Chapter 4

Linear Estimation Theory

- Virtually all branches of science, engineering, and social science for data analysis, system control subject to random disturbances or for decision making based on incomplete information call for estimation theory.
- Many estimation problem can be formulated as a minimum norm problem in Hilbert space.
- The projection theorem can be applied directly to the area of statistical estimation.
- There are a number of different ways to formulate a statistical estimation.
  - ⋄ Least squares.
  - ⋄ Maximum likelihood
  - ⋄ Bayesian techniques.
- When all variables are Gaussian statistics, these techniques produce linear equations.

Preliminaries

- If x is a real-valued random variable,
  - ◇ The *probability distribution* P of the variable x is defined to be
    $$P(\xi) = \text{Prob}(x \leq \xi).$$
  - ◇ The "derivative" $p(\xi)$ of the probability distribution $P(\xi)$ is called the *probability density function (pdf)* of the variable x, i.e.,
    $$P(\xi) = \int_{-\infty}^{\xi} p(x)dx.$$
    - ▷ Note that
      $$\int_{-\infty}^{\infty} p(x)dx = 1.$$
    - ▷ $p(\xi) \geq 0$ for all ξ.
- The expected value of any function g of x is defined to be
  $$\mathcal{E}[g(x)] := \int_{-\infty}^{\infty} g(\xi)p(\xi)d\xi.$$
  - ◇ The *expected value* of x is $\mathcal{E}[x]$.
  - ◇ The *variance* of x is $\mathcal{E}[(x - \mathcal{E}[x])^2]$.

- For random vector $\mathbf{x} = [x_1, \ldots, x_n]^\top$,
  - ◇ There is a *joint probability distribution* P defined by

$$P(\xi_1, \ldots \xi_n) = \text{Prob}(x_1 \leq \xi_1, \ldots, x_n \leq \xi_n).$$

  - ◇ The *covariance matrix* $\text{cov}(\mathbf{x})$ is defined by

$$\text{cov}(\mathbf{x}) = \mathcal{E}\left[(\mathbf{x} - \mathcal{E}[\mathbf{x}])(\mathbf{x} - \mathcal{E}[\mathbf{x}])^\top\right].$$

    - ▷ Two random variables x_i and x_j are said to be uncorrelated or stochastically independent if

$$\mathcal{E}[(x_1 - \mathcal{E}[x_1])(x_2 - \mathcal{E}[x_2])] = \mathcal{E}[x_1 - \mathcal{E}[x_1]]\mathcal{E}[x_2 - \mathcal{E}[x_2]].$$

Least Squares Model

- This is a familiar subject as we have seen in many occasions.
- This problem is not a statistical one.
- It amounts to approximating a vector $\mathbf{y} \in \mathbb{R}^m$ by a vector lying in the column space of $W \in \mathbf{R}^{m \times n}$ and $n < m$.
  - ◇ We assume a linear model that the response $\mathbf{y}$ is related to the input $\boldsymbol{\beta}$ linearly, i.e.,
  $$\mathbf{y} = W\boldsymbol{\beta}.$$
  - ◇ We would like to recover $\boldsymbol{\beta}$ from observed $\mathbf{y}$. (Would it be a linear relationship?)
  - ◇ We are not assuming that the observed $\mathbf{y}$ carries errors.
- It would be interesting to compare the least squares setting with those with random noises.

Least Squares Formulation

- Given
  - ◇ A known matrix $W \in \mathbb{R}^{m\times n}$, $n < m$.
  - ◇ An observation vector $\mathbf{y} \in \mathbb{R}^m$.

  Find $\hat{\boldsymbol{\beta}} \in \mathbb{R}^n$ such that $\|\mathbf{y} - W\hat{\boldsymbol{\beta}}\|$ is minimized over all $\boldsymbol{\beta} \in \mathbf{R}^n$.
- By the projection theorem, the solution exists and is unique.
- The normal equation is given by

$$W^\top(\mathbf{y} - W\hat{\boldsymbol{\beta}}) = 0.$$

- If W has linear independent columns, then

$$\hat{\boldsymbol{\beta}} = \underbrace{(W^\top W)^{-1}W^\top}_{K}\mathbf{y}.$$

  - ◇ Note that the optimal solution $\hat{\boldsymbol{\beta}}$ is related to $\mathbf{y}$ linearly.

Gauss-Markov Model

- A more realistic model in an experiment is

$$\mathbf{y} = W\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

  - ◇ $W \in \mathbb{R}^{m \times n}$ is known.
  - ◇ $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is a random vector with zero mean and covariance $\mathcal{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = Q$.
  - ◇ $\mathbf{y}$ represents the outcome of inexact measurements in $\mathbb{R}^m$.

