Phase-shifting interferometry by a covariance-based method

Abhijit Patil, Pramod Rastogi, and Benny Raphael

A novel generalized approach to phase-shifting interferometry in which phase distribution in an interferogram is evaluated in the presence of nonsinusoidal waveforms and piezoactuator device miscalibration is proposed. The approach is based on the underlying rotational invariance of signal subspaces spanned by two temporally displaced data sets. The advantage of the proposed method lies in its ability to identify arbitrary phase-step values pixelwise from an interference signal buried in noise. The robustness of the proposed method is investigated by addition of white Gaussian noise during the simulations. © 2005 Optical Society of America

OCIS codes: 120.3180, 120.5050.

1. Introduction

Phase-shifting interferometry (PSI) is an important technique that, by permitting phase-angle measurements to be made with high precision, has opened the gateway for automating the process of fringe analysis. Recall that the information that one wishes to measure in interferometry is carried as phase angles in an interferogram.[1–3] Typically, a CCD camera records interferograms for a sequence of reference phases that are produced by means of a piezoactuator device (PZT) placed in the path of a reference beam. Besides the problem of miscalibration by the PZT, other factors such as aging and hysteresis affect the measurement precision that could be obtained by PSI. This problem can be overcome by use of an algorithm that can compute the phase steps pixelwise rather than relying on previously calibrated phase-shifted values. The values of these phase steps can then be used pixelwise for the computation of the interference phase distribution.

In phase shifting, the phase value at each pixel on a data frame is computed from a series of recorded interferograms that have undergone known phase shifts with respect to one another. These techniques have broadly used two approaches, one based on equal phase shifts—typically multiples of $\pi/2$—and the other based on arbitrary phase shifts. These two approaches are usually referred to as conventional and generalized phase shifting interferometry.[4] Progress over the years has seen the development of numerous algorithms with a view to addressing problems arising from systematic[5–18] and random[19–24] sources of error in phase measurement. Furthermore, algorithms have been developed that allow for making accurate measurements in the presence of nonsinusoidal fringe profiles that arise as a direct consequence of multiple reflections inside the laser cavity or detector nonlinearity.[11,14] The last-named algorithms, however, suffer from constraints on phase-step values that could be used depending on the harmonic content of the signal waveform. Moreover, in the generalized phase-shifting approach,[25,26] i.e., that based on arbitrary phase shifts, the algorithms that have been developed are sensitive to nonsinusoidal waveforms and to random sources of error. Thus it would be important to design a robust algorithm that would incorporate salient features found in both conventional and generalized forms of PSI.

In this paper, a new approach to the whole-field estimation of phase steps on a data frame is presented. The approach, based on subspace invariance, allows for compensating for errors caused by the presence of nonsinusoidal fringe profiles and significantly relaxes the restrictions on phase shifting between successive frames. Moreover, the ability to determine the phase shifts on a recorded frame on a pixelwise basis allows for using spherical beams in phase-measurement systems based on the proposed approach. Random noise

The authors are with the Applied Computing and Mechanics Laboratory, Ecole Polytechnique Fédérale de Lausanne, 1015, Switzerland. P. Rastogi's e-mail address is pramod.rastogi@epfl.ch.

Received 15 October 2004; revised manuscript received 24 February 2005; accepted 21 March 2005.

0003-6935/05/275778-08$15.00/0

that affects phase measurement has been shown to follow a Gaussian distribution.[21] We thus investigate the influence of additive white Gaussian noise on phase measurement.

2. Estimation of Phase Steps by Rotational Invariance

The estimation of signal parameters by rotational invariance (ESPRIT) has found wide applications in areas such as radio astronomy, geophysics, sonar signal processing, electronic surveillance, and structural analysis for the estimation of parameters of finite-dimensional signal measurements.[27] It has typically been applied in these fields for the estimation of the direction of arrival of signals, in direction finding of signals, or in the estimation of frequencies buried in white Gaussian noise.[27] This method falls into the category of a high-resolution technique because of its ability to resolve spectral lines separated in frequency $f = \omega/2\pi$ by less than $1/N$ cycles per sampling interval (here, ω and N refer to the angular frequency and the number of samples, respectively), which is the resolution limit for the classical periodogram-based methods, such as a Fourier transform. The approach that we describe assimilates the phase shift applied to the PZT to the frequencies that are present in the spectrum. The method takes advantage of rotational invariance of signal subspaces spanned by two temporally displaced data sets. These data sets are obtained by canonical decomposition of a covariance matrix formed from data frames recorded temporally on the CCD array.

The method is based on a canonical decomposition of positive definite Toeplitz matrices formed from estimated covariance matrices of a sequence; in our case the sequence corresponds to the phase-shifted intensity images acquired temporally. Through this decomposition the signal and noise subspaces are separated. Because the method derives the frequency estimates by using the eigendecomposition of signal covariance matrices and in particular the subspaces associated with them, it belongs in the category of signal subspace methods. Although the spectrum estimation can also be handled efficiently by use of a fast Fourier transform, albeit with limited data samples, the resolution of closely spaced frequencies is troublesome in the presence of noise. In addition, the leakage phenomenon can cause energy in the main lobe to leak into the sidelobes, obscuring and distorting other spectral responses.[28]

