Notes on Optical Feedback Controller

Guillem Megias Homar

2024-06-06

Motivation

The LSST active optics system (AOS) uses wavefront information to correct for errors in the rigid body displacement and rotation of M2 and the camera relative to M1M3 and for bending deformations of both M2 and M1M3; these degrees of freedom will collectively be referred to the optical state of the telescope. A look-up table (LUT) will correct for most of the gravity-induced motion; only the residual due to thermal variations, dome seeing or errors in the gravity look-up table needs to be corrected by AOS. Wavefront information is available from four field points and used to correct motion every 30 seconds.

To determine how to adjust the optical state, so that the wavefront error is minimized, we use the wavefront information to estimate the optical state of the system and derive from there what are the corrections, in the basis of the degrees of freedom, that need to be applied.

Given that we are able to set the degrees of freedom values, one may wonder why we need to estimate them from the wavefront sensing if we have control over them. Thermal variations, gravitational errors, and dome-seeing affect the optical aberrations of the telescope and, therefore, the optical state. The fact that we express the optical state on the basis of the degrees of freedom doesn't mean we are estimating the state of the degrees of freedom at the actuator-level, which we know exactly, but rather the optical state of the system which is due to the actuator-level motions and the optical aberration effects.

Preliminaires

The vector of wavefront measurements $\boldsymbol{y} \in \mathbb{R}^n$ and the optical state vector $\boldsymbol{x} \in \mathbb{R}^m$ are related through the sensitivity matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ as $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$.

We will consider the vector of wavefront measurements $\boldsymbol{y}$ to be a 19-dimensional vector with Zernike coefficients from Z_4 to Z_{22}. However, we could consider the $\boldsymbol{y}$ to be the 19-Zernike vector at different Gaussian quadrature points, in which case $\boldsymbol{y} \in \mathbb{R}^{p\times n}$ where p are the number of quadrature points. If that were the case, then the sensitivity matrix would be $\mathbf{A} \in \mathbb{R}^{p\times n\times m}$.

We could unravel the vectors and end up with an equivalent description to the original one. Therefore, without loss of generality, we will keep the dimensions of $\boldsymbol{y}$ to be n for this summary.

The optical state vector $\boldsymbol{x}$ is a 50-dimensional vetor corresponding to the 50 degrees of freedom of our system, which include by order, the rigid-body motion hexapod degrees of freedom for M2 (dz_{M2}, dx_{M2}, dy_{M2}, Rx_{M2}, Ry_{M2}) and for the Camera hexapod (dz_{Cam}, dx_{Cam}, dy_{Cam}, Rx_{Cam}, Ry_{Cam}), the 20 bending modes of M1M3, and the 20 bending modes of M2.

We will denote the vector of corrections that we apply to the system as $\boldsymbol{u} \in \mathbb{R}^n$. These corrections are also on the basis of the degrees of freedom.

Finally, we will assume the following description of the optical state at time-step $k+1$ holds,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{u} + \boldsymbol{\delta}_k \tag{1}$$

where $\boldsymbol{\delta}_k$ is the average deviation introduced between time-step due to uncontrolled effects.

Optical state estimation

The simplest way to estimate the optical state from the wavefront error measurements –which corresponds to the way we currently do it in `ts_ofc`– is using the Moore-Penrose pseudoinverse ($\mathbf{A}^+$), which is derived as follows,

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{y}$$

$$\mathbf{A}^\top\mathbf{A}\boldsymbol{x} = \mathbf{A}^\top\boldsymbol{y}$$

$$\hat{\boldsymbol{x}} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top\boldsymbol{y} = \mathbf{A}^+\boldsymbol{y}$$

Notably, the sensitivity matrix is a highly-degenerate matrix which has a set of singular vectors with very small singular values. In Figure 1 we plot the singular values for the Sensitivity matrix calculated on 35 Gaussian Quadrature points in the corner wavefront sensors.

The importance of the degeneracy on our optical feedback control comes into play at multiple levels. The first one is the pseudo-inverse. In fact, instead of doing this matrix multiplication above, the pseudo-inverse $\mathbf{A}^+$ is computed *de facto* using the Singular Value Decomposition (SVD) of $\mathbf{A}$.

$$\hat{\boldsymbol{x}} = \mathbf{A}^+\boldsymbol{y} = \mathbf{V}\Sigma^{-1}\mathbf{U}^\top\boldsymbol{y} \tag{2}$$

