










Coronagraphic Data Post-processing Using Projections on Instrumental Modes

Yinzi Xin^{1,2,3,8} , Laurent Pueyo² , Romain Laugier⁴ , Leonid Pogorelyuk¹ , Ewan S. Douglas⁵ , Benjamin J. S. Pope^{6,7} , and Kerri L. Cahoy¹ 

¹ Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; yxin@caltech.edu

² Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

³ Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA

⁴ Université Côte d'Azur, 28 avenue Valrose, Nice, 06 F-06100, France

⁵ University of Arizona, Steward Observatory, 933 N. Cherry Avenue, Tucson, AZ 85721, USA

⁶ School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia

⁷ Centre for Astrophysics, University of Southern Queensland, West Street, Toowoomba, QLD 4350, Australia
Received 2021 February 2; revised 2023 November 30; accepted 2023 December 22; published 2024 March 4

Abstract

Directly observing exoplanets with coronagraphs is impeded by the presence of speckles from aberrations in the optical path, which can be mitigated in hardware with wave front control, as well as in post-processing. This work explores using an instrument model in post-processing to separate astrophysical signals from residual aberrations in coronagraphic data. The effect of wave front error (WFE) on the coronagraphic intensity consists of a linear contribution and a quadratic contribution. When either of the terms is much larger than the other, the instrument response can be approximated by a transfer matrix mapping WFE to detector plane intensity. From this transfer matrix, a useful projection onto instrumental modes that removes the dominant error modes can be derived. We apply this approach to synthetically generated Roman Space Telescope hybrid Lyot coronagraph data to extract “robust observables,” which can be used instead of raw data for applications such as detection testing. The projection improves planet flux ratio detection limits by about 28% in the linear regime and by over a factor of 2 in the quadratic regime, illustrating that robust observables can increase sensitivity to astrophysical signals and improve the scientific yield from coronagraphic data. While this approach does not require additional information such as observations of reference stars or modulations of a deformable mirror, it can and should be combined with these other techniques, acting as a model-informed prior in an overall post-processing strategy.

Unified Astronomy Thesaurus concepts: [Coronagraphic imaging \(313\)](#); [Exoplanet detection methods \(489\)](#); [Exoplanet astronomy \(486\)](#); [Direct imaging \(387\)](#)

1. Introduction

Specialist high-contrast techniques are required to directly observe faint astrophysical objects near brighter objects, such as exoplanets, brown dwarfs, or circumstellar disks orbiting much brighter central stars. High-contrast observations are essential for answering scientific questions involving binary and planetary system population statistics, planet and disk formation and evolution, planetary atmospheres, and planet habitability and the search for biosignatures (Traub & Oppenheimer 2010). Measuring these exoplanet signals is difficult because they often lie at small angular separations from their host star and can be many orders of magnitude fainter. One major obstacle for high-contrast observations is photon noise from the light of the central star. Practical matters such as detector saturation aside, if the star is orders of magnitude brighter than its companion, the photon noise associated with the outer lobes of the star’s point-spread function (PSF) can overwhelm any signal from the companion, even if the on-axis star’s signal is perfectly known. As a result, instruments to directly suppress starlight, such as coronagraphs and nullers, are important in increasing the photon signal-to-noise ratio (S/N) of faint companions.

Another important source of noise is wave front error (WFE), which distorts the signal of the on-axis point source. Sources of WFE include atmospheric turbulence, imperfections in the optics, or thermomechanical changes in the telescope or instrument. At an instant in time, a perturbation to the wave front scatters energy from the core of the PSF into speckles throughout the image that can resemble off-axis sources. If the magnitude of the WFE electric field is smaller than that of the underlying electric field from the PSF, then the speckles are symmetric about zero in detector plane intensity and average out over time. When the WFE is the larger term, as in the case of uncorrected atmospheric turbulence, the speckles are predominantly positive, increasing rather than decreasing the intensity over most of the focal plane, and averaging out to a halo that can obscure off-axis signals. Scattered starlight at larger spatial separations increases the photon noise at those locations in the detector plane, which can also dominate over signals from faint companions.

The goal of high-contrast instruments is to separate the signal of the on-axis star from off-axis sources. Coronagraphs are passive optical elements that spatially filter the light to suppress the signal of an on-axis star, reducing its associated photon noise while letting through off-axis signals (Guyon et al. 2006). Adaptive optics (AO; Tyson 2000) and focal-plane wave front control (Groff et al. 2016) actively correct for WFE to reduce their impact. However, even with suppression from coronagraphs or nullers, the sensitivity to faint astrophysical signals is still limited by residual starlight and its associated photon noise.

⁸ Corresponding author.

Post-processing techniques can use additional available information to further mitigate the effects of WFE and increase sensitivity to real astrophysical signals (see Cantalloube et al. 2022 for a discussion of the state of the art of high-contrast post-processing in the context of a direct imaging data challenge). For example, angular differential imaging exploits observations at different roll angles, taking advantage of azimuthal averaging of the WFE (Marois et al. 2006; Flasseur et al. 2018). Other methods rely on performing principal component analysis (PCA) on reference observations of a calibration star similar to the host star, but without astrophysical companions, to calibrate out residual static or quasi-static starlight (Lafrenière et al. 2007; Soummer et al. 2012; Pueyo 2016). Additional sources of information on residual WFE include telemetry from wave front sensing and control (WFSC) systems such as wave front sensor residuals (Vogt et al. 2011) or focal-plane electric field estimates (Pogorelyuk et al. 2019), data from a self-coherent camera (Baudoz et al. 2006), and data at different wavelengths as exploited in spectral deconvolution (Sparks & Ford 2002).

This work shows that the modeled or measured instrument sensitivity to WFE can be included as an additional source of information in the post-processing of coronagraphic data, information that, in theory, can be combined with the other techniques discussed. This work examines an approach that uses this physical optics model to construct a projection removing the dominant error modes in the appropriate WFE regime, and it finds that this can improve sensitivity to faint companions by up to and over a factor of 2.

2. Coronagraphic Signals

2.1. Data Formation

The model used in this work assumes that the light through the instrument is monochromatic. With a discrete representation of the optical planes of an instrument, a coronagraph can be modeled as a linear operator C , a constant 2D matrix transforming the electric field vector at the pupil plane, E_s , into the electric field vector at the detector plane, E_{det} . If E_{s0} is the electric field vector of the central source (star) at the pupil plane in the absence of aberrations and ΔE_s is a vector of small perturbations to that electric field, representing wave front aberrations (which can be variable in time), then the electric field vector at the detector plane, assuming that the star is the only source of light, is

$$E_{\text{det}} = CE_s = CE_{s0} + C\Delta E_s(t). \quad (1)$$

The intensity measured is the element-wise norm squared of the detector plane electric field (here \bar{x} indicates the element-wise complex conjugate of x and \circ indicates the element-wise product):

$$\begin{aligned} I_s &= |E_{\text{det}}|^2 \\ &= |CE_{s0}|^2 + 2 \operatorname{Re}\{\overline{(CE_{s0})} \circ C\Delta E_s(t)\} + |C\Delta E_s(t)|^2. \end{aligned} \quad (2)$$

The vector of the pupil-plane electric field of a binary companion is given by

$$E_p = \sqrt{c} E_{s0} e^{-ik \cdot x} + \sqrt{c} \Delta E_s(t) e^{-ik \cdot x} = E_{p0} + \Delta E_p(t), \quad (3)$$

where c is the flux ratio between the planet and the star, k is the pupil-plane wavevector indicating the companion's location, and x is the pupil-plane coordinate vector. Namely, the planet's pupil-plane electric field is the star's electric field, but tilted and

scaled by the square root of the flux ratio. The detector plane intensity for the planet can be expressed as

$$I_p = |CE_{p0}|^2 + 2 \operatorname{Re}\{\overline{(CE_{p0})} \circ C\Delta E_p(t)\} + |C\Delta E_p(t)|^2. \quad (4)$$

The total intensity on the detector plane from the star and the planet is the sum of Equations (2) and (4). However, we can make two simplifying assumptions. The first assumption is that the flux of the planet is small relative to the flux of the star, such that $c \ll 1$. The second assumption is that the magnitude of the WFE is small relative to the total magnitude of the electric field, namely $\Delta E_s(t) \ll E_{s0}$, which implies $\Delta E_p(t) \ll E_{p0}$. This is true if we are both in the “small phase regime” (when there is much less than one wave of WFE) and the fractional amplitude error is much less than 1. These assumptions imply that the last two terms of Equation (4) are small relative to the other terms, so we can approximate the total intensity as

$$\begin{aligned} I_{\text{tot}} &\approx |CE_{s0}|^2 + 2 \operatorname{Re}\{\overline{(CE_{s0})} \circ C\Delta E_s(t)\} \\ &\quad + |C\Delta E_s(t)|^2 + |CE_{p0}|^2. \end{aligned} \quad (5)$$

The first term, $|CE_{s0}|^2$, is the residual starlight not blocked by the coronagraph in the case of no aberrations. The second term, $2 \operatorname{Re}\{\overline{(CE_{s0})} \circ C\Delta E_s\}$, is linear in the wave front aberration and corresponds to the interference of the aberration, propagated to the focal plane, with the underlying residual starlight from the coronagraph, analogous to speckle pinning (Bloemhof 2003; Perrin et al. 2003). The third term, $|C\Delta E_s|^2$, is the quadratic term, corresponding to the norm squared of the WFE propagated to the focal plane. The last term, $|CE_{p0}|^2$, is the nominal off-axis signal of interest.

Whether the effects of WFEs at some location in the detector plane are dominated by the linear term or the quadratic term depends on the attenuation of starlight by the coronagraph and the level of the propagated WFE at that location. If the propagated wave front aberrations are smaller in complex amplitude than the residual starlight after the coronagraph with no aberrations, the linear term is dominant. When a coronagraph is not used, this corresponds to the speckle pinning regime, in which the aberrations primarily interfere with the wings of the telescope's PSF (Bloemhof 2003). The same phenomenon occurs with a coronagraph; however, as the amplitudes of the PSF wings are reduced by the coronagraph, the range of WFE over which this occurs is much more limited. Otherwise, if propagated wave front aberrations have relatively larger magnitudes, the quadratic term is dominant. For a given location in the focal plane, the local point of transition between the linear and quadratic regimes occurs when $|2 \operatorname{Re}\{\overline{(CE_{s0})} \circ C\Delta E_s\}| = |C\Delta E_s|^2$, or roughly when $|2CE_{s0}| = |C\Delta E_s|$.