The recorded fringe intensity at a point (x, y) for the tth frame in the presence of white Gaussian noise is given by

$$I(t) = I_{\mathrm{dc}} + \sum_{k=1}^{\kappa} l_k u_k{}^t + \sum_{k=1}^{\kappa} l_k{}^*(u_k{}^*)^t + \eta(t),$$

$$t = 0,\ 1,\ 2,\ \ldots,\ m,\ \ldots,\ N-1, \qquad (1)$$

where $l_k = a_k \exp(jk\varphi)$ and $u_k = \exp(jk\alpha)$; * denotes the complex conjugate, and η, the additive white Gaussian noise with mean zero and variance σ^2. Here a_k is the complex Fourier coefficient, κ is the order of harmonics, $j = \sqrt{-1}$, I_{dc} is the local average value of intensity, and φ and α are phase and phase-shift values, respectively. Coefficient a_k is in fact real and is related to the appropriate choice of phase origin at a point where the intensity reaches a maximum. In Eq. (1), 0 (corresponding to I_{dc}), α, 2α, . . . , $\kappa\alpha$, and $-\alpha$, 2α, . . . , $-\kappa\alpha$ are the frequencies $\{\omega_n\}_{n=0}^{2\kappa}$, and by evaluating them we can identify phase step α. Estimation of location of spectral lines, namely, the sinusoidal frequencies, is a classic problem in spectral estimation. Subspace-based methods such as ESPRIT have been used with success to locate the spectral lines when these are embedded in noise.

The first step consists of forming the covariance or autocorrelation matrix from N phase-shifted sequences recorded at a pixel point (x, y). The autocorrelation function for signal $I(t)$ in Eq. (1) can be defined as[29]

$$r(p) = E[I(t)I^*(t-p)] = \sum_{n=0}^{2\kappa} A_n{}^2 \exp(i\omega_n p) + \sigma^2 \delta_{p,0}, \qquad (2)$$

where $E[\ ]$ represents the expectation operator that averages over the ensemble of two data samples. Details of the derivation of Eq. (2) and A_n can be found in Appendix A. Here $\delta_{p,0}$ is the Kronecker delta function ($\delta_{g,h} = 1$ if $g = h$, and $\delta_{g,h} = 0$ otherwise). The covariance function of $I(t)$ is assumed to depend only on the lag between the two averaged samples. The autocorrelation matrix can thus be written as

$$\mathbf{R}_I = E\left\{ \begin{bmatrix} I^*(t-1) \\ I^*(t-2) \\ \cdot \\ \cdot \\ I^*(t-m) \end{bmatrix} [I(t-1)I(t-2)\ldots I(t-m)] \right\}$$

$$= \begin{bmatrix} r(0) & r^*(1) & \cdot & \cdot & r^*(m-1) \\ r(1) & \cdot & \cdot & \cdot & \cdot \\ r(2) & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & r^*(1) \\ r(m-1) & r(m-2) & \cdot & \cdot & r(0) \end{bmatrix}, \qquad (3)$$

where $r^*(p) = r(-p)$. Matrix $\mathbf{R}_I$ is positive semidefinite for all m and has the form

$$\mathbf{R}_I = \underbrace{\mathbf{A}\mathbf{P}\mathbf{A}^c}_{\mathbf{R}_s} + \underbrace{\sigma^2\mathbf{I}}_{\mathbf{R}_\varepsilon}, \qquad (4)$$

where $\mathbf{R}_s$ and $\mathbf{R}_\varepsilon$ are signal and noise contributions. Matrix $\mathbf{A}_{m\times n}$ can be written as

$$\mathbf{A}_{m\times n} = [\mathbf{a}(\omega_0) \quad \mathbf{a}(\omega_1) \ \ldots \ \mathbf{a}(\omega_{2\kappa})], \qquad (5)$$

where, for instance, the element $\mathbf{a}(\omega_0)$ is an $m \times 1$ matrix with unity entries corresponding to I_{dc}; $\mathbf{a}(\omega_1)$

$= \{1 \exp(i\alpha) \ldots \exp[j\alpha(m-1)]\}^T$, $(\,)^T$ and $(\,)^c$ are the transpose and the conjugate transpose of a vector or a matrix, respectively; and $\mathbf{I}$ is the $m \times m$ identity matrix. Matrix $\mathbf{P}_{n\times n}$ is given by

$$\mathbf{P} = \begin{bmatrix} A_0{}^2 & 0 & . & . & 0 \\ 0 & A_1{}^2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & 0 & . & . & A_{2\kappa}{}^2 \end{bmatrix}. \tag{6}$$

Inasmuch as $\mathbf{R}_I$ is positive semidefinite, its eigenvalues are nonnegative. The eigenvalues of $\mathbf{R}_I$ can be ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$ and $\lambda_{n+1} \geq \lambda_{n+2} \geq \ldots \lambda_m$. Let $\mathbf{S}_{m\times n} = [\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_n]$ be the orthonormal eigenvectors associated with $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$. The space spanned by $\{\mathbf{s}_1, \mathbf{s}_2, \ldots \mathbf{s}_n\}$ is known as the signal subspace. The set of orthonormal eigenvectors $\mathbf{G}_{m\times(m-n)} = [\mathbf{g}_1, \mathbf{g}_2, \ldots \mathbf{g}_{m-n}]$ associated with eigenvalues $\lambda_{n+1} \geq \lambda_{n+2} \geq \ldots \lambda_m$ spans a subspace known as noise subspace. As $\mathbf{APA}^c \in \mathbf{C}^{m\times m}$ ($\mathbf{C}$ represents a complex matrix) has rank $n(n < m)$, it has n eigenvalues, and the remaining $m - n$ eigenvalues are zero. Further, assuming that $(\lambda, \mathbf{r})$ is an eigenpair of $\mathbf{L} \in \mathbf{C}^{m\times m}$, then $(\lambda + \rho, \mathbf{r})$ is an eigenpair of $\mathbf{W}$, where $\mathbf{W} = \mathbf{L} + \rho\mathbf{I}$, with $\rho \in \mathbf{C}$. Hence $\lambda_t = \tilde{\lambda_t} + \sigma^2$, where t spans 1, 2, 3, . . . , m. We observe that

