Linear Analysis of Opto-Mechanical Systems

David C. Redding[1]
Charles Stark Draper Laboratory, Cambridge MA 02139

Mark Milman[2]
Jet Propulsion Laboratory, Pasadena CA 91109

Greg Loboda[3]
Massachusetts Institute of Technology, Cambridge MA 02139

Abstract

We present a general framework for the linear analysis of controlled opto-mechanical systems. Linear optical models are generated in matrix form. These are conjoined with structures and controls linear models to compute system performance as a function of optics, structures and control parameters. System design optimization with respect to optical performance is discussed. Examples illustrate the application of our approach; other examples are available in the literature.

1. Introduction

A general opto-mechanical system consists of a structure on which are mounted optical elements: mirrors, lenses, diffractive elements, sources, detectors and other elements. The optics may include controlled elements such as steering mirrors, deformable mirrors, tilting and decentering secondaries. The structure will experience (quasi)static distortions due to imperfect fabrication, thermal expansion or gravity sag. These static effects can be calibrated and corrected. The structure will also experience dynamic disturbances which can excite structural dynamics, causing vibration of the optical train. These structural dynamic effects may be mitigated by passive or active structural damping or by structural or optical control.

The linearized dynamics of a general opto-mechanical system can be captured in the differential equation of motion:

$$M\ddot{\vec{x}} + D\dot{\vec{x}} + K\vec{x} = B\vec{f} + F\vec{d} \tag{1}$$

The state vector $\vec{x}$ includes the translational and rotational displacements of the structural nodes or bodies that make up the structure, including the optical elements. The control vector $\vec{f}$ includes the forces applied by control actuators. Disturbance forces are represented by the vector $\vec{d}$. The mass matrix M, damping matrix D and stiffness matrix K embody a linear structural model, derived for flexible or rigid body or bodies standard structural analysis techniques[1].

Measurements $\vec{y}$ include optical measurements that are influenced by the structure and other direct measurements of the structure, e.g. accelerometer or strain gauge measurements. They also include optical measurements of external disturbances. These effects are captured in the measurement equation:

$$\vec{y} = C_{optical}\vec{x} + C_{structural}\vec{x} + C_w\vec{w}_{in} = C\vec{x} + C_w\vec{w}_{in} \tag{2}$$

[1]Site Manager, Member SPIE.
[2]Member of the Technical Staff.
[3]Graduate Student.

0-8194-0861-1/92/$4.00

Here $\vec{w}_{in}$ represents an optical target condition to be tracked, such as an incoming wavefront or a target line-of-sight. The matrices $C_{optical}$ and C_w are optical sensitivity matrices of the measurements dependent on the displacement states. The matrix $C_{structural}$ captures the sensitivities of mechanical measurements of the structure to state changes. For simplicity the structural and optical measurements are combined in C.

Controls are computed based on the measurements and applied using the actuators. A general form for feedback control is:

$$\dot{\vec{v}} = F\vec{v} + G\vec{y} \tag{3}$$

$$\vec{u} = -H\vec{v} - K\vec{y} \tag{4}$$

Values for the compensator matrices can be determined using classical or modern control methods[2,3].

Optical output of the system can be represented in many forms. We may be interested in a ray-based metric such as beam direction or beam pathlength or phase distribution. Or we may be interested in an image-quality based metric such as Strehl ratio or encircled energy. As discussed later, we can derive a linear output equation that takes the general form:

$$\vec{w}_{out} = C_{out}\vec{x} \tag{5}$$

Here the C_{out} matrix is another optical sensitivity matrix, derived from the performance metric.

Various analyses can be performed using these equations. For error analysis or tolerancing of the static system, we may work with the output equation only. For instance, the covariance **W** of the output error as a static function of the state error covariance **X**, which includes misalignments of the structure and optics, can be written:

$$\mathbf{W}_{out} = C_{out}\mathbf{X}C_{out}^{T} \tag{6}$$

System error budgets can be developed accurately using this equation.

Calibration of the static system can be performed by taking a series of measurements in different observing geometries: that is, at different pointing angles, field points, focus positions, or by observing multiple objects or performing other excursions in the coordinate space. The nominal geometries of each measurement are recorded and inverted using weighted least-squares estimation techniques[2,4]:

$$\vec{x}_{est} = \begin{bmatrix} C_{obs1} \\ \vdots \\ C_{obsn} \end{bmatrix}^{\text{“}-1\text{”}} \begin{bmatrix} \vec{y}_{obs1} \\ \vdots \\ \vec{y}_{obsn} \end{bmatrix} \tag{7}$$

Estimation of the dynamic system can be performed in an analogous fashion using the system equations and Kalman filtering techniques[2,3].

Frequency-domain analysis is particularly useful for control system design[3]. Open- and closed-loop transfer functions give the frequency response of the optical outputs of the system to optical inputs or to mechanical disturbances, e.g.:

$$\frac{\vec{w}_{out}(s)}{\vec{d}(s)} = -C_{out}(s^2M + sD + K)^{-1}F \tag{8}$$

Finally, we may be interested in performing explicit optimization of the system using our performance metric as part of a cost function. As an example we may wish to compute regulator control gains to minimize our optical performance metric using a Linear Quadratic approach. The cost function takes the form

$$\min J = \int_0^\infty \left(\vec{x}^T C_{out}^T Q C_{out} \vec{x} + \vec{u}^T R \vec{u}\right) dt \tag{9}$$

where Q and R are weighting matrices.

Most of the components of the equations (1-8) are available using standard structural, dynamic and controls techniques. The missing parts are the "C-matrices": the optical sensitivity matrices. The next sections of the paper present a summary derivation of these linear optical models.

2. Coordinate-Free Ray Trace Equations

Ray trace models represent light beams as bundles of rays, which are the trajectories of individual photons. Rays are generally composed of numerous straight-line segments, with direction changes at the reflective or refractive surfaces of optical elements. Ray tracing is the process of propagating a bundle of rays through an optical system to a point of interest such as a detector or the exit pupil (Fig. 1). The ray information can be used to quantify optical outputs of a system, by determining the wavefront error or beam direction error, or by providing phase inputs to diffraction calculations.

This section summarizes a coordinate-free set of ray-trace equations described in more detail in Refs. 5-7. This form was chosen for generality. It also offers an easily differentiable set of equations for deriving linear models, and provides a computationally efficient form for computer calculations[8,9]. This discussion assumes that the optical elements have conic-of-revolution surface shapes: see Ref. 6 for discussion of more general surface types.

Assume (without loss of generality) an element numbering scheme, subscripting each surface in the order that the ray strikes it. Doing this, the ray segment incident on the i^{th} element is the ray segment output from the $i\text{-}1^{th}$ element and is denoted $\hat{r}_{i-1}$. Each succeeding ray segment ($\hat{r}_i$) can be determined from the optical properties at the point of incidence, by applying the law of reflection, Snell's Law, or the grating equation.