This point of transition is different for each pixel and also depends on the coronagraph design, as well as the “nominal” wave front (whether it is flat, as is typical for ground-based coronagraphs, or the wave front corresponding to a dark hole, as is planned for space-based coronagraphs). For this work, we use as an example the Hybrid Lyot Coronagraph (HLC) of the Coronagraph Instrument of the Roman Space Telescope. With the dark hole presented in Section 5.1, which has an average raw contrast (residual stellar intensity divided by unocculted peak intensity) of 5.6×10^{-9} , the point at which $|C\Delta E_s| > |2CE_{s0}|$ for 50% of the pixels in the dark-hole region occurs at roughly 0.1 waves rms of phase error, on average. This means that WFEs less than 0.1 waves rms will primarily be in the

linear regime, while WFEs larger than 0.1 waves rms will primarily be in the quadratic regime, although this is somewhat dependent on the form of the wave front's spatial power spectral density (PSD) that we use in Section 5.4.1.

In this work, robust observables are only formulated for WFE that is predominantly linear or predominantly quadratic throughout the entire focal plane. However, it may be possible to obtain robust observables for when both terms have comparable contributions, a topic that is left for future work.

2.2. Linear Regime

From Equation (5), if we then assume that the linear error term is dominant, then we can drop the quadratic contribution such that the detector plane intensity is approximately

$$\mathbf{I}_{\text{tot},l} \approx |\mathbf{C}\mathbf{E}_{s0}|^2 + 2 \operatorname{Re}\{\overline{(\mathbf{C}\mathbf{E}_{s0})} \circ \mathbf{C}\Delta\mathbf{E}_s(t)\} + |\mathbf{C}\mathbf{E}_{p0}|^2. \quad (6)$$

The contribution of the WFE to the intensity can be expressed as a linear transformation \mathbf{A}_l acting on the WFE:

$$\mathbf{I}_{\text{tot},l} \approx |\mathbf{C}\mathbf{E}_{s0}|^2 + \mathbf{A}_l \Delta\mathbf{E}_s(t) + |\mathbf{C}\mathbf{E}_{p0}|^2. \quad (7)$$

The transfer matrix \mathbf{A}_l can be calculated semianalytically from the coronagraph operator and the unaberrated electric field, as derived from Equation (6):

$$\mathbf{A}_{lkj} = \frac{\partial I_k}{\partial \Delta E_{s_j}} = 2 \operatorname{Re} \left\{ \left(\sum_i C_{ki} E_{s0_i} \right)^* C_{kj} \right\}. \quad (8)$$

The indices i and j label the input basis vectors used to represent the WFE, and the index k labels the detector pixel.

It is desirable to reduce the term dependent on WFE, $\mathbf{A}_l \Delta\mathbf{E}_s(t)$, relative to the terms containing astrophysical signals of interest. This can be achieved by left-multiplying the measured intensities by a matrix \mathbf{K}_l , that projects out the dominant modes of \mathbf{A}_l . Section 3 describes the process of calculating \mathbf{A}_l and finding from it an appropriate \mathbf{K}_l . The observables obtained using projection matrix \mathbf{K}_l are given by

$$\mathbf{O}_l = \mathbf{K}_l \mathbf{I}_{\text{tot},l}. \quad (9)$$

When the WFEs are in the linear regime, this projection is expected to suppress the contribution of WFEs to the measured data. As long as the measurements retain most of the astrophysical signal, then the projection will boost its S/N.

2.3. Quadratic Regime

In the quadratic-dominated regime, we can drop the linear contribution in Equation (5), such that the detector plane intensity is approximately

$$\mathbf{I}_{\text{tot},q} \approx |\mathbf{C}\mathbf{E}_{s0}|^2 + |\mathbf{C}\Delta\mathbf{E}_s(t)|^2 + |\mathbf{C}\mathbf{E}_{p0}|^2. \quad (10)$$

In a discrete numerical model, the contribution of the quadratic term to each pixel labeled k in the detector plane can be expressed as

$$\begin{aligned} |\mathbf{C}\Delta\mathbf{E}_s|_k^2 &= \left(\sum_m C_{km} \Delta E_{s_m} \right)^* \left(\sum_n C_{kn} \Delta E_{s_n} \right) \\ &= \sum_i \sum_j \Delta E_{s_i}^* M_{kij} \Delta E_{s_j}. \end{aligned} \quad (11)$$

The indices m , n , i , and j label the input basis vectors used to represent the WFE (expressed here in terms of perturbation to the complex electric field), and the index k labels the detector pixel. The quantity $\hat{\mathbf{M}}$ with elements M_{kij} is a 3-tensor containing the

second-order partial derivative matrix (Hessian) of each pixel intensity with respect to the WFE and relates each pairwise combination of pupil basis vectors to its effect on each detector plane pixel k . Each entry can be calculated semianalytically from the coronagraph operator using the following formula derived from Equation (11):

$$M_{kij} = \frac{\partial^2 I_k}{\partial \Delta E_{s_i}^* \partial \Delta E_{s_j}} = C_{ki} C_{kj}^* + C_{kj} C_{ki}^*. \quad (12)$$

Assuming that there are N_{pix} pixels of interest on the detector and N basis vectors are used to represent the WFE, then, through a remapping, the 3-tensor $\hat{\mathbf{M}}$ of size $(N_{\text{pix}} \times N \times N)$ can be expanded into a matrix acting on the space of all *pairwise combinations* of pupil basis vectors. Since Hessians are symmetric because partial derivatives commute ($M_{kij} = M_{kji}$), the ordering of each pair of segments does not matter, and the derivatives corresponding to the same pair of original basis vectors can be consolidated into the same entry. This results in a vector space of size $\binom{N+1}{2}$ or the number of pairwise combinations of pupil basis vectors.

The 3-tensor $\hat{\mathbf{M}}$ can thus be represented as an $(N_{\text{pix}} \times \binom{N+1}{2})$ matrix \mathbf{A}_q of second derivatives, acting on a vector β of perturbations defined for each pairwise combination of original basis vectors $\Delta E_{s_i} \Delta E_{s_j}$. This results in the following expression for the quadratic term:

$$|\mathbf{C}\Delta\mathbf{E}_s|^2 = \mathbf{A}_q \beta. \quad (13)$$

The projection is similar to the linear case: the detector intensities can be left-multiplied by a matrix \mathbf{K}_q that projects out the dominant quadratic error modes of \mathbf{M} . The observables with the appropriate projection \mathbf{K}_q are given by

$$\mathbf{O}_q = \mathbf{K}_q \mathbf{I}_{\text{tot},q}. \quad (14)$$

3. Response Matrices and Robust Observables

3.1. Calculating the Response Matrix

This section details the numerical calculation of instrument response matrices and the projection matrices. In this work, the response matrix is calculated with the wave front aberrations represented in the Zernike basis. In this basis, ΔE_{Z_n} is the coefficient of the aberration induced by the n th Noll-ordered Zernike polynomial (Noll 1976), and N is the total number of polynomials chosen to construct the response matrix:

$$\Delta\mathbf{E}_s = \begin{pmatrix} \Delta E_{Z_1} \\ \dots \\ \Delta E_{Z_N} \end{pmatrix}. \quad (15)$$

We define N_{pix} as the total number of detector pixels of the optical model and N_{basis} as the number of Zernike modes to include. The coronagraph operator \mathbf{C} is the $N_{\text{pix}} \times N_{\text{basis}}$ matrix that, when applied to a vector of Zernike coefficients, gives the perturbation they induce in the focal-plane electric field. This operator is typically either already part of the optical model or obtainable by propagating Zernike modes through the optical model and using finite differences to populate its columns. Given the operator \mathbf{C} and the initial unaberrated focal-plane electric field, we can calculate both \mathbf{A}_l and \mathbf{A}_q using Equations (8) and (12). Note that the term $(\sum_k C_{kj} E_{s0_k})$ in Equation (8) is simply the initial unaberrated focal-plane

electric field at pixel k . For more complicated models without simple analytical solutions (such as those that include distortion), automatic differentiation, in which arbitrary exact derivatives can be computed without finite differences, may be useful (Pope et al. 2021).

The linear transfer matrix poses no computational problems, as its size is $N_{\text{pix}} \times N_{\text{basis}}$. However, for the quadratic transfer matrix A_q , the size of the input dimension quickly becomes computationally burdensome for high N_{basis} . For the example system shown in Section 5, an N_{basis} of 528 results in an A_q matrix of width $\binom{N_{\text{basis}} + 1}{2} = 139,656$ (the number of pairwise combinations of pupil basis vectors) and length 5476 (the number of detector pixels of the model). This A_q matrix, when represented as (noncomplex) doubles, is over 6 GB in size. As explained in Section 3, the calculation of the projection matrix involves a singular value decomposition (SVD) of the response matrix. Since calculating the SVD of a matrix of this size is too computationally expensive, we restrict our quadratic transfer matrix to only include the first $N_{\text{redu}} = 100$ Zernikes, which results in an A_q with a width of only $\binom{N_{\text{redu}} + 1}{2} = 5050$. This model is valid only in a smaller area closer to the central star—namely within $\sim 5\lambda/D$, where λ is the wavelength and D the telescope diameter. However, in the Appendix, we explore using an approximation of the quadratic transfer matrix that can extend the area of applicability while circumventing impractical computational costs.

3.2. Calculating the Projection Matrix

Once a response matrix A is obtained, an SVD of $A = USV^T$ is performed, revealing its singular modes and corresponding singular values. Then, a choice of the number of modes to project out (N_m) is made. The remaining $N_{\text{pix}} - N_m$ modes are kept in the post-processing projection K . Accordingly, K is the subset of U that contains the $m + 1$ th and higher left singular modes of A . A pseudo-code summary of the process to find K is given in Algorithm 1. The optimal N_m depends on the signal of interest. For the point-source companion signals explored in this work, N_m is chosen as the cutoff that results in the best detection limit at the separation of interest.

Algorithm 1. Calculate projection matrix K

Input: Transfer matrix A
Input: Cutoff mode (number of modes to project out), N_m
Input: Indices of detector plane pixels in region of interest, idx
Output: Projection matrix K
 $U, S, V^T \leftarrow \text{svd}(A)$
 $K \leftarrow \text{transpose}(U(\text{idx}, N_m + 1: \text{end}))$

If the linear and quadratic projections are used in the appropriate regimes to increase S/N, they could, for example, allow for a binary signal detection with a deeper flux ratio than using the raw intensity data. Detection tests can be performed on both projected and unprojected data to quantify this effect.