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n \geq \sigma^2,$$

$$\lambda_{n+1} = \lambda_{n+2} = \ldots \lambda_m = \sigma^2. \tag{7}$$

Let us set

$$\Lambda = \begin{bmatrix} \lambda_1 - \sigma^2 & 0 & . & . & 0 \\ 0 & \lambda_2 - \sigma^2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & 0 & . & . & \lambda_n - \sigma^2 \end{bmatrix}. \tag{8}$$

From Eqs. (4) and (7) we get

$$\mathbf{R}_I\mathbf{S} = \mathbf{S}\begin{bmatrix} \lambda_1 & 0 & . & . & 0 \\ 0 & \lambda_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 0 & 0 & . & . & \lambda_n \end{bmatrix} = \mathbf{APA}^c\mathbf{S} + \sigma^2\mathbf{S}. \tag{9}$$

We thus obtain

$$\mathbf{S} = \mathbf{A}\underbrace{(\mathbf{PA}^c\mathbf{S}\Lambda^{-1})}_{\Gamma} \tag{10}$$

Equation (10) leads us to the concept of rotation invariance. To understand it better, let us construct matrices $\mathbf{A}_{1_{(m-1)\times n}}$ and $\mathbf{A}_{2_{(m-1)\times n}}$ from $\mathbf{A}$ in Eq. (5) as

$$\mathbf{A}_{1_{(m-1)\times n}} = [\mathbf{I}_{(m-1)\times(m-1)}0_{(m-1)\times 1}]\mathbf{A},$$

$$\mathbf{A}_{2_{(m-1)\times n}} = [0_{(m-1)\times 1} \quad \mathbf{I}_{(m-1)\times(m-1)}]\mathbf{A}, \tag{11}$$

where $\mathbf{I}_{(m-1)\times(m-1)}$ is an identity matrix. Matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ in Eq. (11) are related by unitary matrix $\mathbf{D}_{n\times n}$ as

$$\mathbf{A}_{2_{(m-1)\times n}} = \mathbf{A}_{1_{(m-1)\times n}} \times \underbrace{\begin{bmatrix} \exp(i\omega_1) & & & & \\ & \exp(i\omega_2) & & & \\ & & . & & \\ & & & . & \\ & & & & . \\ & & & & \exp(i\omega_n) \end{bmatrix}}_{D_{n\times n}} \tag{12}$$

Equation (12) shows that the transformation is rotation. This property plays a significant role in spectral estimation. Applying the analogy of Eq. (12), we can derive matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ from matrix $\mathbf{S}$ as

$$\mathbf{S}_{1_{(m-1)\times n}} = [\mathbf{I}_{(m-1)\times(m-1)} \quad 0_{(m-1)\times 1}]\mathbf{S},$$

$$\mathbf{S}_{2_{(m-1)\times n}} = [0_{(m-1)\times 1} \quad \mathbf{I}_{(m-1)\times(m-1)}]\mathbf{S}. \tag{13}$$

We can thus represent matrix $\mathbf{S}_2$ by using Eqs. (10) and (13) as

$$\mathbf{S}_2 = \mathbf{A}_2\Gamma = \mathbf{A}_1\mathbf{D}\Gamma = \mathbf{S}_1\Gamma^{-1}\mathbf{D}\Gamma = \mathbf{S}_1\Upsilon, \tag{14}$$

where $\Upsilon = \Gamma^{-1}\mathbf{D}\Gamma$. It is important to note that, in Eq. (10), both matrices $\mathbf{S}$ and $\mathbf{A}$ have full column rank and hence matrix Γ must be nonsingular. Also from Eq. (10), we can deduce that matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ have full rank. Hence matrix Υ is uniquely given by

$$\Upsilon = (\mathbf{S}_1{}^c\mathbf{S}_1)^{-1}\mathbf{S}_1{}^c\mathbf{S}_2. \tag{15}$$

In Eq. (15), Υ can be estimated from the available sample because Υ and $\mathbf{D}$ have the same eigenvalues. Finally, ESPRIT estimates frequencies $\{\omega_n\}_{n=0}^{2\kappa}$ as arguments of $\hat{\upsilon}_n$, where $\{\hat{\upsilon}_n\}_{n=0}^{2\kappa}$ are the eigenvalues of the following consistent estimate of matrix Υ:

$$\hat{\Upsilon} = (\hat{\mathbf{S}}_1{}^c\hat{\mathbf{S}}_1)^{-1}\hat{\mathbf{S}}_1{}^c\hat{\mathbf{S}}_2. \tag{16}$$

In reality, only the estimate of the covariance matrix in Eq. (3) is available, and hence in Eq. (16), only the estimates for matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, given, respectively, by $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$, are available. To apply the proposed method to the estimation of phase step imparted to the PZT, we first need to determine the number of harmonics that are present in the signal.