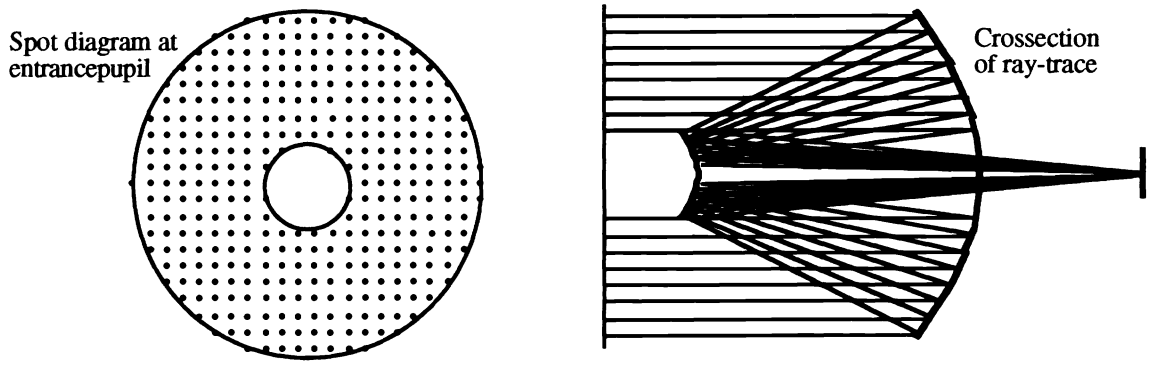

Figure 1. Ray trace of a Cassegrain telescope.

A conic-of-revolution surface can be represented in vector three-space in terms of a "surface dyadic" M and a vector $\vec{N}_0$. These quantities are functions of the surface principal axis direction $\hat{\psi}$ and the eccentricity e and geometric focal length f or conic constant K_c and "radius" K_r (Fig. 2):

$$M \equiv \left(I - e^2 \hat{\psi}\hat{\psi}\right) = \left(I + K_c \hat{\psi}\hat{\psi}\right). \tag{10}$$

$$\vec{N}_0 = -f(1+e)\hat{\psi} = K_r \hat{\psi}. \tag{11}$$

The surface is defined for a point $\vec{\rho}$ incident upon it by the function S

$$S \equiv \vec{\rho} \cdot M \cdot \vec{\rho} + 2\vec{N}_0 \cdot \vec{\rho} = 0 \tag{12}$$

The surface normal is determined from the derivative of S with respect to $\vec{\rho}$:

$$\vec{N} = \vec{N}_0 + M\vec{\rho}. \tag{13}$$

To avoid ambiguity in the discussion of refractive elements, we adopt a convention that the unit normal $\hat{N}$ is pointed away from the incident ray direction $\hat{r}_{i-1}$:

$$\hat{N} = -\text{sign}(\hat{r}_{i-1} \cdot \vec{N})\frac{\vec{N}}{|\vec{N}|} \tag{14}$$

The point of incidence is determined as:

$$\vec{\rho} = \vec{p} + L\hat{r}_{i-1} \tag{15}$$

where L is the geometrical length of the incident ray. Substituting for $\vec{\rho}$ in Eq. 12, L is a solution of

$$(\hat{r}_{i-1} \cdot M \cdot \hat{r}_{i-1})L_g^2 + 2\hat{r}_{i-1} \cdot (M \cdot \hat{p} + \vec{N}_0)L_g + \hat{p} \cdot (M \cdot \hat{p} + 2\vec{N}_0) = 0. \tag{16}$$

The optical pathlength of the ray is Lg scaled by the index of refraction n of the medium that contains it:

$$L = nLg \tag{17}$$

A ray reflected by the surface can be written in terms of the "reflection dyadic" R as

$$\hat{r}_i = R \cdot \hat{r}_{i-1} \tag{18}$$

where R is defined as

$$R = I - 2\hat{N}\hat{N} \tag{19}$$

This dyadic form of the law of reflection is derived in Ref. 5; it apparently first appeared in Ref. 10. A refracted ray is governed by Snell's Law. In coordinate-free form:

$$\hat{r}_i = \mu\hat{r}_{i-1} - \frac{1 - \mu^2}{\sqrt{1 - \mu^2 + \mu^2(\hat{N} \cdot \hat{r}_{i-1})^2} - \mu\hat{N} \cdot \hat{r}_{i-1}}\hat{N} \tag{20}$$

where

$$\mu = n_a/n_b \tag{21}$$

Repeated application of these equations follows each ray from the source to the final element (usually the detector or the exit pupil reference surface). At each succeeding surface, the incident ray direction is the reflected or refracted ray direction of the previous surface, and the incident ray starting position is the point of incidence on the previous surface ($\vec{p}_i = \vec{\rho}_{i-1}$). The pathlength is accumulated as:

$$L_i = \sum_{j=1}^{i} L_j. \tag{22}$$

4. Linear Ray-Trace Models

These ray-trace equations have been differentiated to produce analytic formulas for individual-ray optical sensitivity matrices in terms of a "perturbed ray state," $\vec{w}$. Details are given in Refs. 5-7. The sensitivities give a complete description of each ray valid for "small" perturbations, meaning that perturbed ray directions are close enough to the nominal that the small angle approximation holds. The perturbed ray direction is then:

$$\hat{r}_{pert} = \hat{r}_{nom} + d\hat{r}. \tag{23}$$

The linear models are derived assuming that the transverse aberration (beamwalk) $\vec{\gamma}$ of the perturbed ray away from the nominal ray is small compared to the curvature of each element surface, so that the

movement of the reflection (or refraction) point is confined to a plane tangent to the surface at the nominal reflection point (though the surface normal does . We assume the motions $\vec{\theta}$ and $\vec{\delta}$ of the elements are small. Finally, we neglect second and higher order perturbations. The validity of these approximations is a function of the particular element type and the particular system geometry. An example is presented below.

Given these approximations, perturbed rays can be computed in terms of deviations from exactly-traced nominal rays (Fig. 3). The perturbation of reflected or refracted ray direction is $d\hat{r}_i$. Perturbations of the position of the ray are decomposed into the transverse aberration $\vec{\gamma}_i$, which is perpendicular to nominal ray, and the OPD - denoted dL_i - which is parallel to the nominal ray. These ray perturbations are functions of the incident ray direction perturbation $d\hat{r}_{i-1}$, the incident ray beamwalk $\vec{\gamma}_{i-1}$, and tilt and translation perturbations of the ith element $\vec{\theta}$ and $\vec{\delta}$.

