices.

Remark B.1. Note that for $A, B \in \mathbb{C}_{\leq L}[z]$ the Sylvester matrix (4.8) is nothing but a horizontal stack of multiplication matrices

$$\mathcal{S}_D(A,B) = \begin{bmatrix} \mathbf{M}_{L-D}(\mathbf{a}) & \mathbf{M}_{L-D}(\mathbf{b}) \end{bmatrix}.$$

Hence for the vectors $\mathbf{u} \in \mathbb{C}^{L-D+1}$ and $\mathbf{v} \in \mathbb{C}^{L-D+1}$, the product

$$\mathbf{w} = \mathcal{S}_D(A,B) \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix},$$

corresponds to the coefficients of the polynomial

$$W(z) = A(z)U(z) + B(z)V(z).$$

Proof of Proposition 4.4. We first note that (4.11) is known (see, for example, [72, Theorem 4.7]). Thus, we are left prove the second part, which is somewhat related to [72, Remark 4.8]. We write $A(z) = F(z)Q(z)$, $B(z) = G(z)Q(z)$, so that $\gcd(A,B) = Q(z)$ and $F, G \in \mathbb{C}_{\leq L-K}[z]$. Consider the multiplication matrix

$$\mathbf{M}_{2L-D-K}(\mathbf{q}) \in \mathbb{C}^{(2L-K+1)\times(2L-D-K+1)}.$$

Note that $\mathbf{M}_{2L-D-K}(\mathbf{q})$ is full column rank (equal to $2L-D-K+1$) and therefore, by (4.11) (see also [72, Theorem 4.7]), we have

$$\operatorname{rank} \mathcal{S}_D(A,B) = \operatorname{rank} \mathbf{M}_{2L-D-K}(\mathbf{q}) \tag{B.1}$$

We will show that not only the ranks, but also the ranges of $\mathcal{S}_D(A,B)$ and $\mathbf{M}_{2L-D-K}(\mathbf{q})$ coincide. Note that, by Remark B.1, the range of $\mathcal{S}_D(A,B)$ corresponds to all polynomials $R \in \mathbb{C}_{\leq 2L-D}[z]$ that can be represented as

$$R(z) = U(z)A(z) + V(z)B(z) = Q(z)(U(z)F(z) + V(z)G(z)), \tag{B.2}$$

and therefore any element in the range of $\mathcal{S}_D(A,B)$ belongs to the range of $\mathbf{M}_{2L-D-K}(\mathbf{q})$. Hence, by (B.1), the ranges of the two matrices coincide, which implies that their left kernels coincide as well. In particular, the following equivalence holds true

$$\mathbf{u}^\top \mathcal{S}_D(A,B) = 0 \iff \mathbf{u}^\top \mathbf{M}_{2L-D-K}(\mathbf{q}) = 0.$$

Finally, easy algebraic calculations (see also, for instance, [72, Eq. (33)]) show that

$$\mathbf{u}^\top \mathbf{M}_{2L-D-K}(\mathbf{q}) = \mathbf{q}^\top \mathcal{H}_{K+1}(\mathbf{u}).$$

Therefore, $\mathbf{q}$ is in the left kernel of $\mathbf{H}$ (4.12). Conversely, assume that there is a polynomial $\mathbf{q}'$ such that $\mathbf{q}'^\top \mathbf{H} = 0$. Then the image of $\mathbf{M}_{2L-D-K}(\mathbf{q}')$ must be a subspace of the image of $\mathbf{M}_{2L-D-K}(\mathbf{q})$, which is only possible if $\mathbf{q}$ is a divisor of $\mathbf{q}'$ (which implies that $\mathbf{q}'$ is proportional to $\mathbf{q}$). This shows that the left kernel of $\mathbf{H}$ has dimension 1, hence $\operatorname{rank} \mathbf{H} = K+1-1 = K$. ∎