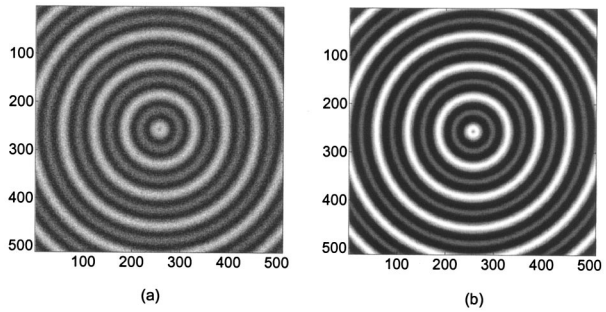

Fig. 1. Fringe maps corresponding to Eq. (18) for $\kappa = 2$ and SNRs of (a) 10 and (b) 60 dB.

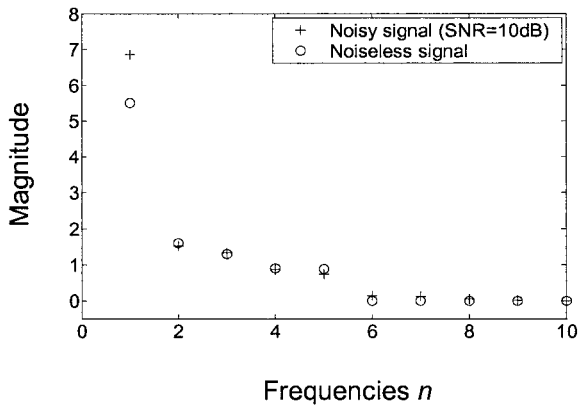

Fig. 2. Estimation of the number of harmonics present in the signal. Diagonal entries are shown for matrix $\mathbf{S}$ obtained from singular-value decomposition of $\mathbf{R}_I$ in Eq. (3). The curve for a noiseless signal shows five significant diagonal values of matrix $\mathbf{S}$. Similarly, the curve for a noisy signal with a SNR of 10 dB shows five significant values in the diagonal of matrix $\mathbf{S}$. The number of harmonics can therefore be reliably estimated from $n = 2\kappa + 1$, where n is the number of frequencies, to be $\kappa = 2$.

3. Determination of the Number of Harmonics in the Signal

Knowledge of the number of harmonics κ that are present in the signal is necessary for designing an appropriate Vandermonde system of equations for the determination of phase φ. Also, once the number of frequencies n that are present in the signal is identified, the value of m for designing a covariance matrix in Eq. (3) can be fixed. Details of the selection of m are explained in Section 4 below.

Consider a noiseless signal consisting of two harmonics ($\kappa = 2$). The signal can be written as

$$I(x, y; t) = I_{dc} + a_{-1}\exp[-i(\varphi + t\alpha)] + a_1\exp[i(\varphi + t\alpha)] + a_{-2}\exp[-2i(\varphi + t\alpha)] + a_2\exp[2i(\varphi + t\alpha)], \quad \text{for } t = 0, 1, \ldots, N-1. \tag{17}$$

We simulate a fringe pattern in which the phase distribution is given by

$$\varphi(x, y) = \Omega\sqrt{(x' - x)^2 + (x' - y)^2}, \tag{18}$$

where (x', x') is the origin for fringe corresponding to phase φ, and Ω is a constant. In Eq. (17) we set $\alpha = \pi/4$ rad. Fringe maps that correspond to Eq. (17) are shown in Figs. 1(a) and 1(b) for noise levels of 10 and 60 dB, respectively. Although the number of harmonics that are present in the signal can be determined by use of the Fourier transform method, they cannot be reliably estimated in the presence of noise and a limited number of data samples.

Interestingly, the singular-value decomposition of $\mathbf{R}_I$ in Eq. (3), represented as $\mathbf{R}_I = \mathbf{USV}^T$, where $\mathbf{S}$ is a diagonal matrix with n nonzero and $m - n$ zero singular values and $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, yields more-reliable information than the Fourier transform method regarding the number of harmonics that are present in the signal.[30] As stated above, m is the autocorrelation length. The singular-value decomposition of covariance matrix $\mathbf{R}_I$ formed by use of Eq. (17) for a noiseless case results in a number of nonzero diagonal entries in $\mathbf{S}$, as $n = 5$ (corresponding to I_{dc}, α, $-\alpha$, 2α, and -2α). In the presence of noise, n principal values of $\mathbf{S}$ would still tend to be larger than the $m - n$ values, which were originally zero. In addition, the n eigenvectors that correspond to the n eigenvalues of $\mathbf{R}_I^T\mathbf{R}_I$ are less susceptible to noise perturbations than are the remaining $m - n$ eigenvectors. Hence the singular-value decomposition of $\mathbf{R}_I$ is the preferred technique for the estimation of the number of harmonics with limited numbers of data frames.

Because the number of frequencies is sometimes not known *a priori*, let us assume that nine frequencies are present in the intensity [Eq. (17)]. The presence of nine frequencies in the figure would indicate that the number of harmonics present in the signal is $\kappa = 4$ (this value was found from the equation $n = 2\kappa + 1$, where n corresponds to the number of frequencies). Figure 2 illustrates typical singular values for $\mathbf{S}$ when $m = 10$ {the minimum value of m is $n + 1$} is selected for the covariance matrix in Eq. (3). In reality, we have only the sample covariance matrix,[29] which is as close as possible to the covariance matrix in Eq. (3). For designing a sample, covariance matrix $N > m$ is required. The details are explained in Section 4 below. For Fig. 2 the number of data frames chosen is 18. The figure shows that, for the noiseless case (open circles) five diagonal entries in $\mathbf{S}$ are significantly larger in magnitude and five diagonal entries are zero. Even in presence of noisy data (crosses), say, a signal-to-noise ratio (SNR) of 10 dB, n (=5) principal values of $\mathbf{S}$ are still larger in magnitude than the $m - 5$ diagonal values. Hence selecting only these significant values of $\mathbf{S}$ (which, in the present example, correspond to $n = 5$) allows for determining the number of harmonics κ (=2) in the signal. A figure such as Fig. 2 thus allows for obtaining a reliable estimation of the number of harmonics that are present in the signal.