- Want to estimate unknown parameter vector $\boldsymbol{\beta} \in \mathbb{R}^n$ from $\mathbf{y} \in \mathbb{R}^m$ using

$$\hat{\boldsymbol{\beta}} := K\mathbf{y}$$

  with K an unknown matrix in $R^{n \times m}$.

- Suppose the approximation is measured by minimizing the expected value of the error, i.e.,

$$\min_{K \in \mathbb{R}^{n \times m}} \mathcal{E}[\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2]$$

  - ◇ Since $\mathbf{y}$ carries random noise, it is a random vector.
  - ◇ Both estimate $\hat{\boldsymbol{\beta}}$ and the difference $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ are random vectors.
  - ◇ The statistics of these random vectors are determined by those of $\boldsymbol{\epsilon}$ and K.

Gauss-Markov Estimate

- Observe that

$$\begin{aligned}\mathcal{E}[\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|^2] &= \mathcal{E}\left[\langle K(W\boldsymbol{\beta}+\epsilon)-\boldsymbol{\beta}, K(W\boldsymbol{\beta}+\epsilon)-\boldsymbol{\beta}\rangle\right] \\ &= \|KW\boldsymbol{\beta}-\boldsymbol{\beta}\|^2+\mathcal{E}[\langle K\epsilon, K\epsilon\rangle].\end{aligned}$$

- Consider unbiased estimation:
  - ◇ Observe

    $$\mathcal{E}[\hat{\boldsymbol{\beta}}]=\mathcal{E}[KW\boldsymbol{\beta}+K\epsilon]=KW\mathcal{E}[\boldsymbol{\beta}].$$

    It is expected that $KW=I_n$.
  - ◇ The problem now becomes, given a symmetric and positive definite matrix Q,

    $$\begin{aligned}\text{minimize}_{K\in\mathbb{R}^{n\times m}} \quad & \text{trace } KQK^\top \\ \text{subject to} \quad & KW=I_n.\end{aligned}$$

  - ◇ This is in the form of a standard minimum norm problem.

- The problem has a closed form solution.
  - The optimal solution is given by

$$K = \left(W^\top Q^{-1} W\right)^{-1} W^\top Q^{-1}.$$

  - The minimum-variance unbiased estimation of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(W^\top Q^{-1} W\right)^{-1} W^\top Q^{-1} \mathbf{y}.$$

  - The special case $Q = I_m$ is the classical least squares problem.
    - The classical least squares solution is providing the unbiased minimum-variance estimate of $\boldsymbol{\beta}$, if the perturbation presented in data is white noise.
- It can be argued that the above solution $\hat{\boldsymbol{\beta}}_i$ is the minimum-variance unbiased estimation of $\boldsymbol{\beta}_i$ for each individual i.
  - This is the true minimum-variance unbiased estimate.

Minimum-Variance Model

- Assume in the linear model
$$\mathbf{y} = W\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$
    - ◇ $W \in \mathbb{R}^{m\times n}$ is known.
    - ◇ $\boldsymbol{\epsilon} \in \mathbb{R}^m$ is a random vector with zero mean and covariance $\mathcal{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = Q$.
    - ◇ $\boldsymbol{\beta}$ is a random vector in $\mathbb{R}^n$ with known statistical information.
    - ◇ $\mathbf{y}$ represents the outcome of inexact measurements in $\mathbb{R}^m$.
- Want to estimate the unknown random vector $\boldsymbol{\beta} \in \mathbb{R}^n$ based on $\mathbf{y} \in \mathbb{R}^m$ using
$$\hat{\boldsymbol{\beta}} := K\mathbf{y}$$
where K is an unknown matrix in $R^{n\times m}$.
- The best approximation is measured by minimizing the expected value of the random error, i.e.,
$$\min_{K\in\mathbb{R}^{n\times m}} \mathcal{E}[\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2]$$

Minimum-Variance Estimate

- Assume $(\mathcal{E}[\mathbf{y}\mathbf{y}^T])^{-1}$ exists. Then the minimum-variance estimator of β is given by

$$\hat{\boldsymbol{\beta}} = \mathcal{E}[\boldsymbol{\beta}\mathbf{y}^T](\mathcal{E}[\mathbf{y}\mathbf{y}^T])^{-1}\mathbf{y}.$$

  - ◇ The estimate is independent of W and ϵ.

