TRACTABILITY OF COMPLEX CONTROL SYSTEMS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
AERONAUTICS & ASTRONAUTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Laurent Lessard

August 2011

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Sanjay Lall, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Matthew West, Co-Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Stephen Boyd

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Stephen Rock

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

Abstract

This thesis is divided into two main parts. In the first part, we consider the problem of efficiently computing wavefront estimates for use in adaptive optics hardware on ground-based telescopes.

Our contribution is to enable wavefront estimation for future large telescopes. To this effect, we have developed a warm-started single-iteration multigrid algorithm that performs as well as conventional vector-matrix-multiplication methods, but at a fraction of the computational cost. We used numerical simulations to compare our algorithm to a variety of other published methods, and validated our findings at the Palomar Observatory.

In the second part, we consider feedback control subject to an information constraint. Such problems are called *decentralized*, and are not always tractable. Using a novel algebraic framework, we are able to prove many structural results, including a new convexity result, in a natural and purely algebraic way. This framework is particularly well-suited for analyzing systems described by rational transfer functions.

We also develop a new condition called *internal quadratic invariance*, a condition under which the controller synthesis can be cast as a convex optimization problem. This describes the most general class of tractable decentralized control problems known to date. The key insight is that the system's representation is not unique; and choosing the right representation can make determining tractability easier.

Both parts of the thesis fit into the broader question of tractability of complex systems. In the first part we look at a practical example which is difficult because of the large number of sensors and actuators. In the second part, we look at decentralized control, which is difficult because of the non-classical information constraint.

Acknowledgments

I would like to thank my principal advisor Sanjay Lall and my co-advisor Matthew West for their continued help and guidance during my time at Stanford. Sanjay's knowledge of control theory is truly impressive. He introduced me to decentralized control, and encouraged me to broaden my mathematical horizons. Not only was this journey incredibly enriching, but Sanjay's calm and friendly attitude made it very enjoyable as well. I consider myself lucky to have been able to collaborate with Sanjay in my research at Stanford.

Matt co-advised me while I did adaptive optics research, and he has been incredibly helpful. This project required tools beyond control theory such as multiscale algorithms and large-scale computation; and Matt was an expert in both. His ability to take a step back from the problem and see connections with other areas of research such as distributed systems or computational fluid dynamics was an inspiration, and instrumental to our success.

I would also like to thank our other collaborators for the adaptive optics project. Doug MacMynowski's (CalTech) expertise in telescope hardware and software was invaluable. Thanks also go out to Antonin Bouchez (CalTech) and Jenny Roberts (JPL), who helped us perform the experiments of Chapter 4 at the Palomar Observatory.

I would like to thank the the other members of my reading committee, Stephen Rock and Stephen Boyd. Prof. Rock's courses on dynamics and control are what got me interested in doing controls research, and I likely would not have chosen this path had it not been for his enthusiastic and captivating teaching style. Prof. Boyd's courses also played a key role. His courses on linear systems and convex optimization formed the theoretical foundation upon which most of my research has been based.

Thanks also go out to Prof. Thomas Weber, who was the chair for my defense committee. Despite the short notice, Thomas took great interest in my research and had very insightful and constructive comments for me both before and after the defense.

My Stanford labmates and colleagues were also an integral part of my research experience. I would like to thank: Sachin Adlakha, John Swigart, Chung-Ching Chang, Jong-Han Kim, Hyung Sik Shin, and Jeff Wu for many useful discussions and lab meetings.

I would like to dedicate this thesis to my parents Barbara and André. They have always encouraged me to find things I am passionate about, and have been incredibly loving and supportive. Thank you so much!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As technology progresses, the engineering systems that surround us become increasingly complex. Many of these systems inherently involve some form of control. Examples include the electronic shock-absorption used in your car, or the electronic fuel injection in a jet engine. Control may also need to happen on a more global scale. Examples include packet routing on the internet, or load balancing on the Western Interconnect power grid. Enabled by increasingly powerful computers and driven by ever-increasing user demand, these systems continue to grow in size and intricacy. This growth can pose a challenge in several different ways, and we examine two particular cases in this thesis.

System Size. Depending on the type of problem we are solving, the best methods currently available do not always scale well with the size of the problem. The classical example is the *traveling salesman* problem, where one must find a tour of n cities that returns to the starting point and minimizes the total distance traveled. The only known way to solve such a problem is to enumerate each possible path and test it. This problem scales very poorly with n. Indeed, if it takes one millisecond to solve the problem for $n = 10$, it will take 6 seconds to solve it for $n = 15$, and roughly 21 years to solve it for $n = 20$.

In Chapters 2–4, we explore the scalability of a particular example: adaptive optics (AO) algorithms for use in ground-based telescopes. In this application, hundreds of measurements are taken in real-time and used to cancel out image distortion caused by atmospheric turbulence. The computation required for the conventional AO algorithm scales as $\mathcal{O}(n^2)$, where n is the number of sensors. This isn't as bad as the traveling salesman problem, which scales as $n!$, but it is nevertheless a problem. We develop an approximate estimation algorithm which scales as $\mathcal{O}(n)$, and we show through numerical simulation and experiment on a real telescope that it achieves performance indistinguishable from that of the conventional algorithm. A more in-depth introduction to adaptive optics can be found at the beginning of Chapter 2.

Decentralization. Complexity can also arise if the controller is decentralized. This means that there are actually several controllers, each with access to a different subset of the available information, and each responsible for a subset of the decisions that must be made. This is an unavoidable feature of very large systems such as wireless networks, where there are delays in transmitting messages between nodes. Often, decisions must be made without full knowledge of the current state of the system.

Decentralized problems are hard in general. Witsenhausen showed in 1968 that even a simple problem involving two agents can be completely intractable. Characterizing tractable architectures is important because it will help guide future decentralized designs and ensure that we are able to analyze the systems we build. While a complete characterization of tractable decentralized problems has yet to be discovered, our research takes important steps towards this goal.

In Chapters 5–7, we develop a new mathematical framework that we use to analyze decentralized control problems. This framework provides an intuitive way to explain existing tractability results, and gives us new results as well. The key insight is that a system may have several different yet equivalent mathematical representations. By choosing the right one, determining tractability is made easier. This observation leads us to a new class of tractable decentralized problems we call *internally quadratically invariant*. This class encompasses all previously known classes, and includes new problems which were previously not known to be tractable. A more in-depth introduction to decentralized control can be found at the beginning of Chapter 5.

Chapter 2

Adaptive Optics

2.1 Introduction

Adaptive optics (AO) is a technology used in ground-based telescopes that greatly improves the image quality seen by the telescope. Even in a ideal telescope using perfectly smooth mirrors and lenses, there are two main sources of image aberration: diffraction and turbulence. Diffraction is an optical effect characterized by the shape of the telescope aperture and the wavelength of the light. In a *diffraction-limited* arrangement, point sources such as stars appear as bright spots surrounded by concentric rings. This is a fundamental limit of telescopes, and can be mitigated by using a larger aperture diameter, or by observing a shorter wavelength of light.

Atmospheric turbulence also leads to image distortion. Earth's atmosphere does not have a uniform temperature distribution, and the local temperature variations are constantly changing and moving due to diffusion, solar heating, and air currents. The temperature differences induce local changes in the index of refraction, and cause distortion. The air in the upper atmosphere is particularly turbulent, and this causes the distortions observed at the telescope aperture to change frequently. Since telescopes typically take long-exposure photographs in order to collect a sufficient amount of light, the distortions get averaged and have a blurring effect. When turbulence dominates over diffraction, we say that the telescope is *seeing-limited*.

Ground-based optical telescopes with apertures larger than 10–20 cm are typically seeing-limited. Due to the fast-changing atmosphere and large exposure times required, one cannot use post-processing techniques to correct for turbulence effects. AO works in real-time by adjusting the image 100–1000 times per second using fast sensors to estimate the distortion, and a deformable mirror to correct the wavefront.

The process is illustrated in Figure 2.1. The purple dashed square contains the feedback control loop. The distorted wavefront enters on the left and gets corrected by the deformable mirror. Part of the corrected light is split off and sent to a wavefront sensor, which sends its measurements to a controller. The controller (K) performs the *reconstruction* step: estimating the shape of the wavefront from measurements. The focus of our research is to make this computationally intensive step more efficient. When reconstruction is complete, the actuator commands are sent back to the deformable mirror. More detail is given in Section 3.2.

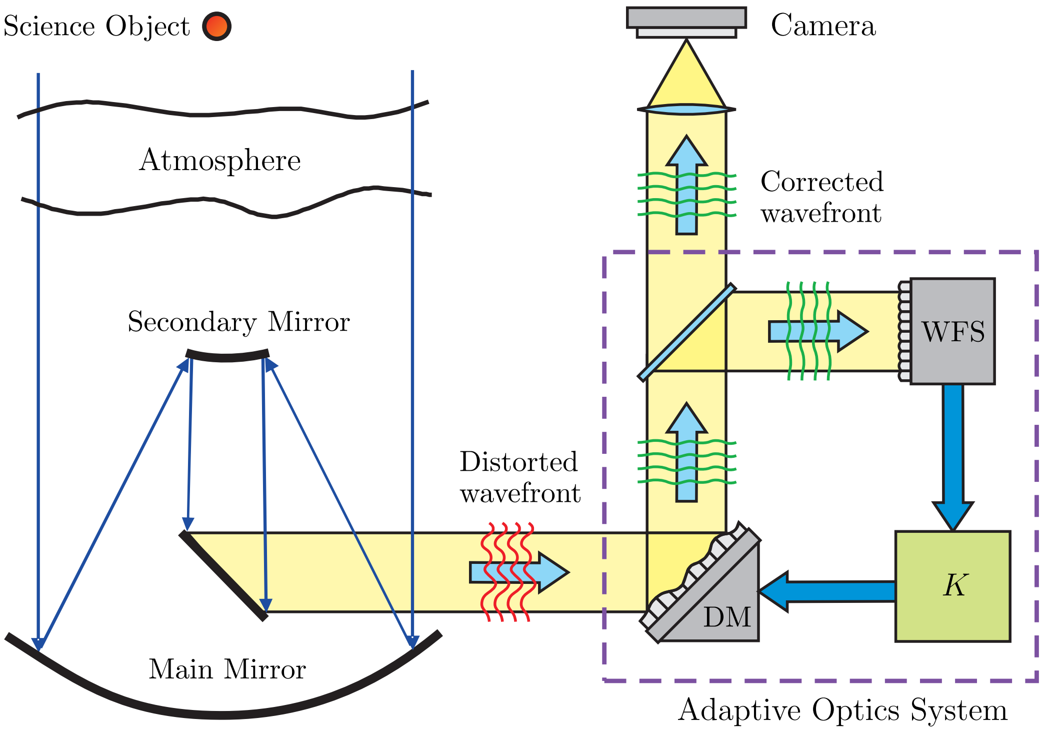

Figure 2.1: Diagram illustrating adaptive optics (AO). The image aberration is estimated using a wavefront sensor (WFS) and passed to a controller (K) which sends the appropriate commands to the deformable mirror (DM).

2.2 Computational Challenge

AO systems have a large number of inputs and outputs. The wavefront sensor (WFS) is an array of small square lenses that provide local gradient measurements of the incoming wavefront. The deformable mirror (DM) is an array of individually actuated mirrors that can be used to reconstruct the shape of the distorted wavefront. In practice a very simple model is used: all atmospheric dynamics are neglected and a static estimation problem is solved at every timestep. Namely, the two-dimensional wavefront shape must be reconstructed from local measurements. The optimal estimation gain matrix is pre-computed offline, and estimation is carried out in real-time by performing a full vector-matrix-multiplication (VMM) at every timestep.

The estimation gain matrix is dense in general, so the VMM step is computationally expensive, and scales with the square of the number sensors and actuators. A further difficulty is that the timesteps must be made short to achieve good performance, typically 1-10 milliseconds. Small AO systems (10^2–10^3 sensors and actuators) have been built and used with great success, but future systems will be much larger (10^4–10^5 sensors and actuators). Future telescopes may also implement so-called multi-conjugate adaptive optics (MCAO) that would further increase the number of sensors, actuators, and computation required.

In this research, we have enabled efficient wavefront estimation for future large AO systems by finding scalable reconstruction algorithms that perform as well as the optimal estimators currently used in practice. In Chapter 3, we compare the leading proposed solutions using computer simulations, and we show that using a *warm-start* technique yields comparable performance with huge computational cost savings [17]. In Chapter 4, we validate our warm-start technique on the 3.1 meter telescope at the Palomar Observatory. We show that our efficient reconstruction algorithm performs as well as the optimal reconstructor used with VMM, and provides a considerable computational savings [16].

Chapter 3

Algorithms

3.1 Literature Review

Since the original paper on AO reconstruction [30], there has been much effort invested in accelerating the estimation task. One way to do this is to observe that the dense estimation gain matrix is (to good approximation) the inverse of a sparse matrix. Thus, there is promise that iterative methods could find the estimates efficiently. Sparse matrix factorization methods [4] have been used with a conjugate-gradient iterative scheme paired with either a multigrid [11, 8] or Fourier [31] preconditioner. This provides convergence in only a small number of iterations. Other methods of acceleration include a Fourier-domain reconstruction [21], a blended Fourier/PCG method [24], and a local control approach [18].

These methods are typically simulated in open-loop using a *quasi-static* assumption: a single phase screen is used to generate a measurement, construct an estimate, and compute the error. In other words, an independent estimation problem is solved for every measurement. Since measurements are obtained at high rates (typically up to 1 kHz), these methods do not take advantage of the fact that the atmosphere has some slow dynamics that do not change from one timestep to the next. In other words, there could be some value in predicting the next estimate given the current one.

More complicated temporal atmospheric models have been developed to address this issue. A popular model is the Taylor frozen-flow approximation, which assumes the atmosphere is composed of stacked translating layers. State-space representations of this model have been proposed, which have led to formulations using the theory of optimal control [7, 20, 22]. However, these methods require the layer wind velocities be either estimated or known a priori.

Another possibility is to use a spectral decomposition. The phase is projected onto a basis such as the Zernike polynomials, and each mode is modeled separately [19, 15]. While the spatiotemporal statistics produced are correct, the associated cost of solving discrete algebraic Riccati equations and storing large dense covariance matrices is very high. Recent work by Poyneer et al. [22] avoids this problem by using a modal decomposition to decouple the Riccati equation, thereby greatly reducing the cost.

We compared the computational performance of 15 iterative reconstructors by running numerical simulations in both open-loop and closed-loop of a large single-conjugate adaptive optics system (SCAO). The sensor sampling rate is chosen by examining the trade-off between sampling rate and minimum achievable estimation error.

We also examined the benefits of warm-starting, where the most recent estimate is stored and used to initialize the subsequent iteration. The alternate approach is to cold-start, where the iterative schemes are initialized at zero. We show that although the iterative reconstructors may converge at different rates in open-loop with a cold-start, they all require a single iteration per timestep in closed-loop. The best methods are the ones with the cheapest cost per iteration.

Similar solution techniques apply to the multi-conjugate (MCAO) case, but the results are different. In MCAO, minimum variance reconstructors (MVR) or some other type of regularization must be implemented to achieve acceptable performance. In SCAO, there is no benefit to using MVR. A simple least-squares reconstructor has a virtually identical performance.

Standard closed-loop techniques give rise to stability problems in MCAO [9, 6]. We tried three different closed-loop architectures in SCAO, and found them all to be stable for every method tested. We also show that the FD-PCG algorithm [31, 10] performs well in the SCAO case.

