Post-coronagraphic PSF sharpening with the vortex coronagraph

G. Orban de Xivry\textsuperscript{a}, O. Absil\textsuperscript{b}, E. Huby\textsuperscript{a, b}, and A. Jolivet\textsuperscript{a}

\textsuperscript{a}Space sciences, Technologies, and Astrophysics Research (STAR) Institute, Université de Liège, Belgium
\textsuperscript{b}Laboratory for Space Studies and Instrumentation in Astrophysics, Observatoire de Paris, France

ABSTRACT

Small-angle coronagraphs are key components for future instruments being developed for the ELTs. Because they deliver high contrast capabilities at small angular separations from the host star, they enable us to address by direct imaging one of the most important science case of the ELTs, the discovery and characterization of exoplanets and proto-planetary systems. Small-angle coronagraphs based on a focal-plane mask such as the vortex phase mask are however quite sensitive to low-order aberrations. They require precise wavefront control to avoid spurious starlight leakage that would decrease the contrast level and possibly mimic the signatures of off-axis companions. Here, we present a low-order wavefront sensor technique based on an iterative process using post-coronagraphic focal images with a vortex coronagraph. Our approach is similar to a PSF sharpening technique, which is sometimes used in classical imaging to reduce non-common path aberrations (NCPA), but applied to post-coronagraphic images with the objective of minimizing the residual light of the on-axis source. We describe the laboratory implementation of this technique on the VODCA coronagraphic bench at the University of Liège, and present experimental results validating our method. Finally, we also present a weak phase solution with the vortex coronagraph and discuss its practical implementation.

Keywords: Vortex; VODCA; small angle coronagraph; non-common path aberrations; PSF sharpening; phase diversity

1. INTRODUCTION

Vortex coronagraphs feature vortex phase masks in their focal plane, formally introducing an optical vortex by generating a phase screw $e^{i\theta}$, with $\theta$ the azimuthal coordinate. The textbook effect is to move the light of an on-axis source outside the geometrical image of the input pupil. Combined with a Lyot stop, it theoretically rejects perfectly the starlight for a clear circular aperture. One implementation of the vector vortex coronagraph is the annular groove phase mask (AGPM) based on subwavelength grating etched onto diamond.\textsuperscript{1} Currently several AGPM have been installed on large telescopes operating in the thermal infrared\textsuperscript{2}: on VLT/NACO and VLT/VISIR in 2012, followed by LBT/LMIRCam in 2013 and Keck/NIRC2 in 2015. This coronagraph is one of the most efficient in term of inner working angle (IWA), throughput, discovery space, contrast, and simplicity.

Small IWA coronagraphs are naturally very sensitive to low-order aberrations producing starlight leakage and degrading contrast, as illustrated in Figure 1. In addition, symmetric leakage (e.g. focus) can also mimics putative circumstellar disk, while asymmetric (large) leakage can mimic off-axis companions. Those considerations make low-order wavefront control a key element of high contrast imaging at small angles. Different factors when choosing such low wavefront sensor are its location, its sensitivity and its complexity. Several locations have been considered; in general the closer to the scientific camera the better. Complexity and related hardware modifications are also a concern potentially hindering their implementation. From those two considerations, techniques that would use directly the scientific images are appealing if their can be made sensitive enough.
As possible solutions, we consider two approaches: the first is a purely empirical one similar to PSF sharpening that is sometimes used to compensate NCPA and increase the Strehl ratio in normal imaging; the second is a phase diversity technique based on the weak phase solution.

The idea of the first technique is simply to optimize some metric by controlling a set of modes on the deformable mirror. This technique relies on a minimization algorithm and on the increased sensitivity to low order aberrations behind the coronagraph. The advantage of this approach is its simplicity: no hardware modification and no prior knowledge. Its disadvantage is the need for a large number of frames from the science camera to perform the minimization.