- Proof Is Interesting!
  - ◇ Write K in rows, i.e., $K = [\mathbf{k}_1^\top, \ldots \mathbf{k}_n^\top]^\top$.
  - ◇ $\mathcal{E}[\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2] = \sum_{i=1}^n \mathcal{E}[(\hat{\beta}_i - \beta_i)^2] = \sum_{i=1}^n \mathcal{E}[(\mathbf{k}_i^\top \mathbf{y} - \beta_i)^2]$.
    - ▷ Suffices to consider each individual term.
  - ◇ Let $f(\mathbf{y}, \beta_i)$ denote the (unknown) joint pdf of $\mathbf{y}$ and β_i.
    - ▷ Define

$$\begin{aligned} g(\mathbf{k}_i) &:= \mathcal{E}[(\mathbf{y}^\top \mathbf{k}_i - \beta_i)^2] \\ &= \int\int (\mathbf{y}^\top \mathbf{k}_i - \beta_i)^2 f(\mathbf{y}, \beta_i)\, d\mathbf{y}\, d\beta_i. \end{aligned}$$

    - ▷ Necessary condition is $\nabla g(\mathbf{k}_i) = 0$.

◇ Easy to see

$$\begin{aligned}\frac{\partial g}{\partial \mathbf{k}_{i,j}} &= \int\int 2(\mathbf{y}^\top \mathbf{k}_i - \beta_i)\mathbf{y}_j f(\mathbf{y}, \beta_i)\, d\mathbf{y}\, d\beta_i \\ &= 2\mathcal{E}[(\mathbf{y}^\top \mathbf{k}_i - \beta_i)\mathbf{y}_j].\end{aligned}$$

◇ Rewrite the necessary condition as

$$\begin{aligned}\mathcal{E}[\mathbf{y}(\mathbf{y}^\top \mathbf{k}_i - \beta_i)] &= 0, \quad \text{(for each } i\text{)} \\ \mathcal{E}[\mathbf{y}\mathbf{y}^\top]K^\top &= \mathcal{E}[\mathbf{y}\beta^\top], \quad \text{(in matrix form)} \\ K &= \mathcal{E}[\beta\mathbf{y}^\top](\mathcal{E}[\mathbf{y}\mathbf{y}^\top])^{-1}.\end{aligned}$$

- The estimate so far is *biased*, unless $\mathcal{E}[\boldsymbol{\beta}] = \mathcal{E}[\mathbf{y}] = 0$.
- In the general case where $\mathcal{E}[\boldsymbol{\beta}] = \boldsymbol{\beta}_0$ and $\mathcal{E}[\mathbf{y}] = \mathbf{y}_0$,

◇ The estimate should assume the form

$$\hat{\boldsymbol{\beta}} := K\mathbf{y} + \mathbf{b}.$$

◇ The minimum-variance estimate is given by

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \mathcal{E}[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)(\mathbf{y} - \mathbf{y}_0)^\top](\mathcal{E}[(\mathbf{y} - \mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0)^\top])^{-1}(\mathbf{y} - \mathbf{y}_0).$$

Equivalent Formulas

- Assume that

$$\begin{aligned} \mathcal{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top] &= Q \in \mathbb{R}^{m\times m}, \\ \mathcal{E}[\boldsymbol{\beta}\boldsymbol{\beta}^\top] &= R \in \mathbb{R}^{n\times n}, \\ \mathcal{E}[\boldsymbol{\epsilon}\boldsymbol{\beta}^\top] &= 0 \in \mathbb{R}^{m\times n}. \end{aligned}$$

- Then the minimum-variance estimate can be written as

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \underbrace{RW^\top(WRW^\top + Q)^{-1}}_{K}\mathbf{y} \\ &= \overbrace{(W^\top Q^{-1}W + R^{-1})^{-1}W^\top Q^{-1}}^{K}\mathbf{y}. \end{aligned}$$

  - ◇ The equivalence can be proved by direct substitution.
  - ◇ Check out the dimension of the inverse matrices involved in the two expressions.

Comparision with Gauss-Markov Estimate

- The Gauss-Markov estimate is

$$\hat{\boldsymbol{\beta}} = \left(W^\top Q^{-1} W\right)^{-1} W^\top Q^{-1} \mathbf{y}.$$

- The more subtle minimum-variance estimate is

$$\hat{\boldsymbol{\beta}} = (W^\top Q^{-1} W + R^{-1})^{-1} W^\top Q^{-1} \mathbf{y}.$$

- If $R^{-1} = 0$, the two estimates are idential.
  - ◇ What is meant by $R^{-1} = 0$?
  - ◇ *Infinite* variance of $\boldsymbol{\beta}$ in the more subtle estimate means that we have *absolutely no a priori knowledge* of $\boldsymbol{\beta}$ at all.
- When $\boldsymbol{\beta}$ is considered as a random variable, the size m of observations $\mathbf{y}$ does not need to be large.
  - ◇ $(WRW^\top + Q)^{-1}$ exists so long as Q is positive positive definite.
  - ◇ Every new measurement simply provides additional information which may modify the original estimate.