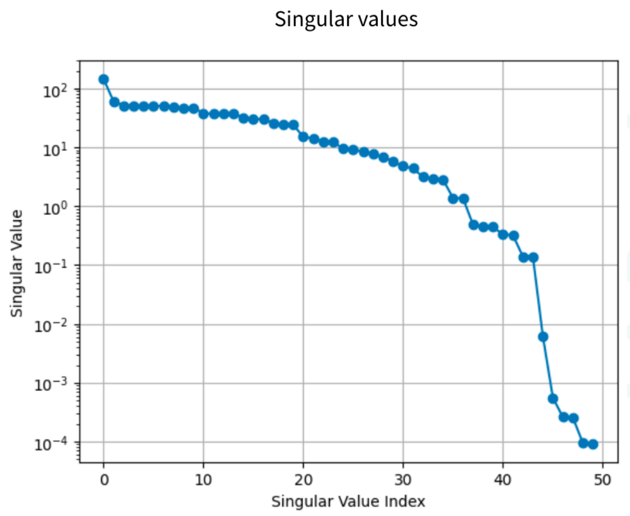

FIGURE 1: Singular values derived from the original sensitivity matrix at the 35 Gaussian Quadrature points of the four wavefront sensors.

The singular vectors from the sensitivity matrix, therefore play a crucial role in our state estimate. Let us rewrite the state estimate based on the real optical state $\boldsymbol{x}$,

$$\hat{\boldsymbol{x}} = \mathbf{A}^{+}\boldsymbol{y} = \mathbf{A}^{+}\mathbf{A}\boldsymbol{x} = \mathbf{V}\Sigma^{-1}\mathbf{U}^{\top}\mathbf{U}\Sigma\mathbf{V}^{\top}\boldsymbol{x} = \mathbf{V}\mathbf{V}^{\top}\boldsymbol{x} \tag{3}$$

Since $\mathbf{A}$ is nearly-degenerate, it doesn't have full column rank, and therefore $\mathbf{V}\mathbf{V}^{\top} \neq \mathbb{I}$. In this case, we can interpret the pseudo-inverse estimate $\hat{\boldsymbol{x}}$ as the orthogonal projection of $\boldsymbol{x}$ onto the row space of $\mathbf{A}$.

Noisy measurements

Let us now depart from the ideal world of physics and delve into reality. In this case, the problem formulation can be rewritten introducing noise to our measurements as,

$$\boldsymbol{y}_{noisy} = \mathbf{A}\boldsymbol{x} + \boldsymbol{w} \tag{4}$$

The first step is to study what the error of our previous pseudo-inverse estimate will look like, that is, what is our *reconstruction* error. The noisy estimate, in this case, would be,

$$\hat{\boldsymbol{x}}_{noisy} = \mathbf{A}^{+}\mathbf{A}\boldsymbol{x} + \mathbf{A}^{+}\boldsymbol{w} \tag{5}$$

So the reconstruction error would be,

$$\|\hat{\boldsymbol{x}}_{noisy} - \hat{\boldsymbol{x}}\|_2^2 = \|\mathbf{A}^{+}\boldsymbol{w}\|_2^2 = \|\mathbf{V}\Sigma^{-1}\mathbf{U}^{\top}\boldsymbol{w}\|_2^2 \tag{6}$$

Where we used the SVD of **A**. Now, let us show that for any vector $\boldsymbol{z} \in \mathbb{R}^n$,

$$\|\mathbf{V}\boldsymbol{z}\|_2^2 = \boldsymbol{z}^{\top}\mathbf{V}^{\top}\mathbf{V}\boldsymbol{z} = \boldsymbol{z}^{\top}\boldsymbol{z} = \|\boldsymbol{z}\|_2^2 \tag{7}$$

where we used the fact that the right-singular vector matrix derived from SVD is such that, $\mathbf{V}^{\top}\mathbf{V} = \mathbb{I}$

Now our reconstruction error can be rewritten as,

$$\|\hat{\boldsymbol{x}}_{noisy} - \hat{\boldsymbol{x}}\|_2^2 = \|\Sigma^{-1}\mathbf{U}^{\top}\boldsymbol{w}\|_2^2 \tag{8}$$

At this point, let us suppose for just a minute that the noise is normalized $\|\boldsymbol{w}\|_2^2 = 1$. Since **U** is orthogonal, we see that,

$$\|\mathbf{U}^{\top}\boldsymbol{w}\|_2^2 \leq \|\boldsymbol{w}\|_2^2 \tag{9}$$

To find the maximum reconstruction error, using the above we see that,

$$\max_{\boldsymbol{w}\in\mathbb{R}^m}\|\Sigma^{-1}\mathbf{U}^{\top}\boldsymbol{w}\|_2^2 = \max_{\boldsymbol{w}\in\mathbb{R}^m}\|\Sigma^{-1}\boldsymbol{w}\|_2^2 \tag{10}$$

Finally, we can see that the worst-case error, in this case, will have an entry equal to 1 in the largest singular value of Σ^{-1}, and the rest will be zero. In this case, we have that,

$$\max_{\boldsymbol{w}\in\mathbb{R}^m}\|\Sigma^{-1}\mathbf{U}^{\top}\boldsymbol{w}\|_2^2 = \max_{\sigma_i}\sigma_i^{-2} = \frac{1}{\sigma_m^2} \tag{11}$$

where σ_m^2 corresponds to the smallest singular value.