In Section 3.2, we describe the system model (geometry, sensors, and noise). In Section 3.3, we discuss least-squares and minimum-variance reconstruction. In Section 3.4, we discuss iterative schemes used to accelerate reconstruction. Finally, in Section 3.5, we analyze computational performance through simulation using a Taylor frozen-flow model and a square 128 × 128 sensor/actuator array.

3.2 System Model

3.2.1 Sensor and Actuator Geometry

In optical telescopes, the aperture is typically annular so the deformable mirror (DM) and sensor array also share this shape. The algorithms we will discuss use sparse matrix operations and do not rely on a particular choice of geometry. Thus, for simplicity, we will assume a square sensor and actuator grid. The actuators lie at the vertices of an $N \times N$ grid and the sensors are aligned with the centers of the faces. This arrangement is called the *Fried geometry*, and is illustrated in Figure 3.1. If the sensors and actuators are collocated, we call it a *Hudgin geometry*. For many implementations, the wavefront shape is estimated at a resolution greater than that afforded by the DM so a *fitting* step is required to find the best DM commands. In this work, we assume the resolutions match so there is no fitting step.

3.2.2 Measurement Equation

The most commonly used wavefront sensors measure either local gradient or local curvature in the incident wavefront. We will assume an array of Shack-Hartmann sensors, which produce gradient measurements in both transverse directions. The measurements can be written as a linear function of the phases at each of the four nearest actuator locations plus Gaussian sensor noise [12]. For example, referring to

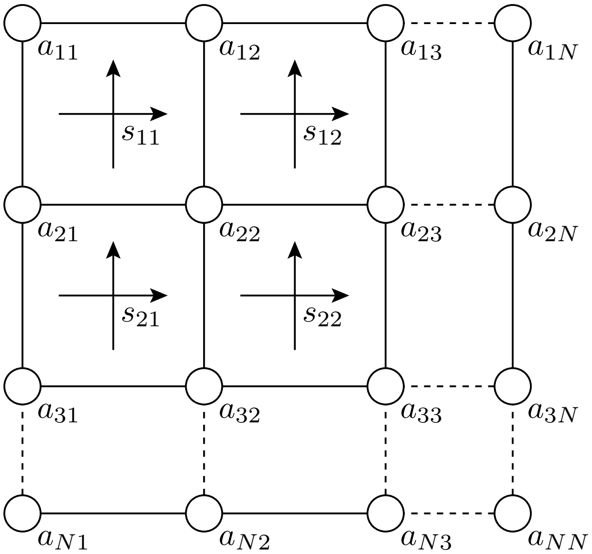

Figure 3.1: Fried Geometry. The sensors (s_{ij}) are at the centers of the faces, and the actuators (a_{ij}) are at the vertices. We measure a noisy gradient of the phase at s_{ij}, and the goal is to estimate the phase at a_{ij}.

Figure 3.1,

$$y(s_{12}) = \frac{1}{2}\begin{bmatrix}\phi(a_{13}) - \phi(a_{12}) + \phi(a_{23}) - \phi(a_{22}) \\ \phi(a_{13}) + \phi(a_{12}) - \phi(a_{23}) - \phi(a_{22})\end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

If we collect all the phases into a vector ϕ arranged in a column-major ordering: $\phi_k = \phi(a_{ij})$ where $k = i + (j-1)N$, and we do the same for y and v, we can write a linear equation relating the phase offsets to the measurements:

$$y = G\phi + v,$$

where $\phi \in \mathbb{R}^{N^2}$, $y \in \mathbb{R}^{2(N-1)^2}$, and v is a vector of zero-mean independent identically distributed Gaussian random variables with $\operatorname{cov} v = \sigma^2 I$. Note that G is a sparse matrix (four nonzero entries per row). From now on, let $n = N^2$ denote the length of the vector ϕ.

The sensor measurements y only depend on relative phase measurements. To take advantage of this, we assume each set of phases is translated such that the average value is zero. In other words, define: $x = (I - \frac{1}{n}11^T)\phi$ where 1 is the $n \times 1$ vector of ones. This is equivalent to removing the *piston mode*, further discussed in Section 3.3. The same linear equations hold:

$$y = Gx + v. \tag{3.1}$$

In the next section, we will discuss how to model the noise variance σ^2.

3.2.3 Noise Model

What follows is a summary of the Shack-Hartmann sensor noise model developed in [12, §5.3]. The sensor noise v is Gaussian to a good approximation, but its variance σ^2 depends on a variety of factors, including the guide star brightness and the noise in the CCD detectors:

$$\sigma^2 = \left(\frac{3\pi^2 K_g}{8}\right)^2 \frac{n_{\mathrm{ph}} + n_{\mathrm{bg}} + N_D\sigma_e^2}{n_{\mathrm{ph}}^2}, \tag{3.2}$$

where n_{ph} and n_{bg} are the expected number of signal and background photo-electrons hitting a single sensor per sampling interval, and σ_e is the RMS read-out error in each of the N_D CCD detector pixels forming each sensor. K_g is a correction factor that accounts for the small gaps between the sensors. Equation (3.2) has this particular form because the number of photons hitting a sensor is distributed as a Poisson process. The number of photo-electrons is related to the sample rate via:

$$n_{\mathrm{ph}} = \frac{\eta\mu_{\mathrm{ph}}A}{b} \qquad n_{\mathrm{bg}} = \frac{\eta\mu_{\mathrm{bg}}A}{b}, \tag{3.3}$$

where b is the sample rate in Hz, and μ_{ph} and μ_{bg} are photon fluxes in photons per square meter per second, A is the portion of the area of the main mirror that projects onto a single sensor subaperture (in m^2), and η is the product of the quantum efficiency of the CCD and the optical efficiency of the various mirrors and filters. Combining

(3.2) and (3.3):

$$\sigma^2 = \left(\frac{3\pi^2 K_g}{8}\right)^2 \left(\frac{\eta(\mu_{\text{ph}} + \mu_{\text{bg}})A}{b} + N_D\sigma_e^2\right)\left(\frac{b}{\eta\mu_{\text{ph}}A}\right)^2. \tag{3.4}$$

This equation relates the sensor noise variance σ^2 to the sample rate b and the photon arrival rates μ_{ph} and μ_{bg}, which are a function of the guide star brightness. Note that increasing the sampling rate makes the sensors noisier, and using a brighter guide star makes the sensors more accurate.

3.3 Wavefront Reconstruction

3.3.1 Least-Squares Reconstruction

The objective is to estimate the phase at each of the actuator locations, so that we may send this information to the deformable mirror and cancel the aberration. Only the relative phase is meaningful, so we can estimate x instead of ϕ. From (3.1):

$$y = Gx + v.$$

Since the system is overdetermined (roughly twice as many sensor measurements as actuators), one can minimize the norm of the residual. This is known as *least-squares reconstruction*. However, a regularization must be performed in part because G is not full-rank: the *piston* (constant) and *waffle* (checkerboard) modes are in the nullspace of G and hence are unobservable. We construct phase estimates with a zero waffle mode because the waffle mode is small in practice, and we also make the piston mode zero because only relative phase offsets matter. If we let V denote the $n \times 2$ matrix whose columns are the normalized piston and waffle modes, the problem becomes to find $\hat{x}$ in order to

$$\begin{array}{ll} \text{minimize} & \|y - G\hat{x}\|^2 \\ \text{subject to} & V^T\hat{x} = 0. \end{array}$$

The solution to this problem is:

$$\hat{x} = \left(G^T G + V V^T\right)^{-1} G^T y.$$

We can compute $\hat{x}$ by solving the linear system:

$$\left(G^T G + V V^T\right) \hat{x} = G^T y. \tag{3.5}$$

3.3.2 Minimum-Variance Reconstruction

In order to better estimate x, one must know something about its prior distribution. The first complete summary of the theory of propagation through atmospheric turbulence is due to Tatarskii [26], based on the assumption of a Kolmogorov power spectral density (PSD) for spatial phase distribution: $\Phi(k) \propto k^{-11/3}$. The model is widely accepted because its predictions agree well with experimental evidence.

The phase ϕ has infinite variance because the PSD is unbounded at zero. However, the piston-removed phase x is normally distributed with zero mean and finite covariance C. The inverse of this covariance matrix can be approximated by a product of sparse matrices:

$$C^{-1} \approx L L^T, \tag{3.6}$$

where L is proportional to a discretization of the Laplacian operator. This procedure was originally applied to a Hudgin geometry [4] but works just as well for Fried geometry [29] as long as we modify the discretized Laplacian accordingly. In both cases, the correct choice is that L be proportional to $G^T G$. The piston and waffle modes are therefore still unobservable. We can estimate x by minimizing the conditional mean square error. We must find $\hat{x}$ in order to

$$\begin{array}{ll} \text{minimize} & \mathrm{E}\left(\|x - \hat{x}\|^2 \,\middle|\, y\right) \\ \text{subject to} & V^T \hat{x} = 0, \end{array}$$

where x is a zero-mean random variable with covariance matrix C. The solution to this problem is:

$$\hat{x} = \left(G^T G + \sigma^2 C^{-1} + V V^T\right)^{-1} G^T y.$$

This is known as *minimum-variance reconstruction*. We can compute $\hat{x}$ by using the approximation in (3.6) and solving the linear system:

$$\left(G^T G + \sigma^2 L L^T + V V^T\right) \hat{x} = G^T y. \tag{3.7}$$

This is very similar to the least-squares solution. Indeed, the solutions are identical in the case of zero sensor noise.

Note that the original formulation of minimum-variance reconstruction [4] takes into account both the fitting and estimation steps. Since we neglect the fitting step and all statistics are Gaussian, the minimum-variance estimate is the same as a maximum a posteriori (MAP) estimate [31, pp. 5282].

3.4 Iterative Methods

3.4.1 MG and MG-PCG Methods

What follows is a brief review of multigrid and preconditioned conjugate-gradient methods. Equations (3.5) and (3.7) are of the form $A\hat{x} = b$, where Au can be computed in $\mathcal{O}(n)$ floating point multiplications for arbitrary $u \in \mathbb{R}^n$. This follows because:

- G and L are sparse so $G^T G u$ and $L L^T u$ cost $\mathcal{O}(n)$.
- $V V^T u = V(V^T u)$ costs $\mathcal{O}(n)$ because $V \in \mathbb{R}^{n \times 2}$.

Such systems can be solved efficiently by using multigrid methods. Multigrid methods use a *smoother*, a cheap linear iterative method that rapidly removes high-frequency content in the error. The residual is projected onto a coarser grid using a *restriction operator*, and the smoother is applied again. The general idea is that low-frequency content in the residual becomes high frequency content when projected onto a coarser grid. This process continues, and the various coarse-level corrections are interpolated back onto the fine grid using a *prolongation operator*. The corrections are then added to the original estimate to improve it. When done repeatedly, this is known as a multigrid (MG) iterative method. If we alternate between a MG iteration and a conjugate gradient iteration, this is known as a multigrid-preconditioned conjugate-gradient iteration (MG-PCG).

Both MG [29] and MG-PCG [11, 8] methods provide $\mathcal{O}(n)$ convergence for the least-squares and minimum-variance reconstructors. The important parameters are:

1. The type of smoother used, typically a weighted-Jacobi (J) or Gauss-Seidel (GS) iteration.

2. The number of smoothing iterations to run on each level before (ν_1) and after (ν_2) applying the coarse-level correction.

3. The choice of restriction and prolongation operator. Here, we use *full weighting* for restriction and *bilinear interpolation* for prolongation [27, §2.3].

4. The cycle pattern, describing how the various levels are visited. Here, we use V-cycles (each level is visited twice per iteration), unless otherwise indicated.

For a comprehensive look at multigrid methods we refer the reader to [27]. Once we have selected the specific multigrid method, the general procedure is:

1. Measurement arrives

2. Run a predetermined fixed number of iterations of the chosen method, which ensures the estimate has converged

3. Send the estimate to the controller which passes the appropriate actuator signals to the DM

4. Go to step 1.

We call this procedure a *cold-start* configuration because every time a new measurement arrives, the iterative process is restarted with an initial guess of $\hat{x}_0 = 0$, the prior mean of the distribution of x.

3.4.2 Warm-Start Configuration

In the warm-start configuration, the most recent phase estimate is used as a guess value for the first iteration whenever a new measurement arrives. This technique is commonly used in numerical linear algebra, and has also been used in the context of adaptive optics [10].

Atmospheric phase offset is strongly correlated in time, so we can expect the most recent estimate to be a good guess for the current phase. As we will see, iterative multigrid reconstruction schemes converge much faster when used in a warm-start configuration.

3.4.3 Computational Cost

We will evaluate various iterative schemes on the basis of computational cost. Here, we chose to count the number of floating point multiplications. This cost includes: smoother iterations, computation of residuals, restriction and prolongation operations required to pass corrections up and down the hierarchy of levels, and conjugate gradient iterations if applicable. All costs were computed analytically to ensure a fair comparison.

In the case of Fourier-based methods, we associated a cost of $\frac{1}{2}n\log_2(n)$ multiplications to perform an FFT on a vector of length n. This is consistent with the cost of a Radix-2 implementation (n is a power of 2). It is worth noting that FFTs can be implemented very efficiently in hardware, so FLOPs may not be an accurate representation of true performance.

There are other possible choices, such as the number of total floating-point operations (multiplications and additions), or memory considerations.

3.5 Simulation

In this section we present our simulation results. Using a Taylor frozen-flow temporal dynamics model, we simulated the open-loop cold-start and warm-start cases, as well as the closed-loop case.

3.5.1 Parameters

The photon flux from the guide star in the visible is:

$$\mu = 0.9405 \times 10^{10-0.4M}, \tag{3.8}$$

where M is the stellar magnitude, and μ is the photon flux measured in photons per square meter per second. We assumed a $30\,\mathrm{m} \times 30\,\mathrm{m}$ square aperture, and a 128×128 array of sensors arranged using Fried geometry. There are therefore 129^2 actuators.

We assumed Shack-Hartmann quad-cell ($N_D = 4$) sensors with a RMS read error of $\sigma_e = 7$ electrons. The gap correction factor was chosen to be $K_g = 1.2$, which is typical of this type of sensor [12]. We also chose a quantum efficiency of 0.8 and an optical efficiency of 0.5, for a total efficiency of $\eta = 0.4$. For the background photons, we chose a value of 20 magnitudes per arcsec^2, which translates to $n_{\mathrm{bg}} \approx 0.01\, n_{\mathrm{ph}}$ in this case.

Phase screens with the proper spatio-temporal correlations were generated using *Arroyo* [1, 2], a C++ library for the simulation of electromagnetic wave propagation through turbulence. We chose seeing conditions consistent with Ellerbroek's Cerro Pachon layered atmospheric model [5] and assumed Kolmogorov statistics for each layer (Taylor frozen flow hypothesis).

Using Arroyo, we generated data consisting of 1000 independent runs for each of 25 logarithmically spaced sample rates between 10 and 10^4 Hz. Each run consists of 10 time-correlated phase screens. Each simulation was generated from these data, and all results were averaged over the 1000 independent runs.

Least-squares and minimum-variance reconstructors were implemented. For the minimum-variance case, we used $L = \gamma G^T G$, where $\gamma^2 = 0.2$ was calculated via Monte-Carlo simulation to minimize the piston-removed mean squared error (MSE).