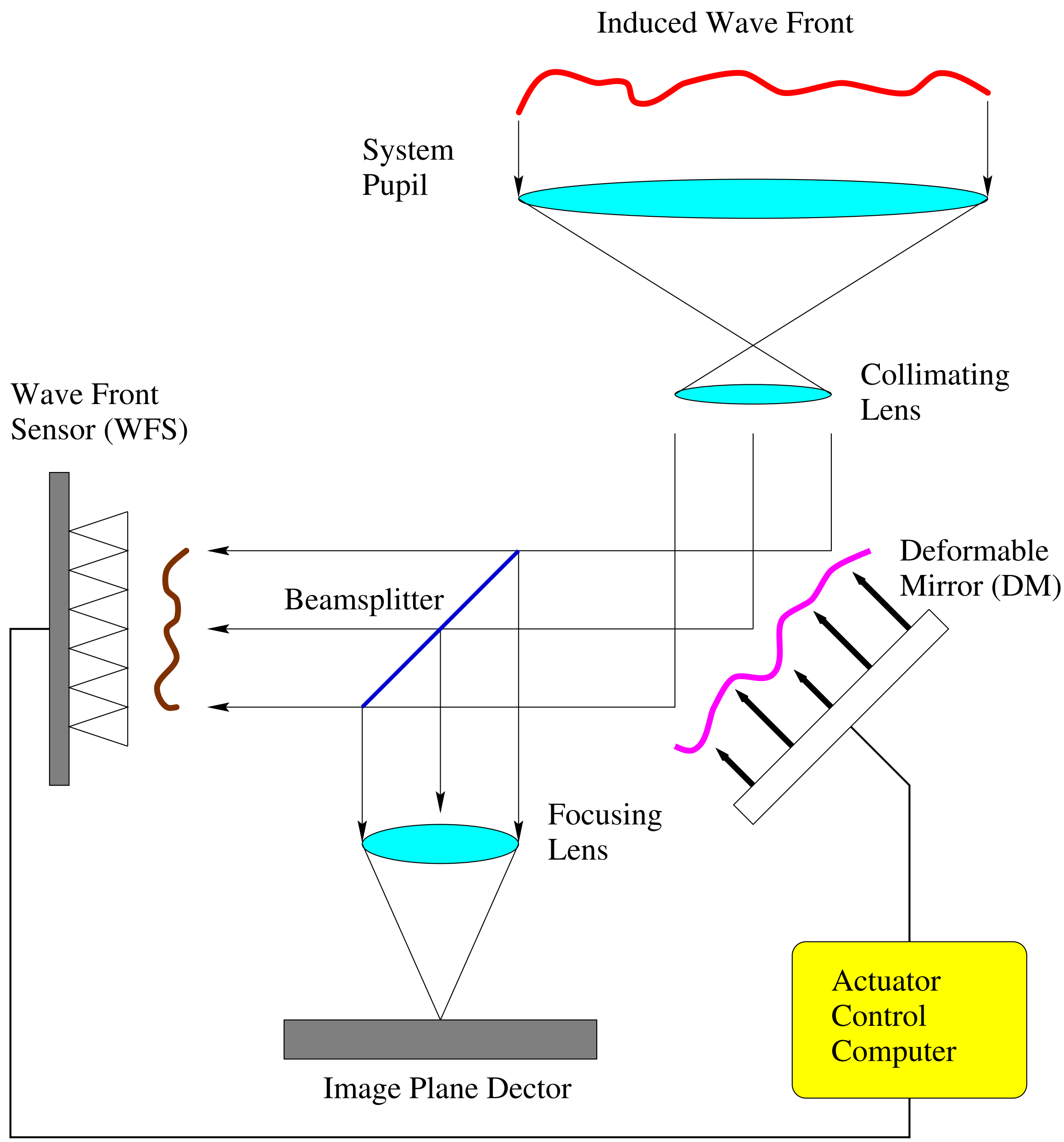
Application to Adaptic Optics

- Imaging through the Atmosphere
- Adaptive Optics System
    - ◇ Basic Relationships
    - ◇ Open-loop Model
    - ◇ Closed-loop Model
- Adaptive Optics Control
    - ◇ An Ideal Control
    - ◇ An Inverse Problem
    - ◇ Temporary Latency
- Numerical Illustration

Atmospheric Imaging Computation

- Purpose:
  - ◇ To compensate for the degradation of astronomical image quality caused by the effects of atmospheric turbulence.
- Two stages of approach:
  - ◇ Partially nullify optical distortions by a deformable mirror (DM) operated from a closed-loop adaptive optics (AO) system.
  - ◇ Minimize noise or blur via off-line post-processing deconvolution techniques (not this talk).
- Challenges:
  - ◇ Atmospheric turbulence can only be measured adaptively.
  - ◇ Need theory to pass atmospheric measurements to knowledge of actuating the DM.
  - ◇ Require fast performance of large-scale data processing and computations.