Appendix C. Cramèr-Rao bound for PPR. Several authors have considered Cramèr-Rao bounds for the classical phase retrieval problem with additive white gaussian noise [1, 2, 55]. These results directly apply to the additive Gaussian noise PPR model (4.1) since it can be equivalently rewritten as a particular one-dimensional noise model (the PPR-1D model introduced in Section 2.4). For completeness, we provide below an alternative derivation of the Cramèr-rao bound described in [55], where we use a full complex-domain approach instead of considering separate Cramèr-Rao bounds on amplitude and phase. Since measurement noise $n_{m,p}$ in (4.1) is i.i.d. Gaussian distributed with variance σ^2, the probability density function of the vector of observations $\mathbf{y}$ is given by

$$p(\mathbf{y}|\boldsymbol{\xi}) = \prod_{m=0}^{M-1}\prod_{p=0}^{P-1} p(y_{m,p}|\boldsymbol{\xi}) \tag{C.1}$$

$$= \prod_{m=0}^{M-1}\prod_{p=0}^{P-1} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\left(y_{m,p} - \boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m,p}\boldsymbol{\xi}\right)^2}{2\sigma^2}\right], \tag{C.2}$$

where we recall that $\mathbf{C}_{m,p} = \mathbf{c}_{m,p}\mathbf{c}_{m,p}^{\mathsf{H}}$ with $\mathbf{c}_{m,p} = \mathbf{b}_p^* \otimes \mathbf{a}_m$ by definition. The log-likelihood of observations reads

$$\log p(\mathbf{y}|\mathbf{x}_{\text{vec}}) = -\frac{MP}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{m=0}^{M-1}\sum_{p=0}^{P-1}\left(y_{m,p} - \boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m,p}\boldsymbol{\xi}\right)^2. \tag{C.3}$$

Since one wants to estimate the complex parameter vector $\boldsymbol{\xi}$, it is necessary to use the complex Fisher Information Matrix (FIM) [73, 46, 53], which reads

$$\mathcal{J}_{\boldsymbol{\xi}} = \begin{bmatrix} \mathcal{I}_{\boldsymbol{\xi}} & \mathcal{P}_{\boldsymbol{\xi}} \\ \mathcal{P}_{\boldsymbol{\xi}}^* & \mathcal{I}_{\boldsymbol{\xi}}^* \end{bmatrix} \in \mathbb{C}^{4N\times 4N}, \tag{C.4}$$

where entries are defined using Wirtinger derivatives [44] since $\boldsymbol{\xi}$ is a complex vector:

$$\mathcal{I}_{\boldsymbol{\xi}} = \mathbf{E}\left[\left(\nabla_{\boldsymbol{\xi}^*}\log p(\mathbf{y}|\boldsymbol{\xi})\right)\left(\nabla_{\boldsymbol{\xi}^*}\log p(\mathbf{y}|\boldsymbol{\xi})\right)^{\mathsf{H}}\right], \tag{C.5}$$

$$\mathcal{P}_{\boldsymbol{\xi}} = \mathbf{E}\left[\left(\nabla_{\boldsymbol{\xi}^*}\log p(\mathbf{y}|\boldsymbol{\xi})\right)\left(\nabla_{\boldsymbol{\xi}^*}\log p(\mathbf{y}|\boldsymbol{\xi})\right)^{\top}\right]. \tag{C.6}$$

Note that the FIM $\mathcal{J}_{\boldsymbol{\xi}}$ defined in (C.4) is isomorphic to the real FIM which would have been obtained by stacking the real and imaginary parts of $\boldsymbol{\xi}$ in a single long vector [46]. This explains why $\mathcal{J}_{\boldsymbol{\xi}}$ has dimensions $4N \times 4N$. Using properties of Wirtinger derivatives, we get

$$\nabla_{\boldsymbol{\xi}^*}\log p(\mathbf{y}|\boldsymbol{\xi}) = -\frac{1}{\sigma^2}\sum_{m=0}^{M-1}\sum_{p=0}^{P-1}(y_{m,p} - \boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m,p}\boldsymbol{\xi})\mathbf{C}_{m,p}\boldsymbol{\xi}\,. \tag{C.7}$$

