High-fidelity phase and amplitude control of phase-only computer generated holograms using conjugate gradient minimisation


1SUPA School of Physics & Astronomy, University of St. Andrews, North Haugh, St. Andrews KY16 9SS, UK
2Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK
3Sorbonne Universités, UPMC Univ Paris 06, UFR 925, 4 place Jussieu, 75252 Paris cedex 05, France
*gdB2@st-andrews.ac.uk

Abstract: We demonstrate simultaneous control of both the phase and amplitude of light using a conjugate gradient minimisation-based hologram calculation technique and a single phase-only spatial light modulator (SLM). A cost function, which incorporates the inner product of the light field with a chosen target field within a defined measure region, is efficiently minimised to create high fidelity patterns in the Fourier plane of the SLM. A fidelity of $F = 0.999997$ is achieved for a pattern resembling an $LG_{01}$ mode with a calculated light-usage efficiency of 41.5%. Possible applications of our method in optical trapping and ultracold atoms are presented and we show uncorrected experimental realisation of our patterns with $F = 0.97$ and 7.8% light efficiency.

© 2017 Optical Society of America


References and links
1. Introduction

Simultaneous control over the amplitude and phase of light has allowed significant advances in optical trapping of microscopic objects [1], microscopy [2] and optical communication [3]. A variety of methods have been developed which allow arbitrary independent control over both. Tandem or cascaded approaches sequentially manipulate the amplitude then phase using either variety of methods have been developed which allow arbitrary independent control over both.

\[ \text{Given} \]

\[ \text{In order to control the light field in a particular plane holographically, we wish to apply a bespoke phase modulation} \]

\[ \phi_{p,q} \]

\[ \text{(with indices} \]

\[ p \]

\[ \text{and} \]

\[ q \]

\[ \text{denoting spatial co-ordinates) to a fixed} \]

\[ \text{incident laser field with amplitude} \]

\[ S_{p,q} \]

\[ \text{in a simple setup with a single phase-only SLM and a single focussing element. The electric field in the plane of the SLM is} \]

\[ E_{p,q}^{\text{in}} = S_{p,q} \exp(i\phi_{p,q}) \]

\[ \text{Given} \]

\[ S_{p,q} \]

\[ \text{and} \]

\[ \phi_{p,q} \]

\[ \text{the electric field in any other plane} \]

\[ E_{n,m}^{\text{out}} \]

\[ \text{(with output plane coordinates denoted by} \]

\[ n \]

\[ \text{and} \]

\[ m \]

\[ \text{is straightforwardly calculated using an appropriate propagator} \]

\[ P \]

\[ \text{such that} \]

\[ E_{n,m}^{\text{out}} = \mathcal{P}(E_{p,q}^{\text{in}}) \]

\[ \text{where} \]

\[ \mathcal{P} \]

\[ \text{is a two-dimensional propagator.} \]
which was previously shown to achieve smooth, accurate and highly-controllable intensity of the value.

The process continues until the cost function stagnates (i.e. when the difference between iterations is below $10^{-5}$) or a predefined maximum number of iterations is reached. We implement the conjugate gradient calculation in Python with the cost function gradient determined using the Theano library [17]. Our codes, and the data presented in this article, are freely available online [18].

$$E_{n,m}^{\text{out}} = \mathcal{P} \left[ E_{p,q}^{\text{in}} \right].$$

For patterns in the far field $\mathcal{P}$ is approximated by a fast Fourier transform [10] such that

$$E_{n,m}^{\text{out}} = \frac{S_{p,q}}{N_T} \sum_{p,q} \exp \left( i \phi_{p,q} \right) \exp \left[ - \left( \frac{2\pi i}{N_T} \right) (pn + qm) \right],$$

$$= \sqrt{I_{n,m}} \exp \left( i \varphi_{n,m} \right),$$

where $N_T = n \times m$, while $I_{n,m}$ and $\varphi_{n,m}$ are the output plane intensity and phase respectively. Calculation of the appropriate phase-only modulation $\phi_{p,q}$ to give an acceptable output field is a well-known inverse problem which, in general, requires numerical solution. Iterative Fourier Transform Algorithms (IFTAs) are commonly used in calculating the phase modulation required to generate a desired intensity distribution, and variants which control both phase and amplitude have been recently demonstrated [11, 12].