3.5.2 Open-Loop Cold-Start Simulation Results

Regardless of the estimation method, there is a tradeoff for choosing the optimal sample rate which depends on the seeing conditions and guide star brightness [12, pg.

74].

The sensors take time-averaged measurements, the iterative scheme takes time to converge, and the control algorithm takes time in responding to the changes in the estimate. These delays result in estimation error, because even if the DM is ideal, it will assume the shape of the measured wavefront, not the current wavefront. As we increase the sample rate of the sensors, this delay error is reduced. However, as seen in Section 3.2.3, increasing the sample rate makes the sensors noisier, resulting in increased estimation error. With this trade-off in mind, we can select the sample rate that minimizes the minimum achievable mean square error.

For each of the $M = 1000$ runs indexed by i, we took two temporally correlated phase screens $x_1^{(i)}$, $x_2^{(i)}$. The first screen is used to generate a noisy measurement and estimate:

$$\begin{aligned} y_1^{(i)} &= G x_1^{(i)} + v_1^{(i)} \\ \hat{x}_1^{(i)} &= A^{-1} G^T y_1^{(i)}. \end{aligned}$$

Note that the noise strength $\operatorname{cov} v_1^{(i)} = \sigma^2 I$ depends on the sample rate as in (3.4). The coefficient matrix to be used is either $A = G^T G + VV^T$ (least-squares), or $A = G^T G + \sigma^2 LL^T + VV^T$ (minimum variance). The estimate is compared with the second screen to compute various normalized MSE measures:

1. Lag error: $\frac{1}{M}\sum_{i=1}^{M} \left\| x_1^{(i)} - x_2^{(i)} \right\|^2 \Big/ \frac{1}{M}\sum_{i=1}^{M} \left\| x_2^{(i)} \right\|^2$

2. Noise error: $\frac{1}{M}\sum_{i=1}^{M} \left\| \hat{x}_1^{(i)} - x_1^{(i)} \right\|^2 \Big/ \frac{1}{M}\sum_{i=1}^{M} \left\| x_2^{(i)} \right\|^2$

3. Total error: $\frac{1}{M}\sum_{i=1}^{M} \left\| \hat{x}_1^{(i)} - x_2^{(i)} \right\|^2 \Big/ \frac{1}{M}\sum_{i=1}^{M} \left\| x_2^{(i)} \right\|^2$

Using the seeing conditions chosen in Section 3.5.1, together with a guide star magnitude of 8, we plotted the three error measures (Figure 3.2). The total error coincides with the lag error at low sample rates, and with the noise error at high sample rates. The optimal sample rate is about 417 Hz for these conditions. As previously noted, we are dealing with open-loop estimation. In a closed-loop configuration, the plot would look similar but would have a different optimal frequency due to the additional errors incurred by the controller dynamics.

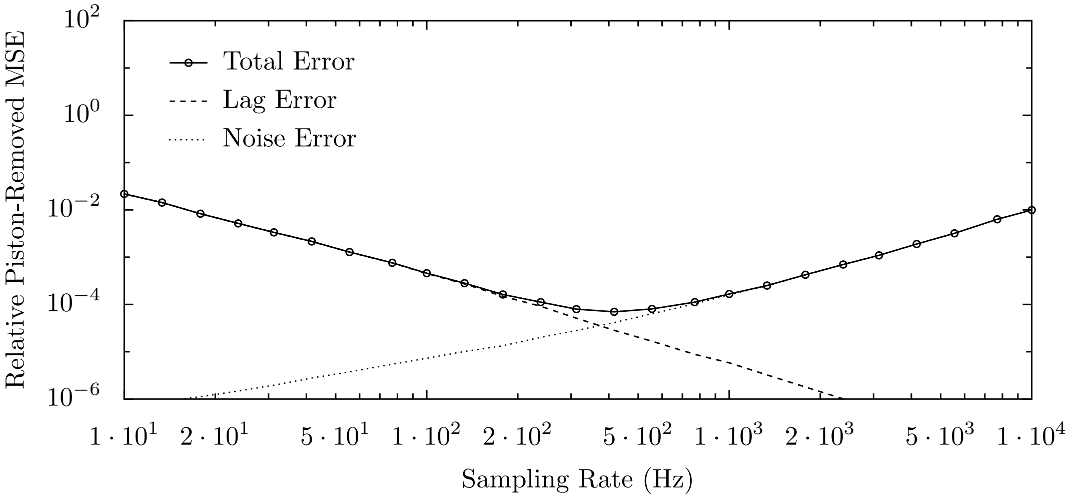

Figure 3.2: Error measures for least-squares reconstruction using a guide star of magnitude $M = 8$ and the seeing conditions from Section 3.5.1. Exact inverse was computed and averaged over 1000 independent runs. In these conditions, the optimal sensor sampling rate is about 417 Hz, and the minimum expected relative error just under 10^{-4}.

One way to quantify the noise level is to use the notion of *signal to noise ratio* (SNR). For consistency with existing literature [11, 15, 24, 29, 31], we use the following definition:

$$\mathrm{SNR} = \left(\frac{\mathrm{E} \, \|Gx\|^2}{\mathrm{E} \, \|v\|^2} \right)^{1/2} .$$

Choosing a brighter star magnitude $M = 5$, and a fainter star magnitude $M = 10$, we can see how SNR varies with sample rate in Figure 3.3. In the literature, SNR values ranging from 1 to 100 are typically assumed, which is consistent with the range obtained in this figure.

Using these three different star brightness values as a way of characterizing different noise levels, we obtained different trade off curves and corresponding optimal sample rates (Figure 3.4).

The minimum-variance reconstructor only outperforms least-squares when we are

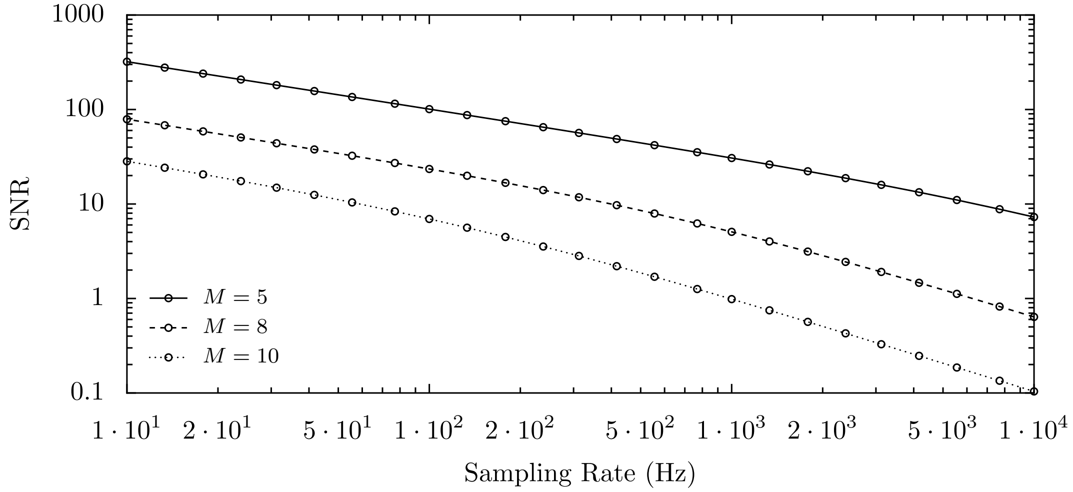

Figure 3.3: SNR variation for various guide star magnitudes M as a function of the sensor sampling rate. A higher SNR is obtained by using a brighter guide star or a lower sampling rate.

noise-dominated (either a faint guide star or an excessively high sample rate). When using the optimal sample rate for large SCAO reconstruction, there is no advantage to using minimum-variance reconstruction. This is consistent with the observation by Ellerbroek that conventional least-squares reconstruction is near optimal for future large AO systems [4].

Next, we compared several existing iterative methods on a basis of computational cost (as described in Section 3.4.3). See Table 3.1 for a complete list. We used both Jacobi and Gauss-Seidel smoothers, with and without PCG, and we varied the number of pre- and post- smoothing steps per iteration. We added some new methods that to our knowledge have not been published specifically for AO: a W-cycle MG scheme (V-cycles were used for all other methods), and asymmetric MG schemes that only perform one smoothing step per V-cycle, such as GS(1,0). Also, the recently proposed Fourier-Domain PCG (FD-PCG) method has been simulated for MCAO systems [31, 10]; here we demonstrate its performance on an SCAO system.

The results are presented in Figure 3.5. With the exception of FD-PCG, the various methods converge to the minimum error from Figure 3.2 in a few iterations with comparable computational effort. The simulation parameters used were the

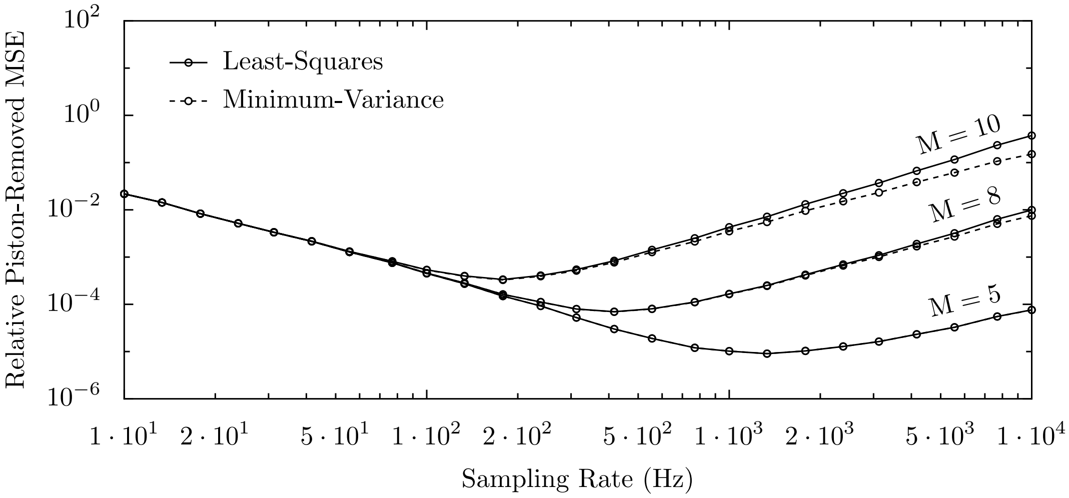

Figure 3.4: Bandwidth-Error trade off curve for various star magnitude values (M) using least-squares and minimum-variance reconstruction. The SNR at the minimum points are 26.2 for $M = 5$, 9.7 for $M = 8$, and 4.5 for $M = 10$. We averaged 1000 independent runs.

same as those used in Figure 3.2, running at the optimal sample rate of 417 Hz. These plots are similar to the ones produced in [11, 29], except that the x-axis counts multiplications required rather than iterations.

We can compare the various cold-start and warm-start schemes by evaluating the total cost to convergence in multiplications. Table 3.2 shows that the most efficient convergence was obtained when we used GS with a W-cycle and the smallest number of smoothing steps possible.

3.5.3 Open-Loop Warm-Start Simulation Results

We now generate the analogous plot to Figure 3.5 but using warm-start. We used the same simulation parameters running at the same optimal rate of 417 Hz. Note that it is not a priori obvious that this is the right choice of sample rate. In a warm-start configuration, if multiple iterations are required to reach convergence, then it may be better to sample more frequently so that newer information is being used in the iteration. In other words, it may be better to *not* iterate to convergence before taking in the next measurement. However, we found that almost every iterative algorithm

Table 3.1: List of iterative schemes

Number	Type	Smoother	Cycle	Cost per iter.
1	MG	GS(1,0)	V	3.406×10^5
2	MG	GS(1,1)	V	4.723×10^5
3	MG	GS(2,2)	V	7.358×10^5
4	MG	J(1,0)	V	3.406×10^5
5	MG	J(1,1)	V	4.723×10^5
6	MG	J(2,2)	V	7.358×10^5
7	MG	GS(1,0)	W	5.050×10^5
8	MG	GS(1,1)	W	7.003×10^5
9	MG-PCG	SGS(1,1)	V	6.377×10^5
10	MG-PCG	SGS(2,2)	V	9.012×10^5
11	MG-PCG	J(1,1)	V	6.377×10^5
12	MG-PCG	J(2,2)	V	9.012×10^5
13	MG-PCG	J(1,1)	W	8.657×10^5
14	MG-PCG	SGS(1,1)	W	8.657×10^5
15	FD-PCG	–	–	12.984×10^5

The cost per iteration is measured in floating-point multiplications. For all methods except FD-PCG, this cost is proportional to the number of actuators.

we tried converged in a single iteration for the parameters used herein. Thus the separation of the choice of algorithm and choice of sample rate remains as in the cold-start case.

We used $\hat{x}_0 = 0$ to start the first iteration when the first measurement arrived. Whenever a new measurement arrived, we used the most recent estimate as an initial guess for the subsequent iteration. We ran each test for 20 sampling intervals to ensure the iterative scheme was operating in steady-state (the transient behavior typically disappeared after 3-6 measurements had been processed). The converged results were then used to compute the relative piston-removed MSE.

For each iterative scheme, we varied the number of iterations executed during each sampling interval and plotted the resulting average relative error. See Figure 3.6. Only one iteration per timestep was required for every multigrid method we tested with the exception of MG-J(1,0) and FD-PCG.

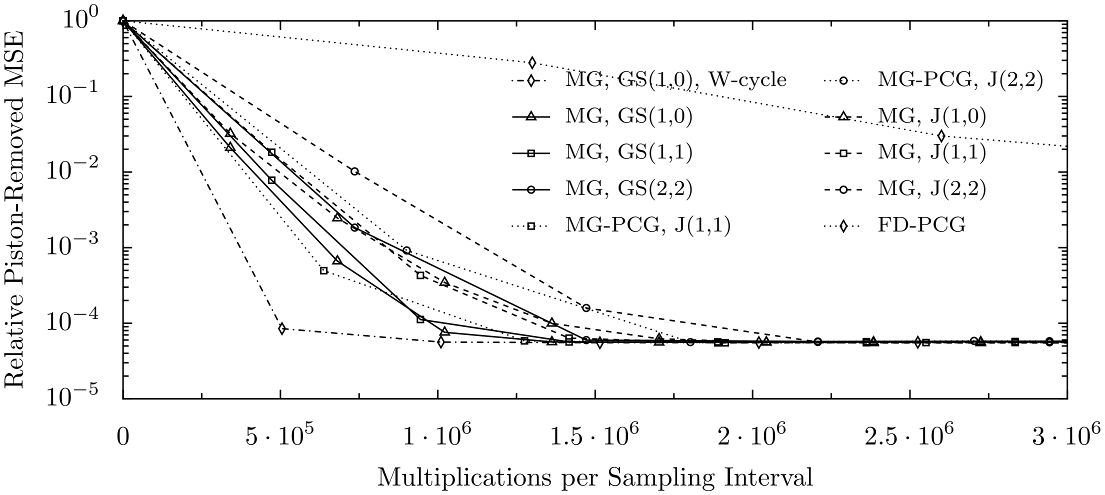

Figure 3.5: Convergence plots comparing simple multigrid (MG), with conjugate gradient methods using either a multigrid (MG-PCG) or Fourier-domain (FD-PCG) preconditioner. MG methods use Gauss-Seidel (GS) or Jacobi (J) smoothers. The pair (ν_1,ν_2) is the number of pre- and post-smoothing steps. Most methods converge in a few iterations with comparable computational effort. We averaged 1000 independent runs.