A Simplified AO System



Basic Notation

- Three quantities:
  - ◇ $\phi(t)$ = turbulence-induced phase profile at time t.
  - ◇ $a(t)$ = deformable mirror (DM) actuator command at time t.
  - ◇ $s(t)$ = wavefront slope sensor (WFS) measurement at time t and with no correction.
- Two transformations:
  - ◇ H := transformation from actuator commands to resulting phase profile adjustments.
  - ◇ G := transformation from actuator commands to slope sensor measurement adjustments.

From Actuator to DM Surface

- H is used to describe the DM surface change due to the application of actuators.
- $r_i(\vec{x})$ = influence function on the DM surface at position $\vec{x}$ with an unit adjustment to the ith actuator.
- Assuming m actuators and linear response of actuators to the command, model the DM surface by

$$\hat{\phi}(\vec{x}, t) = \sum_{i=1}^{m} a_i(t) r_i(\vec{x}).$$

  ◇ Sampled at n DM surface positions, can write

$$\boxed{\hat{\phi}(t) = Ha(t)}$$

  ▷ $H = (r_i(\vec{x}_j)) \in R^{n \times m}$.

  ▷ $\hat{\phi}(t) = [\hat{\phi}(\vec{x}_1, t), \ldots, \hat{\phi}(\vec{x}_n, t)]^T \in R^n$ = discrete corrected phase profile at time t.

From Actuator to WFS Measurement

- G is used to describe the WFS slope measurement associated with the actuator command a.
- Consider the H-WFS model where

$$s_j(t) := -\int d\vec{x}(\nabla W_j(\vec{x})\phi(\vec{x},t), \quad j = 1,\ldots,\ell.$$

  ◇ W_j = given specifications of jth subaperture.

- The measurement corresponding to $\hat{\phi}(\vec{x},t)$ would be

$$\hat{s}_j(t) = \sum_{i=1}^{m} \underbrace{\left(-\int d\vec{x}(\nabla W_j(\vec{x})r_i(\vec{x})\right)}_{G_{ji}} a_i(t).$$