The second idea we consider is to exploit the weak-phase solution of phase diversity applied to the case of the vortex coronagraph. Under the small aberration hypothesis, by applying two phase diversities consecutively, and thus acquiring three images, one can retrieve the real and complex part of the electromagnetic field (E-field) in the Lyot plane. The E-field can then be decomposed into a set of Zernike coefficients that can be analytically back-projected to input pupil Zernike modes exploiting the formalism of Huby et al. [3]. The technique is potentially relatively fast, requiring a limited number of frames, and very sensitive. Its practical implementation has however a number of difficulties that need to be overcome: the necessity to know accurately the E-field diversities introduced in the final image plane (which implies a good modeling of the coronagraph), and the fact that the formalism of Huby et al. is strictly valid for a non-obstructed circular aperture.

In Section 2 we detail the focal sharpening technique applied to vortex coronagraphy and demonstrate the technique by simulation. In the Section 3, we present the results obtained on the Vodca bench. We then move on to the second technique in Section 4, the weak phase solution for the coronagraph, detailing the formalism and discussing a first simulation. We finish by highlighting a few key aspects and discuss future prospects for Vodca.
2. PSF SHARPENING FOR CORONAGRAPHY

The method of image sharpening is based on the optimization of a given metric. It is a simple and empirical technique with a wide range of applications, one of which is the minimization of NCPA unseen by the AO system but affecting the scientific images. In classical imaging, a relevant metric to maximize is the Strehl ratio. But the peak flux can be quite sensitive to various degradations from vibrations, seeing variations, etc. Thus some prefer more robust metrics, see e.g. Bailey et al. [5] using a symmetry argument to correct some low order modes. Whatever the metric, these methods remain limited to a handful of low order aberrations due to the inherent low sensitivity to higher order aberrations provided by the classical PSF.

We propose to implement and test a similar empirical method for the vortex coronagraph. The natural metric to minimize here is the residual star light arriving to the detector, or some metric associated to it. A direct advantage of this approach is the improved sensitivity to aberrations behind the vortex. Indeed as illustrated in Figure 1, the flux leakage is approximately quadratic with respect to the level of aberrations. We can therefore expect to reach higher performance, i.e. higher Strehl, and be able to correct a relatively large number of modes.

Our procedure, based on the post-coronagraphic PSF, is thus as follows:

1. Select a metric and a minimization algorithm. See the two paragraphs below.
2. Select a modal basis (e.g. Zernike or mirror modes), the coefficients of which are the free parameters of the optimization problem.
3. For each iteration, acquire and average a number of images, evaluate the metric and find the next solution of the modal basis

**Metrics.** We consider different metrics which consist in minimizing the total flux, or the total of a power of the flux, within a given region of interest, e.g.:

\[
\text{minimize:} \quad R = \frac{1}{\text{norm}} \sum_{x,y} I(x,y)
\]

\[
\text{minimize:} \quad R = \frac{1}{\text{norm}} \sum_{\text{annulus}} I(x,y)
\]

\[
\text{minimize:} \quad R = \frac{1}{\text{norm}} \sum_{x,y} I^{1/3}(x,y)
\]

The summation region should be chosen according to the range of spatial frequencies one wants to correct (selecting a larger region should only mildly affect the optimization, while selecting a smaller one will lead to some badly seen spatial frequencies). We do not find any noticeable difference in the results obtained with those different metrics. The perfect simulated case always converged to the solution in similar time-wise manners – suggesting that those metrics are similarly sensitive. In practice, on the bench, other factors limit the final performance, see Section 3.