4. Detection Testing

Detections are typically claimed from a statistical hypothesis test (see, e.g., Kasdin & Braems 2006; Jensen-Clem et al. 2017;

Ceau et al. 2019). A test statistic T is calculated from the data and compared to a threshold ξ . A detection is claimed if $T \geq \xi$, and a lack of a detection is claimed otherwise. The fraction of real companions detected is the true-positive rate (TPR). A false positive occurs if there is no companion in the data but the detection test incorrectly claims a detection. The rate at which this occurs is the false-positive rate (FPR).

As the detection threshold ξ is decreased, detecting real companions becomes more likely, but false detections also become more likely (Jensen-Clem et al. 2017). Varying the threshold and plotting the TPR as a function of the FPR results in a receiver operating characteristic (ROC) curve, an example of which is described in Section 5.4.3. ROC curves characterize the performance of a detection scheme and are used in the determination of flux ratio detection limits.

This work uses a simple Delta reduced χ^2 ($\Delta\chi_r^2$) statistic, or the difference in the reduced χ^2 of the data assuming that it contains only noise and the reduced χ^2 of the data assuming that it contains noise and the companion signal. The formula for calculating this test statistic from the data is given by Equation (16) (the bars indicate vector norm, and the divisions are element-wise):

$$\Delta\chi_r^2 = \frac{1}{\nu} \left(\left\| \frac{\mathbf{y}}{\boldsymbol{\sigma}} \right\|^2 - \left\| \frac{\mathbf{y} - \mathbf{x}}{\boldsymbol{\sigma}} \right\|^2 \right). \quad (16)$$

In this formula, \mathbf{y} is the data, which are the synthetically generated realizations of I_{tot} , with or without a planet. Meanwhile, \mathbf{x} is the unaberrated model of the planet signal $I_{p0} = |\mathbf{C}E_{p0}|^2$ (assuming that it is known, such as through a maximum likelihood estimation). The estimated uncertainty of the data is denoted by $\boldsymbol{\sigma}$, and ν is the degrees of freedom (the number of data elements minus the number of free parameters; a binary system's three free parameters are the flux ratio, separation, and position angle). This use of this test statistic is motivated by an assumption that the noise is Gaussian and uncorrelated, under which this quantity is related to the relative log-probabilities of the data containing both the planet signal and noise, versus containing only noise. The noise being uncorrelated and Gaussian is generically not the case. However, the effects of the correlation and non-Gaussianity of the injected noise on the resulting test statistic distributions are properly simulated and captured by the Monte Carlo methods used in this work.

5. Example: Nancy Grace Roman Space Telescope Hybrid Lyot Coronagraph

In this section, the use of robust observables with the Hybrid Lyot Coronagraph of the Roman Space Telescope is analyzed through simulation. However, this approach could also be applied to other coronagraphs, as long as the exposure times are short enough that WFE has not been averaged out. The optical model of CGI is shown in Figure 1 (Kasdin et al. 2020). The optical elements corresponding to the HLC mode (the relevant mode for this work) are depicted in the top row of the three sets of instrument configurations .

5.1. Optical Model

The HLC operates around a dark-hole state, which is obtained using focal-plane wave front control with deformable mirrors (DMs) to measure and minimize the electric field in the detector plane. Such focal-plane wave front control significantly suppresses the amount of starlight in the dark hole and allows for much deeper raw contrasts than with just a flattened

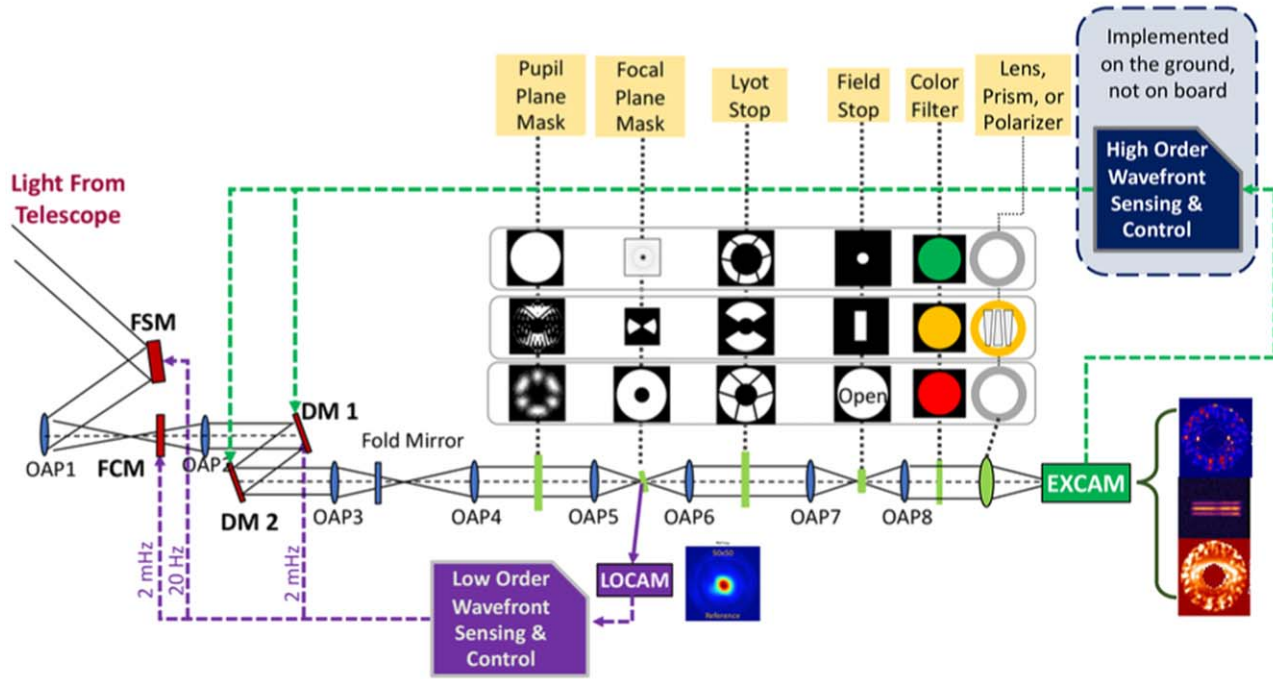


Figure 1. The CGI optical train and wave front sensing and control architecture. The optical elements of the HLC mode of interest are depicted in the top panel. Before an observation, the high-order wave front sensing and control loop is performed on a bright reference star to generate a “dark hole” (an area where starlight is suppressed). Then, the DM shapes are fixed, and the telescope slews to the target star for the observation. During the observation, WFEs accrue as a result of instrumental disturbances and drifts, the effects of which this work aims to mitigate in post-processing. Figure from Kasdin et al. (2020).

wave front. Before an observation, the dark hole is generated using a high-order wave front sensing and control loop on a bright reference star. Then, the DM shapes are fixed, and the telescope slews to the target star for the observation. During the observation, WFEs accrue as a result of instrumental disturbances and drifts. This work aims to mitigate the effects of those WFEs in post-processing. Note that, as a result of the dark-hole generation, the nominal electric field $E_{s,0}$ is not a flat wave front but the pupil-plane electric field obtained at the end of the dark-hole generation sequence.

A Lightweight Space Coronagraph Simulator (LSCS)⁹ derived from the HLC model in the Fast Linearized Coronagraph Optimizer (FALCO; Riggs et al. 2018) toolbox is used for the following simulations. The LSCS relies on the HLC numerical model and focal-plane wave front control algorithm included in FALCO to first generate the initial dark-hole electric field. The numerical model in FALCO is also used to calculate C from the finite-difference sensitivities of the focal-plane electric field to pupil-plane phase error expressed in the Zernike basis (we have made the assumption that the matrix transformation is approximately linear in phase, valid when the phase error is much less than a wave). Although we use finite differences to calculate C , one could also construct it by multiplying together all the matrix transformations of the optical model. These simulations are conducted at a single wavelength of 546 nm.

The average raw contrast of the initial dark hole is 5.6×10^{-9} . The LSCS model takes in Zernike coefficients for phase aberrations, calculates their effect on the focal-plane electric field, and adds them to the initial dark-hole electric field to obtain the focal-plane electric field in the presence of WFEs. The intensity can be calculated as the norm squared of the

focal-plane electric field. Detector noise and photon noise are not simulated. Since the default LSCS models only the first 136 Zernikes, FALCO is first used to extend the LSCS model to 528 Zernikes in order for the entire dark hole to be sampled.

This results in using 528 Zernikes to sample the entire dark hole, or an N_{basis} of 528. The LSCS models a detector that is 74×74 pixels, with 3 pixels per λ/D , for a total pixel number of $N_{\text{pix}} = 5476$. The number of pixels defined to be in the dark hole is $N_{\text{DH}} = 2608$. This model does not consider the effects of amplitude errors and only analyzes phase errors, which, from end-to-end modeling of Roman CGI, are expected to be the dominant form of dynamic aberrations (Krist et al. 2023). However, for a system where dynamic amplitude errors are comparable to dynamic phase errors, both should be included.

5.2. Response Matrices

The Zernike coefficient drift values from the Observing Scenario simulations (OS 9; Krist 2020), based on physical modeling of the telescope, indicate that the WFE expected on Roman will fall within the linear regime of this dark hole. However, the level of WFE may end up being higher than currently expected. Additionally, on ground-based telescopes, WFE from adaptive optics residuals is typically in the quadratic regime. Therefore, for illustrative purposes, both a linearly dominated noise model and a quadratically dominated noise model are examined.