In this paper we propose an alternative iterative method to creating patterns with independent control over the phase and amplitude profiles: using a conjugate gradient minimisation technique which was previously shown to achieve smooth, accurate and highly-controllable intensity patterns [13]. The technique efficiently minimises a specified cost function which can be carefully manipulated to reflect the requirements of the chosen light pattern, such as removing optical vortices from regions of interest. Here, we extend this method to produce a variety of high fidelity and smooth patterns in both phase and intensity, which are designed primarily for optical trapping.

2. Conjugate Gradient Method

Conjugate gradient minimisation is a widely-used, efficient numerical method to optimise high-dimensional functions, which is intuitively described in [14]. Our original conjugate gradient optimisation routine, which tailors the output-plane amplitude of phase-only holograms, is presented in more detail in [13]. The main advantage of this approach is the high level of control it gives over any feature of interest in the output plane, provided that the feature can be encapsulated within an analytical cost function $C$. This defines an effective error to be minimised, and judicious choice of the cost function terms can allow precise guiding of the hologram optimisation process. For holograms, the cost function is based on the difference between the calculated electric field and a chosen target, and the parameter space for the optimisation encompasses all the different phase distributions that the SLM can generate.

Figure 1 shows a diagrammatic representation of the calculation. The initial position in the parameter space of $C$ is determined by $S_{p,q}$ (a two-dimensional Gaussian profile with $1/e^2$-radius $\sigma$) and a guess phase $\phi_{p,q} = R \left( p^2 + q^2 \right) + D \left( p \cos \theta + q \sin \theta \right)$. The two terms in $\phi_{p,q}$ respectively control the size and position of the envelope of the output plane intensity. This combination of phase patterns is known to suppress the formation of optical vortices during hologram calculation, which can otherwise cause premature stagnation and low accuracy [15, 16].

As an initial step, we calculate $\partial C / \partial \phi_{p,q}$ for each pixel to determine the direction of steepest descent $g_i$ and minimise $C$ along this direction to change $\phi_{p,q}$. For subsequent iterations $i$ of the process, the descent direction $\alpha_i$ is the conjugate direction

$$\alpha_i = g_i + \left( \frac{g_i \cdot g_i}{(g_{i-1} \cdot g_{i-1})} \right) \alpha_{i-1}.$$

The process continues until the cost function stagnates (i.e. when the difference in the value of the cost function between iterations is below $10^{-5}$) or a predefined maximum number of iterations is reached. We implement the conjugate gradient calculation in Python with the cost function gradient determined using the Theano library [17]. Our codes, and the data presented in this article, are freely available online [18].
This cost function minimisation approach to hologram calculation is fundamentally different than the more commonly-used IFTA. While both start from an initial phase guess, in the IFTA the phase is evolved by performing multiple successive Fourier transforms between SLM and output planes, imposing the known amplitude (SLM plane) or desired electric field (output plane) at each step. Importantly, the IFTA is not a minimization routine, and thus provides no guarantee of convergence. Furthermore, the flexibility afforded by the freedom in cost function choice allows one to prioritize whichever features of the output plane are deemed to be important. In the work here, we prioritise accuracy of both amplitude and phase, as well as smoothness of the amplitude, within a subset of the output plane at the expense of light utilization efficiency. In order to find a hologram which gives acceptable amplitude and phase, we find that a good choice of cost function is

\begin{equation}
C = 10^d \left( 1 - \sum_{n,m} \text{Re} \left( \bar{\tau}_{n,m} \tau_{n,m}^* \right) \right)^2, \tag{4}
\end{equation}

\begin{equation}
C = 10^d \left( 1 - \sum_{n,m} \sqrt{I_{n,m}} T_{n,m} \cos (\Phi_{n,m} - \varphi_{n,m}) \right)^2, \tag{5}
\end{equation}

where \( \tau_{n,m} = \sqrt{I_{n,m}} \exp (i\Phi_{n,m}) \) is the target electric field, and the over-tilde denotes normalisation over a specified region of interest, which is small compared to the total output plane.
Similar to the MRAF method [15], we choose this region of interest to encompass regions of non-zero amplitude in the target pattern (known as the measure region) plus a surrounding area of zero intensity. Experimentally, the light which the algorithm places outside the region of interest can be spatially filtered. The multiplicative prefactor $10^d$ is used to increase the steepness of the cost function within the parameter space to improve convergence time and accuracy.