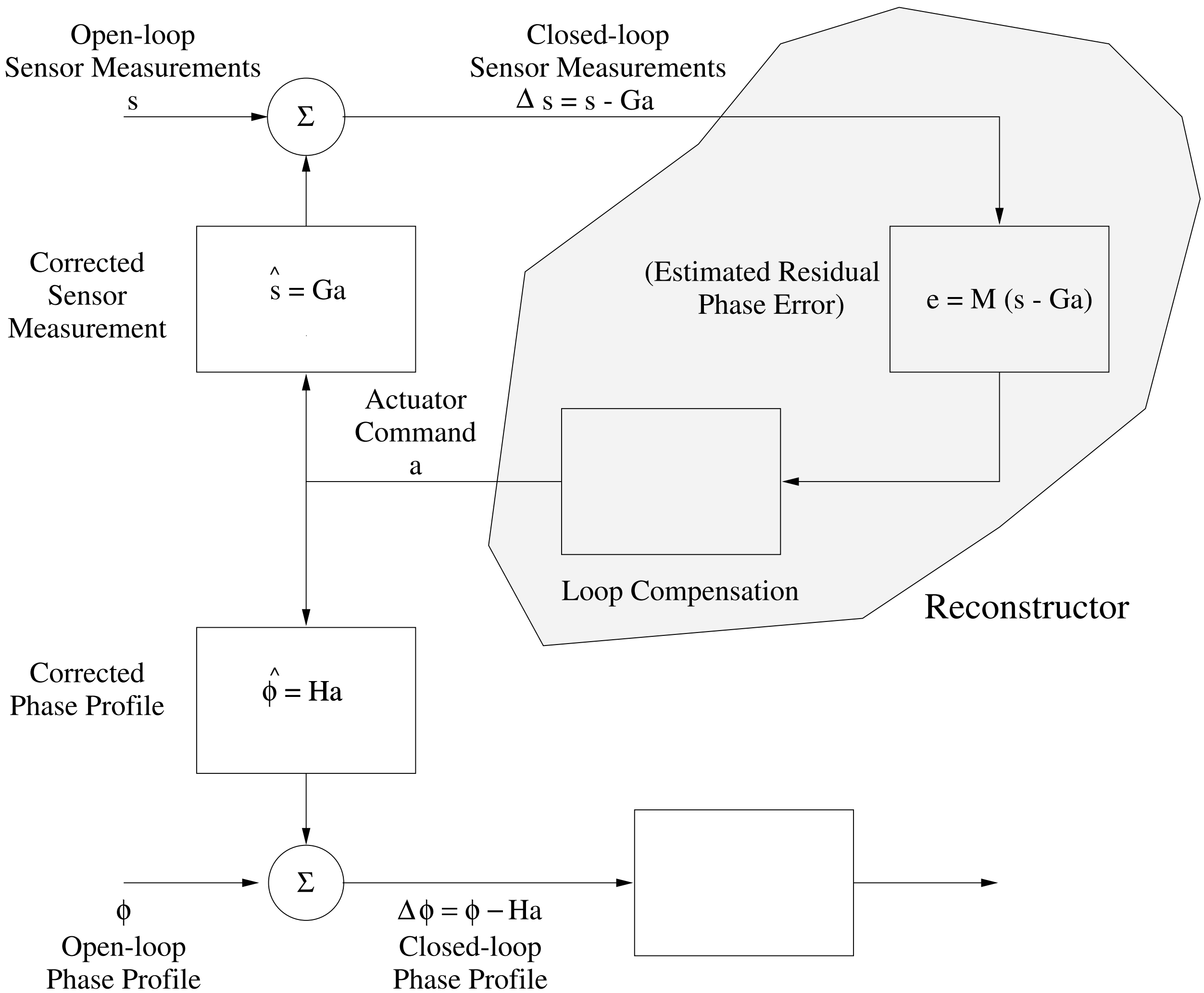
  ◇ Can write

$$\boxed{\hat{s}(t) = Ga(t)}$$

  where $G = [G_{ij}] \in R^{\ell \times m}$.

  ◇ The DM actuators are *not* capable of producing the exact wavefront phase $\phi(\vec{x},t)$ due to its finiteness of degrees of freedom. So $\hat{s} = Ga$ is not an exact measurement.

A Closed-loop AO Control Model



What is Available?

- Two residuals that are available in a *closed-loop* AO system:
  - ◇ $\Delta\phi(t) := \phi(t) - Ha(t)$
    - ▷ Represents the residual phase error remaining after the AO correction.
    - ▷ Also means instantaneous closed-loop wavefront distortion at time t.
  - ◇ $\Delta s(t) := s(t) - Ga(t)$
    - ▷ Represents feedback applied to $s(t)$ by DM actuator adjustment.
    - ▷ Also means *observable* wavefront sensor measurement at time t.
- In practice, there is a servo lag or delay in time Δt, i.e., it is likely
  - ◇ $\Delta\phi(t) := \phi(t) - Ha(t - \Delta t)$.
  - ◇ $\Delta s(t) := s(t) - Ga(t - \Delta t)$.

  Thus the data collected are not perfect.

Open-loop Model

- Assume a linear relationship between open-loop WFS measurement s and turbulence-induced phase profile ϕ:
$$\boxed{s = W\phi + \epsilon}. \tag{1}$$
  - ϵ = measurement noise with mean zero.
  - In the H-WFS model, W represents a quadrature of the integral operator evaluated at designated positions $\vec{x}_j$, $j = 1, \dots n$.
- Want to estimate ϕ using $\tilde{\phi}$ from the model
$$\tilde{\phi} = E_{open} s$$
so that the variance
$$\mathcal{E}[\|\phi - \tilde{\phi}\|^2]$$
is minimized.
  - The wave front reconstruction matrix E_{open} is given by
$$E_{open} = \mathcal{E}[\phi s^T](\mathcal{E}[ss^T])^{-1}.$$
  - For unbiased estimation, need to enforce the condition that $E_{open}W = I$.

Closed-loop Model

- For the H-WFS model, it is reasonable to assume the relationship

$$WH = G. \tag{2}$$

- Then

$$\begin{aligned} s &= W\phi + \epsilon \\ &= W(Ha + \Delta\phi) + \epsilon \\ &= WHa + (W\Delta\phi + \epsilon). \end{aligned}$$

It follows that

$$\boxed{\Delta s = W\Delta\phi + \epsilon}. \tag{3}$$

  - ◇ The closed-loop relationship (3) is identical to the open-loop relationship (1).