This allows to compute explicitly the block terms $\mathcal{I}_{\boldsymbol{\xi}}$ and $\mathcal{P}_{\boldsymbol{\xi}}$ that define $\mathcal{J}_{\boldsymbol{\xi}}$. Using noise independence, one gets

$$\mathcal{I}_{\boldsymbol{\xi}} = \frac{1}{\sigma^4}\mathbf{E}\left[\left(\sum_{m,p}(y_{m,p} - \boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m,p}\boldsymbol{\xi})\mathbf{C}_{m,p}\boldsymbol{\xi}\right)\left(\sum_{m',p'}(y_{m',p'} - \boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m',p'}\boldsymbol{\xi})\boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m',p'}\right)\right] \tag{C.8}$$

$$= \frac{1}{\sigma^4}\sum_{m,p,m',p'}\mathbf{E}\left[n_{m,p}n_{m',p'}\right]\mathbf{C}_{m,p}\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m',p'} \tag{C.9}$$

$$= \frac{1}{\sigma^2}\sum_{m,p}\mathbf{C}_{m,p}\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{H}}\mathbf{C}_{m,p} \tag{C.10}$$

$$= \frac{1}{\sigma^2}\sum_{m,p}|\mathbf{c}_{m,p}^{\mathsf{H}}\boldsymbol{\xi}|^2\mathbf{c}_{m,p}\mathbf{c}_{m,p}^{\mathsf{H}}. \tag{C.11}$$

Similar calculations leads to:

$$\mathcal{P}_{\boldsymbol{\xi}} = \frac{1}{\sigma^2}\sum_{ij}\mathbf{C}_{m,p}\boldsymbol{\xi}(\boldsymbol{\xi})^{\top}\mathbf{C}_{m,p}^{\top} = \frac{1}{\sigma^2}\sum_{m,p}\left(\mathbf{c}_{m,p}^{\mathsf{H}}\boldsymbol{\xi}\right)^2\mathbf{c}_{m,p}\mathbf{c}_{m,p}^{\top}. \tag{C.12}$$

A key result [53] is that the inverse of the complex FIM (C.4) provides a lower bound on the covariance and pseudo-covariance of any unbiased estimator $\hat{\boldsymbol{\xi}}$ of the complex parameter $\boldsymbol{\xi}$:

$$\begin{bmatrix} \operatorname{cov}\hat{\boldsymbol{\xi}} & \operatorname{pcov}\hat{\boldsymbol{\xi}} \\ \operatorname{pcov}\hat{\boldsymbol{\xi}}^* & \operatorname{cov}\hat{\boldsymbol{\xi}}^* \end{bmatrix} \succeq \mathcal{J}_{\boldsymbol{\xi}}^{-1}. \tag{C.13}$$

When the complex FIM is singular—as in phase retrieval [1, 2]—,one can show its pseudo-inverse remains a valid lower bound for the MSE; following the discussion in [55], we still refer to the resultant bound as the CRB with little abuse. In particular, we obtain the following bound on the MSE on any unbiased PPR estimator $\hat{\mathbf{X}}$ for the model (4.1):

$$\text{MSE}(\hat{\mathbf{X}}) = \mathbf{E}\|\hat{\mathbf{X}} - \mathbf{X}\|_F^2 = \mathbf{E}\|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_2^2 = \operatorname{Tr}\operatorname{cov}\hat{\boldsymbol{\xi}} \geq \operatorname{Tr}\left(\left[\mathcal{J}_{\boldsymbol{\xi}}^{\dagger}\right]_{[:2N,:2N]}\right) \tag{C.14}$$

where the subscript $_{[:2N,:2N]}$ denotes the restriction to the upper-left block of $\mathcal{J}_{\boldsymbol{\xi}}^{\dagger}$.