3. Numerical Results

![Fig. 2. The far-field results from the conjugate gradient optimisation showing normalised intensity $\tilde{I}$ (colour) and phase $\phi$ (grey) in the region of interest. (a) Laguerre-Gaussian mode, (b) ring lattice with azimuthally-varying phase, (c) square lattice with azimuthally-varying phase, (d) graphene lattice with alternating phase, (e) flat-top intensity with inverse-square power-law phase, (f) Gaussian line with linear phase gradient, (g) chicken intensity with egg phase. The flat top pattern (e) has the light outside the measure region removed for clarity. The error metrics for each pattern are shown in Table 1.](image)

We test our method on a range of target patterns particularly chosen with applications in optical trapping in mind. Independent spatial control over both the amplitude and phase of trap light is also increasingly desirable in the field of ultracold atoms, for example in the transfer of orbital angular momentum from light to atoms [19], and in the creation of artificial gauge fields [20–22]. In the particular case of trapping ultracold atoms in continuous geometries [15, 23–27], accuracy and smoothness of the intensity are vital to avoid fragmentation.

We calculate a pattern of phase values between 0 and $2\pi$ for the SLM plane of $256 \times 256$ pixels (with a pixel size of $24 \mu m$). This is padded with zeros in the border such that the plane is $512 \times 512$ pixels, such that there is no loss of resolution in the resulting $512 \times 512$ output plane. The patterns are diagonally offset from the center of the plane by 85 pixels to avoid the zeroth order (undiffracted light) that would appear due to the finite efficiency of the SLM. This constrains two of the initialisation parameters to $D = -\pi/2$ and $\theta = \pi/4$.

We show the region of interest of the calculated intensity and phase for each of our target patterns in Fig. 2. Pattern (a) is similar to a Laguerre-Gaussian (LG) mode, which have a
wide variety of uses [28], including in ultracold atom experiments to induce circulation states [19]. Patterns (b) and (c) are ring and square lattices with underlying phase windings, which have potential applications for quantum simulation of magnetic flux in solid state systems [20]. Ultracold atoms confined in a honeycomb lattice with alternating phase between nearest neighbouring sites [Pattern (d)] have also been shown to experience an artificial gauge field in a graphene quantum simulator [21], while a trapping potential comprising a flat intensity profile and an inverse square power-law phase [Pattern (e)] has been proposed for investigations on sonic horizons and artificial black holes [22]. Pattern (f) is a Gaussian line with a phase gradient across it which can be used to trap particles in optical tweezers, but at the same time cause them to flow [29]. As a test of our method’s versatility, we have also chosen the more arbitrary Pattern (g) of a chicken and eggs [30] in the amplitude and phase respectively.

This gallery of targets, as well as being relevant for the aforementioned scientific goals, also tests the algorithm’s capability over a wide variety of pattern features. Firstly, Pattern (a) provides a good benchmark for our method as the LG-mode is a well-defined solution to the Maxwell equations. Patterns (b) and (c) show that we can retain the phase structure of LG modes but with arbitrary amplitude profiles. Moreover, Patterns (a), (e) and (f) probe the ability to generate very smooth, continuous geometries, while patterns (b)-(d) probe more discrete geometries.

We also test the production of target patterns with Cartesian [Patterns (c) and (f)], azimuthal [Patterns (a), (b) and (e)], and triangular symmetry [Pattern (d)], as well as patterns with no underlying symmetry [Pattern (g)]. Finally, Patterns (e) and (g) have uncorrelated requirements in the intensity and phase patterns.