It is worth noting that by warm-starting, we get no change in the expected steady-state error, just faster convergence to that error. Thus, the trade-offs from Figures 3.2 and 3.4 still hold when we warm-start.

The data from Figure 3.6 are summarized in Table 3.3, where we compare the various iterative schemes once again. This table tells us how much computation would be required to implement the algorithms in a warm-start configuration rather than iterating to convergence every time a new measurement arrives.

3.5.4 Closed-Loop Simulation Results

In a closed-loop setting, the DM corrects the incident phase *before* the sensors take measurements.

In this simulation, we began with a sequence of time-correlated piston-removed phase screens generated using Arroyo: $\{x_1, x_2, \ldots\}$. We closed the loop in two different

Table 3.2: Cost comparison (iterating to convergence)

Iterative scheme	Iteration cost (multiplications)	Iterations to convergence	Total cost (multiplications)
FD-PCG	1.30×10^6	9	1.17×10^7
MG, J(1,0)	3.41×10^5	6	2.04×10^6
MG-PCG, J(2,2)	9.01×10^5	2	1.80×10^6
MG, GS(2,2)	7.36×10^5	2	1.47×10^6
MG, GS(1,1)	4.72×10^5	3	1.42×10^6
MG, GS(1,0)	3.41×10^5	4	1.36×10^6
MG-PCG, J(1,1)	6.38×10^5	2	1.27×10^6
MG, GS(1,0), W-cycle	5.05×10^5	2	1.01×10^6

The iterations to convergence were extracted from Figure 3.5. The cheapest method using this metric is GS(1,0) using a W-cycle.

ways: a standard loop closure (left) and a pseudo-open loop (POLC) [6] implementation (right).

$$\begin{aligned} y_t &= G(x_t - u_{t-1}) + v_t & \qquad y_t &= G(x_t - u_{t-1}) + v_t \\ \hat{e}_t &= K y_t & \qquad \hat{x}_t &= K(y_t + G u_{t-1}) \\ u_{t+1} &= u_t + \beta \hat{e}_t & \qquad u_{t+1} &= u_t + \beta(\hat{x}_t - u_{t-1}) \end{aligned}$$

In the standard case, the iterative scheme is used to find the estimated error, whereas in POLC, the previous input is used to compute an equivalent open-loop measurement, and the iterative scheme is used to find an estimate of the actual phase screen x_t. In both cases, K is *one iteration* of the least-squares iterative scheme of our choice. In practice, a wide range of control gains lead to stable systems with the minimum error. We chose $\beta = 0.5$ for all our simulations.

Both models assume two timesteps of delay: a 1-step delay to process the measurements since y_t depends on the past input u_{t-1}, and a 1-step delay to compute the estimate since the future input u_{t+1} depends on y_t. The relative error at timestep t is computed using the formula: $\|x_t - u_{t-1}\|^2 \Big/ \|x_t\|^2$. This is analogous to the way we computed error for the cold-start and warm-start open-loop simulations. Note that

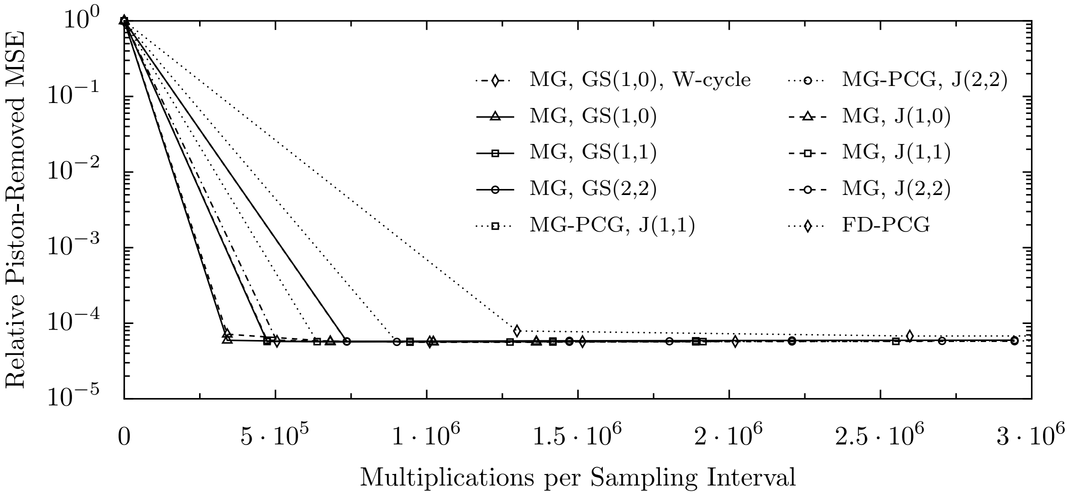

Figure 3.6: Plots comparing converged values of various methods using the open-loop warm-start technique. For every method except MG-J(1,0) and FD-PCG, only one iteration per measurement is required for minimum error. The best method to choose is simply the one that has the smallest iteration cost. We averaged 1000 independent runs.

the average error in closed-loop will be higher than the average error in open-loop, because our closed-loop implementations have a 2-step delay.

We simulated the standard case using both cold-start and warm-start, and we simulated the POLC case using warm-start. In cold-start, POLC requires at least a few iterations per timestep to converge to the minimum error, as in Figure 3.5. We found that all three loop closure schemes are stable for every method we tested, and that every one can be implemented with one timestep per iteration.

Results are presented in Figure 3.7. For the standard closed-loop case using cold-start, every iterative scheme with the exception of MG-J(1,0) and FD-PCG converged to a minimum error floor. These are the only two methods we tested that require more than one iteration per timestep to achieve the minimum error. Note that the iterative scheme is estimating the *error*, and not the actual phase. Since we expect the error to be small, an initial guess value of 0 (cold-starting) produces fast convergence in only one iteration.

We can also use warm-start with the standard case. After iterating, we store the error vector, and use it to warm-start our iteration at the next timestep. With this

Table 3.3: Cost comparison (fewest possible iterations)

Iterative scheme	Iteration cost (multiplications)	Iterations per timestep	Total cost (multiplications)
FD-PCG	1.30×10^6	2	2.60×10^6
MG-PCG, J(2,2)	9.01×10^5	1	9.01×10^5
MG, GS(2,2)	7.36×10^5	1	7.36×10^5
MG, J(1,0)	3.41×10^5	2	6.81×10^5
MG-PCG, J(1,1)	6.38×10^5	1	6.38×10^5
MG, GS(1,0), W-cycle	5.05×10^5	1	5.05×10^5
MG, GS(1,1)	4.72×10^5	1	4.72×10^5
MG, GS(1,0)	3.41×10^5	1	3.41×10^5

The iterations to convergence were extracted from Figure 3.6. In a warm-start configuration, we do not need to iterate to convergence at every timestep in order to achieve the minimum error. Note that the methods are ordered differently here than in Table 3.2. The cheapest method here is GS(1,0).

small change, all the methods we tested converged to the minimum error floor with virtually identical error.

Finally, we tested the POLC case. In this case, we store our previously applied input, and use it to convert our closed-loop measurement into an open-loop measurement. The iterative scheme is then used to estimate the actual phase. This method benefits greatly from warm-starting, as one might expect from the open-loop results in Figure 3.6. Once again, all methods converged to the minimum error. We also tried this method with cold-start, and found that 3-4 iterations per timestep were typically required, as in Figure 3.5.

This is the same conclusion we drew from the open-loop warm-start plot, which means that we can compute the cost required for closed-loop implementation directly from Table 3.3. To find the required computer speed, we multiply the total cost by the sample rate.

3.6 Conclusion

We have explored the effectiveness of using warm-started iterative methods for adaptive optics reconstruction. In open-loop or POLC, warm-start provides a significant benefit. When applied to most iterative methods, convergence is achieved in only one iteration, reducing the number of multiplications required by a factor of about three.

For a standard closed-loop implementation, the advantage of warm-starting is less significant because most iterative algorithms already achieve the optimal performance with a single cold-started iteration per timestep. The control removes much of the (large-amplitude) correlated information between timesteps, so that the residual error is both smaller and less correlated. Nevertheless, when we warm-start in this case, every algorithm we tested provides the same minimal error, and convergence in one iteration per timestep. This shows that iteration cost is the most meaningful performance metric.

In principle, any iterative scheme can be warm-started, and should always yield a computational speedup. For methods such as Fourier-based reconstruction, which do not afford an iterative implementation, it is not clear how to take advantage of the warm-start approach.

In the next chapter, we implement some of these reconstructors and validate our findings on a real adaptive optics system.

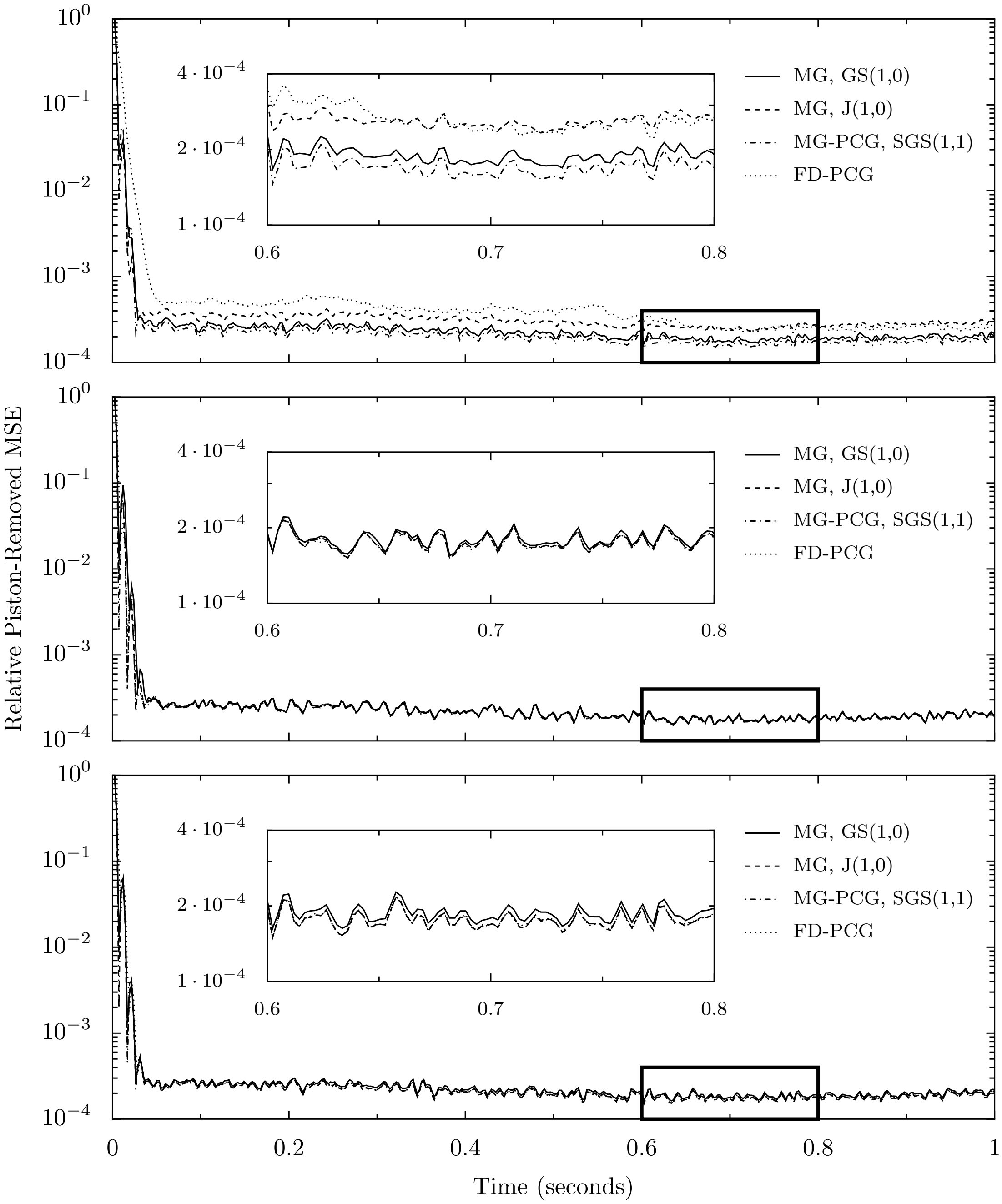

Figure 3.7: Closed-loop time series in cold-start (top), warm-start (middle), and warm-started POLC (bottom). 500 timesteps were simulated, performing one iteration per timestep. We tested 15 different iterative schemes (only 4 are plotted). All were stable and performed equally well, with the exception of MG-J(1,0) and FD-PCG in the cold-start standard case, which converged to a slightly higher error.

Chapter 4

Experiments

4.1 Introduction

In Chapter 3, we showed via numerical simulation that warm-started iterative methods achieve fast and accurate wavefront reconstruction. In this chapter, we validate our findings on the Palomar adaptive optics installation at Palomar mountain in Southern California.

In general, wavefront reconstruction is implemented in hardware using vector-matrix multiplication (VMM). If we have n actuators and $2n$ sensor measurements, a dense $n \times 2n$ reconstructor matrix is precomputed and stored. Every time a new measurement arrives, VMM is performed between the stored reconstructor matrix and the vector of measurements. This is inadequate for future large (10^4–10^5 actuator) systems, since computation scales as $\mathcal{O}(n^2)$.

Many faster methods have been proposed and analyzed using computer simulations. Examples include: a conjugate-gradient (CG) method [11], a Fourier-domain (FD) method [21], a blended FD/CG method [24], and a sparse method [25]. The work presented in Chapter 3 is also published in [17], and shows that all of the above methods are equally effective when implemented using warm-start with a single iteration (SIMG). A similar conclusion was reached for the multiconjugate (MCAO) case [10].

Two computationally efficient methods that have previously been tested on-sky are the FD method [23] and a sparse method [25]. These methods are $\mathcal{O}(n \log n)$. In this chapter, we detail experimental validation of the SIMG method on a real SCAO system. Our results were published in [16]. SIMG is $\mathcal{O}(n)$, and shows no performance degradation when compared to the least-squares reconstructor.

4.2 Reconstruction Summarized

We will now give a brief overview of wavefront reconstruction, and make note of a couple facts relevant to physical implementation. For more detail, please refer to Section 3.3. A good model for the wavefront sensor (WFS) is (3.1):

$$y = Gx + v,$$

where x is the wavefront phase, y are the measurements, v is white noise, and G is a sparse influence matrix. The least squares reconstruction matrix is found by taking the pseudoinverse. In practice, we can compute it by evaluating $K = (G^T G + \epsilon I)^{-1} G^T$ for a small ϵ. This ensures that unobservable modes such as piston and waffle are zeroed out. The SIMG method uses a single multigrid sweep with an initial guess of 0 to obtain an approximate solution to the equation

$$(G^T G + \epsilon I)\hat{x} = (G^T y).$$

If the measurements are taken in open-loop, $\hat{x}$ is the wavefront phase. In this case, $\hat{x}_0 = 0$ is a bad guess, so multiple iterations are required to achieve acceptable convergence. However, we showed in Chapter 3 that when we operate in closed-loop, only one iteration is required. In this case, $\hat{x}$ is the *change* in wavefront phase between successive timesteps, and using an estimate of 0 is a good guess.