The ray perturbation state vector at the i^{th} element can be written as a 7¥1 column matrix $\vec{w}$, where:

$$\vec{w}_i = \begin{bmatrix} d\hat{r}_i \\ \vec{\gamma}_i \\ dL_i \end{bmatrix} \tag{24}$$

Similarly,$\vec{u}_i$ is a 6¥1 column matrix consisting of the small translation and rotation of the i^{th} element:

$$\vec{u}_i = \begin{bmatrix} \vec{\theta}_i \\ \vec{\delta}_i \end{bmatrix} \tag{25}$$

The propagation of the perturbed ray state from element to element can be written as a state difference equation:

$$\vec{w}_i = \frac{\partial \vec{w}_i}{\partial \vec{w}_{i-1}} \vec{w}_{i-1} + \frac{\partial \vec{w}_i}{\partial \vec{u}_i} \vec{u}_i \tag{26}$$

Here $\partial\vec{w}_i/\partial\vec{w}_{i-1}$ is the transition matrix from the i-1^{th} to the i^{th} element, and $\partial\vec{w}_i/\partial\vec{u}_i$ is the influence matrix at the i^{th} element

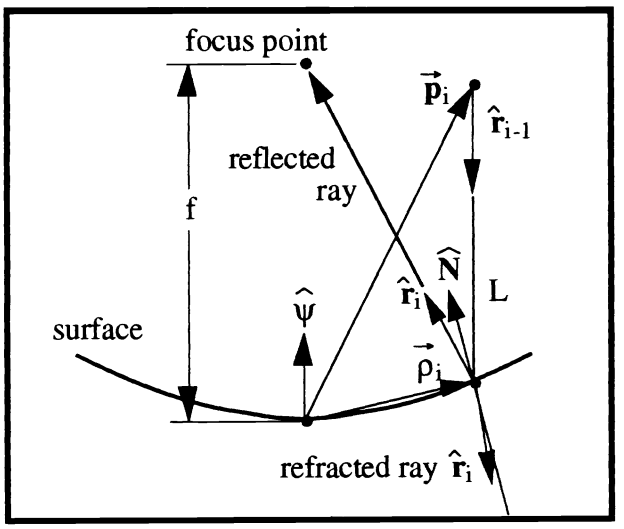

Figure 2. Exact Ray-Trace Geometry

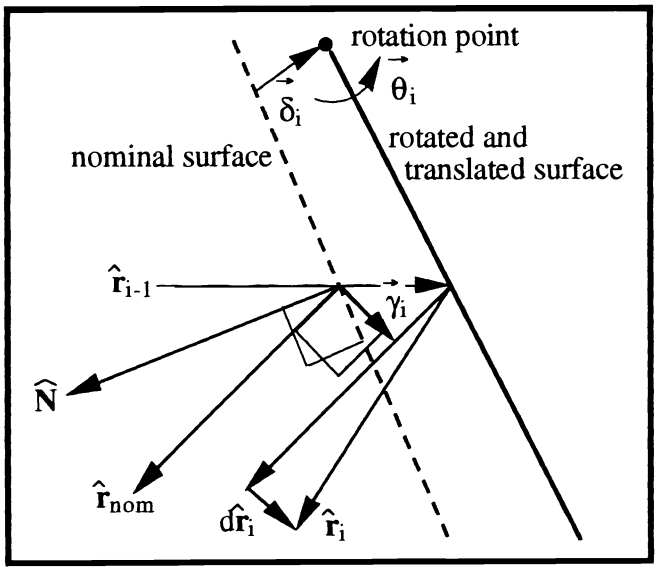

Figure 3. Perturbation Geometry

Sensitivities of the ray perturbation state at one element to ray or element perturbations at another element are easily calculated as products of these matrices. The sensitivity of the ray at the n^{th} element to

the ray at the i^{th} element is:

$$\frac{\partial \vec{w}_n}{\partial \vec{w}_i} = \frac{\partial \vec{w}_n}{\partial \vec{w}_{n-1}} \cdots \frac{\partial \vec{w}_{i+1}}{\partial \vec{w}_i}. \tag{27}$$

The sensitivity of the ray at the n^{th} element to perturbations of the i^{th} element is:

$$\frac{\partial \vec{w}_j}{\partial \vec{u}_i} = \frac{\partial \vec{w}_j}{\partial \vec{w}_{j-1}} \cdots \frac{\partial \vec{w}_{i+1}}{\partial \vec{w}_i} \frac{\partial \vec{w}_i}{\partial \vec{u}_i}. \tag{28}$$

Again, Refs. 5-7 provide more details.

A complete linear model of a beam train combines the effects of perturbations of the rays and elements as seen at a reference surface. Tracing m rays through a system of n elements, the combined linear model is the C-matrix linear optical model:

$$\begin{bmatrix} \vec{w}_{n_{ray\ 1}} \\ \vdots \\ \vec{w}_{n_{ray\ m}} \end{bmatrix} = C \begin{bmatrix} \vec{w}_0 \\ \vec{u}_1 \\ \vdots \\ \vec{u}_n \end{bmatrix} = \begin{bmatrix} \left[\frac{\partial \vec{w}_n}{\partial \vec{x}_0} \frac{\partial \vec{w}_n}{\partial \vec{u}_1} \cdots \frac{\partial \vec{w}_n}{\partial \vec{u}_n} \right]_{ray\ 1} \\ \vdots \\ \left[\frac{\partial \vec{w}_n}{\partial \vec{x}_0} \frac{\partial \vec{w}_n}{\partial \vec{u}_1} \cdots \frac{\partial \vec{w}_n}{\partial \vec{u}_n} \right]_{ray\ m} \end{bmatrix} \begin{bmatrix} \vec{w}_0 \\ \vec{u}_1 \\ \vdots \\ \vec{u}_n \end{bmatrix} \tag{29}$$

Computing the C-matrix in the structural model coordinate frame, the $\vec{u}_i$'s generally correspond directly to particular structural model outputs, and the integration of linear structural and optics models requires only a matrix multiplication, as per Eqs. (1-9).

Often the complete C-matrix model is unnecessary. C-matrix models of pathlength only, or ray direction only, or other combinations of the ray states can be derived from the full C-matrix by using transform matrices to pick off only those states of interest. For instance, a pathlength-only model is

$$\vec{w}_L = \begin{bmatrix} L_1 \\ \vdots \\ L_n \end{bmatrix} = C_L \vec{u} = T_L C \begin{bmatrix} \vec{w}_0 \\ \vec{u}_1 \\ \vdots \\ \vec{u}_n \end{bmatrix} \tag{30}$$

where TL is the $(m \times 7m)$ matrix

$$T_L = \begin{bmatrix} [0\,0\,0\,0\,0\,0\,1] & & \\ & \ddots & \\ & & [0\,0\,0\,0\,0\,0\,1] \end{bmatrix} \tag{31}$$

Similarly, a model of ray transverse-aberration as projected onto the n^{th} surface is

$$\vec{w}_\gamma = \begin{bmatrix} \vec{\gamma}_1 \\ \vdots \\ \vec{\gamma}_v \end{bmatrix} = C_\gamma \vec{u} = T_\gamma C \begin{bmatrix} \vec{w}_0 \\ \vec{u}_1 \\ \vdots \\ \vec{u}_n \end{bmatrix} \tag{32}$$

where

$$T_\gamma = \begin{bmatrix} T_{\gamma 1} & & 0 \\ & \ddots & \\ 0 & & T_{\gamma n} \end{bmatrix} \tag{33}$$

and (the y refers to the principal axis of the n^{th} surface)

$$T_{\dot{\gamma}i} = \left[\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right] \tag{34}$$