4. Evaluation of the Algorithm

We tested the subspace-based method by simulating the fringe pattern in Eq. (17) with phase step $\alpha = \pi/4$ and phase φ defined by Eq. (18). In practice, only the estimate $\hat{\mathbf{R}}_I$ of a covariance matrix is known, and the sample covariance matrix is designed with[29]

$$\hat{\mathbf{R}}_I = \frac{1}{N}\sum_{t=m}^{N} \mathbf{I}(t)\mathbf{I}^c(t), \tag{19}$$

which is as close as possible to $\hat{\mathbf{R}}_I$ in Eq. (3) in a least-squares sense. In Eq. (19), vector $\mathbf{I}(t)$ represents $\{I(0), I(1), I(2), I(3), \ldots, I(N-1)\}$. To design a sample covariance matrix we select an appropriate value of m in Eq. (19). As the value of n is determined by the method described in Section 3 to be $n = 5$, the value of m is chosen such that $m > n$. It is observed that $m > n$ increases the accuracy of phase step estimates. However, this increase in accuracy is obtained at increased computational cost. In addition, m too close to n and N does not yield covariance matrix $\hat{\mathbf{R}}_I$ similar to $\mathbf{R}_I$, which in turn results in spurious phase-step estimates. Performing eigendecomposition of $\hat{\mathbf{R}}_I$ gives estimates for eigenvectors $\hat{\mathbf{S}}$ from which matrices $\hat{\mathbf{S}}_1$ and $\hat{\mathbf{S}}_2$ are determined. Finally, using Eq. (16), we estimate phase step α.

We now study the influence of additive white Gaussian noise at SNRs of 0 to 100 dB, on the estimation of phase steps. Figure 3(a) shows the case for $m = 6$ when the number of data frames N is 8. Phase step α at any arbitrary pixel location on the data frame cannot be estimated from this plot for low SNR values. Figures 3(b)–3(d) show typical plots for $m = 6, 7$, where we have increased the number of data frames N to 10. Figures 3(b) and 3(c) show an increase in performance in the determination of the phase step at low SNRs. Figure 3(d) shows a value of m that is too close to N and thus yields spurious results. That accuracy in the determination of a phase step increases with large data frames is shown in Figs. 3(e) and 3(f), where $N = 14$. Figures 3(e) and 3(f) show typical plots for $m = 9, 11$, respectively. From these figures we observe that $m = 9$ yields a better result than $m = 11$ at low SNRs. From the study of the above cases it can be concluded that phase step α can be reliably estimated even at low SNRs with increase in the number of data frames N and for values of m not too close to n and N.

5. Measurement of Phase Distribution

Parameter l_k can be solved by use of the linear Vandermonde system of equations obtained with Eq. (1) once the value of phase step α has been estimated from the proposed method. A Vandermonde matrix usually arises in polynomial least-squares fitting, in Lagrange interpolating[31,32] polynomials, or in the statistical distribution of the distribution moments.[31,32] A Vandermonde matrix of order n can be written as

$$\begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & \cdot & 1 \\ x_1 & x_2 & x_3 & \cdot & \cdot & \cdot & \cdot & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdot & \cdot & \cdot & \cdot & x_n^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdot & \cdot & \cdot & \cdot & x_n^{n-1} \end{bmatrix}. \tag{20}$$

In some references, the transpose of the matrix in Eq. (20) is also known as the Vandermonde matrix. The solution of the $n \times n$ Vandermonde matrix requires n^2 operations. The advantage of a Vandermonde matrix is that its determinant is always nonzero (hence invertible) for different values of $x = (x_1, x_2, x_3, \ldots, x_n)$. In our case the matrix can be written in the following convenient form:

$$\begin{bmatrix} \exp(i\kappa\alpha_0) & \exp(-i\kappa\alpha_0) & \exp[i(\kappa-1)\alpha_0] & \cdot & \cdot & \cdot & \cdot & 1 \\ \exp(i\kappa\alpha_1) & \exp(-i\kappa\alpha_1) & \exp[i(\kappa-1)\alpha_1] & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \exp[i\kappa\alpha_{(N-1)}] & \exp[-i\kappa\alpha_{(N-1)}] & \exp[i(\kappa-1)\alpha_{(N-1)}] & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} l_\kappa \\ l_k^* \\ l_{\kappa-1} \\ \cdot \\ I_{\mathrm{dc}} \end{bmatrix} = \begin{bmatrix} I_0 \\ I_1 \\ I_2 \\ \cdot \\ I_{N-1} \end{bmatrix}, \tag{21}$$

where $\alpha_0, \alpha_1, \alpha_2, \ldots$, and α_{N-1} are the phase steps for frames $I_0, I_1, I_2, \ldots$, and I_{N-1}, respectively. Phase φ is subsequently computed from the argument of l_1. Figure 4(a) shows typical plots for error obtained in the computation of phase φ in the presence of noise, assumed to be 30 dB. Our simulation was done for $m = 9$ and 14 data frames. The proposed concept [Fig. 4(a)] is compared to a well-known algorithm proposed by Surrel[11] that is also reputed to be insensitive to calibration error and nonsinusoidal waveforms. A typical absolute error in computation of phase φ obtained with Surrel's algorithm is shown in Fig. 4(b). We obtained the plot by simulating six data frames and by keeping the phase step as $\pi/2$ to minimize second-order harmonics [the algorithm advocates selection of a $2\pi/(\kappa + 2)$ phase step for minimizing κth-order harmonics]. A linear calibration error of 5% is assumed while we study the response of the algorithm to white Gaussian noise with a SNR of 30 dB.