REFERENCES

[1] R. Balan, *Reconstruction of signals from magnitudes of redundant representations: The complex case*, Foundations of Computational Mathematics, 16 (2016), pp. 677–721, https://doi.org/10.1007/s10208-015-9261-0.

[2] A. S. Bandeira, J. Cahill, D. G. Mixon, and A. A. Nelson, *Saving phase: Injectivity and stability for phase retrieval*, Applied and Computational Harmonic Analysis, 37 (2014), pp. 106–125, https://doi.org/10.1016/j.acha.2013.10.002.

[3] A. S. Bandeira, Y. Chen, and D. G. Mixon, *Phase retrieval from power spectra of masked signals*, Information and Inference: a Journal of the IMA, 3 (2014), pp. 83–102, https://doi.org/10.1093/imaiai/iau002.

[4] A. H. Barnett, C. L. Epstein, L. Greengard, and J. Magland, *Geometry of the Phase Retrieval Problem: Graveyard of Algorithms*, Cambridge University Press, 1 ed., Apr. 2022, https://doi.org/10.1017/9781009003919.

[5] A. Baroni, V. Chamard, and P. Ferrand, *Extending Quantitative Phase Imaging to Polarization-Sensitive Materials*, Physical Review Applied, 13 (2020), p. 054028, https://doi.org/10.1103/PhysRevApplied.13.054028.

[6] A. Baroni and P. Ferrand, *Reference-free quantitative microscopic imaging of coherent arbitrary vectorial light beams*, Optics Express, 28 (2020), p. 35339, https://doi.org/10.1364/OE.408665.

[7] A. Beck, *First-order methods in optimization*, SIAM, 2017, https://doi.org/10.1137/1.9781611974997.

[8] R. Beinert, *Non-negativity constraints in the one-dimensional discrete-time phase retrieval problem*, Information and Inference: A Journal of the IMA, 6 (2017), pp. 213–224, https://doi.org/https://doi.org/10.1093/imaiai/iaw018.

[9] R. Beinert, *One-dimensional phase retrieval with additional interference intensity measurements*, Results in Mathematics, 72 (2017), pp. 1–24, https://doi.org/10.1007/s00025-016-0633-9.

[10] R. Beinert and G. Plonka, *Ambiguities in one-dimensional discrete phase retrieval from fourier magnitudes*, Journal of Fourier Analysis and Applications, 21 (2015), pp. 1169–1198, https://doi.org/10.1007/s00041-015-9405-2.

[11] R. Beinert and G. Plonka, *Enforcing uniqueness in one-dimensional phase retrieval by additional signal information in time domain*, Applied and Computational Harmonic Analysis, 45 (2018), pp. 505–525, https://doi.org/10.1016/j.acha.2016.12.002.

[12] T. Bendory, R. Beinert, and Y. C. Eldar, *Fourier Phase Retrieval: Uniqueness and Algorithms*, in Compressed Sensing and its Applications, H. Boche, G. Caire, R. Calderbank, M. März, G. Kutyniok, and R. Mathar, eds., Springer International Publishing, Cham, 2017, pp. 55–91, https://doi.org/10.1007/978-3-319-69802-1_2. Series Title: Applied and Numerical Harmonic Analysis.

[13] T. Bendory and D. Edidin, *Algebraic Theory of Phase Retrieval*, Notices of the American Mathematical Society, 69 (2022), p. 1, https://doi.org/10.1090/noti2540.

[14] T. Bendory, Y. C. Eldar, and N. Boumal, *Non-convex phase retrieval from STFT measurements*, IEEE Transactions on Information Theory, 64 (2017), pp. 467–484, https://doi.org/10.1109/TIT.2017.2745623.

[15] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 199–225, https://doi.org/10.1137/110848074.