**Optimization algorithm.** Since the Zernike modes (or in general any other modal basis) will not be independent from each other relative to the flux leakage, a sequential least-square minimization on each mode may not be the optimum method. Therefore, we consider different algorithms (see e.g. Press et al. [6] for detailed descriptions, and Scipy\(^7\) for their implementations in Python), less sensitive to local minima and more adapted to our problem:

- The Nelder-Mead simplex method. The algorithm works by refining a simplex that expands, contracts, and generally adapts itself to the landscape, thus being more resilient to local minima. It does not rely on computing gradients.
- Powell’s method. The method uses a set of directions and performs one-dimensional minimization at a time. It also does not rely on computing gradients. In our case, when optimizing a large set of parameters, those directions generally end up being each mode.
- L-BFGS-B algorithm, the limited memory BFGS bounded algorithm. It computes gradients on-the-fly, and refines them at each step.

All three algorithms, with appropriate initial conditions, can converge correctly to the solution. While we do not perform a detailed comparison of those minimization routines, the Powell’s method appears to be the simplest, most robust, and fastest algorithm to the problem at hand.

2.1 Simulation

In order to test the method for the vortex coronagraph, we develop a FFT-based code to simulate the vortex. Since we do not consider polarization here, the electromagnetic field is expressed in a scalar form. The vortex is thus simulated with a pure vortex term given by $e^{i\theta}$ where the topological charge is $l = 2$ in our case, and a leakage term, i.e. a fraction of the E-field not affected by the phase ramp. This leakage term is thus the equivalent of the PSF attenuated by the peak rejection ratio of the simulated vortex phase mask.

This numerical treatment is then fed to our optimization routine, which for each iteration modifies the input Zernike aberrations, recalculates the vortex PSF and calculates its residual metric. One simulation is illustrated in Figure 2. The radial profiles show the off-axis PSF (black), a strongly aberrated post-vortex PSF (blue), and the result after optimization (green). The initial aberrations are spread randomly over 100 modes following a powerlaw distribution $\propto n^2$ with $n$ the Zernike radial order. The total injected wavefront error is 200nm, for an operating wavelength at 3.75$\mu$m. The final vortex PSF is obtained after running our optimization routine on all 100 modes and subsequently adding a residual aberration of 20nm rms wavefront uniformly distributed over the same 100 modes. This last step is performed for the purpose of comparison to lab measurements, which will become clear in the next section. We observe that in those conditions, at $\lambda/D > 1.2$, the raw contrast is $\sim 2 \times 10^{-5}$.

In this example, we have optimized all modes involved in the input aberrations. However the routine (esp. with the Powell’s algorithm) works equally well when only a fraction of the modes are optimized.

![Figure 2](image_url)

Figure 2. Simulation of an aberrated vortex PSF and the result after the sharpening technique. The off-axis PSF is given in black. The blue radial profile is the initial aberrated vortex PSF with $\sim$200nm of aberrations spread randomly (weighted by a powerlaw distribution, function of the modal order) over 100 Zernike modes. The green curve is the radial profile obtained after optimization, to which we have added a residual aberration of 20nm rms distributed across 100 modes. We add this residual error to obtain a more realistic curve comparable to lab measurements.
3. APPLICATION ON VODCA

In a next step, we implement the post-coronagraphic PSF sharpening technique on our bench. VODCA is the Vortex Optical Demonstrator for Coronagraphic Applications. It is an achromatic coronagraphic test bench developed at the University of Liège currently designed to operate in the NIR from ∼1.5 to 4 µm (H to L band), developed mainly to assess the performance of vortex phase masks.

A detailed description of the bench is out of the scope of this paper, but its main characteristics are:

- the optical train is fully reflective, except for the vortex phase mask itself,
- the source is a supercontinuum laser delivering a single-mode broadband beam from approx. 1.5 to 4 µm,
- the detector, a FLIR camera, is an infrared InSb camera cooled to 77 K operating at wavelengths ranging from approx. 1.5 to 5.5 µm,
- the deformable mirror is an ALPAO DM with 97 actuators, which currently also acts as a pupil stop.
- the hardware devices are controlled by a multi-threaded Python code with a Qt interface. This allows fast operations and online monitoring of the performance.