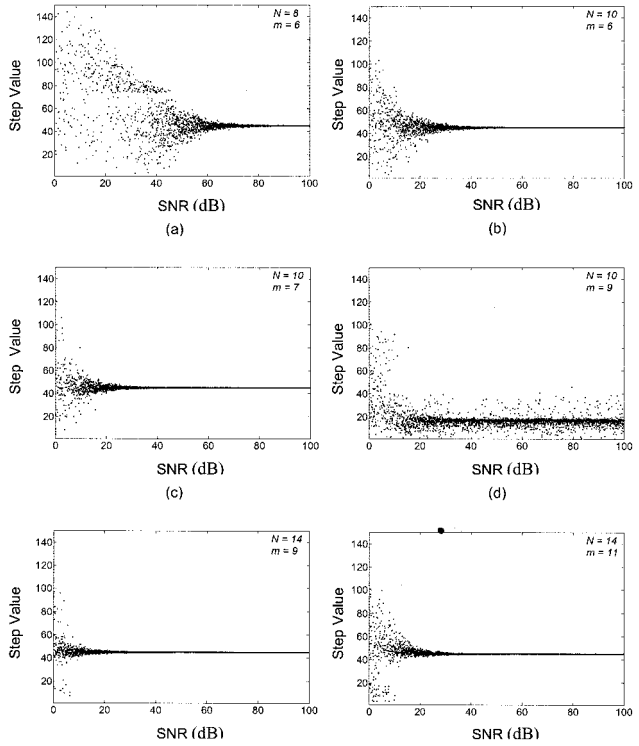

Fig. 3. Phase step α (in degrees) versus SNR at an arbitrary pixel location for several values of N and m.

A study of Figs. 4(a) and 4(b) suggests that the two algorithms perform similarly in the presence of noise. Our proposed concept requires more data frames than suggested by Surrel. However, our algorithm does not impose any restriction on the selection of phase-step values, which can be chosen arbitrarily. In addition, our algorithm does not rely on the exactness of phase-step values obtained by calibration because it functions by computing the phase-step values after acquiring the data frames. As the phase steps are extracted pixelwise, our concept allows for the use of noncollimated beams. A unique feature of our algorithm lies in its ability to accommodate multiple PZTs. The use of two PZTs in an optical setup has been proposed by Rastogi for simultaneous extraction of two orthogonal components.[33,34] In such a case our proposed concept can easily be adapted to facilitate

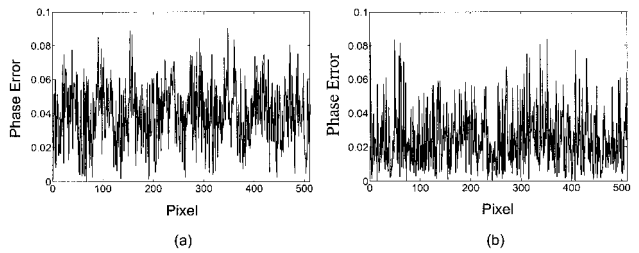

Fig. 4. Typical absolute errors in computation of phase φ (in radians) at a SNR of 30 dB. (a) Phase obtained with our proposed method with the phase-step value obtained from Fig. 3(e). (b) Phase obtained with the algorithm proposed by Surrel.[11]

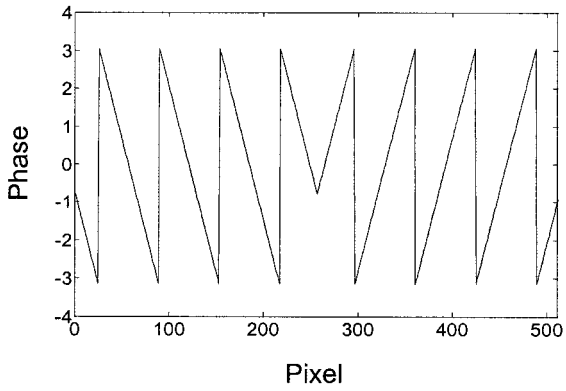

Fig. 5. Typical wrapped phase φ (in radians) obtained for the phase-step value determined from Fig. 3(e) for a SNR of 30 dB.

use of multiple PZTs in an optical configuration. It is difficult to adapt the algorithms mentioned in Refs. 33 and 34 to configurations that contain multiple PZTs and at the same time exhibit the possibility of minimizing the systematic and random sources of errors.

Figure 5 shows a typical plot of the corresponding wrapped phase φ. The algorithm requires typically 4–5 min of computation time to obtain the phase distribution over 512 × 512 pixels for second-order harmonics on a desktop PC (Pentium IV processor with 2.66 GHz).