[16] E. J. Candès, X. Li, and M. Soltanolkotabi, *Phase retrieval from coded diffraction patterns*, Applied and Computational Harmonic Analysis, 39 (2015), pp. 277–299, https://doi.org/10.1016/j.acha.2014.09.004.
[17] E. J. Candès, X. Li, and M. Soltanolkotabi, *Phase Retrieval via Wirtinger Flow: Theory and Algorithms*, IEEE Transactions on Information Theory, 61 (2015), pp. 1985–2007, https://doi.org/10.1109/TIT.2015.2399924.
[18] H. N. Chapman and K. A. Nugent, *Coherent lensless X-ray imaging*, Nature photonics, 4 (2010), pp. 833–839, https://doi.org/10.1038/nphoton.2010.240.
[19] Z. Chen, Y. Zhou, Y. Liang, and Z. Lu, *Generalized-smooth nonconvex optimization is as efficient as smooth nonconvex optimization*, 2023, https://arxiv.org/pdf/2303.02854.pdf.
[20] R. A. Chipman, G. Young, and W. S. T. Lam, *Polarized light and optical systems*, Optical sciences and applications of light, Taylor & Francis, CRC Press, Boca Raton, 2018, https://doi.org/10.1201/9781351129121.
[21] R. M. Corless, P. M. Gianni, B. M. Trager, and S. M. Watt, *The singular value decomposition for polynomial systems*, in Proceedings of the 1995 international symposium on Symbolic and algebraic computation - ISSAC '95, Montreal, Quebec, Canada, 1995, ACM Press, pp. 195–207, https://doi.org/10.1145/220346.220371.
[22] X. Dai, S. Xu, X. Yang, K. C. Zhou, C. Glass, P. C. Konda, and R. Horstmeyer, *Quantitative Jones matrix imaging using vectorial Fourier ptychography*, Biomedical Optics Express, 13 (2022), p. 1457, https://doi.org/10.1364/BOE.448804.
[23] V. Elser, *Phase retrieval by iterated projections*, Journal of the Optical Society of America A, 20 (2003), pp. 40–55, https://doi.org/10.1364/JOSAA.20.000040.
[24] V. Elser, T.-Y. Lan, and T. Bendory, *Benchmark problems for phase retrieval*, SIAM Journal on Imaging Sciences, 11 (2018), pp. 2429–2455, https://doi.org/10.1137/18M1170364.
[25] A. Fannjiang and T. Strohmer, *The numerics of phase retrieval*, Acta Numerica, 29 (2020), pp. 125–228, https://doi.org/10.1017/S0962492920000069.
[26] P. Ferrand, M. Allain, and V. Chamard, *Ptychography in anisotropic media*, Optics Letters, 40 (2015), p. 5144, https://doi.org/10.1364/OL.40.005144.
[27] P. Ferrand, A. Baroni, M. Allain, and V. Chamard, *Quantitative imaging of anisotropic material properties with vectorial ptychography*, Optics Letters, 43 (2018), p. 763, https://doi.org/10.1364/OL.43.000763.
[28] J. R. Fienup, *Reconstruction of an object from the modulus of its Fourier transform*, Optics Letters, 3 (1978), pp. 27–29, https://doi.org/10.1364/OL.3.000027.
[29] J. R. Fienup, J. C. Marron, T. J. Schulz, and J. H. Seldin, *Hubble space telescope characterized by using phase-retrieval algorithms*, Applied Optics, 32 (1993), pp. 1747–1767, https://doi.org/10.1364/AO.32.001747.
[30] J. J. Gil and R. Ossikovski, *Polarized light and the Mueller matrix approach*, CRC press, Boca Raton, 2 ed., 2022, https://doi.org/doi.org/10.1201/9780367815578.
[31] D. H. Goldstein, *Polarized light*, CRC press, Boca Raton, 3 ed., 2017, https://doi.org/10.1201/b10436.
[32] T. Goldstein, C. Studer, and R. Baraniuk, *A field guide to forward-backward splitting with a FASTA implementation*, 2014, https://doi.org/10.48550/arXiv.1411.3406.
[33] J. W. Goodman, *Introduction to Fourier optics*, Roberts and Company publishers, Englewood, CO, 3 ed., 2005.
[34] J. P. Gordon and H. Kogelnik, *PMD fundamentals: Polarization mode dispersion in optical fibers*, Proceedings of the National Academy of Sciences, 97 (2000), pp. 4541–4550, https://doi.org/10.1073/pnas.97.9.4541.
[35] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, *HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere*, The Astrophysical Journal, 622 (2005), p. 759, https://doi.org/10.1086/427976.
[36] P. Grohs, S. Koppensteiner, and M. Rathmair, *Phase retrieval: uniqueness and stability*, SIAM Review, 62 (2020), pp. 301–350, https://doi.org/10.1137/19M1256865.
[37] S.-M. Guo, L.-H. Yeh, J. Folkesson, I. E. Ivanov, A. P. Krishnan, M. G. Keefe, E. Hashemi, D. Shin, B. B. Chhun, N. H. Cho, M. D. Leonetti, M. H. Han, T. J. Nowakowski, and S. B.