Element-based coordinates can be used to transform motions in actuator coordinates into the C-matrix frame or to eliminate degrees of freedom from the model. For example, including element coordinates in a pathlength model:

$$\vec{w}_L = T_L \begin{bmatrix} \left[\frac{\partial \vec{w}_n}{\partial \vec{w}_0}T_0 \;\; \frac{\partial \vec{w}_n}{\partial \vec{u}_1}T_1 \cdots \frac{\partial \vec{w}_n}{\partial \vec{u}_n}T_n\right]_{ray1} \\ \vdots \\ \left[\frac{\partial \vec{w}_n}{\partial \vec{w}_0}T_0 \;\; \frac{\partial \vec{w}_n}{\partial \vec{u}_1}T_1 \cdots \frac{\partial \vec{w}_n}{\partial \vec{u}_n}T_n\right]_{raym} \end{bmatrix} \begin{bmatrix} \vec{w}_0^0 \\ \vec{u}_1^1 \\ \vdots \\ \vec{u}_n^n \end{bmatrix} \tag{35}$$

5. Images and Linear Diffraction Models

For most optical systems, point images can be computed in the following way. First, a bundle of rays is traced from a point source in the field, through the system to a spherical reference surface located at the exit pupil. This reference surface is centered at the nominal focal point on the detector and so coincides with the wavefront for a perfect point image (Fig. 4). The pathlength of each ray to the reference surface is differenced with the chief ray (j=1) pathlength. These optical path differences (OPDs) define the phase of the light, so the pupil complex amplitude G_e at the ray *j* incidence point (x_e, y_e) is

$$G_e(x_e,y_e) = \begin{cases} \exp(ik\,(L_j - L_1)) & \textit{if not obscured} \\ 0 & \textit{if obscured or outside aperture} \end{cases} \tag{36}$$

Here $k = 2\pi/\lambda$. The rays also determine the pupil projections of obscurations, where G_e is set to 0.

The complex amplitude G_d at the detector is determined as[11,12]

$$G_d(x_d,y_d) = \exp(ikz)\exp\left[\frac{k}{2z}(x_d^2+y_d^2)\right]\int_s G_e(x_e,y_e)\exp\left(-i\,\frac{2\pi}{\lambda z}(x_e x_d - y_e y_d)\right)dx_e dy_e \tag{37}$$

Here z is the propagation distance; the integral is the Fourier transform of G_e. This expression is subject to the assumptions of Fresnel diffraction[11,12]. As they do not affect the intensity, we will drop the phase factors multiplying the integral and write

$$G_d(x_d,y_d) = \int_s G_e(x_e,y_e)\exp\left(-i\,\frac{2\pi}{\lambda z}(x_e x_d - y_e y_d)\right)dx_e dy_y = \mathcal{F}\{G_e(x_e,y_e)\} \tag{38}$$

in place of Eq. (37). The image intensity $\mathfrak{I}$ is the modulus of G_d

$$\mathfrak{I}(x_d,y_d) = G_d(x_d,y_d)\,G_d^*(x_d,y_d) \tag{39}$$

To compute images, it is convenient to use a regular square grid of rays, so that each ray j samples an equal area $A_j = A/m$ of the beam cross-section (A is the total beam cross-sectional area and m is the number of rays). The ray grid will preserve its regularity at the exit pupil, even if the system is significantly aberrated, provided the exit pupil reimages a point (e.g. the entrance pupil) where the ray grid is still regular. Then a direct correspondence can be set up between the individual rays and entries in the complex amplitude matrix G_e. Optical pathlength for ray j then provides the phase information for a corresponding entry in G_e. The regularity of the sampling of the grid points allows the use of fast Fourier transform (FFT) techniques to compute images using Eq. (38). Implicit in our discussion here are two needed mappings: of rays to complex amplitudes; and of complex amplitude to location in physical space.

Linear models of the complex amplitude can be obtained by differentiating Eq. (38):

$$\frac{\partial}{\partial \vec{u}} G_d(x_d,y_d) = ik \int_s \frac{\partial \vec{w}_L}{\partial \vec{u}} \exp(ik\, \vec{w}_L) \exp\left(-i \frac{2\pi}{\lambda z}(x_e x_d - y_e y_d)\right) dx_e dy_e \tag{40}$$

6. Optical Performance Metrics

A performance metric is picked to match a particular objective, such as "maximize image quality." Functions that are extremized at the conditions where best image quality is obtained include Strehl ratio, encircled energy and wavefront error. Beam pointing systems or beamwalk control systems have objectives like "minimize pointing jitter." Metrics such as ray direction error or ray transverse aberration capture these effects. The wavefront error, ray direction error and transverse aberration metrics are derived from geometric optics only. Strehl ratio and encircled energy are derived from physical optics.

The geometric performance metrics are assembled by combining components of $\vec{w}$. For instance, a quadratic cost function that penalizes total wavefront error is:

$$J_{Lttl} = \frac{1}{2}\vec{w}_L^T \vec{w}_L = \frac{1}{2}\vec{u}^T C_L^T C_L \vec{u} = \frac{1}{2}\vec{u}^T C^T T_L^T T_L C \vec{u} \tag{41}$$

Note that this metric includes the wavefront effects of piston and tilt errors, which do not contribute to image quality in general. Also, it can be desirable to penalize some rays more than others. This can be done, for instance using the Q weighting matrix in Eq. (9). A similar cost function for transverse aberration is

$$J_\gamma = \frac{1}{2}\vec{w}_\gamma^T \vec{w}_\gamma = \frac{1}{2}\vec{u}^T C_\gamma^T C_\gamma \vec{u} \tag{42}$$

Performance metrics based on image quality can similarly be expressed in terms of the C-matrix. An ideal image for a telescope is simply a diffracted aperture. The image takes the form (for an unobscured circular aperture) of the well-known Airy pattern. Aberrations cause the magnitude of the peak intensity to decrease below its ideal value and the extent of the spot to spread. "Strehl ratio" is the ratio of the peak intensity at the focal point of the actual aberrated system to the peak intensity at the same point for a perfect, unaberrated system. "Encircled energy" relates to the amount of energy that falls into a particular region surrounding the focal point[12]. The discussion here follows Ref. 13.

We first derive the Strehl ratio. We denote the intensity at the center of the focal plane by h

$$h = |G_d(0,0)|^2 = \left| \int_s \exp(ik\, \vec{w}_L) dx_e dy_e \right|^2 \tag{43}$$

For a *nominally unaberrated system* $\vec{w}_L$ is zero and the intensity at the focal spot is the peak intensity hnom

$$h_{nom} = \left| \int_s dx_e dy_e \right|^2 = A_e^2 \tag{44}$$

The Strehl ratio *SR* is defined

$$SR = h/h_{nom} \tag{45}$$

Now consider the effect of wavefront aberrations. Expanding h about the unaberrated nominal:

$$h = h_{nom} + \frac{\partial h}{\partial \vec{u}}\vec{u} + \vec{u}^T \frac{\partial^2 h}{\partial \vec{u}^2}\vec{u} + O(u^3) \tag{46}$$

The gradient term vanishes:

$$\frac{\partial h}{\partial \vec{u}} = \frac{\partial G_d(0,0)}{\partial \vec{u}} G_d^*(0,0) + G_d(0,0)\frac{\partial G_d^*(0,0)}{\partial \vec{u}}$$
$$= ik\ A \int_s \frac{\partial \vec{w}_L}{\partial \vec{u}} dx_e dy_e - ik\ A \int_s \frac{\partial \vec{w}_L}{\partial \vec{u}} dx_e dy_e = 0 \tag{47}$$