The matrices A_l and A_q are calculated according to Section 3.1. The linear matrix includes all Zernikes present in the optical model and thus has an input dimension of $N_{\text{basis}} = 528$. The quadratic matrix includes only the first 100 Zernikes and thus has an input dimension of $\binom{N_{\text{redu}} + 1}{2} = 5050$. The relevant dimensions of the objects used in this analysis are

⁹ <https://github.com/leonidprinceton/LightweightSpaceCoronagraphSimulator>

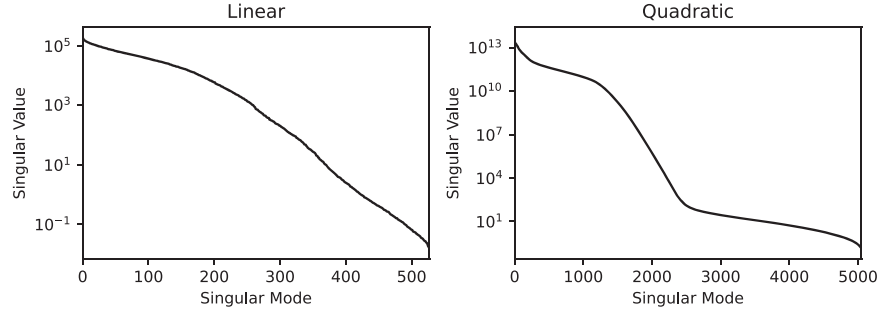


Figure 2. The singular values of A_l (left) and A_q (right). Note that the transfer matrices are rectangular and have $N_{\text{pix}} = 5476$ total singular modes, but the singular values beyond the size of the input dimension are all 0.

Table 1
Quantities and Dimensions for Analysis of the Roman Space Telescope HLC

Quantity	Description	Dimension (Dependency)	Dimension (Value)
E_{pup}	Vector of electric in pupil plane	N_{basis}	528
E_{det}	Vector of detector plane electric field	N_{pix}	5476
I_{det}	Vector of detector plane intensity	N_{pix}	5476
E_{DH}	Vector of detector plane electric field in dark hole	N_{DH}	2608
I_{DH}	Vector of detector plane intensity in dark hole	N_{DH}	2608
A_l	Linear regime instrument response matrix	$N_{\text{pix}} \times N_{\text{basis}}$	5476×528
U_l	Left singular matrix of A_l	$N_{\text{pix}} \times N_{\text{pix}}$	5476×5476
S_l	Singular value matrix of A_l	$N_{\text{pix}} \times N_{\text{basis}}$	$5,476 \times 528$
V_l	Right singular matrix of A_l	$N_{\text{basis}} \times N_{\text{basis}}$	528×528
K_l	Linear regime projection matrix	$(N_{\text{pix}} - N_m) \times N_{\text{DH}}$	$(N_{\text{pix}} - N_m) \times 2,608$
O_l	Vector of linear regime observables	$(N_{\text{pix}} - N_m)$	$(N_{\text{pix}} - N_m)$
A_q	Quadratic regime instrument response matrix	$N_{\text{pix}} \times \binom{N_{\text{redu}} + 1}{2}$	5476×5050
U_q	Left singular matrix of A_q	$N_{\text{pix}} \times N_{\text{pix}}$	5476×5476
S_q	Singular value matrix of A_q	$N_{\text{pix}} \times \binom{N_{\text{redu}} + 1}{2}$	5476×5050
V_q	Right singular matrix of A_q	$\binom{N_{\text{redu}} + 1}{2} \times \binom{N_{\text{redu}} + 1}{2}$	5050×5050
K_q	Quadratic regime projection matrix	$(N_{\text{pix}} - N_m) \times N_{\text{DH}}$	$(N_{\text{pix}} - N_m) \times 2608$
O_q	Vector of quadratic regime observables	$(N_{\text{pix}} - N_m)$	$(N_{\text{pix}} - N_m)$

listed in Table 1. Note that the cutoff number N_m is a variable to optimize over.

5.3. Projection Matrices

According to Algorithm 1, an SVD of $A = USV^T$ is performed for each transfer matrix, revealing their singular modes and corresponding singular values. The singular values of the transfer matrices are shown in Figure 2. The first 10 singular modes of each transfer matrix as represented in the detector plane intensity basis (with pixels not in the dark hole masked) are plotted in Figure 3.

From both the linear and the quadratic transfer matrix, model-based projection matrices with a range of cutoff modes are calculated according to Algorithm 1. To rule out the effect of dimensionality alone on the data set, random projection matrices of the same size are also generated. This is done by taking the SVD of a matrix the same size as the A matrices, but populated with values drawn uniformly from -1 to 1 , and then removing the same number of dominant modes as is done with A . These matrices are applied to synthetically generated data to quantify their effect on the detectability of binary companion signals.

5.4. Synthetic Data Analysis

5.4.1. Synthetic Data Generation

FALCO is used to generate a library of off-axis PSFs corresponding to the dark-hole state, which can be injected as binary companions. These off-axis PSFs do not incorporate any WFE that is added on top of the dark-hole state. However, the effect of WFE on the off-axis signal is expected to be much smaller than its effect on the on-axis stellar signal, so not modeling the effects of WFE on the off-axis signal should have a negligible impact on the data.

The optical system is first initialized in the dark-hole state. Two noise models are considered: one in the linear regime, and one in the quadratic regime. Each data set thus consists of 20 instantaneous frames of independent noise realizations. For each frame, the spatial PSD given in Equation (17) is used to generate the WFE:

$$\text{PSD}(n_z) = an_z^b. \quad (17)$$

In this equation, n_z is the Noll-ordered index of the Zernike coefficient. The normalization parameter a is chosen to be 10 nm for the linear regime and 130 nm for the quadratic regime. The power-law exponent b is chosen to be -2 . These PSDs correspond to an average WFE (calculated over 100

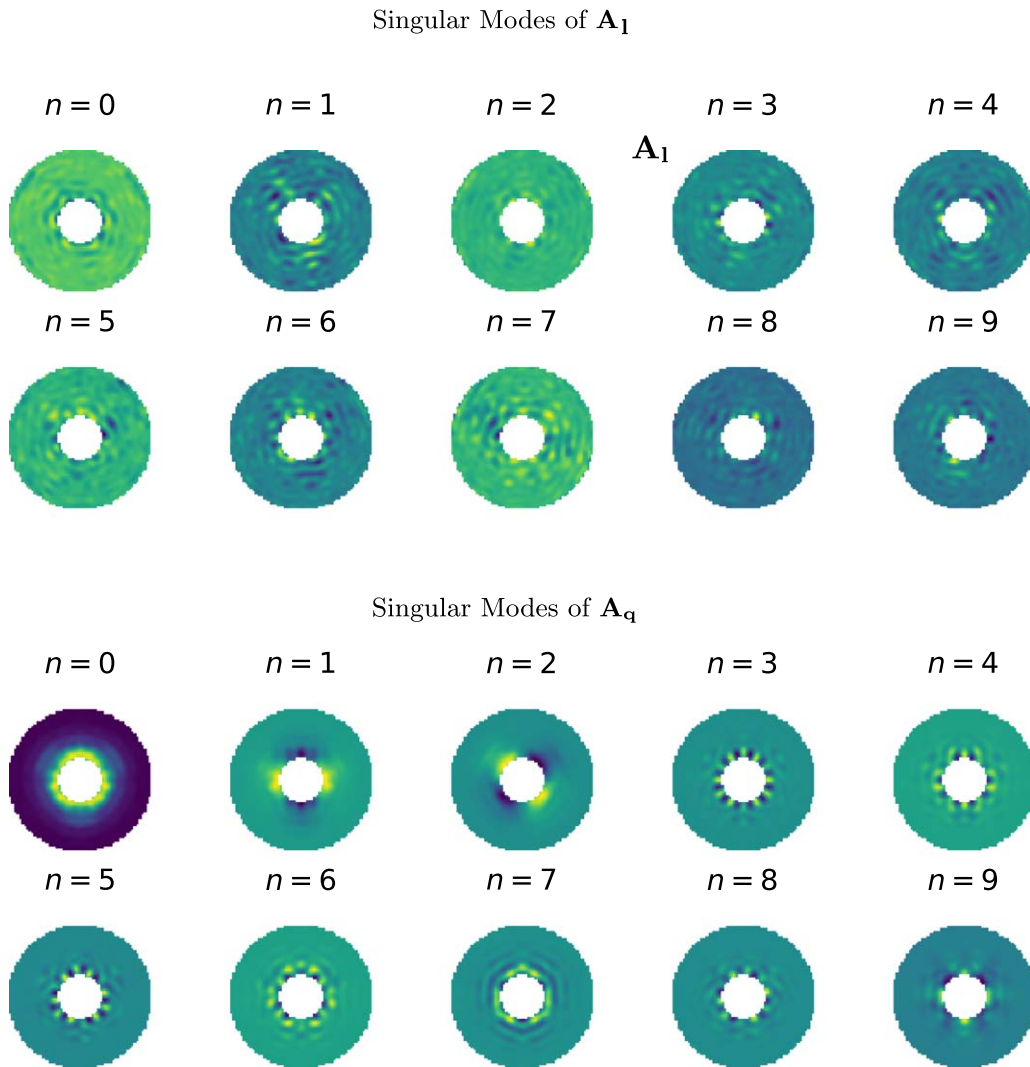


Figure 3. Top: the first 10 singular modes of A_1 as represented in the detector plane intensity basis (linear scale). Bottom: the first 10 singular modes of A_q as represented in the detector plane intensity basis (linear scale). The HLC design is nearly circularly symmetric, broken only by the six secondary mirror struts (which can also be seen in the Lyot stop). Because the quadratic transfer matrix depends only on the coronagraph operator C , its singular modes exhibit cosine- and sine-like azimuthal behavior associated with circularly symmetric operators. However, the linear transfer matrix depends on both C and the focal-plane electric field at the end of dark-hole creation, which is random and not circularly symmetric. Thus, its singular modes show no such symmetry structures. These singular modes correspond to the intensity patterns most likely to be attributed to WFE. Meanwhile, the companion's intensity pattern (the PSF at its location) overlaps very little with these dominant modes, so its signal is mostly retained when the dominant modes are projected out.

realizations) of about 7 nm (0.013 waves) rms for the linear regime data and about 110 nm (0.2 waves) rms for the quadratic regime data. As discussed in Section 2.1, the linear-quadratic transition occurs at approximately 0.1 waves rms. Although 110 nm rms of dynamic WFE is unrealistically high for the Roman HLC, we include this regime for demonstration purposes, as this level of WFE would be relevant on ground-based telescopes.