The main metric for accuracy is the fidelity, which is defined as $F = \sum_{n,m} |I_{n,m}^a - I_{n,m}^o|^2$ [9] and is evaluated over non-zero amplitude within the measure region. The light efficiency ($\eta$) is the fraction of light in the output plane that is within the region of interest. A relative phase error $\epsilon_\phi$ within the measure region and the non-uniformity error $\epsilon_{nu}$ for regions in the patterns that have a flat intensity [12] are defined as:

$$\epsilon_\phi = \frac{\sum_{n,m} |(\Phi_{n,m} - \varphi_{n,m} + \tilde{\phi})|^2}{\sum_{n,m} |\Phi_{n,m}|^2},$$

$$\epsilon_{nu} = \frac{\sum_{n,m} |M_{n,m} (I_{n,m} - I_o)|^2}{\sum_{n,m} |M_{n,m} I_{n,m}|^2},$$

where $P$ is a correction term to account for the cyclical nature of the phase, $M_{n,m}$ is a binary mask which is equal to one where the target intensity is approximately uniform and zero everywhere else and $I_o = (1/N) \sum_{n,m} M_{n,m} I_{n,m}$ is the average output intensity in the uniform region ($N$ is the total number of pixels in the measure region).

![Fig. 3. (a) Evolution of fidelity $F$ for the Gaussian Line pattern shown in Fig. 2(f) with $\sigma = 1.5$ mm and $R = 3.5$ mrad px$^{-2}$. At low values of the steepness $d$ of the cost function, the algorithm stagnates earlier and returns a lower fidelity hologram. (b) The final fidelity and the time per iteration $t$ both improve as $d$ is increased. (c) Fidelity and (d) efficiency $\eta$ as a function of incident laser beam size $\sigma$ and quadratic guess phase curvature $R$. Small beam sizes and reduced guess phase curvature give highest efficiency but lowest fidelity.](image-url)
Table 1. Error metrics for the calculated patterns in Fig. 2, with optimal values of $\sigma$, $R$ and region of interest diameter ROI.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>$\sigma$ (mm)</th>
<th>$R$ (mrad px$^{-2}$)</th>
<th>ROI px</th>
<th>$1 - F$</th>
<th>$\eta$</th>
<th>$\epsilon_\Phi$</th>
<th>$\epsilon_{nu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Laguerre Gauss</td>
<td>1.0</td>
<td>4.5</td>
<td>42</td>
<td>$3.0 \times 10^{-6}$</td>
<td>41.5</td>
<td>0.0003</td>
<td>0.005</td>
</tr>
<tr>
<td>(b) Square Lattice</td>
<td>1.2</td>
<td>4.5</td>
<td>124</td>
<td>$1.6 \times 10^{-5}$</td>
<td>10.6</td>
<td>0.0009</td>
<td>0.02</td>
</tr>
<tr>
<td>(c) Ring Lattice</td>
<td>1.4</td>
<td>3.9</td>
<td>71</td>
<td>$1.5 \times 10^{-6}$</td>
<td>24.6</td>
<td>0.00006</td>
<td>0.001</td>
</tr>
<tr>
<td>(d) Graphene</td>
<td>1.4</td>
<td>2.7</td>
<td>78</td>
<td>$4.4 \times 10^{-4}$</td>
<td>13.1</td>
<td>0.0003</td>
<td>0.010</td>
</tr>
<tr>
<td>(e) Flat Top</td>
<td>1.0</td>
<td>4.5</td>
<td>63</td>
<td>$1.8 \times 10^{-4}$</td>
<td>11.3</td>
<td>0.2</td>
<td>0.007</td>
</tr>
<tr>
<td>(f) Gaussian Line</td>
<td>1.4</td>
<td>2.9</td>
<td>45</td>
<td>$1.4 \times 10^{-5}$</td>
<td>20.4</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>(g) Chicken &amp; Egg</td>
<td>1.6</td>
<td>4.5</td>
<td>128</td>
<td>$7.1 \times 10^{-2}$</td>
<td>2.0</td>
<td>1.3</td>
<td>-</td>
</tr>
</tbody>
</table>