The second-order term depends on the hessian matrix $H = \partial^2 h/\partial \vec{u}^2$. Solving,

$$H = \frac{\partial^2 h}{\partial \vec{u}^2} = \frac{\partial^2 G_d(0,0)}{\partial \vec{u}^2} G_d^*(0,0) + 2\frac{\partial G_d(0,0)}{\partial \vec{u}}\frac{\partial G_d^*(0,0)}{\partial \vec{u}} + G_d(0,0)\frac{\partial^2 G_d^*(0,0)}{\partial \vec{u}^2}$$
$$= 2k^2\left(\int_s \frac{\partial \vec{w}_L}{\partial \vec{u}} dx_e dy_e\right)\left(\int_s \frac{\partial \vec{w}_L}{\partial \vec{u}} dx_e dy_e\right) - Ak^2 \int_s \left(\frac{\partial \vec{w}_L}{\partial \vec{u}}\right)\left(\frac{\partial \vec{w}_L}{\partial \vec{u}}\right)^T dx_e dy_e \tag{48}$$

(the second derivatives of $\vec{w}$ cancel out). Substituting into Eq. (46),

$$h = h_{nom} + k^2\left(\int_s \frac{\partial \vec{w}_L}{\partial \vec{u}}\vec{u} dx_e dy_e\right)^2 - Ak^2 \int_s \left(\frac{\partial \vec{w}_L}{\partial \vec{u}}\vec{u}\right)^2 dx_e dy_e \tag{49}$$

Note that, as the hessian H is nonpositive definite,

$$h \le h_{nom} \tag{50}$$

and *SR* is always less than or equal to 1. Writing $d\vec{w}_L$ for $(\partial \vec{w}_L/\partial \vec{u})\vec{u}$,

$$SR = 1 + k^2\left\{\frac{1}{A^2}\left(\int_s d\vec{w}_L dx_e dy_e\right)^2 - \frac{1}{A}\int_s (d\vec{w}_L)^2 dx_e dy_e\right\} + O(u^3) \tag{51}$$

which is the usual form for the Strehl ratio approximation in the case of small aberrations.

If (as discussed in Section 5) the ray incidence points in the pupil form a uniform square grid, a simple zeroth order quadrature scheme can be used to approximately solve these integrals. Each ray samples a sub-aperture of area A/m, over which $\vec{w}_L$ is assumed constant. Substituting $C_L = \partial \vec{w}_L/\partial \vec{u}$,

$$\left(\int_s \frac{\partial \vec{w}_L}{\partial \vec{u}}\vec{u} dx_e dy_e\right)^2 = \frac{A^2}{m^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{m} u_i C_{ki} C_{lj} u_j$$
$$= \frac{A^2}{m^2}\vec{u}^T M \vec{u} \tag{52}$$

where

$$M_{i,j} = \sum_{k=1}^{m}\sum_{l=1}^{m} C_{ki} C_{lj} \tag{53}$$

Also,

$$\int_s \left(\frac{\partial \vec{w}_L}{\partial \vec{u}}\vec{u}\right)^2 dx_e dy_e = \frac{A}{m}\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{j=1}^{n} u_i C_{ki} C_{kj} u_j$$
$$= \frac{A}{m}\vec{u}^T C_L^T C_L \vec{u} \tag{54}$$

Thus, to the accuracy of the quadrature,

$$H = -2A^2k^2\left[\frac{1}{m}C_L^T C_L - \frac{1}{m^2}M\right], \tag{55}$$

so that the Strehl ratio is

$$SR = 1 + \frac{1}{A_e^2}\vec{u}^T H \vec{u} + O(u^3) \tag{56}$$

It is interesting to note that the solution of the integrals of Eq. 51 gives us a convenient form for the *mean-square* wavefront error:

$$J_{Lmsq} = \vec{u}^T\left[\frac{1}{m}C_L^T C_L - \frac{1}{m^2}M\right]\vec{u} + O(u^3) \tag{57}$$

Encircled energy metrics can be posed in 2 ways. One (to be maximized) is "the amount of energy falling within a circle of radius r of the focal point." Another (to be minimized) is "the radius r about the focal point that contains m% of the total energy E." The first is computed as

$$E_r = \int_r |G_d(x_d,y_d)|^2 dx_d dy_d \tag{58}$$

The second is a function r defined implicitly by the equation

$$E_r = \int_r |G_d(x_d,y_d)|^2 dx_d dy_d - \mu \int_s |G_d(x_d,y_d)|^2 dx_d dy_d$$

Here the second term gives the percentage of the total energy. The function r exists by the intermediate value theorem and is precisely the "100 x m percent" encirclement of energy criterion about the nominal focal point of the system.

Assuming sufficient smoothness in the function E, by differentiating r implicitly with respect to u the sensitivity of the encirclement of energy criterion can also be obtained. These calculations lead to the gradient expression

$$\frac{\partial r}{\partial \vec{u}} = \frac{\displaystyle\int_{|\omega| \le 1} \frac{\partial}{\partial \vec{u}} |G_d(x_d,y_d)|^2 d\omega}{\pi r \displaystyle\int_{|\omega| = 1} |G_d(x_d,y_d)|^2 d\omega}.$$

In general it can be shown that this gradient vanishes for unaberrated configurations. Thus an unaberrated wavefront is stationary for this metric.

7. Optimization Problems

The metrics defined above can be applied in a variety of ways. They provide measures of performance for error analyses and simulation studies, as in Eq. (5). They provide cost functions for the derivation of alignment or control laws, or for feedback control gains. They can also be used in general system design, to optimize sensitivities with respect to manufacturing error or disturbance sources.

The first problem is to correct a static deformation, call it x, of the surface of a reflector to improve the received signal. The structural model for this problem is simply

$$u = K^{-1}Bf, \tag{59}$$

where f denotes the applied force, B is the control influence matrix, K is the stiffness matrix, and u is the resulting deflection. One obvious way to correct the system is to null as best as possible the deformation. A least squares problem that directly minimizes the rms surface error is posed as

$$\min_{f \in \Delta} \langle x - K^{-1}BF,\ x - K^{-1}Bf \rangle \tag{60}$$

where Δ is the actuator force constraint set. This minimization does not take into account an optical quality

measure, and thus does not distinguish between the optical sensitivities of different surface deformations. An alternative approach that incorporates these sensitivities is to use the quadratic form defined by the Hessian in the above minimization. The resulting weighted least squares problem has the form

$$\min_{f \in \Delta} \langle x - K^{-1}BF,\ H(x - K^{-1}Bf) \rangle, \tag{61}$$

and directly optimizes the optical sensitivity. This particular problem is a quadratic programming problem which will be worked out in greater detail in the following section for a segmented reflector example.