The resulting 528 Zernike coefficients are propagated through the LSCS to calculate the resulting dark-hole intensities. In order to create data with an injected companion planet, the off-axis PSF at the desired separation is scaled by the companion's flux ratio and then added to the dark-hole intensity. The separation of the injected companion is set to be $6.5\lambda/D$ in the linear case (which is the middle of the dark hole) and $4.0\lambda/D$ in the quadratic case (since the model is only valid within $\sim 5\lambda/D$). The position angles of both are set to be 0. Frames without the injected companion are used for the control case. Figure 4 shows example data frames: the initial dark hole, example frames with the aberrations from both

noise models applied, and the same frames with injected companion signals. The flux ratio of the companion is 2×10^{-7} for the frame with linear regime errors and 5×10^{-6} for the frame with quadratic regime errors. These flux ratios correspond to particularly bright planets chosen to be visible by eye.

It is worthwhile to examine how well the response matrices calculated in Section 5.2 can reconstruct the intensity errors present in the synthetic data. Figure 5 compares the intensity error resulting from WFE as calculated from the optical model with the intensity error calculated by multiplying the WFE by the appropriate response matrix, for example, frames in both the linear and quadratic regimes. In both regimes, the response matrices largely reproduce the spatial structure of the intensity error from the optical model.

5.4.2. Processing Synthetic Data

The quantity $|CE_{s0}|^2$ is the initial dark-hole intensity without any extra WFE applied (e.g., as determined from the data at the

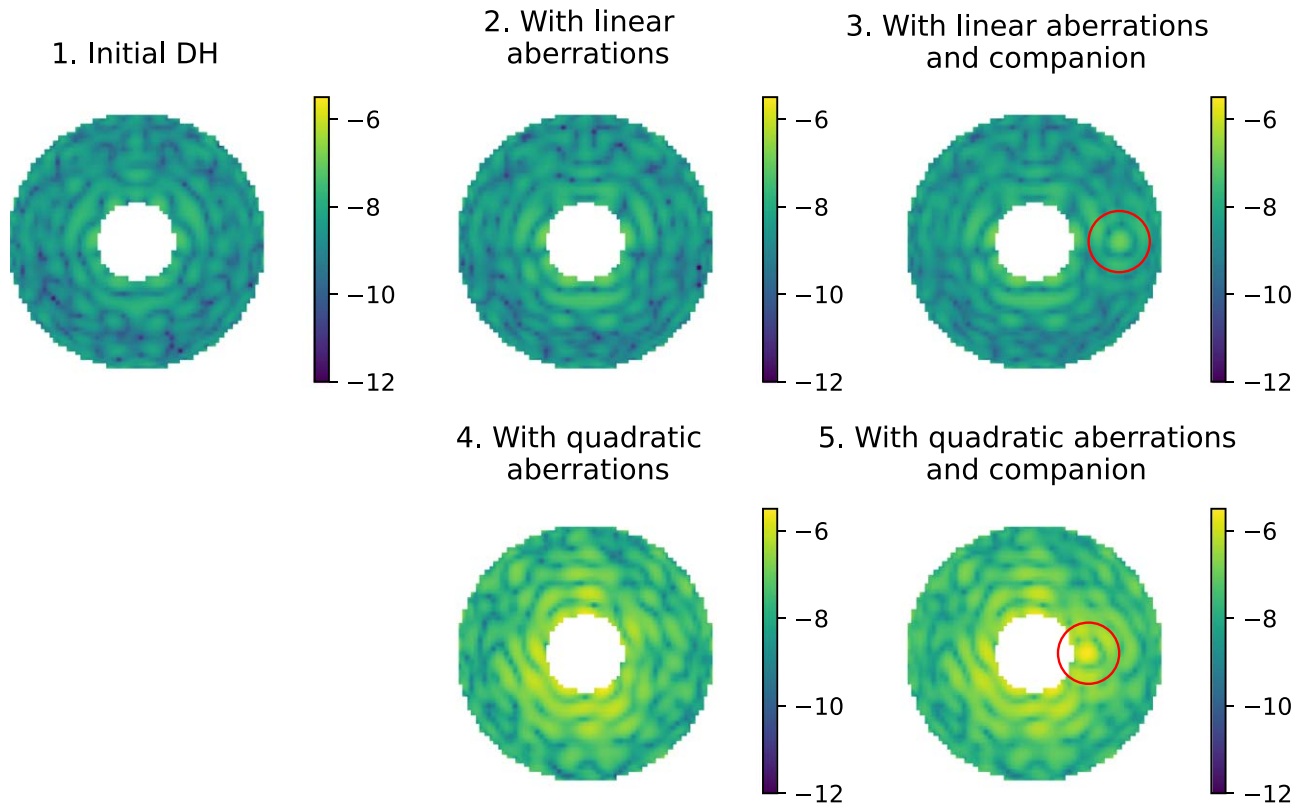


Figure 4. (1) Initial dark-hole intensity achieved using electric field conjugation with the HLC. (2) A single snapshot with linear regime wave front aberrations. (3) The same snapshot with an injected companion with a flux ratio of 2×10^{-7} at $6.5\lambda/D$ (indicated with a red circle). (4) A single snapshot with quadratic regime wave front aberrations. (5) The same snapshot, but with an injected companion with a flux ratio of 5×10^{-6} at $4\lambda/D$ (indicated with a red circle). All intensities are shown in \log_{10} of raw contrast.

end of the dark-hole digging sequence on the reference star). This nominal signal is first subtracted from each frame. Then, the pixels within the defined dark hole are gathered into the vector $\Delta \mathbf{I}_{\text{DH}}$. The data are left-multiplied by the appropriate \mathbf{K} matrix to obtain the observables $\mathbf{O} = \mathbf{K} \Delta \mathbf{I}_{\text{DH}}$. The data are also left-multiplied by the random matrix of the same size as \mathbf{K} to obtain data whose dimension has been reduced randomly. For each case, the average of the data over the 20 frames is used as the final measurement, while the standard deviation of the frames is used as the measurement uncertainty. Note that the process outlined does not rely on reference stars or dithering by deformable mirrors and can be used even on observations for which reference observations or wave front diversity is unavailable.

5.4.3. Flux Ratio Detection Limits

Detection tests are applied to these measured intensities and observables in order to characterize the detectability of a companion with these measurements. Detection limits are determined using the Monte Carlo method. One thousand random data sets are generated for each noise model with a given flux ratio. Each data set is processed as raw intensity data, and with each projection matrix with a different cutoff mode, and the $\Delta \chi_r^2$ values are calculated for each case. Figure 6 shows example histograms of the resulting $\Delta \chi_r^2$ values for a $c = 5.4 \times 10^{-7}$ at $4.0\lambda/D$ companion with the quadratic noise model, as well as the corresponding ROC curves, for the projection matrix with cutoff mode $N_m = 70$

(which, as shown in Figure 7, is the optimal cutoff at this spatial separation). The ROC curve shows that while using the robust observables results in an FPR = 0.01 and TPR = 0.9 detection of the injected companion, both the raw intensity and the randomly dimensionally reduced data remain very far from detectability.

This process is repeated for a range of flux ratios (to a precision of two significant figures). The resulting FPR = 0.01 and TPR = 0.9 detection limits for both regimes, as a function of cutoff mode N_m , are shown in Figure 7. Note that these flux ratio detection limits are not based on any statistical assumptions or extrapolations, but rather real FPRs and TPRs calculated by analyzing 1000 synthetically generated data sets, with injected companions of the given flux ratios and separations. The results show that with the linear regime noise model the robust observables increase the detectability of a companion at $6.5\lambda/D$ by 28%. With the quadratic regime noise model, using robust observables increases the detectability of a companion at $4.0\lambda/D$ by over a factor of two, and the improvement is not particularly sensitive to N_m beyond the first few modes. For the linear regime, this approach can also easily be extended to companions throughout the entire dark hole, though significant computation would be required to optimize N_m at all separations. For the quadratic regime, our model is only valid within $\sim 5\lambda/D$, though the Appendix discusses a method that can be used to extend the spatial coverage without incurring impractical computational costs.