4.3 Telescope Description

Our tests were performed on the Palomar Adaptive Optics (PALAO) system on the Hale 5.1 meter telescope [28], pictured in Figure 4.1. The PALAO system has a deformable mirror (DM) with 241 active actuators and a Shack-Hartmann wavefront sensor (WFS) array with 256 subapertures, producing a total of 512 measurements.

Figure 4.1: View of the Palomar Observatory.

The optics bench is shown in Figure 4.2, and a diagram of the mirror geometry is shown in Figure 4.3. Note that the DM and WFS are aligned in a Fried geometry, which is the same as the one analyzed in Chapter 3 (Figure 3.1).

The AO system collects measurements y at up to 2 kHz. Tip and tilt are removed from the wavefront using a fast-steering mirror (FSM) and proportional-integral (PI) controller. The rest of the wavefront offset $\hat{x}$ is reconstructed by VMM: $\hat{x} = Ky$. This estimate is fed back through a second PI loop to the DM. The closed-loop corrected wavefront is split using a dichroic mirror. The visible part of the signal is sent to the WFS, while the near-infrared portion is sent to the Palomar High Angle Resolution Observer (PHARO) camera [13] for imaging. A block diagram is shown in Figure 4.4. Note the resemblance between this diagram and the one from Chapter 3 (Figure 2.1).

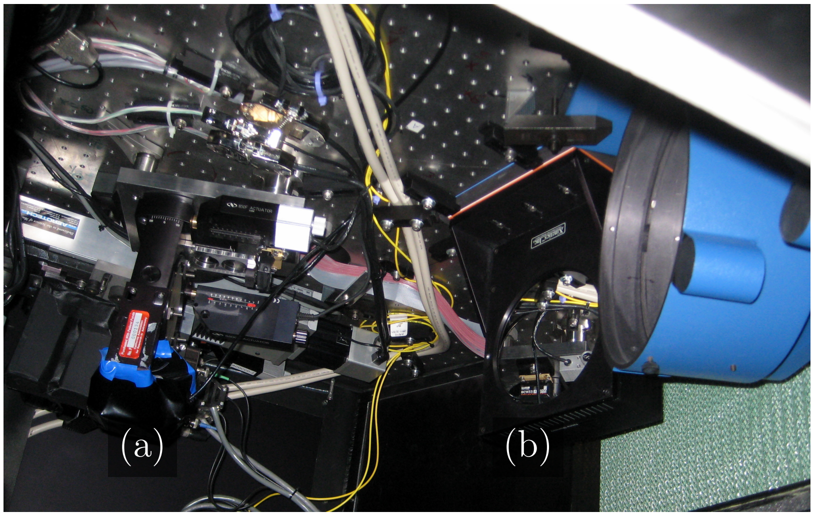

Figure 4.2: Adaptive Optics bench at the Palomar Observatory. Note the wavefront sensor (a) and the deformable mirror (b). The individual segments are not visible because of the high smoothness and reflectivity of the mirror surface.

4.3.1 Implementation Details

Since the PALAO actuators are in a circular arrangement rather than the convenient square arrangement assumed in Chapter 3, some tweaks were required. We circumscribed the circular grid with a 17 × 17 square grid and filled the extra space with *virtual* actuators. Thus, the influence matrix G was augmented appropriately with 0's, as in [24]. The new system is $y = \bar{G}\bar{x}$, where $\bar{G}$ is 512×289 instead of 512×241. This new system is equivalent to the original one, so no approximation was made. We can now use the SIMG method [17] on this system to reconstruct all the actuator commands, and we simply truncate the virtual actuators once we're done.

It is worth noting that this trick destroys the *block-toeplitz with toeplitz block* structure of G^TG. Although this does not affect SIMG in any way since G is still sparse, methods such as Fourier-domain reconstruction [21] or Fourier-based preconditioning [31] rely on a shift-invariant structure. In order to adapt to a circular aperture, they must either use a heuristic to correct for edge effects [21] or use an enlarged computational domain [11, 31]. No special provisions were made to account for the central obscured region; these measurements are simply zeroed.

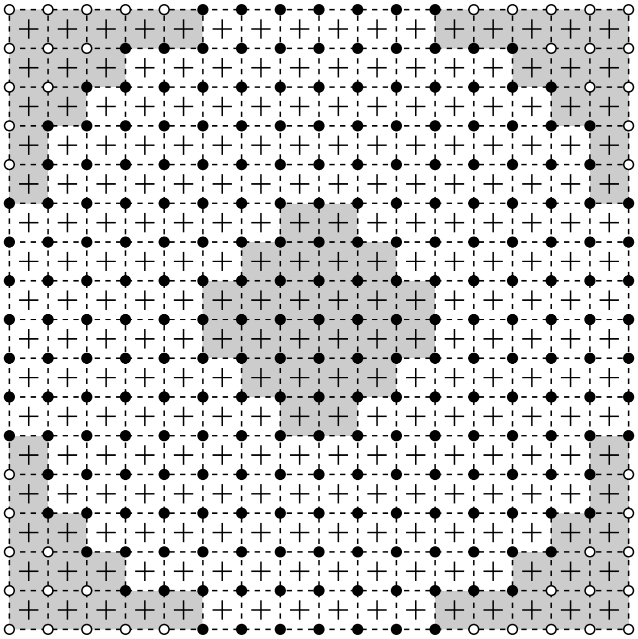

Figure 4.3: Diagram of the PALAO system showing how the actuators and sensors line up. The black dots are the 241 active actuators, the white dots are virtual actuators and the +s are the 256 Shack-Hartmann sensors. The shaded region shows which sensors are obscured because of the aperture shape. Note that there are active actuators inside the central obscuration.

The PALAO hardware is only equipped for VMM, so we implemented our SIMG method by using an equivalent VMM reconstructor. This is possible since SIMG is a sequence of linear operations: Jacobi/Gauss-Seidel iteration, restriction/prolongation, residual computation, and multiplication by G^T. The equivalent VMM reconstructor has estimation error identical to that of the SIMG method, but does not benefit from the computational speedup.

A notable benefit of using equivalent VMM reconstructors for testing is that they can be loaded into PALAO in a matter of seconds. This allows us to perform rapid sequential testing of different algorithms, and thereby average out the highly variable atmospheric conditions. The same approach was used in [23, 25]. The 3217-actuator PALM-3000 system currently in development [3] is being designed with sufficient computation to allow full VMM reconstructors; this will permit similar experiments on a much larger system to be conducted in the future.

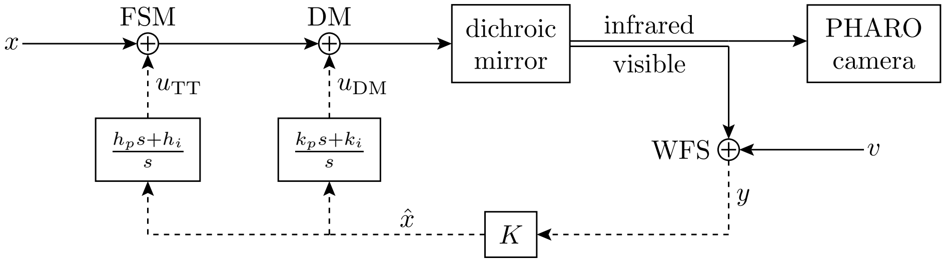

Figure 4.4: Block diagram representing the PALAO system. Solid lines indicate the optical path, while dashed lines indicate the signal path. Note that the fast-steering mirror mirror (FSM), deformable mirror (DM) and wavefront sensor (WFS) are represented as summation junctions

4.3.2 Method and Results

On May 19, 2008, we ran tests on three stars that ranged from bright (magnitude 8) to dim (magnitude 13.5). The sky was exceptionally calm and clear during our experiment, so we chose a very dim star for our final test. For each star, we adjusted the WFS sample rate, tip/tilt control gains, and DM control gains in order to maximize the Strehl ratio using the baseline (least-squares) reconstructor; see Table 4.1. The signal-to-noise ratio (SNR) per subaperture was measured by comparing the average flux per subaperture to the variance while the wavefront sensor was recording the sky background frames [12]. We used a Kshort filter (1.99 to 2.30 microns) plus a neutral density filter appropriate for the star brightness: 0.1% for the two brighter stars, and no filter for the faintest star.

Table 4.1: System parameters used for each star

Number	Catalog designation	Brightness	Sample Rate	T/T	SNR
1	Tycho-2 #2563-170-1	$V = 8.10$	1 kHz	0.30	14.9
2	Tycho-2 #2580-2328-1	$V = 10.01$	500 Hz	0.40	11.8
3	USNO-B1.0 #1204-0241816	$R = 13.5$	50 Hz	0.40	3.5

The sample rates and tip/tilt integral gains (T/T) above were chosen to optimize the instantaneous Strehl ratio. The optimal DM gains were the same for each star: a proportional gain of 0.25 and an integral gain of 0.01.

We tested the baseline least-squares reconstructor, and three multigrid schemes. One of them was GS(1,0)-V, meaning we used a Gauss-Seidel smoother in a V-cycle, with 1 pre-smoothing iteration and no post-smoothing. The other two were GS(1,1)-V, and J(1,0)-V, where the J indicates a Jacobi smoother. Each V-cycle is run only once per measurement. Note that J(1,0)-V is the simplest possible multigrid method. More complicated variations such as GS(2,2)-W using a W-cycle are also possible. Alternatively, one can execute multiple iterations of a chosen method per measurement. In the limit, these more complicated (and costlier) variations approach the baseline least-squares reconstructor. In our experiment, we tested the three simplest reconstructors and found that their performance was indistinguishable from that of the baseline least-squares reconstructor. The images produced for a typical reconstruction look like Figure 4.5.

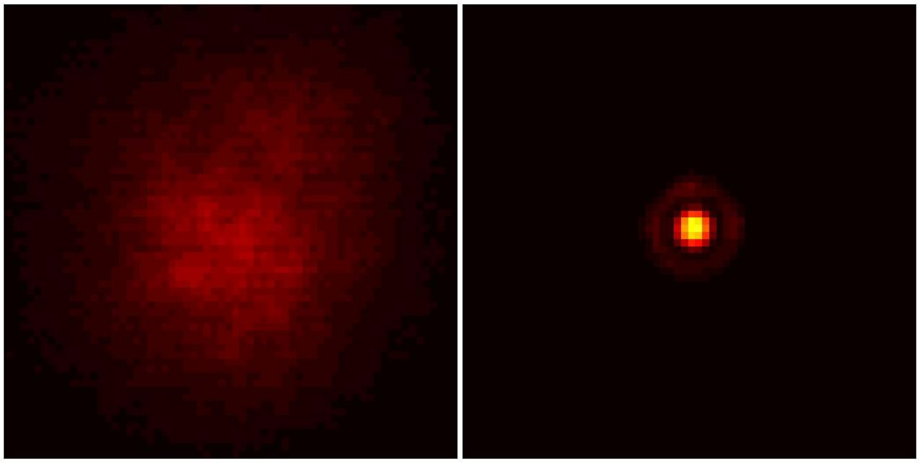

Figure 4.5: Images of the star Tycho-2 #2563-170-1 (Star number 1 in Table 4.1) with AO off (on the left) and on (on the right). The Airy disc [12] is visible on the right, indicating that the atmospheric disturbance has been canceled and the image is diffraction-limited.

Each reconstructor was tested four times per star, and each test consisted of acquiring three consecutive 10-second exposure images. This cyclical testing pattern allows us to average performance over the variable atmospheric conditions, and was inspired by a similar experiment to test a Fourier-based reconstructor at Palomar [23]. We also collected images of the sky background by pointing the telescope 60 arcseconds away from the target star, before and after the reconstructors were tested.

For each image, we computed the Strehl ratio by first subtracting the median sky background from each frame (eliminating sky photons and detector bias). We then measured the ratio of the peak brightness of the star to that of a theoretical diffraction-limited point-spread function with the same total flux and pixel position. Refer to Figure 4.6 for plots showing the Strehl ratio for three stars. The gaps in time during the first test are due to restarting the AO system. The four methods tested performed equally well. This agrees with recent theoretical predictions [17].

The seeing measured by the MASS and DIMM systems at the Palomar Observatory [14] is plotted over the time of this experiment in Figure 4.7. The higher variability in Strehl ratio measured for stars 1 and 3 is consistent with the observed variability in seeing. The average and Strehl ratio for each method is compared for the second star in Figure 4.8. These data have the lowest variance, so if there is any appreciable difference between the methods, it would show up here.

4.4 Conclusion and Discussion

This experiment has shown that SIMG methods perform as well as least-squares reconstruction on a real AO system for both bright and dim guide stars. The major benefit of using SIMG is reduced computation. For a system with n actuators and $2n$ sensor measurements, $2n^2$ multiplications per timestep are required to process the measurements using VMM. In contrast, about $27n$ are required for MG-J(1,0)-V or MG-GS(1,0)-V, and $34n$ for MG-GS(1,1)-V. This includes the multiplication by G^T and the cost of smoothing, residual computation, restriction, and prolongation on every level. For PALAO, this results in fewer multiplications by a factor of about 17.

For a 3217-actuator system such as the PALM-3000 [3], using SIMG would reduce reconstruction computation by a factor of about 220 compared to VMM. Furthermore, if we use a Jacobi smoother, every step of the reconstruction is highly parallelizable; even if the sensor measurements are read sequentially, we can perform all the fine-grid computations, or roughly 3/4 of the work, *while* the measurements are being read in.

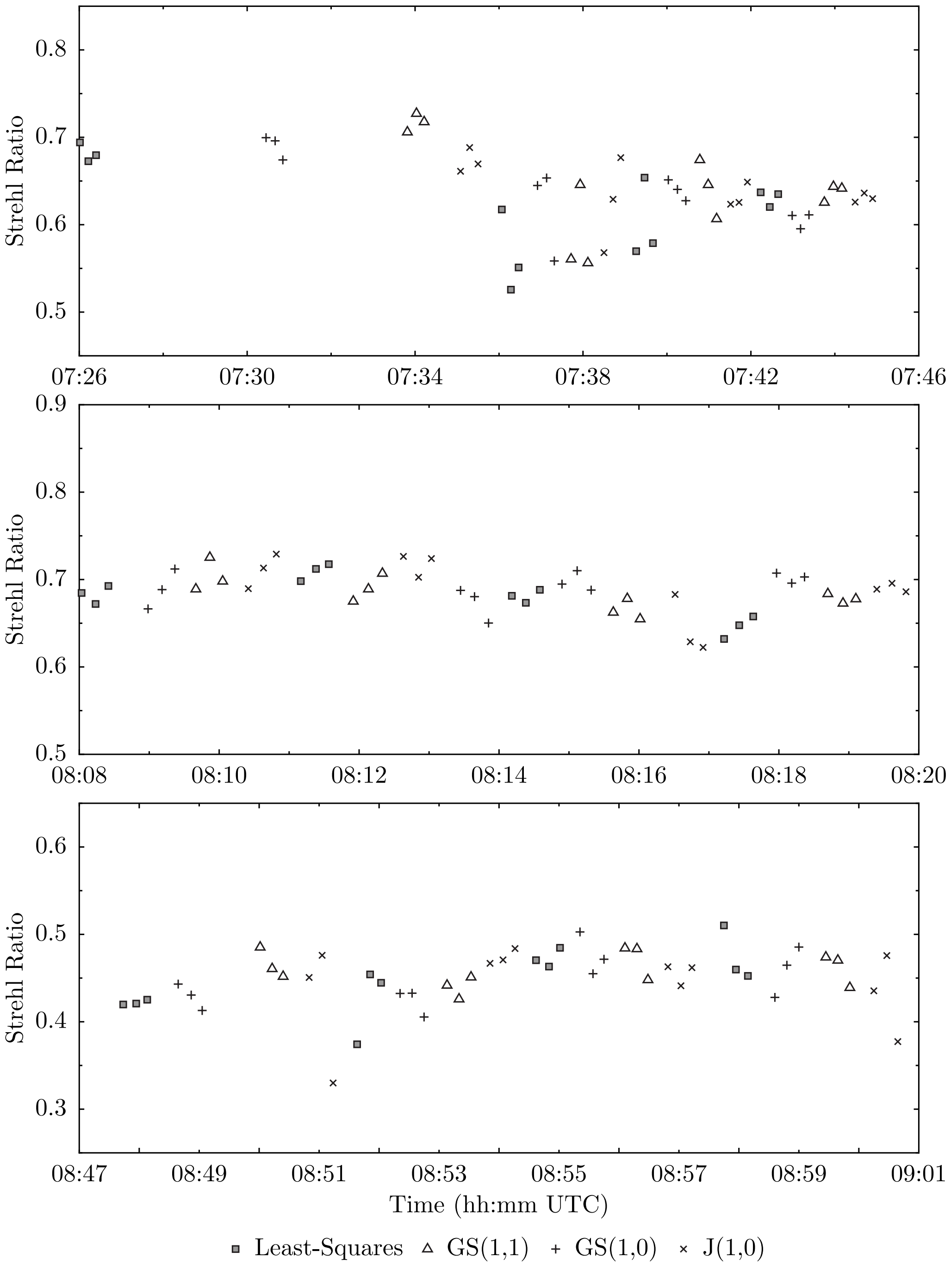

Figure 4.6: Plot comparing the Strehl ratio of various reconstructors on Star 1,2,3 (top, middle, and bottom, respectively). Each point represents a 10-second exposure image.