This static problem has a dynamical counterpart that fits well into the standard state space theory of linear systems. The analogous dynamical model for the system is

$$M\ddot{x} + D\dot{x} + Kx = Bd, \tag{62}$$

where M denotes the mass matrix, D is the damping matrix, K again denotes the stiffness matrix, B is a nodal influence matrix, and d is a disturbance which we assume to be white noise. (A colored noise disturbance with rational spectrum is easily introduced by augmenting the system dynamics.) The output equation that captures the optical performance metric is

$$z = H^{1/2}x. \tag{63}$$

Under standard ergodic assumptions, $E(|z|^2)$ (E= expectation operator) gives the time--average intensity degradation at the focal point due to the input disturbance d. With Eqs. 62-63 serving as the open loop system model, feedback control design or design optimization problems may be posed to attenuate the transmissibility from the disturbance d to the controlled output z.

Next we consider a design problem in which the goal is to choose the (unaberrated) nominal optical prescription u^* that minimizes the sensitivity of the intensity at the focal point with respect to anticipated errors arising from manufacturing tolerances, misalignments due to disturbances, etc. A formulation for this problem is the following:

$$\min_{u^* \in M} - E \langle H(u^*)v, v \rangle. \tag{64}$$

Here M is the set of prescriptions that yield unaberrated systems, and v is the random vector containing the misalignments, manufacturing error, etc. The statistical information required to solve this problem is the covariance matrix $V=E(vv^T)$. Then (2.9) is equivalent to

$$\min_{u^* \in M} \mathrm{tr}(H(u^*)V). \tag{65}$$

Solving the optimization problem (Eq. 65) requires computing the differential of H over the surface M. After some work we arrive at

$$\frac{\partial H_{st}}{\partial \bar{u}_i} = -2Ak^2 \int_S W_{is}W_t + W_{it}W_s d\sigma. \tag{66}$$

An interesting thing to observe here is that modulo the approximation of the integral above, the gradient of H coincides with the gradient of C^TC, where C is the COMP C-matrix. Thus although CTC is an incomplete approximation to H, the missing term, which essentially is related to an average wavefront error, does not enter into the gradient calculations. Using finite differences and COMP, the partials in Eq. 66 can be approximated. Coupling these approximations with quadrature schemes the Hessian sensitivities can be computed, and thus problems of the type Eq. 65 can be worked.

8. Linear Model Accuracy: a Segmented Mirror Example

As a first example to illustrate the methods discussed in the text we consider a Schmidt-Cassegrain telescope as sketched in Figs. 4-5. The primary mirror is a parabolic reflector made up of 6 hexagonal segments, each with a diameter of 1.3m. The secondary has a diameter of 0.78 m, and is assumed fixed in this study. The primary has an f number of 2.0; the system f number is 18.0. The primary mirror segments of this telescope are deformable for figure control purposes. A study of this system for direct detection of dim binary companion stars is presented in Ref. 14; control issues for similar systems are also discussed in Ref. 15.

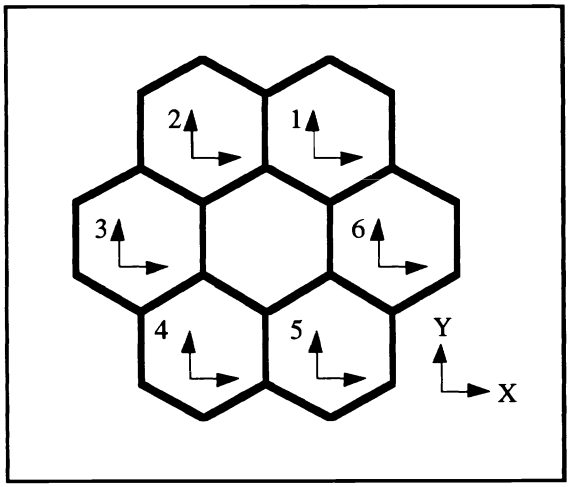

Figure 4. Low Order Panel Layout.

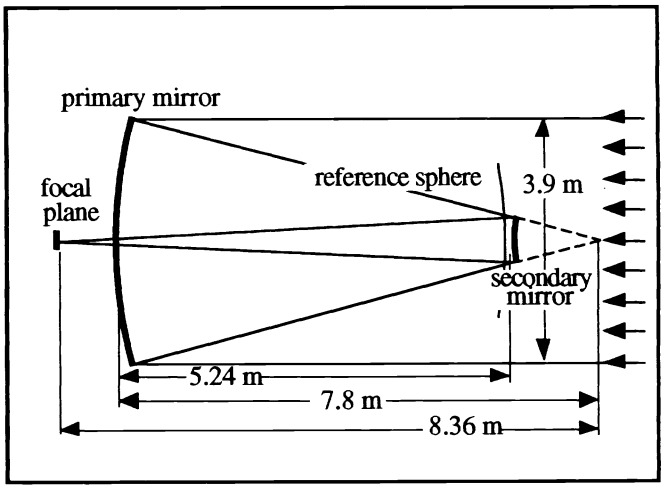

Figure 5. Optical Layout.

In this section, we illustrate the accuracy of linear models developed for this system. The analysis was performed assuming the panels were rigid, with six degrees of freedom per segment. The C-matrix for the hessian (Eq. 55) of the system was constructed using the differential ray trace capability of the COMP code[8]. Spectral analysis of the hessian was performed to identify best and worst-case deformation modes.

Excursions in the direction of the worst-case eigenvector were done to compare the relative accuracies of the linear C-matrix model with exact ray-trace calculations for a wavefront error metric. As indicated in Fig. 6, the comparison is quite precise over the limited range of the figure. Experience with this and other systems shows that linear models can be accurate to several decimal places over even fairly large perturbations - several millimeters in the case of typical 3 meter class telescopes. Similar comparative results for diffraction-based Strehl ratio ("true" Strehl computed as a far-field diffraction based on geometric phase in the exit pupil as per Section 5 above) are shown in Fig. 7. Again, over the small range plotted the linear and exact models are indistinguishable.

Results comparing the "true" Strehl to the approximate Strehl of Eq. 56 are plotted in Fig. 8 for the same worst-case eigen mode. As expected, the two compare reasonably well up to about 80% Strehl ($\lambda/14$). Above that point the approximate numbers deviate sharply. This result is consistent with the truncation of the series of Eq. 46, and is well known in standard practice.[12]

9. Composite Models: a Tolerancing Example

Once a telescope or other optical instrument is assembled, including (inevitably) the introduction of alignment errors, standard procedure calls for it to be aligned. Alignment can be accomplished using any of a number of techniques, ranging from lining up spots on a screen to full interferometric testing. These techniques can be said to be optimization procedures and explicitly modelled as such. For instance, a tilt mirror may be adjusted to line up pupils in a camera. This procedure places a calibration spot at some ideal position on a screen placed at an appropriate location in the beam train, subject to the limits of the ability of a technician to measure spot positions. It can be modelled as the minimization of transverse aberration of the original (perturbed) system at the screen using one or more rays.

Now the question is: what is the *post-alignment residual* error? This question can be addressed by building up composite matrix models of the beam train: linear models with embedded alignments or other optimizations. Such a procedure is illustrated in this section for the case of the Hubble Space Telescope/Wide Field and Planetary Camera (HST/WFPC). The problem we address is the creation of a linear model of residual pupil shear, assuming normally distributed fabrication errors in the This example is taken from Ref. 16 which analyzes the detected motion of calibration spots to ascertain probable alignment shifts incurred during launch.