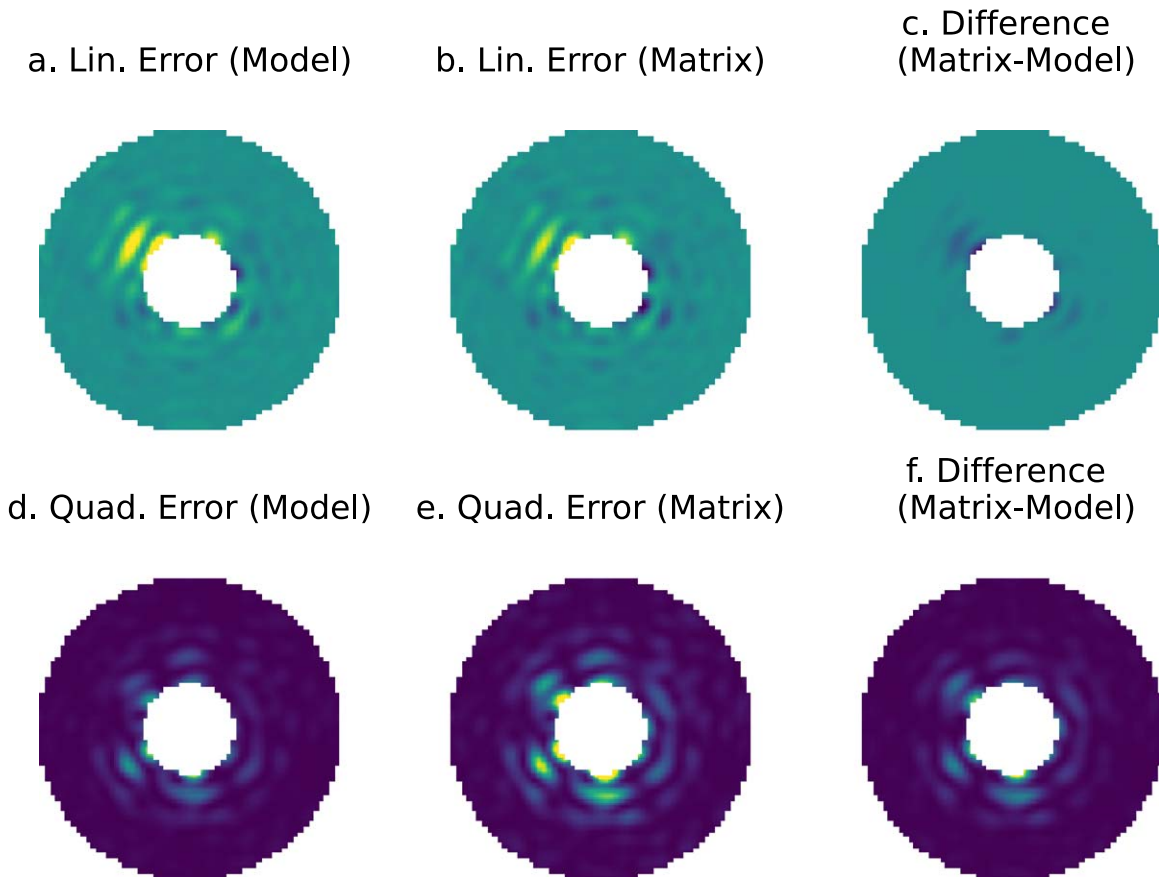


Figure 5. (a) Example linear regime intensity error from the optical model. (b) Corresponding linear regime intensity error reconstructed by response matrix A_l , plotted on the same scale as panel (a). (c) The difference between the response matrix prediction and the optical model prediction, plotted on the same scale as panels (a) and (b). (d) Example quadratic regime intensity error from the optical model. (e) Corresponding quadratic regime intensity error reconstructed by the response matrix A_q , plotted on the same scale as panel (d). (f) The difference between the response matrix prediction and the optical model prediction, plotted on the same scale as panels (d) and (e). Slight differences arise because the model includes both the linear and quadratic error terms while the matrix predictions only include one or the other, i.e., the linear matrix prediction neglects the contribution of the quadratic term and the quadratic matrix prediction neglects the contribution of the linear term (as well as the influence of any Zernikes past the first 100). While the linear matrix prediction is biased low near the peaks and the quadratic matrix prediction is biased high overall, our method depends only on how well the *spatial structure* of the errors is reproduced. A relevant metric for characterizing the spatial overlap is the normalized inner product between the optical model prediction and the transfer matrix prediction, where a value of 1 indicates perfect spatial overlap and a value of 0 indicates perfect spatial orthogonality. In this case, the normalized inner product is 0.936 for the linear regime example and 0.985 for the quadratic regime example, sufficient for providing a quantifiable improvement in detection sensitivity.

6. Discussion

6.1. Temporally Correlated WFE and Compatibility with Other Post-processing Techniques

This work aims to characterize the effect of using robust observables in isolation. Thus, only noise models in which the WFE is uncorrelated in time are examined, since additional post-processing techniques are typically used to handle time-correlated data. Robust observables are compatible with these other post-processing techniques and can serve as an instrument-motivated prior in the overall post-processing strategy. For example, random errors can first be reduced by projecting the data into a subspace that is robust to WFE. Then, reference observations, along with PCA-based methods such as KLIP (Soummer et al. 2012), can be used to calibrate static and quasi-static errors and decorrelate the frames in time. This is similar to the calibration approach used in nonredundant aperture masking interferometry or kernel-phase interferometry, in which data are projected onto closure phases or kernel phases, respectively, which are then calibrated based on reference observations (Martinache 2010; Ireland 2013; Pope et al. 2021). A more sophisticated approach would be to

formulate post-processing as a statistical inference problem, where a least-squares fit with the reference frames makes up one term in the cost function and a prior over the instrumental modes (e.g., weighted by the singular value spectrum) makes up another term.

Ygouf et al. (2016) show that for the time-varying WFE expected on the Roman Space Telescope HLC, classical PSF subtraction with a reference observation increases the contrast gain by a factor of a few to about 10, depending on the scenario. Future work includes investigating how much overall post-processing gain can be achieved when robust observables and calibration strategies are combined, as well as which hybrid strategies maximize the sensitivity that can be obtained with all available information.

6.2. PSD Engineering

The robust observables derived in this work are agnostic to the actual temporal or spatial PSD of the static and dynamical WFEs and are intended to be applied when these PSDs are not well-known or imperfectly characterized. As of today, this is the case for all ground-based instruments (as predictions of the

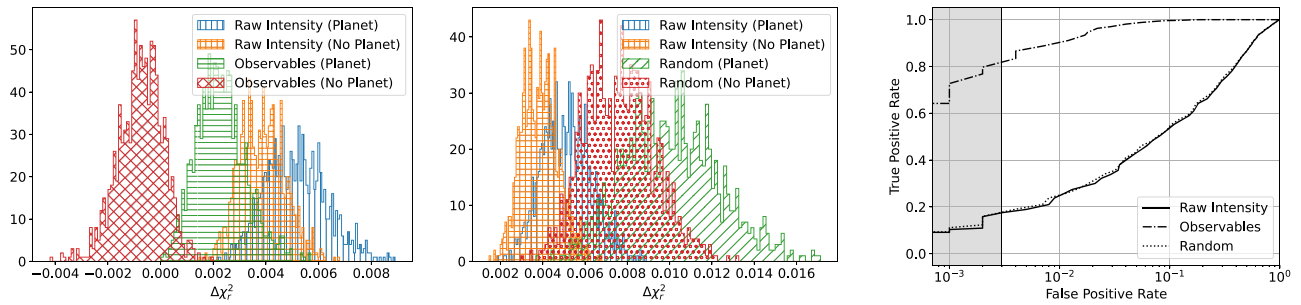


Figure 6. Detection test results for the quadratic regime noise model. The companion planet considered has a flux ratio of 5.4×10^{-7} and is located at $4.0\lambda/D$. Left: histograms from using raw intensities compared to those from using quadratic robust observables with the optimal cutoff of $N_m = 70$. The histograms using raw intensity overlap significantly, making it difficult to distinguish between a model with a planet and a model without one, while the histograms using the robust observables are further separated and more distinguishable. Middle: histograms for using raw intensities and a random projection matrix of the same size as the instrumentally motivated projection. Both sets of histograms overlap significantly, and the random projection does not improve the distinguishability of the two models. Right: ROC curves corresponding to the histograms. The gray area indicates FPRs that are not well sampled, as they involve fewer than three data sets with false detections. The ROC curve shows that while using the robust observables results in an FPR = 0.01 and TPR = 0.9 detection of the injected planet, both the raw intensity and the randomly dimensionally reduced data remain very far from detectability.

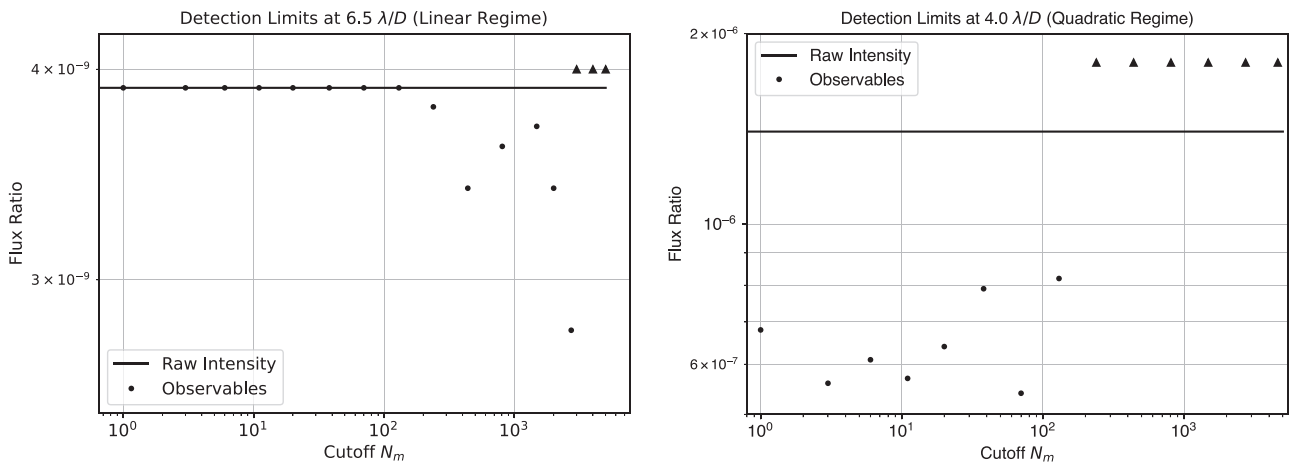


Figure 7. Flux ratio detection limits (FPR = 0.01, TPR = 0.90) for a binary companion (to two significant figures) as a function of cutoff mode. Upward-pointing triangles indicate that a projection matrix with the specified cutoff mode performs worse than using the raw intensity, which occurs when the modes the majority of the planet signal overlaps with have also been projected out. Left: linear regime with a companion at $6.5\lambda/D$. The optimal cutoff mode is 2727, which results in a detection limit of 2.8×10^{-9} . Unshowable in log–log scale is the detection limit with $N_m = 0$, which, with observables, is 3.9×10^{-9} . This is, as expected from the fact that no error modes are removed, equal to the raw intensity detection limit. Right: quadratic regime with a companion at $4.0\lambda/D$. The optimal cutoff mode is 70, which results in a detection limit of 5.4×10^{-7} . Unshowable in log–log scale is the detection limit with $N_m = 0$, which, with observables, is 1.4×10^{-6} . This is, as expected from the fact that no error modes are removed, equal to the raw intensity detection limit.

influence of the atmosphere are quite imperfect) and space-based missions (as Hubble Space Telescope and JWST observatory level key metrics for requirements are expressed in terms of encircled energy, not contrast). However, it has been proposed that for future space telescope coronagraphs the telescope WFE PSD must comply with stringent requirements in order to facilitate exoplanet detection (Nemati et al. 2020).

For instance, the PASTIS approach (Lebouilleux et al. 2018; Lajinja et al. 2019) considers the effects of the quadratic response on the *average* intensity contrast over the entire dark hole (or region of interest), calculating which modes the coronagraph is most sensitive to in order to determine stability tolerances for the segments accordingly. Calculating robust observables for post-processing is akin to doing PASTIS backward, where the modes the coronagraph is most sensitive to are calculated in order to project them out of the data. For such telescopes that have PSDs engineered based on the instrument response, the additional gain from using robust observables will depend on how well the error modes are suppressed in hardware, as well as the timescales at which

power in those modes leaks through. To some extent, robust observables will remain applicable to such future telescopes and instruments in the spatial and temporal subspaces in which they do not meet their requirements.

6.3. Model Accuracy

In this analysis, the model used to generate the instrument response matrices is exactly the same model that is used to generate the synthetic data. In real observations, the instrument model will not exactly match the behavior of the actual instrument, and one future avenue to explore is how well a model must match the instrument in order for robust observables to work on real data. This technique’s robustness can be investigated by first calculating the response matrices using one model and then changing the parameters of the model (e.g., the coronagraphic mask size and displacement, the DM alignment, the detector pixel scale) before generating synthetic data and examining how well the robust observables work in the presence of model mismatch.