- Can estimate the residual phase error $\Delta\phi(t)$ using $\Delta\tilde{\phi}(t)$ from the model

$$\Delta\tilde{\phi} = E_{closed}\Delta s$$

  - ◇ E_{closed} = wavefront reconstruction matrix.
  - ◇ For unbiased estimation, it requires that $E_{closed}W = I$. Hence

$$E_{closed}G = e_{closed}(WH) = H.$$

Actuator Control

- An Ideal Control:
  - $\Delta\phi$ = residual error after DM correction by current command a_c.
  - New command a_+ should reduce the residual error, i.e., want to

$$\min_{a} \|Ha - \phi\|.$$

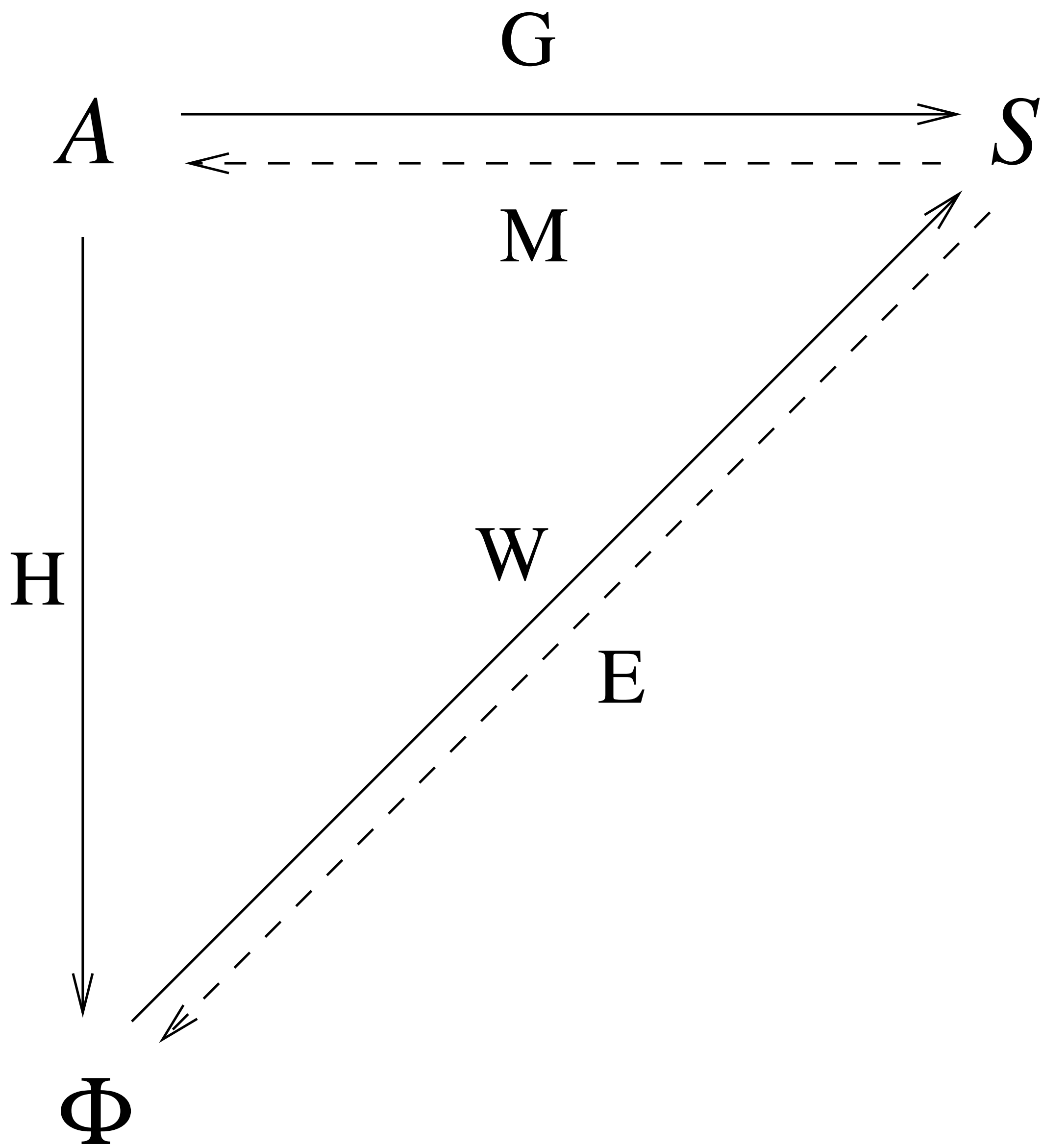
  - Define $\Delta a := a_+ - a_c$, then want to

$$\min_{\Delta a} \|H\Delta a - \Delta\phi\|.$$

  - But $\Delta\phi$ is not observable directly. It has to be estimated from Δs.
- Estimating Δa directly from Δs:

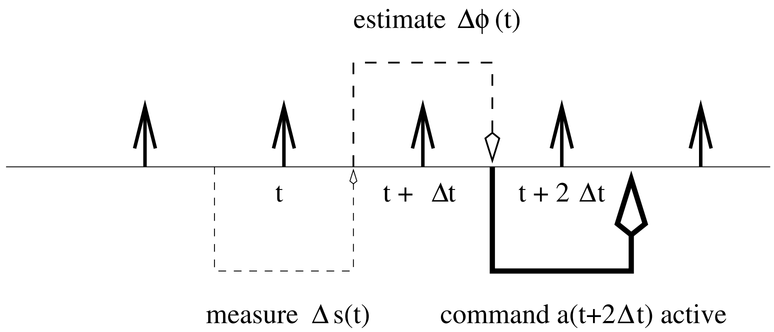
$$\boxed{\Delta a = M\Delta s} \tag{4}$$

An Inverse Problem



Actuator Control with Temporary Latency

- Due to finite bandwidth of the control loop, Δs is not immediately available.
- Time line for the scenario of a 2-cycle delay,



- ARMA control scheme:

$$a(t+2\Delta t) := \sum_{k=0}^{p} c_k a(t+(1-k)\Delta t) + \sum_{j=0}^{q} b_j M_j \Delta s(t - j\Delta t).$$

$$\boxed{a^{(r+2)} = \sum_{k=0}^{p} c_k a^{(r+1-k)} + \sum_{j=0}^{q} b_j M_j \Delta s^{(r-j)}, r = 0, 1, \ldots}$$

Expected Effect on the AO System

- Suppose
  - ⋄ $\exp[s(t)]$ is independent of time t throughout the cycle of computation.
  - ⋄ Matrix $\sum_{j=0}^{q} b_j M_j$ is of full column rank.
- Then
  - ⋄ The WFS feedback measurement $\Delta s^{(n)}$ is eventually nullified by the actuators, i.e.,

$$\exp[s] = G \lim_{n\to\infty} \exp[a^{(n)}].$$

  - ⋄ The expected residual phase error is inversely related to the expected WFS measurement noise ϵ via

$$0 = W \lim_{n\to\infty} \exp[\Delta\phi^{(n)}] + \exp[\epsilon].$$

- Compare with the ideal control:
  - ⋄ Even if $\exp[\epsilon] = 0$, not necessarily $\exp[\| \lim_{n\to\infty} \Delta\phi_n \|^2]$ will be small because W has non-trivial null space.

Almost Sure Convergence

- Each control $a^{(r+j)}$ is a random variable $\Longrightarrow$ The control scheme is a stochastic process.
- Each control $A^{(r+j)}$ is also a realization of the corresponding random variable $\Longrightarrow$ The control scheme is a deterministic iteration.
- Convergence of deterministic iteration on independent random samples $\Longrightarrow$ Almost sure convergence of stochastic process.
- Need fast convergence:
  - ◇ Stationary statistic is not realistic.
  - ◇ Atmospheric turbulence changes rapidly.
  - ◇ Can only assume stationary statistic for a short period of time.

- Define

$$\begin{aligned} \mathbf{a}_{r+2} &:= [a^{(r+2)}, a^{(r+1)}, \ldots a^{(r-q+1)}]^T, \quad r = 0, 1, \ldots \\ \mathbf{b} &:= [\sum_{j=0}^{q} b_j M_j G s', 0, \ldots, 0]^T. \end{aligned}$$

- The ARMA scheme becomes

$$\mathbf{a}_{r+2} = A\mathbf{a}_{r+1} + \mathbf{b}$$

where A is the $m(q+2) \times m(q+2)$ matrix

$$A := \begin{bmatrix} c_0 I_m & c_1 I_m - b_0 M_1 G & \ldots & & c_{q+1} I_m - b_q M_q G \\ I_m & 0 & & & 0 \\ 0 & I_m & & & \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \ldots & I_m & 0 \end{bmatrix}$$

- Almost convergence $\Longleftrightarrow$ Spectral radius $\rho(A)$ of A is less than one.
- Asymptotic convergence factor is precisely $\rho(A)$.

Numerical Simulation

- Consider the 2-cycle delay scheme

$$a(t+2\Delta t) = a(t+\Delta t) + 0.6H^{\dagger}W^{\dagger}\Delta s(t).$$

- Test data:

surface positions n	$=$	5
number of actuators m	$=$	4
number of subapertures ℓ	$=$	3
size of random samples z	$=$	2500
H	$=$	$rand(n,m)$
W	$=$	$rand(\ell,n)$
G	$=$	WH
L_ϕ	$=$	$rand(n,n)$
L_ϵ	$=$	$diag(rand(\ell,1))$
μ_ϕ	$=$	$zeros(n,1)$
μ_ϵ	$=$	$zeros(\ell,1)$

- Random samples:

$$\begin{aligned} \phi &= \mu_\phi * ones(1,z) + L_\phi * randn(n,z), \\ \epsilon &= \mu_\epsilon * ones(1,z) + L_\epsilon * randn(\ell,z). \end{aligned}$$