The f-number at the AGPM is f# ∼ 50, which means the PSF size on the AGPM is ≤ 200µm. On the detector, the diffraction limit (1λ/D) corresponds to ∼10 pixels, i.e. a well sampled PSF.

For the optimization, we choose to use mirror modes - instead of Zernike polynomials - to mitigate fitting error, but we keep tip and tilt in our control basis. The results of several minimizations obtained on the bench are illustrated for a L-band AGPM in Figure 3. We perform the optimization for different number of modes and compare the raw contrast curves obtained. We can see the control radius receding, in particular from 20 to 50 modes from ∼2 - 3 λ/D, and the raw contrast improving. The right side of the panel in Figure 3 illustrates the final off-axis and on-axis PSFs (arcsinh color scale). In the on-axis PSF image we can distinguish the first Airy ring of coronagraphic PSF (leakage term).

The raw contrast at >1.5λ/D is about 2×10⁻⁵ or about 11.7mag. The speckle noise shows a plateau up to ∼3λ/D, with a level of 2×10⁻⁵, this is about 20nm residual WF rms error as deduced from our earlier simulation.

We can identify several sources of limitation for the final performance:

- the supercontinuum source presents an observable scintillation which decreases with increased power. This variation is not a total intensity variation but is pronounced when using narrow band filters (which is our typical operational mode). This is thus related to the output spectrum of the source and is probably a consequence of variation of spectral broadening efficiency. In our operating condition, the peak-to-valley flux variation can be as high as ∼10%. This is largely mitigated by averaging about 40 frames per iteration.
- Vibrations and turbulence on the bench also affect the starlight suppression. We measured a PtV motion of ∼0.4 pixels for the off-axis PSF, which corresponds roughly to ∼0.05λ/D. Behind the vortex, we recorded leakage variation with up to 3×10⁻⁴ relative to the off-axis PSF flux (normal events are on the order of 1×10⁻⁴). With a very good AGPM as shown in Figure 3, this can represent up to +40% (or +20% respectively) of flux variation.
- a third effect that is relatively difficult to characterize and to compensate for with our setup is the “creep effect”, or more generally the shape stability provided by the DM. It is acknowledged that the ALPAO DM presents such “creep effect”, i.e. a relative slow drift away from the newly applied shape that depends on the history of past mechanical stresses. It has been characterized by, e.g., Bitenc et al. [8]. Our strategy to mitigate this problem is an initial warmup of the DM and its electronics for about half an hour, and then concerning the minimization we limit the aberrations range that is explored by the optimization routine to balance a good behavior of the algorithm and limit mechanical stress. It is also our impression that the relative lengthy optimization will tend to diminish the creep effect.
Figure 3. (Left) Coronagraphic profiles measured in the $L$-band (wide $L$-band filter, 3.5 - 4.0$\mu$m). The peak-to-peak attenuation is $\sim 8 \times 10^{-4}$. The black and grey curves represent the off-axis PSF, the dashed black curve is the profile of a 2D-Airy function fit. Color curves are measured on-axis coronagraphic PSF with different degree of correction with the deformable mirror on VODCA. The estimated level of speckle noise (DM flat wavefront error) on the bench at this wavelength is $\sim 2 \times 10^{-5}$ in the 1 - 3 $\lambda/D$ regime. (Right) Off-axis and coronagraphic PSF with all modes corrected. Images are represented in arcsinh scale. The coronagraphic image is scaled by a factor 100 and the off-axis image is saturated in the core.

As illustrated in Figure 3, this post-coronagraphic PSF sharpening works well. It is fairly robust and reproducible. The main disadvantage is its time consumption: depending on the number of modes that we want to correct the minimization can require 1000 to 10000 function evaluations. At 60Hz frame rate and 40 frames per iterations to mitigate the different noises described above, the optimization can take up to about one hour.
4. WEAK PHASE SOLUTION WITH THE VORTEX

The second approach we propose consists of two steps, sketched in Figure 4:

1. Recover the electric field in the Lyot plane by phase diversity. For this step, we propose to use the weak phase solution for the case of the vortex. This allows an analytical treatment and requires three images.