For instruments equipped with wave front modulating devices such as deformable mirrors, however, the instrument response matrix may also be calculated experimentally. If a perturbation within the linear regime is applied, the difference in measured intensity can be directly registered into the appropriate column of the linear response matrix. The technique for experimentally building the quadratic response matrices is equivalent to the approach used for PASTIS (Laginja et al. 2019), with the difference that the measurements are not averaged over a dark hole, but rather maintained for every pixel. Additionally, some wave front and control schemes, such as implicit electric field conjugation (Haffert et al. 2023), already involve an empirical measurement of the instrument response, which can be used to derive linear regime robust observables without having to set aside additional calibration time. Experimentally building instrument response matrices circumvents the need to have a well-matched numerical model and allows for the response matrices to capture effects in the real instrument.

7. Conclusions

A coronagraph model with linear and quadratic contributions of WFE to detector plane intensity is developed, and when either term is dominant, the coronagraph response can be approximated by a transfer matrix. A useful projection can be found from this transfer matrix that removes the dominant error modes, resulting in observables that are more robust to WFE in the regime of interest. These robust observables are extracted from synthetically generated data with the Hybrid Lyot Coronagraph of the Roman Space Telescope in both the linear and quadratic regimes. The performance of the robust observables is compared to that of the raw intensity data through calculations of their respective binary companion flux ratio detection limits. In these examples, using the robust observables significantly increases the sensitivity to the signal of a binary companion. A projection onto a robust subspace can in theory be combined with other families of post-processing algorithms. Hybrid post-processing approaches would incorporate information on the instrument response alongside the other available information (such as angular diversity, spectral diversity, reference observations, or WFC telemetry) to fully maximize the sensitivity to astrophysical signals in coronagraphic data; however, the approach outlined in this work can be applied to observational data and result in post-processing gains even if such additional information is unavailable.

Acknowledgments

We thank the anonymous reviewers for the feedback on improving this manuscript. We thank Frantz Martinache, Mamadou N’Diaye, and Alban Ceau for very helpful discussions on this topic. We also thank Dimitri Mawet for additional perspectives.

This work is supported by the National Science Foundation Graduate Research Fellowship under grant No. 1122374 and the WFIRST Science Investigation team prime award under grant No. NNG16PJ24C

This research made use of NASA’s Astrophysics Data System.

B.J.S.P. would like to acknowledge the traditional owners of the land on which the University of Queensland is situated, the Turrbal and Jagera people. We pay respect to their Ancestors

and Descendants, who continue cultural and spiritual connections to Country.

Software: This research made use of FALCO, the Fast Linearized Coronagraph Optimizer (Riggs et al. 2018); the Lightweight Space Coronagraph Simulator (<https://github.com/leonidprinceton/LightweightSpaceCoronagraphSimulator>); Astropy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018); NumPy (Harris et al. 2020); SciPy (Virtanen et al. 2020); and Matplotlib (Hunter 2007).

Appendix Quadratic Model Approximation and Extension

As explained in Section 3, the calculation of the projection matrix involves an SVD of the response matrix, but a quadratic response matrix that includes all 528 Zernikes needed to span the dark hole would have a size of $5476 \times 139,656$. Since calculating the SVD of a matrix of this size is too computationally burdensome, we explore an approximation of the quadratic response that models only the impact of norm squared of each input basis vector while neglecting the effects of the pairwise combinations. Namely, we use an approximate response matrix A'_q with elements

$$A'_{q_{kj}} = C_{kj}^* C_{kj}. \quad (\text{A1})$$

The index j labels the input basis vector, and the index k labels the detector pixel. The size of A'_q scales linearly with the number of Zernike models, and in our case it would be of size 5476×528 , which is easily decomposable.

Note that A'_q cannot be used to accurately reproduce quadratic regime intensity error. However, A'_q is nevertheless useful for identifying a subspace robust to quadratic regime WFEs, leading to increased S/N. We can observe this by comparing the detection test results with and without using the approximation for a model with 100 Zernikes. We calculate the approximation A'_q using Equation (A1) and use the original A_q from Section 5.3. Detection tests on quadratic regime synthetic data similar to the one from Section 5.4 are performed, using projection matrices derived from both A_q and A'_q . The resulting flux ratio detection limits as a function of cutoff mode are shown in Figure A1.

The full matrix achieves the best results with a cutoff mode of 70, leading to a detection limit of 5.4×10^{-7} , while the approximate matrix achieves the best results with a cutoff mode of 2727, also leading to a detection limit of 5.4×10^{-7} . These results show that the approximation performs as well as the full model.

To understand why this is the case, we analyze the subspaces spanned by the identified optimal projection matrices. We define \mathbf{P} as the projection onto the dominant modes of A_q , \mathbf{P}' as the projection onto the dominant modes of A'_q , and \mathbf{P}_r as a random projection matrix the same shape as \mathbf{P}' . We also define \mathbf{K} as the projection onto the remaining modes (the robust subspace) of the full model, \mathbf{K}' as the projection onto the robust subspace of the approximate model, and \mathbf{K}_r as a random projection matrix the same shape as \mathbf{K}' . We then calculate the subspace angles (Jordan 1875) between each of these projection matrices and \mathbf{P} using the function `scipy.linalg.subspace_angles`. These subspace angles provide an indication of how much the subspace spanned by each of these projection matrices overlaps with the subspace spanned by the dominant

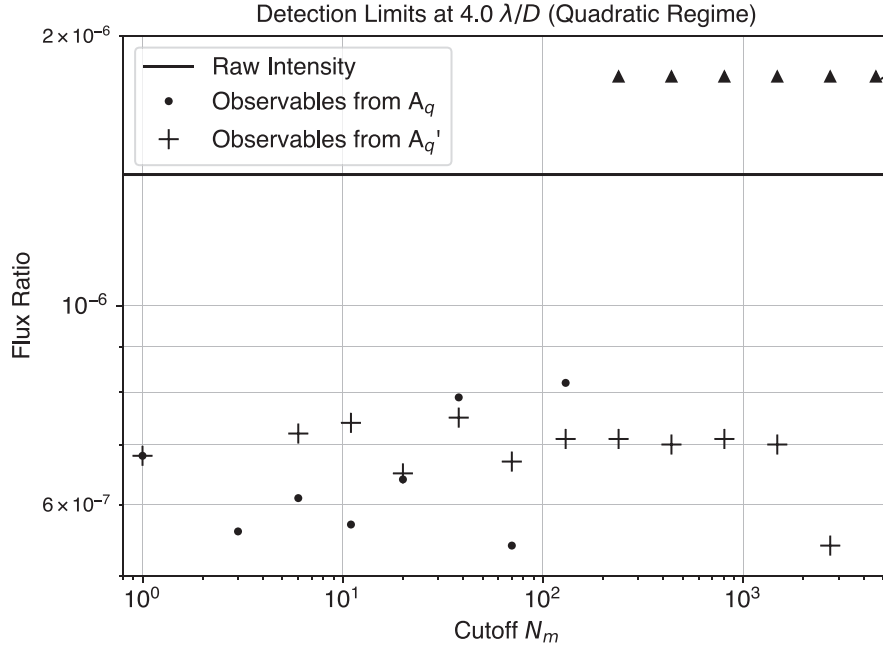


Figure A1. Quadratic regime flux ratio detection limits (FPR = 0.01, TPR = 0.90) to two significant figures, as a function of cutoff mode, for a companion at $4.0\lambda/D$. Only the first 100 Zernikes are used in the model used to calculate the full and approximate quadratic transfer matrices, but WFEs up to 538 Zernikes are included in the synthetic data. Upward-pointing triangles or spikes indicate that a projection matrix with the specified cutoff mode performs worse than using the raw intensity, which occurs when the modes the majority of the planet signal overlaps with have also been projected out. The full matrix achieves the best results with a cutoff mode of 70, leading to a detection limit of 5.4×10^{-7} while the approximate matrix achieves the best results with a cutoff mode of 2727, also leading to a detection limit of 5.4×10^{-7} . Unshowable in log–log scale is the detection limit with $N_m = 0$, which, with observables, is 1.4×10^{-6} . This is, as expected from the fact that no error modes are removed, the same as the raw intensity detection limit of 1.4×10^{-6} . These results indicate that the approximation performs as well as the full model.

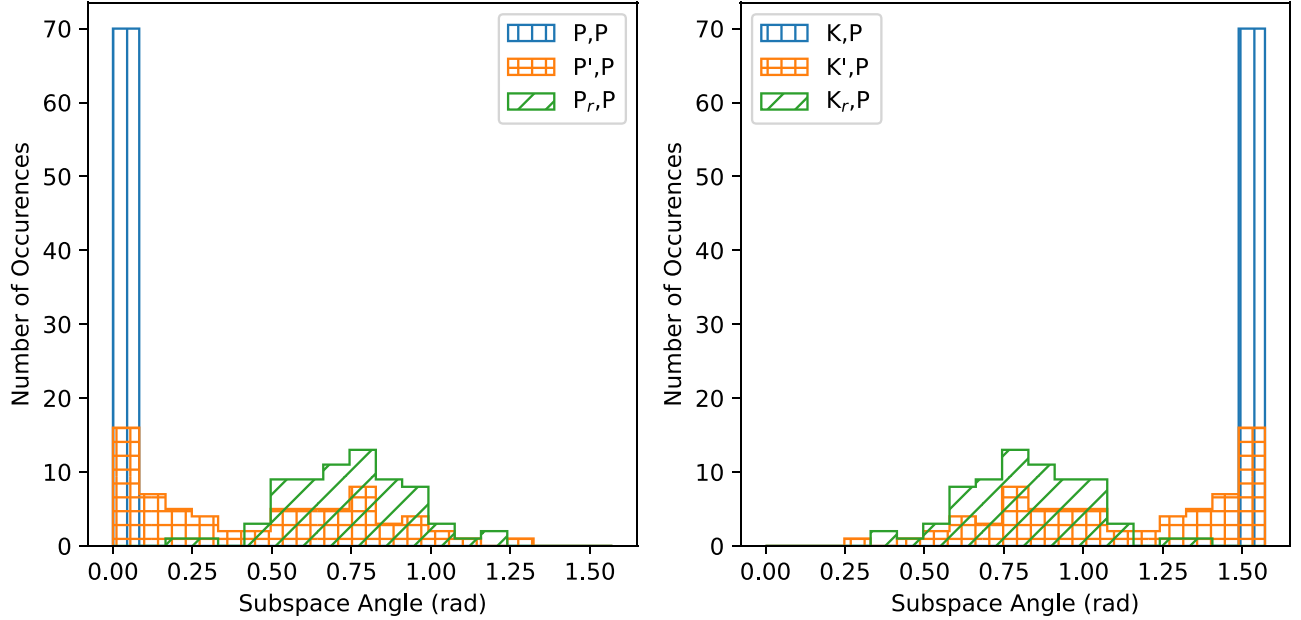


Figure A2. The subspace angles between various projection matrices (onto dominant modes on the right, onto a robust subspace on the left) and \mathbf{P} , the projection onto the dominant error modes determined from the full quadratic model. The number of principle angles with value 0 is the dimension of overlap between the subspaces. Angles with value $\pi/2$ indicate overlap with the subspace orthogonal to \mathbf{P} . \mathbf{P}' (the space of dominant modes derived from the approximate model) overlaps with \mathbf{P} (the space of dominant modes derived from the full model) significantly more than random. Crucially, \mathbf{K}' (the robust subspace from the approximate model) overlaps with the subspace orthogonal to \mathbf{P} significantly more than random, which is why data projected onto this subspace is still robust to WFE.

modes identified by the full model. The results are shown in Figure A2.