If we now remove the assumption of having a normalized error, we can rewrite our reconstruction error as,

$$\|\hat{\boldsymbol{x}}_{noisy} - \hat{\boldsymbol{x}}\|_2^2 = \|\Sigma^{-1}\mathbf{U}^{\top}\boldsymbol{w}\|_2^2 \leq \frac{1}{\sigma_m^2}\|\boldsymbol{w}\|_2^2 \tag{12}$$

If the smallest singular value is very small, then the reconstruction error can be very bad. We, therefore, need to truncate the SVD and remove those terms with small singular value. The python function `np.linalg.pinv` currently being used does that automatically given a tolerance `rcond` that sets the smallest singular value allowed.

Noise covariance for state estimation

We now want to explore the possibility of using the knowledge of the noise covariance estimate that we can get from real data, to improve our estimate of the state. This differs from the previous noisy measurement section in that we use information from the real noise, apart from removing those degenerate modes that could impact the state estimate due to the noise.

For this derivation, we will use the Minimum Variance Unbiased Estimator (MVUE), which for our linear model and assuming Gaussian statistics corresponds to the conditional mean E$[X|Y]$. To derive the expression of this conditional expectation, we will start from the joint distribution of the optical state ($\boldsymbol{x}$) and noise ($\boldsymbol{w}$). Let us denote the state covariance $Cov(\boldsymbol{x}, \boldsymbol{x})$ with **X** and the noise covariance $Cov(\boldsymbol{w}, \boldsymbol{w})$ with **W**.

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{w} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu_x} \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{X} & 0 \\ 0 & \mathbf{W} \end{bmatrix} \right) \tag{13}$$

where we used the fact that the noise is independent of the state to determine the off-diagonal terms.

Now let us find the joint distribution of the optical state ($\boldsymbol{x}$) and the wavefront measurement ($\boldsymbol{y}$).

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu_x} \\ \mathbf{A}\boldsymbol{\mu_x} \end{bmatrix}, \begin{bmatrix} \mathbf{X} & \mathbf{X}\mathbf{A}^\top \\ \mathbf{A}\mathbf{X} & \mathbf{A}\mathbf{X}\mathbf{A}^\top + \mathbf{W} \end{bmatrix} \right) \tag{14}$$

Let us work out explicitly how we derived the term $Cov(\boldsymbol{y}, \boldsymbol{y})$,

$$\begin{aligned} Cov(\boldsymbol{y}, \boldsymbol{y}) &= Cov(\mathbf{A}\boldsymbol{x} + \boldsymbol{w}, \mathbf{A}\boldsymbol{x} + \boldsymbol{w}) \\ &= Cov(\mathbf{A}\boldsymbol{x}, \mathbf{A}\boldsymbol{x}) + Cov(\boldsymbol{w}, \boldsymbol{w}) + Cov(\boldsymbol{w}, \mathbf{A}\boldsymbol{x}) + Cov(\mathbf{A}\boldsymbol{x}, \boldsymbol{w}) \\ &= \mathbf{A}\mathbf{X}\mathbf{A}^{\top} + \mathbf{W} \end{aligned} \tag{15}$$

where we used the independence of the optical state and noise, and the standard covariance property of $Cov(\mathbf{A}\boldsymbol{x}, \mathbf{A}\boldsymbol{x}) = \mathbf{A}\mathbf{X}\mathbf{A}^{\top}$. The off-diagonal term can be obtained similarly.

Now, the mean of the conditional distribution of two bivariate normal random variables can be shown to be,

$$E[\boldsymbol{x}|\boldsymbol{y}] = \boldsymbol{\mu_x} + \Sigma_{\mathbf{xy}}{\Sigma_{\mathbf{yy}}}^{-1}(\boldsymbol{y} - \boldsymbol{\mu_y}) \tag{16}$$

Now, we will assume that the mean of the optical state distribution is zero $\boldsymbol{\mu_x} = 0$. Then, substituting the covariance from Equation 15, we have,

$$E[\boldsymbol{x}|\boldsymbol{y}] = \mathbf{X}\mathbf{A}^{\top}(\mathbf{A}^{\top}\mathbf{X}\mathbf{A} + \mathbf{W})^{-1}\boldsymbol{y} \tag{17}$$

Overall, our MVUE estimator when we introduce the term of noise is,

$$\hat{\boldsymbol{x}} = \mathbf{X}\mathbf{A}^{\top}(\mathbf{A}^{\top}\mathbf{X}\mathbf{A} + \mathbf{W})^{-1}\boldsymbol{y} \tag{18}$$

We can check dimensions now, $\mathbf{A} : (n, m)$, $\mathbf{W} : (m, m)$, $\boldsymbol{y} : (n, 1)$, and $\mathbf{X} : (n, n)$. Note that this estimate has been proposed but is not implemented at all in `ts_ofc`.