Mehta, *Revealing architectural order with quantitative label-free imaging and deep learning*, eLife, 9 (2020), p. e55502, https://doi.org/10.7554/eLife.55502.

[38] K. Huang, Y. C. Eldar, and N. D. Sidiropoulos, *Phase Retrieval from 1D Fourier Measurements: Convexity, Uniqueness, and Algorithms*, IEEE Transactions on Signal Processing, 64 (2016), pp. 6105–6117, https://doi.org/10.1109/TSP.2016.2601291, http://ieeexplore.ieee.org/document/7547374/ (accessed 2020-04-03).

[39] K. Jaganathan, Y. Eldar, and B. Hassibi, *Phase retrieval with masks using convex optimization*, in 2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015, pp. 1655–1659, https://doi.org/10.1109/ISIT.2015.7282737.

[40] K. Jaganathan, Y. C. Eldar, and B. Hassibi, *STFT phase retrieval: Uniqueness guarantees and recovery algorithms*, IEEE Journal of selected topics in signal processing, 10 (2016), pp. 770–781, https://doi.org/10.1109/JSTSP.2016.2549507.

[41] K. Jaganathan and B. Hassibi, *Reconstruction of Signals From Their Autocorrelation and Cross-Correlation Vectors, With Applications to Phase Retrieval and Blind Channel Estimation*, IEEE Transactions on Signal Processing, 67 (2019), pp. 2937–2946, https://doi.org/10.1109/TSP.2019.2911254.

[42] K. Jaganathan, S. Oymak, and B. Hassibi, *Sparse phase retrieval: Uniqueness guarantees and recovery algorithms*, IEEE Transactions on Signal Processing, 65 (2017), pp. 2402–2410, https://doi.org/10.1109/TSP.2017.2656844.

[43] X. Jiang, S. Rajan, and X. Liu, *Wirtinger Flow Method With Optimal Stepsize for Phase Retrieval*, IEEE Signal Processing Letters, 23 (2016), pp. 1627–1631, https://doi.org/10.1109/LSP.2016.2611940.

[44] K. Kreutz-Delgado, *The Complex Gradient Operator and the CR-Calculus*, June 2009, http://arxiv.org/abs/0906.4835.

[45] B. Leshem, R. Xu, Y. Dallal, J. Miao, B. Nadler, D. Oron, N. Dudovich, and O. Raz, *Direct single-shot phase retrieval from the diffraction pattern of separated objects*, Nature Communications, 7 (2016), p. 10820, https://doi.org/10.1038/ncomms10820.

[46] B. Loesch and B. Yang, *Cramér-rao bound for circular and noncircular complex independent component analysis*, IEEE transactions on signal processing, 61 (2012), pp. 365–379, https://doi.org/10.1109/TSP.2012.2226166.

[47] A. M. Maiden and J. M. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, Ultramicroscopy, 109 (2009), pp. 1256–1262, https://doi.org/10.1016/j.ultramic.2009.05.012.