6. Conclusions

We have proposed a novel approach to recovering phase distribution in the presence of nonsinusoidal waveforms. A significant advantage of the proposed method lies in its ability to measure phase steps and without the need to rely on a previously made phase-shift calibration. The method enables one to choose arbitrarily phase shifts from 0 to π. The proposed technique works well with both diverging and converging beams because it relies on retrieving the phase-step values pixelwise before applying them to the Vandermonde system of equations. The accuracy in the measurement of the phase steps in the presence of additive white Gaussian noise has been shown to increase with large data frames.

Appendix A

Let us consider a signal

$$I(t) = I_{\mathrm{dc}} + \sum_{k=1}^{K} a_k \exp(ik\varphi)\exp(i\alpha kt) + \sum_{k=1}^{K} a_k \exp(-ik\varphi)\exp(-i\alpha kt) + \eta(t),$$

$$t = 0,\ 1,\ 2,\ \ldots\ m,\ \ldots,\ N-1. \qquad \text{(A1)}$$

Here $\eta(t)$ represents white Gaussian noise. The covariance of a function $I(t)$ is defined as[29]

$$r(p) = E[I(t)I^*(t-p)], \qquad \text{(A2)}$$

For simplicity, let us consider $\kappa = 1$ and rewrite Eq. (A1) as

$$I(t) = I_{dc} + a_1 \exp(i\varphi)\exp(i\alpha t) + a_1 \exp(-i\varphi)\exp(-i\alpha t) + \eta(t). \quad \text{(A3)}$$

Similarly, let us write $I^*(t - p)$ for $\kappa = 1$ as

$$I^*(t-p) = I_{dc} + a_1 \exp(i\varphi)\exp[i\alpha(t-p)] + a_1 \times \exp(-i\varphi)\exp[-i\alpha(t-p)] + \eta^*(t-p). \quad \text{(A4)}$$

Substituting Eqs. (A3) and (A4) into Eq. (A2), we obtain

$$\begin{aligned} r(p) &= E[I(t)I^*(t-p)] \\ &= E\{I_{dc}{}^2 + I_{dc}a_1 \exp(-i\varphi)\exp(-i\alpha t) \\ &\quad + I_{dc}a_1 \exp(i\varphi)\exp(i\alpha t) + \exp(i\alpha p)[a_1{}^2 \\ &\quad + I_{dc}a_1 \exp(-i\varphi)\exp(-i\alpha t) + a_1{}^2 \exp(-2i\varphi) \\ &\quad \times \exp(-2i\alpha t)] + \exp(-i\alpha p)[a_1{}^2 \\ &\quad + I_{dc}a_1 \exp(i\varphi)\exp(i\alpha t) + a_1{}^2 \exp(2i\alpha t)] \\ &\quad + \eta(t)\eta^*(t-p)\}. \end{aligned} \quad \text{(A5)}$$

Equation (A5) can be written in the following compact form:

$$r(p) = E[I_{dc}{}^2 + c_1 + \exp(i\alpha p)(a_1{}^2 + c_2) + \exp(-i\alpha p) \times (a_1{}^2 + c_3) + \eta(t)\eta^*(t-p)], \quad \text{(A6)}$$

where $c_1 = I_{dc}d_1 \exp(-i\varphi)\exp(-i\alpha t) + I_{dc}a_1 \exp(i\varphi) \times \exp(i\alpha t)$, $c_2 = I_{dc}a_1 \exp(-i\varphi)\exp(-i\alpha t) + a_1{}^2 \exp(-2i\varphi)\exp(-2i\alpha t)$, and $c_3 = I_{dc}a_1 \exp(i\varphi)\exp(i\alpha t) + a_1{}^2 \exp(2i\varphi)\exp(2i\alpha t)$. Let $E\{I_{dc}{}^2 + c_1\} = A_0{}^2$, $E\{a_1{}^2 + c_2\} = A_1{}^2$, and $E\{a_1{}^2 + c_3\} = A_2{}^2$. Therefore

$$r(p) = A_0{}^2 + A_1{}^2 \exp(i\alpha p) + A_2{}^2 \exp(-i\alpha p) + \sigma^2\delta_{p,0}. \quad \text{(A7)}$$

In Eq. (A7), $\sigma^2\delta_{p,0}$ is the expectation for Gaussian noise $\eta(t)$ and is given by

$$E[\eta(k)\eta^*(j)] = \sigma^2\delta_{k,j}. \quad \text{(A8)}$$

In practice, one computes expectation E in Eq. (2) by averaging over a finite number of frames. If a large number of frames is taken for averaging, the exponential terms that contain t in c_1, c_2, and c_3 will oscillate uniformly from 0 to 2π. In this limit, the expectation of c_1, c_2, and c_3 will approach zero because

$$\int_0^{2\pi} \exp(i\psi)d\psi = 0. \quad \text{(A9)}$$

However, if a finite number of frames is taken for averaging, expectation c_1, c_2, and c_3 will have a small finite value different from zero. Hence, for κ harmonics in the intensity, the final derivation of covariance of $I(t)$ is given by

$$r(p) = E[I(t)I^*(t-p)] = \sum_{n=0}^{2\kappa} A_n{}^2 \exp(i\omega_n p) + \sigma^2\delta_{p,0}. \quad \text{(A10)}$$

This research is funded by the Swiss National Science Foundation.