2. Project the E-field of the Lyot plane to the input pupil plane to obtain the entrance phase map. This relies on the formalism developed by Huby et al. [3] which provides a complex linear relationship from entrance pupil Zernike modes to aberrations in the Lyot plane.

We detail the method, first by describing the small aberrations case with the vortex, then the weak phase solution formalism, and then the practical implementation illustrated by a simulation.

Small aberrations with the vortex. Under the small aberration hypothesis, the wavefront in the pupil plane and the Lyot plane can be written to first order as

$$E_{\text{pupil}} \approx 1 + i\phi = 1 + i \sum_{j=2}^{\infty} a_j Z_j(r)$$

$$E_{\text{lyot}} \approx \alpha + i\beta$$

(1)

The formalism of Huby et al. [3] describes the Lyot plane field $E_{\text{lyot}}$ as a function of polynomials $\zeta_j$. Those polynomials, after the Lyot stop, can be expressed as a complex function of Zernike mode $\zeta_j = \sum_k C_{jk} Z_k(r)$, where $C$ is the conversion matrix illustrated in Tab. A. 1. of 3. Therefore, the field in the Lyot plane can be expressed as follows

$$E_{\text{lyot}} \approx i \sum_{j=2}^{\infty} a_j \zeta_j(r)$$

$$\approx -a_j C_3,jk Z_k(r) + ia_j C_{\Re,jk} Z_k(r)$$

(2)

where $C_3 = \Im(C)$ and $C_{\Re} = \Re(C)$, and we have dropped the summation symbols.

Weak-phase solution for the Lyot plane. Similarly to classical imaging, we can express a general weak-phase solution for the vortex coronograph. Assuming that three images are taken, among which two with known phase-diversity, the images can be written:

$$p_1 = |E_1|^2 = |\mathcal{F}(\alpha + i\beta)|^2 = |x + iy|^2$$

$$p_2 = |E_2|^2 = |\mathcal{F}((\alpha + \alpha_{d1}) + i(\beta + \beta_{d1}))|^2 = |(x + x_{d1}) + i(y + y_{d1})|^2$$

$$p_3 = |E_3|^2 = |\mathcal{F}((\alpha + \alpha_{d2}) + i(\beta + \beta_{d2}))|^2 = |(x + x_{d2}) + i(y + y_{d2})|^2$$

(3)

where $p_1$ is the original image in the coronagraphic focal plane, and $p_2$ and $p_3$ are the images obtained with added known phase-diversity, labelled $d_1$ and $d_2$. Under known phase-diversity, one can solve the linear equations:
\[
\begin{bmatrix}
2x_{d1} & 2y_{d1} \\
2x_{d2} & 2y_{d2}
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix}
= \begin{bmatrix}
p_2 - p_1 - x_{d1}^2 - y_{d1}^2 \\
p_3 - p_1 - x_{d2}^2 - y_{d2}^2
\end{bmatrix}
\] (4)

The phase diversities should be linearly independent to invert the matrix and find \( x \) and \( y \). Since the right side of Eq. 4 involves the subtraction of two similar PSFs, the coefficients calculated will be noisy and the focal plane field \( E_1 = x + iy \) may be more robustly estimated with

\[
E_1 = |p_1|^{0.5} \exp(i \arg(x + iy))
\] (5)

Finally, the Lyot E-field is derived by inverse Fourier transform \( E_{lyot} = \mathcal{F}^{-1}\{E_1\} \).