Optimal integral controller (OIC)

We derive the corrections to be applied by minimizing a particular cost function J. In our case, optimizing image quality, i.e. minimizing the variance of FWHM is an obvious goal. Furthermore, we want to limit large actuator swings to avoid damage to the glass, as well as to ensure smooth transitions between iterations. This additional goal can be achieved by incorporating the variance of the actuator commands ($\boldsymbol{u}$) in the cost function with some weight ρ. The matrix $\mathbf{H}$ defines the distribution of control authority among various actuator groups (rigid body and shape actuators) such that 1 μm or 1 arcsec rigid body displacement corresponds to 1 N actuator force.

The cost function that we minimize to derive the corrections are,

$$J = \boldsymbol{y}_{k+1}^{\top}\mathbf{Q}\boldsymbol{y}_{k+1} + \rho\boldsymbol{u}^{\top}\mathbf{H}\boldsymbol{u} \tag{19}$$

note that this J is not related to the J introduced in the previous section, it is just a convention to denote a cost function. Here we introduce the subindex k to denote the timestep. We will want to derive the corrections at timestep $k+1$ based on the previous state k.

Matrix definitions

Let us first refresh the definitions of the penalty matrix **H** and the image quality matrix **Q**. Based on the explanations of the first paragraph, this matrices are defined as,

$$\mathbf{H} = \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{50} \end{pmatrix} \tag{20}$$

where each diagonal term is the standard deviation of the actuator change due to that degree of freedom from the influence matrix times how many microns of piston does 1N force corresponds to.

Similarly, for the image quality matrix, we have,

$$\mathbf{Q} = \left(\frac{2\pi}{\lambda}\right)^2 \mathtt{diag}(\alpha)^2 \tag{21}$$

where α is the alpha value of the normalized point source sensitivity (PSSN) in the basis of Zernikes (Z_4-Z_{22}), and λ is the wavelength of the corresponding filter.

Derivation of corrections

Now, that we have set these definitions, we want to derive what is the best corrections to apply at timestep k. To derive them, we will minimize the cost function J, meaning that we want to minimize both the penalty term, which corresponds to penalizing large swings in our degrees of freedom, and the first term, which corresponds to minimizing the wavefront error.

To minimize the cost function with respect to $\boldsymbol{u}$, we need to rewrite the cost function in terms of x_k, we will leverage our initial definition $\boldsymbol{y}_k = \mathbf{A}\boldsymbol{x}_k$

$$J = \boldsymbol{x}_{k+1}^\top \mathbf{A}^\top \mathbf{Q} \mathbf{A} \boldsymbol{x}_{k+1} + \rho \boldsymbol{u}^\top \mathbf{H} \boldsymbol{u} \tag{22}$$

Now recall the basic assumption of our system in equation 1. We will assume for now that there is no uncontrolled change between timesteps, so we will have,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{u} \tag{23}$$

With this assumption, the cost function becomes,

$$J = (\boldsymbol{x}_k + \boldsymbol{u})^\top \mathbf{A}^\top \mathbf{Q} \mathbf{A} (\boldsymbol{x}_k + \boldsymbol{u}) + \rho \boldsymbol{u}^\top \mathbf{H} \boldsymbol{u} \tag{24}$$

Now that we have the cost function written in terms of the corrections $\boldsymbol{u}$, let us derive the optimal set of corrections by minimizing it,

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{u}} &= 2\rho \mathbf{H} \boldsymbol{u} + 2\mathbf{A}^\top \mathbf{Q} \mathbf{A} \boldsymbol{u} + 2\mathbf{A}^\top \mathbf{Q} \mathbf{A} \boldsymbol{x}_k \\ &= 2(\mathbf{A}^\top \mathbf{Q} \mathbf{A} + \rho \mathbf{H}) \boldsymbol{u} + 2\mathbf{A}^\top \mathbf{Q} \mathbf{A} \boldsymbol{x}_k \end{aligned} \tag{25}$$

Equating the above to zero, we find that the corrections to be applied are,

$$\boldsymbol{u} = -(\mathbf{A}^\top \mathbf{Q} \mathbf{A} + \rho \mathbf{H})^{-1} \mathbf{A}^\top \mathbf{Q} \mathbf{A} \boldsymbol{x}_k = -\mathbf{F} \boldsymbol{x}_k \tag{26}$$

We normally introduce a gain term α that determines the speed of convergence. With that we can derive an expression for the state of our system at timestep $k+1$,

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha \mathbf{F} \boldsymbol{x}_k + \boldsymbol{\delta}_k \tag{27}$$