The HST/WFPC system is sketched in Fig. 9. Note that there are 8 separate cameras sharing the light from the HST main telescope. The beam train for each includes the main HST primary and secondary mirror, the WFPC pickoff mirror, two lenses and a pyramid mirror. The pyramid mirror splits the light in 4 directions; it has two settings (clocked by 45°) allowing it to feed all 8 cameras. After the pyramid mirror, each camera has its own separate repeater telescope. Pupil shear is measured as the transverse aberration of the chief ray at the repeater secondary.

The alignment procedure is to tilt the pickoff mirror so as to minimize pupil shear in all 8 cameras simultaneously.

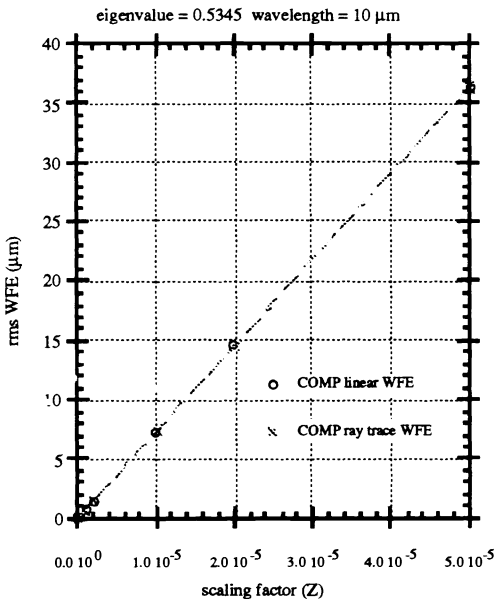

Figure 6. WFE vs. Perturbation Magnitude.

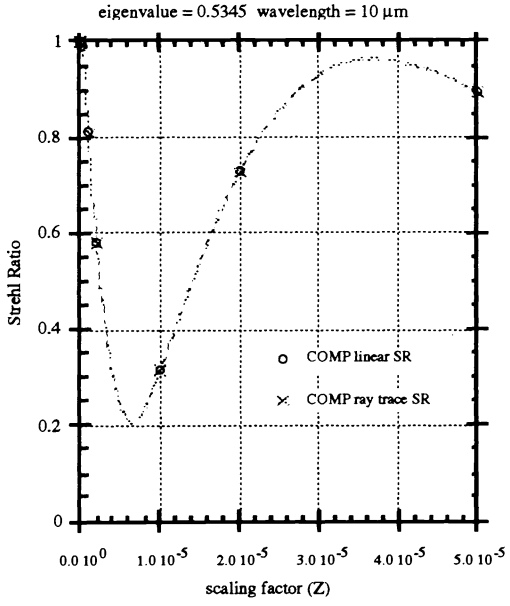

Figure 7. Strehl Ratio vs. Perturbation Magnitude.

The first step in constructing the composite model is to compute C_Y-matrices for each camera, giving the motion of each chief ray as a function of element displacements. These C_Y-matrices are then combined into a single matrix. The linear model for relative transverse aberration (pupil shear) at all of the repeater secondaries as a function of the various element perturbations takes the form:

$$\vec{x}_{SM} = \frac{\partial \vec{x}_{SM}}{\partial \vec{u}} \vec{u}$$

The sensitivity of the relative transverse aberrations to element perturbations $\frac{\partial \vec{x}_{SM}}{\partial \vec{u}}$ is the linear model output from COMP. It is a 16¥108 matrix. $\vec{x}_{SM}$ is the vector of relative transverse aberrations for all 8 cameras:

$$\vec{x}_{SM} = \begin{bmatrix} \vec{s}_1 \\ \vdots \\ \vec{s}_8 \end{bmatrix}$$

Here $\vec{s}_i = \begin{bmatrix} s_x \\ s_y \end{bmatrix}$ is a 2-vector of the ith camera relative transverse aberration at the repeater secondary in local coordinates. $\vec{u}$ is the vector of element perturbations:

$$\vec{u} = \begin{bmatrix} \vec{u}_{PO} \\ \vec{u}_{PYR} \\ \vec{u}_{F1} \\ \vec{u}_{R1} \\ \vdots \\ \vec{u}_{F8} \\ \vec{u}_{R8} \end{bmatrix}$$

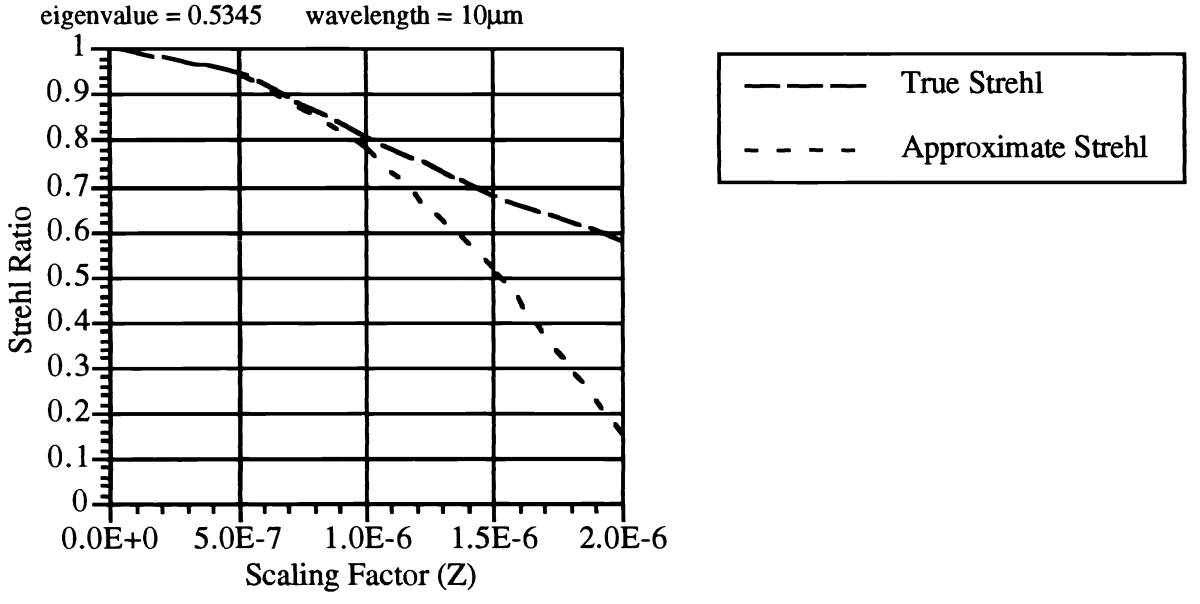

Figure 8. True vs. Approximate Strehl Ratio for Example 1.