The number of principle angles with value 0 is the dimension of overlap between the subspaces. As expected, the subspace angles between \mathbf{P} and itself are all 0, meaning that it overlaps completely with itself. Also as expected, the angles between \mathbf{K}

and \mathbf{P} are all $\pi/2$, as \mathbf{K} is orthogonal to \mathbf{P} . Both of the random matrices have a random distribution of angles with \mathbf{P} centered around $\pi/4$. Meanwhile, \mathbf{P}' (the space of dominant modes derived from the approximate model) overlaps with \mathbf{P} (the space of dominant modes derived from the full model) significantly more than random. Crucially, \mathbf{K}' (the robust

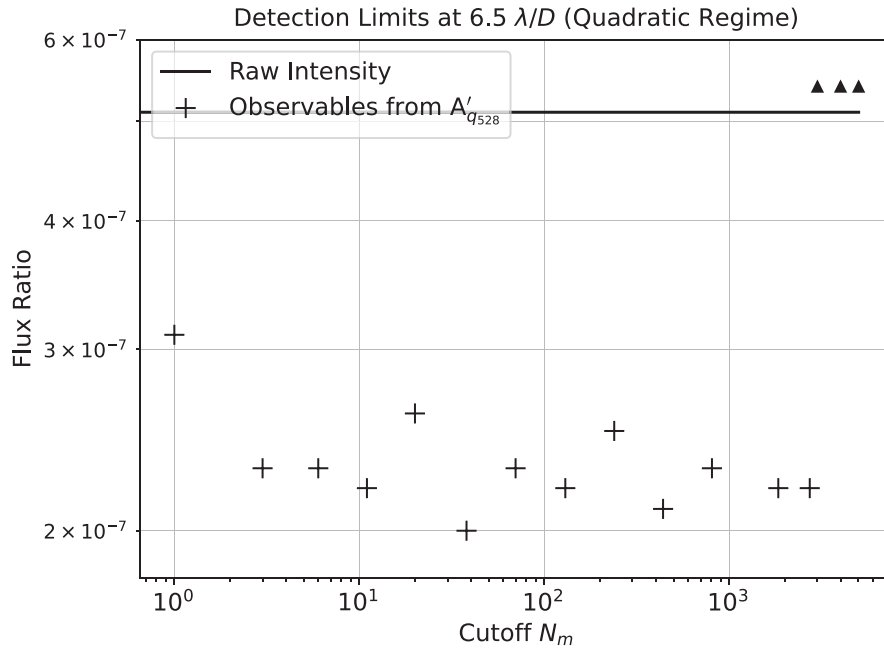


Figure A3. Quadratic regime flux ratio detection limits (FPR = 0.01, TPR = 0.90) to two significant figures, as a function of cutoff mode, for a companion at $6.5\lambda/D$. Upward-pointing triangles indicate that a projection matrix with the specified cutoff mode performs worse than using the raw intensity, which occurs when the modes the majority of the planet signal overlaps with have also been projected out. The optimal cutoff mode is 38, which results in a detection limit of 2.0×10^{-7} . Unshowable in log–log scale is the detection limit with $N_m = 0$, which, with observables, is 4.9×10^{-7} . This is, as expected from the fact that no error modes are removed, close to the raw intensity detection limit of 5.1×10^{-7} .

subspace from the approximate model) overlaps with \mathbf{P} significantly less than random, and with the subspace orthogonal to \mathbf{P} significantly more than random, which is why data projected onto this subspace are still robust to WFE. This result shows why the approximate model, despite poorly predicting the detector intensity response, is nevertheless useful for identifying a subspace that overlaps significantly with the robust subspace of the full model.

We can thus use this approximation with all 528 Zernikes in our model to analyze spatial separations beyond the $\sim 5\lambda/D$ spanned by the first 100 Zernikes. To demonstrate this, we build $A'_{q_{528}}$ according to Equation (A1) and perform detection tests at a separation of $6.5\lambda/D$. The results are shown in Figure A3.

Our tests show that the approximation $A'_{q_{528}}$ can successfully increase S/N at spatial separations beyond the original regime of validity of A_q . Thus, even though the input dimension of the quadratic model scales clumsily with the number of basis vectors, an approximation considering only norm-squared terms can still be used to find observables that are robust to quadratic WFE and thus provide detection gains at farther spatial separations of interest.

ORCID iDs

Yinzi Xin <https://orcid.org/0000-0002-6171-9081>
 Laurent Pueyo <https://orcid.org/0000-0003-3818-408X>
 Romain Laugier <https://orcid.org/0000-0002-2215-9413>
 Leonid Pogorelyuk <https://orcid.org/0000-0001-6387-9444>
 Ewan S. Douglas <https://orcid.org/0000-0002-0813-4308>
 Benjamin J. S. Pope <https://orcid.org/0000-0003-2595-9114>
 Kerri L. Cahoy <https://orcid.org/0000-0002-7791-5124>

References

- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Baudoz, P., Boccaletti, A., Baudrand, J., & Rouan, D. 2006, in IAU Colloq. 200: Direct Imaging of Exoplanets: Science & Techniques, ed. C. Aime & F. Vakili (Cambridge: Cambridge Univ. Press), 553
- Bloemhof, E. E. 2003, *ApJL*, **582**, L59
- Cantalloube, F., Christiaens, V., Cantero, C., et al. 2022, *Proc. SPIE*, **12185**, 1218505
- Ceau, A., Mary, D., Greenbaum, A., et al. 2019, *A&A*, **630**, A120
- Flasseur, O., Denis, L., Thiébaud, E., & Langlois, M. 2018, *A&A*, **618**, A138
- Groff, T. D., Riggs, A. J. E., Kern, B., & Kasdin, N. J. 2016, *JATIS*, **2**, 1
- Guyon, O., Pluzhnik, E. A., Kuchner, M. J., Collins, B., & Ridgway, S. T. 2006, *ApJS*, **167**, 81
- Haffert, S. Y., Males, J. R., Ahn, K., et al. 2023, *A&A*, **673**, A28
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Ireland, M. J. 2013, *MNRAS*, **433**, 1718
- Jensen-Clem, R., Mawet, D., Gonzalez, C. A. G., et al. 2017, *AJ*, **155**, 19
- Jordan, C. 1875, *Bulletin de la Société Mathématique de France*, **3**, 103
- Kasdin, N. J., Bailey, V. P., Mennesson, B., et al. 2020, *Proc. SPIE*, **11443**, 114431U
- Kasdin, N. J., & Braems, I. 2006, *ApJ*, **646**, 1260
- Krist, J. 2020, Roman CGI OS9 Time Series Simulations (Hybrid Lyot Coronagraph, Band 1), https://roman.ipac.caltech.edu/sims/Coronagraph_public_images.html#CGI_OS9
- Krist, J. E., Steeves, J. B., Dube, B. D., et al. 2023, *JATIS*, **9**, 045002
- Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É. 2007, *ApJ*, **660**, 770
- Laginja, I., Lebouilleux, L., Pueyo, L., et al. 2019, *Proc. SPIE*, **11117**, 1111717
- Lebouilleux, L., Sauvage, J.-F., Pueyo, L., et al. 2018, *JATIS*, **4**, 035002
- Marois, C., Lafreniere, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, *ApJ*, **641**, 556
- Martinache, F. 2010, *ApJ*, **724**, 464
- Nemati, B., Stahl, H. P., Stahl, M. T., Ruane, G. J. J., & Sheldon, L. J. 2020, *JATIS*, **6**, 039002
- Noll, R. J. 1976, *JOSAA*, **66**, 207
- Perrin, M. D., Sivaramakrishnan, A., Makidon, R. B., Oppenheimer, B. R., & Graham, J. R. 2003, *ApJ*, **596**, 702

- Pogorelyuk, L., Kasdin, N. J., & Rowley, C. W. 2019, [ApJ](#), **881**, 126
- Pope, B. J. S., Pueyo, L., Xin, Y., & Tuthill, P. G. 2021, [ApJ](#), **907**, 40
- Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, [AJ](#), **156**, 123
- Pueyo, L. 2016, [AJ](#), **824**, 117
- Riggs, A. J. E., Ruane, G., Sidick, E., et al. 2018, [Proc. SPIE](#), **10698**, 106982V
- Soummer, R., Pueyo, L., & Larkin, J. 2012, [ApJL](#), **755**, L28
- Sparks, W. B., & Ford, H. C. 2002, [ApJ](#), **578**, 543
- Traub, W. A., & Oppenheimer, B. R. 2010, in *Exoplanets*, ed. S. Seager (Tucson, AZ: Univ. Arizona Press), 111
- Tyson, R. K. 2000, in *Introduction to Adaptive Optics*, ed. A. R. Weeks, Jr. (Washington, DC: SPIE)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, [NatMe](#), **17**, 261
- Vogt, F. P. A., Martinache, F., Guyon, O., et al. 2011, [PASP](#), **123**, 1434
- Ygouf, M., Zimmerman, N. T., Pueyo, L., et al. 2016, [Proc. SPIE](#), **9904**, 99045M