For the example of the Gaussian line pattern (with $\sigma = 1.5$ mm and $R = 3.5$ mrad px$^{-2}$), Fig. 3(a) shows the evolution of the fidelity through the calculation for different values of the steepness parameter $d$ in Eq. (5). Lower values of $d$ cause early stagnation of the algorithm into poor quality local minima. The maximum iteration number was reached for $d > 6$, whilst the fidelity would increase at approximately the same rate for $d > 4$ (only $d = 1, 2, 3, 6$ and 9 are shown in Fig. 3(a) for clarity). We restrict $d \leq 9$ to allow the computation to be performed in 32-bit floating point representation. We found that, within this limit, a steeper cost function would not only lead to improved fidelities in the patterns, but also faster calculation times per iteration $t$ (as shown in Fig. 3(b)). Thus, for all patterns shown in this article, we have used $d = 9$. A typical minimization routine converges in $< 200$ iterations at a total duration of $< 75$ s with a standard desktop computer (2.5 GHz processor).

For each pattern we perform an optimisation over the initialization conditions $\sigma$ and $R$ (see Figs. 3(c) and 3(d)). It was found that smaller incident laser beam sizes and reduced curvature in the guess phase led to higher light efficiency at a reduced fidelity. The beam size and curvature for the patterns in Fig. 2 were chosen to provide both good light efficiency whilst maintaining a high fidelity. The optimal values of calculated holograms are shown in Table 1.

The authors of [12] recently developed an IFTA for full-plane control of amplitude and phase, which they compared to a previous regionally-constrained algorithm [11]. They find that the regionally-constrained algorithm is more accurate at the cost of light-utilisation efficiency, which has also been seen in amplitude-only control algorithms [13, 15] and in the present work. For far-field holograms of lines of continuous intensity with phase gradients, Wu, et al., report in [12] that the regional algorithm gives $\epsilon_{nu} = 0.04\%$, $\epsilon_\Phi = 1.63\%$ and $\eta = 3.48\%$, while the full-plane IFTA is less accurate ($\epsilon_{nu} = 3.48\%$ and $\epsilon_\Phi = 3.77\%$) but achieves higher efficiency ($\eta = 77.84\%$). For our chosen cost function in Eq. (5), the comparable continuous patterns amongst our range of targets (i.e. the Gaussian Line and Flat Top) are significantly smoother: we find $\epsilon_{nu}$ is lower by a factor 6-20 and $\epsilon_\Phi$ is lower by one or two orders of magnitude than the regional IFTA results reported in [12]. The light-utilisation of our conjugate gradient optimised patterns is a factor 3-11 times higher than their regional IFTA results, but between 15-53% of their full-plane IFTA results. We note that the freedom in choice of the cost function terms and their relative weightings could be exploited to prioritise the efficiency of light usage at the expense of accuracy or smoothness if this is of greater importance to a particular application. Finally, we note that the results of [12] show that the IFTA has already achieved its highest accuracy after $< 40$ iterations, which is fewer than the $\leq 200$ required by conjugate gradient minimisation.
Fig. 4. (a) Experimental Setup. The laser light is separated into two beams with a polarising beam splitter (PBS), using a half-waveplate (λ/2) to control the relative power in each beam. One beam is phase-modulated by the SLM, and focused onto a CCD camera with an achromatic doublet lens. The second beam is co-polarized and overlapped with the first using a non-polarizing beam splitter (BS). To image only intensity, we block the second beam. The interference pattern produced on the CCD camera when both beams are unblocked is used to detect the phase. (b) - (d) Measured intensity (left) and phase (right) for (b) Gaussian line, (c) Laguerre-Gaussian and (d) graphene lattice. The white scalebar in (b) denotes 300 µm, and is common to all images. Color scaling as in Fig. 2.