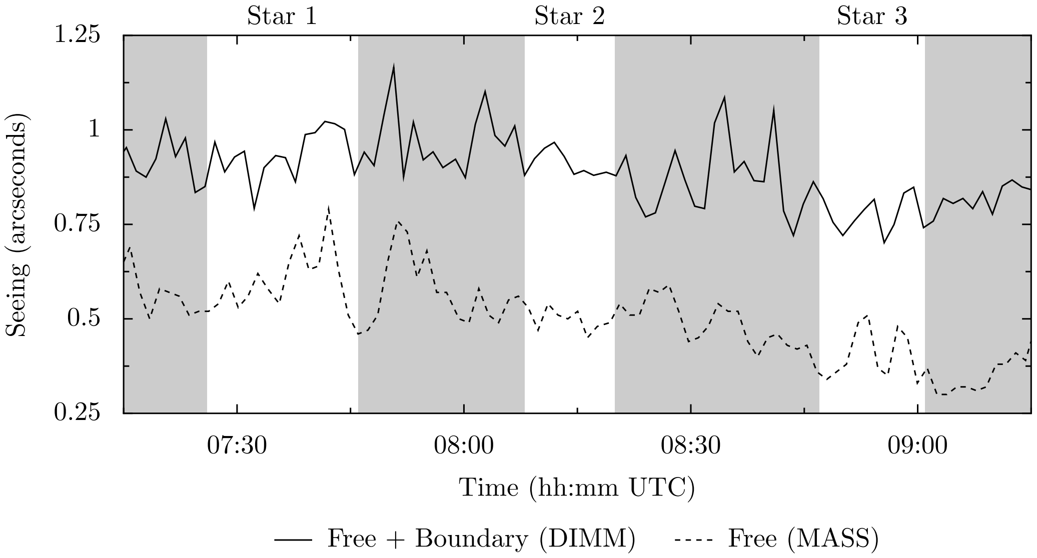

Figure 4.7: MASS-DIMM seeing at 500 nm measured throughout the night. The white regions indicate when data were being collected.

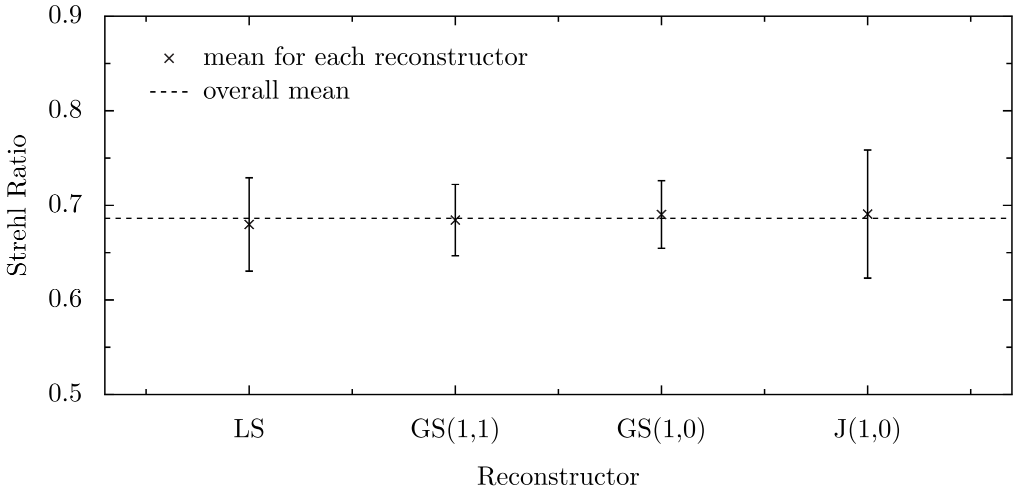

Figure 4.8: Plot showing the mean performance of each reconstructor on Star 2. The error bars indicate the 95% confidence interval for the true mean. There is no statistically discernible difference in performance between the four reconstructors we tested. The two other stars have similar plots, but with higher variances.

Chapter 5

Decentralized Systems

5.1 Introduction

In decentralized control problems, each controller has access to some subset of the measurements and must control some subset of the actuators. Such situations are of practical interest because it is often infeasible to have a single computer process all the information and make all the decisions. For example, we may be trying to design an auto-pilot for a swarm of vehicles flying in formation, where each vehicle only has access to noisy local measurements of the positions of its nearest neighbors. Another example is packet routing in networks. Each switch must make decisions based on local information, but the goal is to optimize the efficiency of the whole network.

It has been shown that in some cases, the information constraint imposed by decentralization can make the control synthesis problem intractable [34, 54]. Much work has gone into characterizing which decentralized problems are tractable.

When the system to be controlled has a linear plant, quadratic cost, and Gaussian noise (LQG), the optimal *centralized* controller is linear, and can be computed efficiently. However, in 1968, Witsenhausen [54] provided a now famous counterexample showing that for *decentralized* control, the optimal LQG controller is not linear in general. Subsequently, Blondel and Tsitsiklis [34] proved that a certain class of decentralized control problems is NP-hard.

This led to an effort to characterize which decentralized problems have optimal controllers that are linear. Radner [45] showed that this was true for a special class called *static team decision problems*. Ho and Chu [37] generalized Radner's result by showing that the larger class of *partially nested* systems could be converted into static team decision problems and hence solved easily.

More recently, Rotkowitz and Lall identified the largest known class of tractable decentralized control problems, which they called *quadratically invariant* (QI) [49, 50, 47]. Computational tractability and linearity of the optimal controller arise because in these cases the set of achievable closed-loop maps is convex.

The QI class is broad, but does not cover all tractable decentralized control problems, nor all the problems for which the optimal controller is linear. For example, Bansal and Basar [32] showed that by using a different quadratic cost function in the Witsenhausen counter-example, the problem is still not QI, but has a linear optimal solution.

In the subsequent chapters, we generalize the notion of quadratic invariance to include systems defined by multidimensional rational functions, and show that QI is largely an algebraic concept. We also show that some non-QI problems can be transformed to QI problems and thereby solved. We call this new class of systems *internally quadratically invariant* (IQI).

In the remainder of this chapter we will review relevant mathematical concepts and show a new converse convexity result: the set $\{K(I - P_{22}K)^{-1} \mid K \in S\}$ is convex if and only if K is quadratically invariant with respect to S. This is important because $K(I - P_{22}K)^{-1}$ figures prominently in the formula for the closed-loop map.

5.2 Preliminaries

We now introduce a mathematical framework for representing linear systems. The fundamental vector space used here is a Banach space, which is consistent with existing literature [49]. We will introduce our own framework in subsequent chapters.

If $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces, we denote by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ the set of all bounded linear operators $A : \mathcal{X} \to \mathcal{Y}$. We abbreviate $\mathcal{L}(\mathcal{X}, \mathcal{X})$ to $\mathcal{L}(\mathcal{X})$. A map $A \in \mathcal{L}(\mathcal{X})$ is called ***invertible*** if there exists $B \in \mathcal{L}(\mathcal{X})$ such that $AB = BA = I$. Define the ***resolvent***

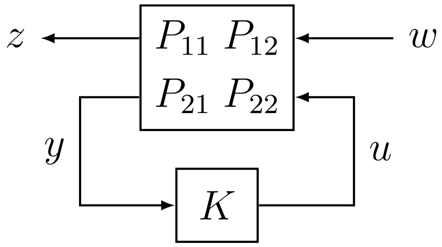

Figure 5.1: Closed-loop interconnection between a plant P and a controller K

set $\rho(A) = \{\lambda \in \mathbb{C} \mid (\lambda I - A) \text{ is invertible}\}$. This set is always open, and possibly disconnected, though it contains all sufficiently large $\lambda \in \mathbb{C}$. We will denote by $\rho_{uc}(A)$ the unbounded connected component of $\rho(A)$.

Suppose $\mathcal{U}, \mathcal{W}, \mathcal{Y}, \mathcal{Z}$ are Banach spaces over $\mathbb{R}$. We can think of these spaces as the controlled inputs, disturbances, measurements, and regulated outputs, respectively. Suppose we also have the plant $P \in \mathcal{L}(\mathcal{W} \times \mathcal{U}, \mathcal{Z} \times \mathcal{Y})$ and controller $K \in \mathcal{L}(\mathcal{Y}, \mathcal{U})$. We partition P in the conventional way, and connect the controller to the plant as in Figure 5.1. Since P_{22} is used frequently, we define the shorthand notation: $G = P_{22}$.

Note that Figure 5.1 is simply a graphical representation of the equations

$$\begin{bmatrix} z \\ y \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix}, \qquad u = Ky.$$

The resulting closed-loop map (from w to z) is given by the linear fractional transform $f(P, K) \in \mathcal{L}(\mathcal{W}, \mathcal{Z})$, where

$$f(P, K) = P_{11} + P_{12}K(I - GK)^{-1}P_{21}.$$

This interconnection is well-posed whenever $I - GK$ is invertible. More formally, we may define the set of admissible controllers $M \subset \mathcal{L}(\mathcal{Y}, \mathcal{U})$ as:

$$M = \{K \in \mathcal{L}(\mathcal{Y}, \mathcal{U}) \mid (I - GK) \text{ is invertible}\}.$$

Here, we only consider controllers which are bounded linear operators. Define the set

$N \subset M$, which we will need later:

$$N = \{K \in \mathcal{L}(\mathcal{Y},\mathcal{U}) \mid 1 \in \rho_{uc}(GK)\}.$$

5.2.1 Optimization

It is convenient to define the function $h_G : M \to M$,

$$h_G(K) = -K(I - GK)^{-1}.$$

We will often omit the subscript G when it is clear by context. Note that h is an ***involution***. That is, h is its own inverse: $h(h(K)) = K$ for all $K \in M$. It follows that h is a bijection from M to M.

Our goal is to solve the optimization problem

$$\begin{array}{ll} \text{minimize} & \left\|P_{11} + P_{12}K(I - GK)^{-1}P_{21}\right\| \\ \text{subject to} & K \in S \cap M \end{array} \tag{5.1}$$

where $S \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is a closed subspace. This formulation has a simple subspace constraint on K, but the objective function is potentially nonconvex. Define the parameter $Q = h(K)$ and use the involution property of h to rewrite (5.1) as

$$\begin{array}{ll} \text{minimize} & \left\|P_{11} - P_{12}QP_{21}\right\| \\ \text{subject to} & Q \in h(S \cap M). \end{array} \tag{5.2}$$

This new formulation has a convex objective function, but a potentially nonconvex constraint on K.

In the next section, we define quadratic invariance, a property that ensures that the set $h(S \cap M)$ is convex.

5.2.2 Quadratic Invariance

We now review the concept of quadratic invariance [49], which will serve as a starting point for subsequent chapters.

Definition 1. *The subspace $S \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is said to be* ***quadratically invariant*** *with respect to G if $KGK \in S$ for all $K \in S$.*

Theorem 2 (from [49])**.** *Suppose that $G \in \mathcal{L}(\mathcal{U},\mathcal{Y})$, and $S \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is a closed subspace. Further suppose that $N \cap S = M \cap S$. Then S is quadratically invariant with respect to G if and only if $h(S \cap M) = S \cap M$.*

So if S is quadratically invariant with respect to G, the optimization problem (5.2) is equivalent to

$$\begin{aligned} \text{minimize} \quad & \|P_{11} - P_{12}QP_{21}\| \\ \text{subject to} \quad & Q \in S \cap M. \end{aligned} \tag{5.3}$$

If Q^* solves this problem, then the K^* that solves (5.1) is found via $K^* = h(Q^*)$. In most practical problems of interest, well-posedness requirements force the optimal Q to lie within M, and so we may find it by solving the convex optimization problem:

$$\begin{aligned} \text{minimize} \quad & \|P_{11} - P_{12}QP_{21}\| \\ \text{subject to} \quad & Q \in S. \end{aligned} \tag{5.4}$$

5.3 Converse Result

We now present a new converse convexity result. Suppose that h is well-defined for all $K \in S$. Quadratic invariance ensures that the set of achievable closed-loop maps is convex by providing a necessary and sufficient condition under which $h(S) = S$. Thus,

$$\{P_{11} + P_{12}h(K)P_{21} \mid K \in S\} = \{P_{11} + P_{12}KP_{21} \mid K \in S\}.$$

In principle, one could also achieve convexity if $h(S)$ is a convex set other than S. Our result is that this never occurs [43]. In other words: if $h(S) = T$ where T is convex, then $T = S$. We begin with some definitions.

Definition 3. *Suppose $\mathcal{X}$ is a Banach space over $\mathbb{R}$, and $S \subset \mathcal{X}$. We call S a **double-cone** if for all $x \in S$ and $\alpha \in \mathbb{R}$, we have $\alpha x \in S$.*

Note that every subspace is a double-cone, but not all double-cones are subspaces.

Definition 4. *Suppose $\mathcal{X}$ is a Banach space over $\mathbb{R}$, and $T \subset \mathcal{X}$. We call T a **star-set** if for all $x \in T$ and $\alpha \in [0, 1]$, we have $\alpha x \in T$.*

Note that every convex set is a star-set, but not all star-sets are convex.

Theorem 5. *Suppose $S \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is a closed double-cone, $T \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is a star-set, and $h(S \cap M) = T \cap M$, then $T \cap M = S \cap M$.*

Proof. Fix some $K \in S \cap M$. Since $K \in M$, $I - GK$ is invertible, and $1 \in \rho(GK)$. The resolvent set of a bounded linear operator is an open set, so there exists a sufficiently small $\varepsilon > 0$ such that $1 - \alpha \in \rho(GK)$ for all $0 \leq \alpha < \varepsilon$. For any such α, it follows that $I - (1 - \alpha)GK$ is invertible. It follows that $(I - (1 - \alpha)GK)\,(I - GK)^{-1}$ is invertible as well. Expanding this expression, we find that it is equal to $I - \alpha Gh(K)$. Thus $\alpha h(K) \in M$.