[48] I. Markovsky, *Structured low-rank approximation and its applications*, Automatica, 44 (2008), pp. 891–909, https://doi.org/10.1016/j.automatica.2007.09.011.

[49] J. Miao, P. Charalambous, J. Kirz, and D. Sayre, *Extending the methodology of X-raycrystallography to allow imaging of micrometre-sized non-crystalline specimens*, Nature, 400 (1999), pp. 342–344, https://doi.org/10.1038/22498.

[50] R. P. Millane, *Phase retrieval in crystallography and optics*, Journal of the Optical Society of America A, 7 (1990), pp. 394–411, https://doi.org/10.1364/JOSAA.7.000394.

[51] R. Monteiro et al., *First-and second-order methods for semidefinite programming*, Mathematical Programming, 97 (2003), pp. 209–244, https://doi.org/10.1007/s10107-003-0451-1.

[52] K. Nagasaka, *Toward the best algorithm for approximate gcd of univariate polynomials*, Journal of Symbolic Computation, 105 (2021), pp. 4–27.

[53] E. Ollila, V. Koivunen, and J. Eriksson, *On the Cramér-Rao bound for the constrained and unconstrained complex parameters*, in 2008 5th IEEE Sensor Array and Multichannel Signal Processing Workshop, IEEE, 2008, pp. 414–418, https://doi.org/10.1109/SAM.2008.4606902.

[54] O. Pedatzur, A. Trabattoni, B. Leshem, H. Shalmoni, M. Castrovilli, M. Galli, M. Lucchini, E. Månsson, F. Frassetto, L. Poletto, et al., *Double-blind holography of attosecond pulses*, Nature Photonics, 13 (2019), pp. 91–95, https://doi.org/10.1038/s41566-018-0308-z.

[55] C. Qian, N. D. Sidiropoulos, K. Huang, L. Huang, and H. C. So, *Phase retrieval using feasible point pursuit: Algorithms and Cramér–Rao bound*, IEEE Transactions on Signal Processing, 64 (2016), pp. 5282–5296, https://doi.org/10.1109/TSP.2016.2593688.

[56] J. Ranieri, A. Chebira, Y. M. Lu, and M. Vetterli, *Phase retrieval for sparse signals: Uniqueness*

conditions, 2013, https://arxiv.org/abs/1308.3058.

[57] O. Raz, N. Dudovich, and B. Nadler, *Vectorial Phase Retrieval of 1-D Signals*, IEEE Transactions on Signal Processing, 61 (2013), pp. 1632–1643, https://doi.org/10.1109/TSP.2013.2239994.

[58] O. Raz, B. Leshem, J. Miao, B. Nadler, D. Oron, and N. Dudovich, *Direct phase retrieval in double blind Fourier holography*, Optics Express, 22 (2014), p. 24935, https://doi.org/10.1364/OE.22.024935.

[59] O. Raz, O. Schwartz, D. Austin, A. S. Wyatt, A. Schiavi, O. Smirnova, B. Nadler, I. A. Walmsley, D. Oron, and N. Dudovich, *Vectorial Phase Retrieval for Linear Characterization of Attosecond Pulses*, Physical Review Letters, 107 (2011), p. 133902, https://doi.org/10.1103/PhysRevLett.107.133902.

[60] A. Saba, J. Lim, A. B. Ayoub, E. E. Antoine, and D. Psaltis, *Polarization-sensitive optical diffraction tomography*, Optica, 8 (2021), p. 402, https://doi.org/10.1364/OPTICA.415343.

[61] D. Sayre, *Some implications of a theorem due to Shannon*, Acta Crystallographica, 5 (1952), pp. 843–843, https://doi.org/10.1107/S0365110X52002276.

[62] B. Schaefer, E. Collett, R. Smyth, D. Barrett, and B. Fraher, *Measuring the Stokes polarization parameters*, American Journal of Physics, 75 (2015), p. 6, https://doi.org/10.1119/1.2386162.