Each individual $\vec{u}$-vector is a 6-vector made up of rotational $\vec{\theta}$ and translational $\vec{\delta}$ perturbation 3-vectors:

$$\vec{u}_i = \begin{bmatrix} \vec{\theta}_i \\ \vec{\delta}_i \end{bmatrix}$$

The element perturbations are organized so as to be compatible with a NASTRAN structure model. The subscript abbreviations are: PO = pick-off mirror; PYR = pyramid mirror; F_i = i^{th} camera fold mirror; R_i = i^{th} camera relay barrel.

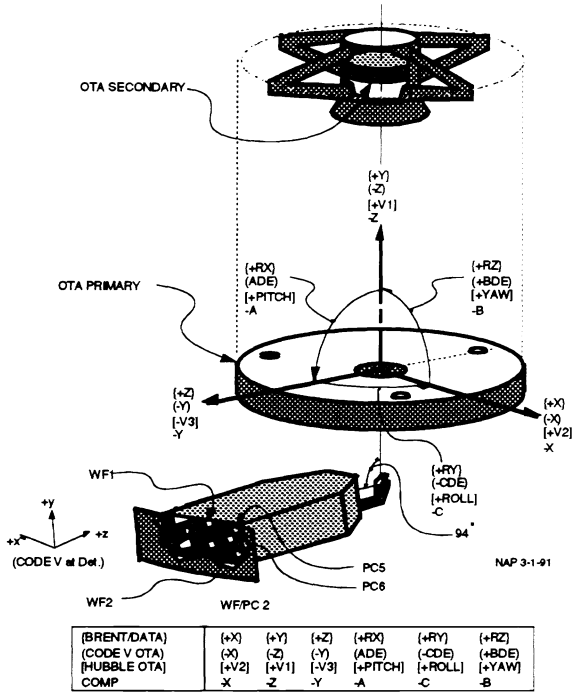

Figure 2. OTA/WFPC coordinate systems (courtesy of Norm Page).

The pickoff mirror tilt angles needed to minimize the relative transverse aberrations are denoted $\vec{\theta}_{POc}$. Including them in the pupil shear equation, we get

$$\vec{x}_{SM} = \frac{\partial \vec{x}_{SM}}{\partial \vec{u}} \vec{u} + \frac{\partial \vec{x}_{SM}}{\partial \vec{\theta}_{PO}} \vec{\theta}_{POc}$$

$\vec{\theta}_{POc}$ is determined as the set of angles which minimizes $\vec{x}_{SM}^T \vec{x}_{SM}$:

$$\vec{\theta}_{POc} = -G\,\vec{u}$$

where the gain matrix G is:

$$G = \left[\left(\frac{\partial \vec{x}_{SM}}{\partial \vec{\theta}_{PO}} \right)^T \left(\frac{\partial \vec{x}_{SM}}{\partial \vec{\theta}_{PO}} \right) \right]^{-1} \left(\frac{\partial \vec{x}_{SM}}{\partial \vec{\theta}_{PO}} \right)^T \frac{\partial \vec{x}_{SM}}{\partial \vec{u}}$$

The combined linear model for the post-alignment pupil shear is:

$$\vec{x}_{pupil} = C\vec{u}$$

Thus the linear model C-matrix, including the embedded alignment, is:

$$C = \frac{\partial \vec{x}_{SM}}{\partial \vec{u}} - \frac{\partial \vec{x}_{SM}}{\partial \vec{\theta}_{PO}} G$$

The magnitude of the pupil shear in each camera is RSSed from its components.

Several cases were run using NASTRAN-generated perturbations. Results for a case that models the effects of gravity release are provided in Table 1. The linear model results are compared with results from a nonlinear optimization performed using Code V. The results agree very well, with all differences in the hundredth-pixel range. The main difference between the results is the time taken to compute them. The linear results were computed in a few seconds on a Macintosh personal computer. The Code V results required several minutes on a VAX minicomputer.

Table 1. Comparison of results for gravity release case.

	Pupil shear magnitude results: all channels							
	WF 1	**WF 2**	**WF 3**	**WF 4**	**PC 5**	**PC 6**	**PC 7**	**PC 8**
Linear	8.6790e-06	1.2606e-05	1.3261e-05	1.3566e-05	1.0024e-05	1.1612e-05	1.2026e-05	1.1530e-05
Code V	8.63e-06	1.27e-05	1.33e-05	1.36e-05	1.00e-05	1.17e-05	1.20e-05	1.15e-05

Note: All results are in meters.

10. Conclusion

This paper has presented a framework for applying standard linear analysis techniques to opto-mechanical systems. These techniques, including covariance analysis, optimization, estimation and simulation, provide extremely powerful tools for optical tolerancing, calibration, controller design, mechanical structure design, system-level error analysis and optimization.

11. Acknowledgement

This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

12. References

1. Przemieniecki, J., **Theory of Matrix Structural Analysis**, McGraw-Hill, 1968.
2. Bryson, A. and Ho, Y., **Applied Optimal Control**, Halsted Press, 1975.
3. Kwakernack and Sivan, **Linear Optimal Control Systems**, Wiley-Intersciences, 1977.
4. Press, W., Flannery, B., Teukolsky, S. and Vettering, W., **Numerical Recipes**, Cambridge University Press, 1986.
5. Redding, D. and Breckenridge, W., "Optical Modelling for Dynamics and Control Analysis," Journal of Guidance, Control and Dynamics, Vol. 14, No. 5, September, 1991.
6. Redding, D. and Breckenridge, W., "Coordinate-Free Raytrace Equations for Aspheric and Deformed Optical Surfaces," Memo DCR-91-38, July 13, 1991.
7. Redding, D. and Breckenridge, W., "Linearized Ray-Trace Analysis," Proceedings of the 1990 International Lens Design Conference, G. Lawrence, Ed., Proc. SPIE 1354, June 1990.
8. Redding, D., Needels, L., Wallace, J.K., Yu, J. and Levine, B.M., "Controlled Optics Modelling Package User Manual, Release 0.9w," Memo DCR-91-47, October 22, 1991.
9. Redding, D., Levine, B.M., Yu, J. and Wallace, J.K., "A Hybrid Ray Trace and Diffraction

Propagation Code for Analysis of Optical Systems," 1992 OELase Symposium, January 1992.
10. Silberstein, L., "Simplified Method of Tracing Rays Through Any Optical System," Longmans, Green and Co., London, 1918.
11. Goodman, J., "Introduction to Fourier Optics," McGraw Hill, 1968.
12. Born, M. and Wolf, E., Principles of Optics, Pergamon Press, 1980 (sixth edition).
13. Milman, M., "Working Paper on Optical Metrics," informal memo, January 1992.
14. Redding, D., Breckenridge, W., Lau, K., Sevaston, G., Levine, B.M. and Shaklan, S., "Segmented Mirror Figure Control for a Space-Based Far-IR Astronomical Telescope," Structures, Sensing and Control, J. Breakwell, V. Varadan, editors, Proc. SPIE 1489, April 1991.
15. Loboda, G., Masters thesis (in work).
16. Redding, D., "COMP Models of the WFPC-2," Memo DCR-91-25, May 27, 1991.
17. Wette, M., Milman, M. and Redding, D., "Experiences in Integrated Structure/Optics/Control Design Optimization," Joint NASA/SDIO Control/Structures Interaction Conference, Orlando, 1990.