**Retrieving the small aberrations in the pupil plane. Procedure & simulation.** Based on the proposed formalism, we implement an iterative procedure with the following steps:

1. Acquire three images with linearly independent phase diversity – possibly limiting spatial frequencies by windowing –, solve the linear system Eq. 4 and compute the field in the Lyot plane.
2. Decompose the real \( \alpha \) and imaginary \( \beta \) part of the \( E_{lyot} \) in Zernike polynomials, which gives the polynomial coefficients \( b_r \) for the real and \( b_i \) for the imaginary part.
3. Using Eq. 1 and Eq. 2, and inverting we can express the Zernike coefficients \( a \) vector in the input pupil plane as

\[
a = (b_r + ib_i)[-C_3 + iC_9]^{-1}.
\] (6)
4. We enforce real Zernike coefficients \( a \) for the pupil plane.
5. The \( a \) vector is finally used in a proportional integrator (possibly with a leaky gain) to correct the wavefront iteratively.

The weak phase solution requires the exact knowledge of the diversity introduced, i.e. the E-field diversities in the focal plane \( x_{d1}, x_{d2}, y_{d1}, y_{d2} \). This implies a proper modelling of the vortex coronagraph.

To simulate and test the method, we use our coronagraphic model, see Section 2.1, and implement the presented iterative procedure. An illustration of the results is given in Figure 5 for a circular pupil and a vortex coronagraph with a peak rejection ratio of 1000:1. We inject 30 random Zernike modes with a total WFE rms of \( \sim 0.03 \) wave, or about 100nm in L-band. In this example, we use amplitude diversities by masking an edge of the entrance pupil : 0.25% for \( p_2 \) and 1% of the pupil area for \( p_3 \). Similar results are obtained with phase diversities, with e.g. 0.1 wave rms of \( Z_5 \) and \( Z_6 \) respectively. Those diversities can be arbitrarily small in the case of infinite SNR.

In Figure 5, we illustrate the results of the simulation. We can see that the method converges rapidly in only a few iterations and that the Zernike coefficients are almost perfectly recovered.
Figure 5. Upper row: (left) input initial wavefront, (middle) the final subtracted wavefront (WFE derived by the algorithm), (right) the input Zernike coefficients in blue and the final computed coefficients in orange. Bottom row: (left) initial vortex PSF, (middle) final vortex PSF, (right) flux and WFE as a function of iteration 10 iterations in total.

5. CONCLUSIONS

In this proceeding, we present two methods to minimize low-order aberrations responsible for starlight leakage after the vortex coronagraph.

The following comments and conclusions can be made:

- the focal plane sharpening technique on the VODCA bench has brought a noticeable gain on measured raw contrast curves by improving wavefront control. Indeed, current contrasts in L-band reach about $10^{-5}$ at $\sim 1.5\lambda/D$, a factor approx. 10 times better than previously reported with our infrared vortex phase masks. This improvement and the repeatability of those measurements are largely accountable to the wavefront control (which is especially important for low aberrations such as focus and astigmatism). Higher spatial frequencies (above the DM control range) are however generally degraded.

- the focal plane sharpening technique is expensive in terms of data frame requirement and access to the deformable mirror. This makes it impractical for on-sky operation. But it can be envisaged for day time calibration of NCPA. In particular it can easily bootstrap large aberrations. It could possibly be used in combination with other techniques such as speckle nulling, e.g. : a first step with the focal sharpening technique to correct 20-30 modes, and then applying speckle nulling to further clean up the aberrations at $\geq 1 - 2\lambda/D$.

- the weak phase solution can converge fairly rapidly and needs only small diversities. As other phase diversity algorithms, it however relies on a precise modeling of the coronagraph. In addition, the weak phase solution assumes implicitly that all the phase aberrations are upstream of the vortex. Practical implementation would require a detailed analysis of the stability and operating range of the technique.

Future works will explore the robustness of the focal plane sharpening method at shorter wavelength and with less sensitive vortex, i.e. on AGPM with a larger leakage term thus reducing the sensitivity to aberrations. We are also currently implementing speckle nulling procedures on our bench, which will also allow us to test the efficiency of a joint focal plane sharpening and speckle nulling method.
REFERENCES