[63] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, *Phase Retrieval with Application to Optical Imaging: A contemporary overview*, IEEE Signal Processing Magazine, 32 (2015), pp. 87–109, https://doi.org/10.1109/MSP.2014.2352673.

[64] O. Smirnova, S. Patchkovskii, Y. Mairesse, N. Dudovich, D. Villeneuve, P. Corkum, and M. Y. Ivanov, *Attosecond Circular Dichroism Spectroscopy of Polyatomic Molecules*, Physical Review Letters, 102 (2009), p. 063601, https://doi.org/10.1103/PhysRevLett.102.063601.

[65] Q. Song, A. Baroni, R. Sawant, P. Ni, V. Brandli, S. Chenot, S. Vézian, B. Damilano, P. de Mierry, S. Khadir, et al., *Ptychography retrieval of fully polarized holograms from geometric-phase metasurfaces*, Nature communications, 11 (2020), p. 2651, https://doi.org/10.1038/s41467-020-16437-9.

[66] Q. Song, X. Liu, C.-W. Qiu, and P. Genevet, *Vectorial metasurface holography*, Applied Physics Reviews, 9 (2022), p. 011311, https://doi.org/10.1063/5.0078610.

[67] S. Song, J. Kim, S. Hur, J. Song, and C. Joo, *Large-area, high-resolution birefringence imaging with polarization-sensitive fourier ptychographic microscopy*, ACS Photonics, 8 (2021), pp. 158–165, https://doi.org/10.1021/acsphotonics.0c01695.

[68] S. Song, J. Kim, T. Moon, B. Seong, W. Kim, C.-H. Yoo, J.-K. Choi, and C. Joo, *Polarization-sensitive intensity diffraction tomography*, Light: Science & Applications, 12 (2023), p. 124, https://doi.org/10.1038/s41377-023-01151-0.

[69] R. Trebino, *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses*, Springer Science & Business Media, New York, 2000, https://doi.org/10.1007/978-1-4615-1181-6.

[70] J. S. Tyo, D. L. Goldstein, D. B. Chenault, and J. A. Shaw, *Review of passive imaging polarimetry for remote sensing applications*, Applied optics, 45 (2006), pp. 5453–5469, https://doi.org/10.1364/AO.45.005453.

[71] K. Usevich, J. Flamant, M. Clausel, and D. Brie, *On factorization of rank-one auto-correlation matrix polynomials*. preprint, 2023, https://hal.science/hal-04062934.

[72] K. Usevich and I. Markovsky, *Variable projection methods for approximate (greatest) common divisor computations*, Theoretical Computer Science, 681 (2017), pp. 176–198, https://doi.org/10.1016/j.tcs.2017.03.028.

[73] A. Van den Bos, *A Cramér-Rao lower bound for complex parameters*, IEEE Transactions on Signal Processing, 42 (1994), https://doi.org/10.1109/78.324755.

[74] L. Vandenberghe and S. Boyd, *Semidefinite programming*, SIAM review, 38 (1996), pp. 49–95, https://doi.org/10.1137/1038003.

[75] I. Waldspurger, A. d'Aspremont, and S. Mallat, *Phase recovery, maxcut and complex semidefinite programming*, Mathematical Programming, 149 (2015), pp. 47–81, https://doi.org/10.1007/s10107-013-0738-9.

[76] I. A. Walmsley and C. Dorrer, *Characterization of ultrashort electromagnetic pulses*, Advances in Optics and Photonics, 1 (2009), pp. 308–437, https://doi.org/10.1364/AOP.1.000308.

[77] R. Xu, M. Soltanolkotabi, J. P. Haldar, W. Unglaub, J. Zusman, A. F. J. Levi, and R. M. Leahy, *Accelerated Wirtinger Flow: A fast algorithm for ptychography*, June 2018, http://arxiv.org/

abs/1806.05546.