Also, $K \in S$, $h(K) \in T$, and so $\alpha h(K) \in T$ whenever $0 \leq \alpha \leq 1$, because T is a star-set. It follows that for $\alpha \in [0, \varepsilon)$, $\alpha h(K) \in T \cap M$.

Now apply h to both sides: $h(\alpha h(K)) \in h(T \cap M) = S \cap M$, where we made use of the involutive property of h. Expanding $h(\alpha h(K))$, we find that it is equal to $\alpha K(I - (1 - \alpha)GK)^{-1}$. Since S is a double-cone, we may multiply this expression by $-1/\alpha$, and the result will still lie in S. Thus, $-K(I - (1 - \alpha)GK)^{-1} \in S$. Now define the function $g : [0, \varepsilon) \to \mathcal{L}(\mathcal{Y},\mathcal{U})$ by

$$g(\alpha) = -K(I - (1 - \alpha)GK)^{-1}.$$

Since S is closed, and $g(\alpha) \in S$ for $\alpha \in [0, \varepsilon)$, then

$$\lim_{\alpha \to 0^+} g(\alpha) \in S.$$

Since $(I - GK)$ is invertible, g is right-continuous at 0. So we may take the limit $\alpha \to 0^+$ by simply evaluating g at $\alpha = 0$. Thus, we conclude that $h(K) \in S$. Now h

is a bijection from M to M, and so we actually have $h(K) \in S \cap M$. Since K was an arbitrary element of $S \cap M$, it follows that $h(S \cap M) \subset S \cap M$. Using the involutive property of h once more, $h(S \cap M) = S \cap M$, as required. ■

Since every convex set is a star-set, and every subspace is a double-cone, we can state the following corollary to Theorem 5:

Corollary 6. *Suppose $S \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is a closed subspace, $T \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is convex, and $h(S \cap M) = T \cap M$. Then $h(S \cap M) = S \cap M$.*

Finally, we may combine the existing quadratic invariance result (Theorem 2) with the above result, and obtain a strong result connecting convexity to quadratic invariance in the Banach space case.

Corollary 7. *Suppose $G \in \mathcal{L}(\mathcal{U},\mathcal{Y})$ and $S \subset \mathcal{L}(\mathcal{Y},\mathcal{U})$ is a closed subspace such that $I - GK$ is invertible for all $K \in S$, and $M = N$. Then the set*

$$\left\{K(I-GK)^{-1} \;\middle|\; K \in S\right\}$$

is convex if and only if S is quadratically invariant with respect to G.

Proof. Note that $N \cap S = M \cap S = S$ in this case. Sufficiency is immediate from Theorem 2. Necessity holds because Theorem 5 implies that if $h(S)$ is convex, then $h(S) = S$. Then, Theorem 2 implies that S must be quadratically invariant with respect to G. ■

In the following chapter, we give an algebraic version of both the basic QI result and our new convexity result.

Chapter 6

Algebraic Framework

6.1 Introduction

We saw in Chapter 5 that subject to some technical conditions, quadratic invariance is a necessary and sufficient condition under which the set $h(S)$ is convex. We treated the Banach space case, but the QI result holds in more generality for causal maps on extended spaces [50]. This encompasses continuous and discrete systems, stable or unstable, and even systems with delays.

In both the extended case and the Banach case, the results are proven using tools from analysis. Since the maps in question are potentially infinite-dimensional, questions of convergence arise. One must also take care in defining appropriate topologies so that the notion of convergence is the correct one. Both frameworks require S to be a closed subspace, which is problematic when we seek controllers expressible as rational transfer functions.

Since the QI result holds in a very broad sense, and the QI condition is algebraic in nature, it encourages one to seek an algebraic framework in which the results can be expressed naturally. In this chapter, we present such a framework; we consider plants and controllers to be matrices whose entries belong to a commutative ring. A similar framework was suggested [53], which generalizes the notion of a transfer function matrix and applies it to feedback stabilization. Algebraic systems theory has a long and rich history, dating back to the 1960's [39]. Our work appears in [41].

In Section 6.2, we cover some required mathematical background. In Section 6.3, we present results that hold in the general commutative ring case. In Section 6.4, we consider a more specific ring; multidimensional rational functions.

6.2 Rings

A commutative ring is a tuple $(R, +, \cdot)$ consisting of a set R, and two binary operations which we call *addition* and *multiplication*, respectively. The following properties hold for all $a, b, c \in R$. First, $(R, +)$ is an abelian group:

i) Closure: $a + b \in R$

ii) Commutativity: $a + b = b + a$

iii) Associativity: $a + (b + c) = (a + b) + c$

iv) Additive identity: there exists $0_R \in R$ such that $a + 0_R = 0_R + a = a$

v) Additive inverse: there exists $-a \in R$ such that $a + (-a) = (-a) + a = 0_R$

Next, $(R, \cdot)$ is a commutative monoid:

vi) Closure: $a \cdot b \in R$

vii) Commutativity: $a \cdot b = b \cdot a$

viii) Associativity: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

ix) Multiplicative identity: there exists $1_R \in R$ such that $a \cdot 1_R = 1_R \cdot a = a$

Finally, the addition and multiplication operations satisfy two distributive properties:

x) $a \cdot (b + c) = a \cdot b + a \cdot c$

xi) $(a + b) \cdot c = a \cdot c + b \cdot c$

We will often omit the multiplication symbol, and simply concatenate the variables. So ab should be interpreted as $a \cdot b$. We will use R to denote an arbitrary commutative ring satisfying the axioms above.

If some element $a \in R$ has a multiplicative inverse in R, a is called a ***unit***. The set of all units of R forms a group under multiplication, and is denoted $U(R)$. If $U(R) = R \setminus \{0_R\}$, then R is a field.

We will often arrange elements of R into a matrix, and specify dimensions as a superscript. For example, $R^{m \times n}$ denotes the set of $m \times n$ matrices where each entry is an element of R. Some real matrix concepts extend naturally to matrices over R. The most basic are matrix addition and matrix multiplication. The determinant $\det : R^{n \times n} \to R$ is well defined, since for any $A \in R^{n \times n}$, $\det(A)$ is a polynomial in the entries $A_{ij} \in R$. The classical adjoint $\operatorname{adj} : R^{n \times n} \to R^{n \times n}$ also makes sense, because its definition is in terms of determinants of submatrices. For any $A \in R^{n \times n}$, the fundamental property of adjoints extends to the commutative ring case

$$A \operatorname{adj}(A) = \operatorname{adj}(A) A = \det(A) I_R.$$

where the matrix I_R is the identity matrix in $R^{n \times n}$. That is, the matrix whose diagonal and off-diagonal entries are 1_R and 0_R, respectively. We will use $0_R^{n \times n}$ to denote the $n \times n$ matrix whose entries are all 0_R.

The ***characteristic polynomial*** of a matrix $A \in R^{n \times n}$ is the function $p_A : R \to R$ defined by $p_A(x) = \det(A - xI_R)$. In general, p_A is a polynomial of degree n:

$$p_A(x) = p_0 + p_1 x + \cdots + p_n x^n,$$

where $p_i \in R$.

We will also use a notion that generalizes that of a subspace. An ***R-module*** consists of an abelian group $(H, +)$ and an operation $R \times H \to H$ (called scalar multiplication), such that for all $r, s \in R$ and $x, y \in H$,

i) $r(x + y) = rx + ry$

ii) $(r + s)x = rx + sx$

iii) $(rs)x = r(sx)$

iv) $1_R x = x$

In particular, a subset $S \subset R^{n\times m}$ is an R-module if it is closed under addition and satisfies the property $rX \in S$ for all $X \in S$ and $r \in R$.

The most important fact about commutative rings that we will use is the Cayley-Hamilton theorem. This well-known result for real matrices also holds in $R^{n\times n}$.

Lemma 8. *(Cayley-Hamilton) Suppose $A \in R^{n\times n}$. Define the function $\tilde{p}_A : R^{n\times n} \to R^{n\times n}$ as*

$$\tilde{p}_A(X) = p_0 I_R + p_1 X + \cdots + p_n X^n,$$

where p_i are the coefficients of the characteristic polynomial $p_A(x)$. Then $\tilde{p}_A(A) = 0_R^{n\times n}$. In other words, A satisfies its own characteristic polynomial.

Proof. See for example, [44, p. 7-8]. ■

The concept of a matrix inverse can also be extended to matrices over R. A matrix $A \in R^{n\times n}$ is ***invertible*** if $\det(A) \in U(R)$. In this case, the inverse is unique and equal to

$$A^{-1} = (\det(A))^{-1} \operatorname{adj}(A).$$

We can use the Cayley-Hamilton to express the adjoint and hence the inverse as finite sums as well.

Lemma 9. *Suppose $A \in R^{n\times n}$ is invertible. There exist $p_1, \ldots, p_n \in R$ such that*

$$-\operatorname{adj}(A) = p_1 I_R + p_2 A + \cdots + p_n A^{n-1}.$$

Proof. Using Lemma 8, we know that

$$p_0 I_R + p_1 A + \cdots + p_n A^n = 0_R^{n\times n},$$

where the p_i satisfy $\det(A - xI_R) = p_0 + p_1 x + \dots p_n x^n$. Setting $x = 0_R$, we have $p_0 = \det(A)$. Multiply by $\operatorname{adj}(A)$ on the right, and obtain

$$\det(A)\operatorname{adj}(A) + p_1 A \operatorname{adj}(A) + \dots p_n A^n \operatorname{adj}(A) = 0_R^{n\times n}.$$

Applying the fundamental property of the classical adjoint: $\operatorname{adj}(A)A = A\operatorname{adj}(A) = \det(A)I_R$, we have

$$\det(A)\left(\operatorname{adj}(A) + p_1 I_R + p_2 A + \dots + p_n A^{n-1}\right) = 0_R^{n\times n}.$$

Since A is invertible, $\det(A) \in U(R)$ and so $\det(A)$ has a multiplicative inverse. Multiply by this inverse and obtain

$$-\operatorname{adj}(A) = p_1 I_R + p_2 A + \dots + p_n A^{n-1},$$

as required. ■

6.3 QI for Rings

In this section, we take the algebraic notion of quadratic invariance [49, 50], and show how it fits into the framework of matrices over commutative rings. Our definition of quadratic invariance for rings is similar to the definition for Banach spaces.

Definition 10. *Suppose $G \in R^{m\times n}$ and $S \subset R^{n\times m}$ is an R-module. S is* ***quadratically invariant*** *with respect to G if for all $K \in S$, we have $KGK \in S$.*

We also define M and h in a manner analogous to the one presented in Chapter 5. For a particular $G \in R^{m\times n}$, define the set $M \subset R^{n\times m}$ as:

$$M = \left\{ K \in R^{n\times m} \;\middle|\; (I_R - GK) \text{ is a unit } \right\}.$$

Also define the function $h_G : M \to M$ as

$$h_G(K) = -K(I_R - GK)^{-1}.$$

Our main approach is to apply the Cayley-Hamilton theorem to show that $h(K)$ can be expressed as a finite sum of terms. When S is quadratically invariant under G, each term in the sum belongs to S and so $h(S \cap M) = S \cap M$.

Lemma 11. *Suppose $G \in R^{m \times n}$ and $S \subset R^{n \times m}$ is an R-module. Further suppose that $2_R \in U(R)$. If S is quadratically invariant with respect to G, then for all $K \in S$:*

$$K(GK)^i \in S \qquad \text{for } i = 1, 2, \ldots$$

Proof. The result follows by induction using the identity [49]:

$$\begin{aligned} K(GK)^{i+1} = 2_R^{-1} \big[&(K + K(GK)^i) G (K + K(GK)^i) \\ &- KGK - \left(K(GK)^i\right) G \left(K(GK)^i\right)\big], \end{aligned}$$

where 2_R^{-1} is the multiplicative inverse of $1_R + 1_R$, which exists by assumption. ■

Theorem 12. *Suppose $G \in R^{m \times n}$ and $S \subset R^{n \times m}$ is an R-module. Further suppose that $2_R \in U(R)$. If S is quadratically invariant with respect to G, then*

$$h(S \cap M) = S \cap M$$

Proof. Suppose $K \in S \cap M$. Using Lemma 9, write:

$$\begin{aligned} h(K) &= -\left(\det(I_R - GK)\right)^{-1} K \operatorname{adj}(I_R - GK) \\ &= \left(\det(I_R - GK)\right)^{-1} \sum_{i=1}^{m} p_i K (I_R - GK)^{i-1} \\ &= \sum_{i=1}^{m} h_i K (GK)^{i-1}, \end{aligned}$$

where the $h_i \in R$ are obtained by expanding each $(I_R - GK)^{i-1}$ term and collecting like powers of GK. All terms in the sum are in S, via Lemma 11. Since S is an R-module, it follows that $h(K) \subset S \cap M$. Using the involutive property of h, we have $h(S \cap M) = S \cap M$. ■

Counterexample. We will now show that the requirement $2_R \in U(R)$ is necessary. Consider the ring of integers $\mathbb{Z}$, and define:

$$S = \left\{ \begin{bmatrix} 2x & y & z \\ y & z & 0 \\ z & 0 & 0 \end{bmatrix} \;\middle|\; x, y, z \in \mathbb{Z} \right\}, \qquad G = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

It is easy to check that S is a $\mathbb{Z}$-module, and is quadratically invariant with respect to G. Now consider a particular element of S:

$$K_0 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Note that $\det(I - GK_0) = 1$, so $K_0 \in S \cap M$. However, $h(K_0) \notin S$, so Theorem 12 does not hold. Indeed, $K_0(GK_0)^2 \notin S$, so Lemma 11 does not hold either. The requirement that $2_R \in U(R)$ can be dropped if we strengthen our notion of quadratic invariance. One way to do this is to require that $K_1 G K_2 \in S$ for all $K_1, K_2 \in S$.

This result shows that in a purely algebraic setting, quadratic invariance implies that the set of achievable closed-loop maps is affine. For the remainder of the chapter, we turn our attention to a more specialized commutative ring: rational functions.

6.4 Rational Functions

We now turn our attention to rational functions of multiple variables. This leads to quadratic invariance results without any closure requirement on S. Furthermore, the framework is flexible enough to allow systems with delays or spatiotemporal systems (see Section 6.6).

Let $\mathbb{R}(\mathbf{s})$ be the set of rational functions in the variables $\mathbf{s} = (s_1, s_2, \ldots, s_k)$, with coefficients in $\mathbb{R}$. We say that $r \in \mathbb{R}(\mathbf{s})$ is ***proper*** if for every i, the degree of s_i in the numerator is less than or equal to the degree of s_i in the denominator. The set of proper rationals will be denoted $\mathbb{R}(\mathbf{s})_p$. For example, the rational function

$$\frac{s_1 s_2 s_3}{s_1^2 + 2s_2 + s_3}$$

is proper. Similarly, we define $\mathbb{R}(\mathbf{s})_{sp}$ to be the set of strictly proper rationals. Finally, $\mathbb{R}(\mathbf{s})_n$ is the set of rationals that are proper but not strictly proper. That is, each variable s_i has the same degree in the numerator and denominator. As a convention, $0 \in \mathbb{R}(\mathbf{s})_{sp}$. We may alternatively characterize properness by using limits. We state the following lemma without proof.