4. Experimental Verification

We verify the calculated holograms experimentally using the setup shown in Fig. 4(a). The output of a 1070 nm fiber laser (IPG YLP-5-1070-LP) is expanded to an experimentally-convenient $1/e^2$ waist of 3.0 mm and split using a polarising beam splitter. One path is phase-modulated as it is reflected (14° AOI) by a liquid crystal SLM (BNS P1920) and focused onto a CCD camera (Thorlabs DCU200 Series) using an $f = 150$ mm achromatic doublet. The other path gives a reference beam which is optionally recombined with the modulated beam after the focusing optic to produce interference fringes. These fringes are used to extract the phase of the modulated light via the Fourier transform fringe analysis method [31].

We note that the SLM used in the experiment has more (1920 × 1152) and smaller (9.2 × 9.2 µm) pixels than the calculated holograms in Section 3. In order to minimize calculation time, we chose to keep a small hologram of 256 × 256 pixels. To do this, we separate the SLM into bins of 3 × 3 pixels (the remaining pixels are not used). We calculate new optimal holograms for the effective pixel size of 27 µm and experimentally-convenient beam waist of 3 mm. These are presented in the Theory columns of Table 2.

As shown in Figs. 4(b)–4(d) and detailed in Table 2, the measured fidelities are lower than the numerical predictions. There are a number of experimental issues that are commonly seen in such experiments, such as imperfect response of the spatial light modulator including a diffraction efficiency which varies across the output plane, and also optical aberrations such as finite-aperture effects. These are commonly improved by the addition of feedback [23, 26] or the characterisation of wavefront aberration in the optical system [32, 33]. Due to the 45% diffraction
efficiency of the SLM (i.e. the percentage of light diffracted into the first order by a linear phase gradient) we include the rescaled efficiency $\eta^* = \eta / 0.45$, which gives fairer comparison between the numerical and experimental results. Higher diffraction efficiencies could be obtained by replacing the SLM with a micro-fabricated diffractive optical element.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>$F$</th>
<th>$\eta^*$</th>
<th>$\epsilon_\Phi$</th>
<th>$\epsilon_{nu}$</th>
<th>$F$</th>
<th>$\eta^*$</th>
<th>$\epsilon_\Phi$</th>
<th>$\epsilon_{nu}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian Line</td>
<td>0.99993</td>
<td>8.3</td>
<td>0.005</td>
<td>0.004</td>
<td>0.97</td>
<td>7.8</td>
<td>1.85</td>
<td>0.48</td>
</tr>
<tr>
<td>Laguerre Gauss</td>
<td>0.99999</td>
<td>8.4</td>
<td>0.0004</td>
<td>0.004</td>
<td>0.97</td>
<td>7.8</td>
<td>2.59</td>
<td>0.52</td>
</tr>
<tr>
<td>Graphene</td>
<td>0.9996</td>
<td>7.0</td>
<td>0.0004</td>
<td>0.015</td>
<td>0.96</td>
<td>6.2</td>
<td>2.76</td>
<td>0.42</td>
</tr>
</tbody>
</table>

5. Conclusion

In summary, we have demonstrated that smooth, high fidelity light patterns with independent control over the amplitude and phase can be generated with a single phase-only SLM. The holograms calculated with the conjugate gradient minimisation approach surpass the accuracy and smoothness of previous IFTA approaches. We note that our approach achieves comparable results in $F$ and $\eta$ for image-quality holograms to the super-pixel method for DMDs [9], and improved $F$ for the Laguerre Gaussian mode, at the expense of constraining the pattern to a subset of the output plane.

This approach to hologram calculation is compatible with existing methods for the generation of multi-wavelength holographic optical traps [25]. In this work we have concentrated on using a fast Fourier transform as the propagator $P$. However, we find that near-field patterns calculated using Angular Spectrum Wavefront Propagation [10] achieve comparable fidelity, efficiency and smoothness. The accurate control over amplitude and phase will be crucial to a future research direction in the design of axially-structured light fields.

Funding

Leverhulme Trust (RPG-2013-074); EPSRC (EP/G03673X/1; EP/L015110/1).

Acknowledgments

We thank L. Walker and T. Doherty for useful discussions and T. Scrivener and P. Collins for the loan of the SLM.