References

1. P. Carré, "Installation et utilisation du comparateur photoélectrique et interférentiel du Bureau International des Poids et Mesures," Metrologia **2,** 13–23 (1966).
2. J. H. Bruning, D. R. Herriott, J. E. Gallagher, D. P. Rosenfeld, A. D. White, and D. J. Brangaccio, "Digital wavefront measuring interferometer for testing optical surfaces and lenses," Appl. Opt. **13,** 2693–2703 (1974).
3. K. Creath, "Phase-shifting holographic interferometry," in *Holographic Interferometry,* P. K. Rastogi, ed., Vol. 68 of Springer Series in Optical Sciences (Springer-Verlag, 1994), pp. 109–150.
4. T. Kreis, *Holographic Interferometry: Principles and Methods* (Akademie Verlag, 1996), pp. 101–170.
5. J. E. Greivenkamp and J. H. Bruning, "Phase shifting interferometry," in *Optical Shop Testing*, D. Malacara, ed. (Wiley, 1992), pp. 501–598.
6. J. Schwider, R. Burow, K. E. Elssner, J. Grzanna, R. Spolaczyk, and K. Merkel, "Digital wave-front measuring interferometry: some systematic error sources," Appl. Opt. **22,** 3421–3432 (1983).
7. Y. Zhu and T. Gemma, "Method for designing error-compensating phase-calculation algorithms for phase-shifting interferometry," Appl. Opt. **40,** 4540–4546 (2001).
8. P. Hariharan, B. F. Oreb, and T. Eiju, "Digital phase-shifting interferometry: a simple error-compensating phase calculation algorithm," Appl. Opt. **26,** 2504–2506 (1987).
9. J. Schwider, O. Falkenstorfer, H. Schreiber, and A. Zoller, "New compensating four-phase algorithm for phase-shift interferometry," Opt. Eng. **32,** 1883–1885 (1993).
10. Y. Surrel, "Phase stepping: a new self-calibrating algorithm," Appl. Opt. **32,** 3598–3600 (1993).
11. Y. Surrel, "Design of algorithms for phase measurements by the use of phase stepping," Appl. Opt. **35,** 51–60 (1996).
12. J. van Wingerden, H. J. Frankena, and C. Smorenburg, "Linear approximation for measurement errors in phase shifting interferometry," Appl. Opt. **30,** 2718–2729 (1991).
13. K. G. Larkin and B. F. Oreb, "Design and assessment of symmetrical phase-shifting algorithms," J. Opt. Soc. Am. A **9,** 1740–1748 (1992).
14. K. Hibino, B. F. Oreb, D. I. Farrant, and K. G. Larkin, "Phase shifting for nonsinusoidal waveforms with phase-shift errors," J. Opt. Soc. Am. A **12,** 761–768 (1995).
15. Y.-Y. Cheng and J. C. Wyant, "Phase-shifter calibration in phase-shifting interferometry," Appl. Opt. **24,** 3049–3052 (1985).
16. B. Zhao, "A statistical method for fringe intensity-correlated error in phase-shifting measurement: the effect of quantization error on the *N*-bucket algorithm," Meas. Sci. Technol. **8,** 147–153 (1997).
17. B. Zhao and Y. Surrel, "Effect of quantization error on the computed phase of phase-shifting measurements," Appl. Opt. **36,** 2070–2075 (1997).
18. R. Józwicki, M. Kujawinska, and M. Salbut, "New contra old wavefront measurement concepts for interferometric optical testing," Opt. Eng. **31,** 422–433 (1992).

19. P. de Groot, "Vibration in phase-shifting interferometry," J. Opt. Soc. Am. A **12,** 354–365 (1995).
20. P. de Groot and L. L. Deck, "Numerical simulations of vibration in phase-shifting interferometry," Appl. Opt. **35,** 2172–2178 (1996).
21. C. Rathjen, "Statistical properties of phase-shift algorithms," J. Opt. Soc. Am. A **12,** 1997–2008 (1995).
22. P. L. Wizinowich, "Phase-shifting interferometry in the presence of vibration: a new algorithm and system," Appl. Opt. **29,** 3271–3279 (1990).
23. C. Joenathan and B. M. Khorana, "Phase measurement by differentiating interferometric fringes," J. Mod. Opt. **39,** 2075–2087 (1992).
24. K. Kinnnstaetter, A. W. Lohmann, J. Schwider, and N. Streibl, "Accuracy of phase shifting interferometry," Appl. Opt. **27,** 5082–5089 (1988).
25. C. J. Morgan, "Least squares estimation in phase-measurement interferometry," Opt. Lett. **7,** 368–370 (1982).
26. J. E. Greivenkamp, "Generalized data reduction for heterodyne interferometry," Opt. Eng. **23,** 350–352 (1984).
27. R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust. Speech Signal Process. **37,** 984–995 (1989).
28. J. J. Fuchs, "Estimating the number of sinusoids in additive white noise," IEEE Trans. Acoust. Speech Signal Process. **36,** 1846–1853 (1988).
29. T. Söderström and P. Stoica, "Accuracy of high-order Yule–Walker methods for frequency estimation of complex sine waves," IEE Proc. F **140,** 71–80 (1993).
30. R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," IEEE Trans. Aerosp. Electron. Syst. **19,** 134–139 (1983).
31. M. Marcus and H. Minc, *A Survey of Matrix Theory and Matrix Inequalities* (Dover, 1992), pp. 15–16.
32. K. M. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed. (Prentice-Hall, 1971).
33. P. K. Rastogi, "Phase shifting applied to four-wave holographic interferometers," Appl. Opt. **31,** 1680–1681 (1992).
34. P. K. Rastogi, "Phase-shifting holographic moiré: phase-shifter error-insensitive algorithms for the extraction of the difference and sum of phases in holographic moiré," Appl. Opt. **32,** 3669–3675 (1993).