Lemma 13. *Suppose $h \in \mathbb{R}(\mathbf{s})$. For every $i \in \{1, \dots, k\}$, Let*

$$\bar{s}^i = \{\bar{s}_1, \dots, \bar{s}_{i-1}, \bar{s}_{i+1}, \dots, \bar{s}_k\} \subset \mathbb{R}$$

be some assignment of the remaining $k-1$ variables. Define

$$c_i(\bar{s}^i) = \lim_{s_i \to \infty} h(\bar{s}_1, \dots, \bar{s}_{i-1}, s_i, \bar{s}_{i+1}, \dots, \bar{s}_k).$$

We have:

$$h \in \mathbb{R}(\mathbf{s})_p \iff \begin{cases} \textit{for all } i \in \{1, \dots, k\}, \\ c_i(\bar{s}^i) \textit{ is finite for almost all } \bar{s}^i \end{cases}$$

$$h \in \mathbb{R}(\mathbf{s})_{sp} \iff \begin{cases} \textit{for all } i \in \{1, \dots, k\}, \\ c_i(\bar{s}^i) = 0 \textit{ for almost all } \bar{s}^i \end{cases}$$

The definition of invertibility follows from the definition used with R. Since $U(\mathbb{R}(\mathbf{s})) = \mathbb{R}(\mathbf{s}) \setminus \{0\}$, a matrix $A \in \mathbb{R}(\mathbf{s})^{n \times n}$ is invertible if $\det(A)$ is not identically zero. It follows that $\mathbb{R}(\mathbf{s})$ is in fact a field. The set $\mathbb{R}(\mathbf{s})_p \subset \mathbb{R}(\mathbf{s})$ is closed under addition and multiplication, but not inversion. It is therefore a subring of $\mathbb{R}(\mathbf{s})$. The invertible proper elements are precisely the set $\mathbb{R}(\mathbf{s})_n = U(\mathbb{R}(\mathbf{s})_p)$. The remaining elements are strictly proper, $\mathbb{R}(\mathbf{s})_{sp} \subset \mathbb{R}(\mathbf{s})_p$, and are an ideal of $\mathbb{R}(\mathbf{s})_p$.

Lemma 14. *Suppose* $G \in \mathbb{R}(\mathbf{s})_{sp}^{m\times n}$ *and* $K \in \mathbb{R}(\mathbf{s})_{p}^{n\times m}$. *Then* $(I - GK)$ *is invertible, and* $(I - GK)^{-1} \in \mathbb{R}(\mathbf{s})_{p}^{m\times m}$.

Proof. By Lemma 13, we have that for any i,

$$\lim_{s_i\to\infty} \det(I - GK) = \det(I) = 1$$

for almost any assignment of the remaining variables $\{s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_k\}$. This holds because G is strictly proper and K is proper. Applying Lemma 13 once more, we conclude that $\det(I - GK) \in \mathbb{R}(\mathbf{s})_n$ and so $(I - GK)$ is invertible. Furthermore, $(I - GK) \in \mathbb{R}(\mathbf{s})_p$. Consequently, $\operatorname{adj}(I - GK) \in \mathbb{R}(\mathbf{s})_p$ because $\mathbb{R}(\mathbf{s})_p$ is a subring. It follows that $(I - GK)^{-1} \in \mathbb{R}(\mathbf{s})_p$. ■

6.5 QI for Rationals

In this section, we prove our main result: for rational functions, quadratic invariance is a necessary and sufficient condition under which $h(S)$ is equal to S.

Lemma 15. *Suppose* $G \in \mathbb{R}(\mathbf{s})_{sp}^{m\times n}$ *and* $K \in \mathbb{R}(\mathbf{s})_{p}^{n\times m}$. *Then there exist* $h_1, \ldots, h_m \in \mathbb{R}(\mathbf{s})_n$ *such that:*

$$h(K) = \sum_{i=1}^{m} h_i K(GK)^{i-1}.$$

Proof. By Lemma 14, $(I - GK)$ is invertible, and so $h(K)$ is always well-defined. We may express it in terms of the classical adjoint:

$$h(K) = -K(I - GK)^{-1} = \frac{-1}{\det(I - GK)} K \operatorname{adj}(I - GK),$$

and apply Lemma 9 to express the adjoint as a finite sum:

$$\begin{aligned} h(K) &= \frac{1}{\det(I-GK)}\sum_{i=1}^{m} p_i K(I-GK)^{i-1} \\ &= \frac{1}{\det(I-GK)}\sum_{i=1}^{m} p_i \sum_{j=1}^{i}(-1)^{j-1}\binom{i-1}{j-1}K(GK)^{j-1} \\ &= \sum_{j=1}^{m}\underbrace{\left[\frac{(-1)^{j-1}}{\det(I-GK)}\sum_{i=j}^{m}\binom{i-1}{j-1}p_i\right]}_{h_j}K(GK)^{j-1}. \end{aligned}$$

The next step is to show that $h_j \in \mathbb{R}(\mathbf{s})_n$. We will do this via Lemma 13 by showing that the limits $s_i \to \infty$ are finite and nonzero for almost all assignments of the remaining variables. Recall that the p_i are defined in terms of a determinant:

$$p(x) = \det(I - GK - xI) = p_0 + p_1 x + \cdots + p_m x^m.$$

Now apply Lemma 13. For every i, and any $x \in \mathbb{R}$,

$$\lim_{s_i\to\infty} \det(I - GK - xI) = (1-x)^m$$

for almost all $\bar{s}^i$. Equating coefficients, we find

$$\lim_{s_i\to\infty} p_i(\bar{s}_1,\ldots,\bar{s}_{i-1},s_i,\bar{s}_{i+1},\ldots,\bar{s}_k) = (-1)^i\binom{m}{i}$$

for almost all $\bar{s}^i$. Using this fact, we may now evaluate the limit of each h_j as $s_i \to \infty$.

$$\begin{aligned} &\lim_{s_i\to\infty} h_j(\bar{s}_1,\ldots,\bar{s}_{i-1},s_i,\bar{s}_{i+1},\ldots,\bar{s}_k) \\ &\qquad = (-1)^{j-1}\sum_{i=j}^{m}\binom{i-1}{j-1}\binom{m}{i}(-1)^i \\ &\qquad = -1 \end{aligned}$$

for almost all $\bar{s}^i$, and we conclude that $h_j \in \mathbb{R}(\mathbf{s})_n$, as required. ■

Theorem 16. *Suppose $G \in \mathbb{R}(\mathbf{s})_{sp}^{m\times n}$, and $S \subset \mathbb{R}(\mathbf{s})_{p}^{n\times m}$ is an $\mathbb{R}(\mathbf{s})_p$-module.*

$$S \text{ is QI with respect to } G \qquad \Longleftrightarrow \qquad h(S) = S.$$

Proof. ($\Longrightarrow$) Choose $K \in S$. Using Lemma 15, write:

$$h(K) = \sum_{j=1}^{m} h_j K(GK)^{j-1}$$

where $h_j \in \mathbb{R}(\mathbf{s})_n$. By Lemma 11, $K(GK)^j \in S$. Since S is an $\mathbb{R}(\mathbf{s})_p$-module, the finite sum also belongs to S, and we conclude that $h(S) \subset S$. By the involutive property of h, it follows that $h(S) = S$.

($\Longleftarrow$) Suppose conversely that S is not QI with respect to G. So there must exist some $K_0 \in S$ with $K_0GK_0 \notin S$. Let $r \in \mathbb{R}$ and define $K = rK_0$. Note that $K \in S$, since S is an $\mathbb{R}(\mathbf{s})_p$-module. Now write:

$$h(K) = \sum_{j=1}^{m} h_j K(GK)^{j-1} = \sum_{j=1}^{m} r^j h_j(r) \underbrace{K_0(GK_0)^{j-1}}_{F_j}.$$

Note that $h_j(r)$ depends on r. We have $F_j \in \mathbb{R}(\mathbf{s})_{sp}^{n\times m}$ as well. Assume that $h(K) \in S$ for every r. Then in particular, for any choice of $r_1, r_2, \ldots, r_m \in \mathbb{R}$ such that the quantities $h_i(r_j)$ are well-defined, we have:

$$\begin{gathered} h_1(r_1)F_1 + r_1 h_2(r_1)F_2 + \cdots + r_1^{m-1} h_m(r_1)F_m \in S \\ \vdots \\ h_1(r_m)F_1 + r_m h_2(r_m)F_2 + \cdots + r_m^{m-1} h_m(r_m)F_m \in S. \end{gathered} \tag{6.1}$$

Any linear combination of the sums on the left-hand side must also belong to S.

Consider the matrix

$$C = \begin{bmatrix} h_1(r_1) & r_1 h_2(r_1) & \cdots & r_1^{m-1} h_m(r_1) \\ h_1(r_2) & r_2 h_2(r_2) & \cdots & r_2^{m-1} h_m(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(r_m) & r_m h_2(r_m) & \cdots & r_m^{m-1} h_m(r_m) \end{bmatrix}$$

By Lemma 15, it is a matrix of rational functions. In fact, $C \in \mathbb{R}(\mathbf{s})_p^{m\times m}$. We would like to verify that C is invertible, so we apply Lemma 13. This is straightforward since we already know the limits of the h_j from Lemma 15. For every i and for almost all $\bar{s}^i$,

$$\begin{aligned} &\lim_{s_i\to\infty} \det(C(\bar{s}_1, \ldots, \bar{s}_{i-1}, s_i, \bar{s}_{i+1}, \ldots, \bar{s}_k) \\ &\qquad = (-1)^m \det \begin{bmatrix} 1 & r_1 & \cdots & r_1^{m-1} \\ 1 & r_2 & \cdots & r_2^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & r_m & \cdots & r_m^{m-1} \end{bmatrix} \\ &\qquad = (-1)^m \prod_{1\le i<j\le m} (r_j - r_i), \end{aligned}$$

where we used a property of Vandermonde matrices to evaluate the determinant. As long as we choose distinct r_i, $\det(C)$ tends to a finite and nonzero limit, and so C is invertible, and $C^{-1} \in \mathbb{R}(\mathbf{s})_p^{m\times m}$. If we treat the rows of C^{-1} as coefficients, and compute the corresponding linear combinations of (6.1), we obtain m equations:

$$F_i \in S \qquad i = 1, \ldots, m.$$

In particular, we have $F_2 \in S$. But $F_2 = K_0 G K_0 \notin S$, a contradiction. We conclude that our assumption was incorrect, so there exists some K for which $h(K) \notin S$. ■

6.6 Examples

6.6.1 Sparse Controllers

The simplest class of systems that we can analyze are systems with rational transfer functions subject to controllers with sparsity constraints. It is clear that if every nonzero entry in the controller is required to be a proper rational function in $\mathbb{R}(s)_p$, the set S of allowable controllers is an $\mathbb{R}(s)_p$-module.

6.6.2 Network with Delays

Consider a distributed system where the subsystems affect one another via delay constraints. We wish to design a decentralized controller subject to communication delay constraints between subcontrollers.

Introduce the delay operator d that represents a delay of one time unit. The plant G and controller K are therefore rational functions in s and d. The constraint $K \in \mathbb{R}(s,d)_p$ naturally guarantees that negative delays are forbidden, thus enforcing causality.

Define the ***delay*** of a transfer function as the difference between the degree of d in its denominator and numerator. For example,

$$\text{delay}\left(\frac{1}{sd+2}\right) = 1 \quad \text{and} \quad \text{delay}\left(\frac{s+d^2}{s^2d+d^5}\right) = 3.$$

As a convention, $\text{delay}(0) = \infty$. We can impose delay constraints on the controller using a set of the form

$$S = \{K \in \mathbb{R}(s,d)_p \mid \text{delay}(K_{ij}) \geq a_{ij}\},$$

where $a_{ij} \geq 0$ is the minimum delay between subcontrollers i and j. One can verify that S is an $\mathbb{R}(s,d)_p$-module, and so we may apply Theorem 16 to derive conditions under which the problem is convex. Similar results proved using very different methods can be found in [48].

6.6.3 Multidimensional Systems

Rational functions in multiple variables with mixed properness constraints are valid in our framework. For example, suppose our transfer functions depend on two sets of variables: $R = \mathbb{R}(s_1, \ldots, s_m, z_1, \ldots, z_n)$. Further suppose that we impose a properness constraint on $s_1, \ldots, s_m$, but not on $z_1, \ldots, z_n$. This might occur, for example, if some of the variables are spatial, and it doesn't make sense to impose a properness constraint on them. This framework is used to represent spatiotemporal dynamics in a variety of important papers [46, 36, 33, 35].

The set R is indeed a commutative ring, and so we may apply Theorem 12. Theorems 16 and 24 hold as well, with appropriate modifications to the notation.

Chapter 7

Internal Quadratic Invariance

7.1 Introduction

In the previous two chapters, we extended the notion of quadratic invariance to a general algebraic framework. When the decentralization constraint set is quadratically invariant, the set of achievable closed-loop maps is affine. In this chapter, we develop the notion of *internal quadratic invariance* (IQI). When a constraint set is IQI, it may be transformed into an equivalent QI set even though it is not initially QI. We show that the IQI property is easy to test, and is more general than QI [40, 42, 41].

7.2 Preliminaries

In this chapter, we use the same the rational function framework as in Chapter 6, but we will only treat the case of rational functions of a single variable. To simplify notation, we will let $\mathcal{R}$ denote the set of rational functions in z. We begin with some additional definitions.

If $A \in \mathcal{R}_p^{m\times n}$, define the range and nullspace of A as

$$\begin{aligned} \operatorname{range} A &= \left\{Ax \mid x \in \mathcal{R}_p^n\right\} \\ \operatorname{null} A &= \left\{x \in \mathcal{R}_p^n \;\middle|\; Ax = 0\right\}. \end{aligned}$$

These sets are both $\mathcal{R}_p$-modules. Also, we call a matrix $W \in \mathcal{R}^{n\times n}$ a ***projector*** if $W^2 = W$. We will also require the concept of ***normal rank***. A matrix $A \in \mathcal{R}^{m\times n}$ has normal rank k if $A(z)$ is a rank-k matrix for all but finitely many $z \in \mathbb{C}$. For an introduction to normal rank and related concepts, see [38].

7.3 Main Result

Our main result is that subject to a condition we call *internal quadratic invariance*, the set of achievable closed-loop maps is affine, and thus amenable to convex search.

Definition 17. *Let $P \in \mathcal{R}_{sp}^{m\times n}$ and $S \subset \mathcal{R}_p^{n_2\times m_2}$ be an $\mathcal{R}_p$-module. Let W_1 and W_2 be any projectors such that*

$$\begin{aligned} \operatorname{range} W_1 &= \operatorname{range} \begin{bmatrix} P_{21} & P_{22} \end{bmatrix} \\ \operatorname{null} W_2 &= \operatorname{null} \begin{bmatrix} P_{12} \\ P_{22} \end{bmatrix}. \end{aligned} \tag{7.1}$$

*We say S is **internally quadratically invariant** (IQI) with respect to P if $W_2 S W_1$ is QI with respect to P_{22}.*

Proper projectors satisfying (7.1) always exist. Furthermore, internal quadratic invariance does not depend on the particular choice of projectors W_i. In other words, internal quadratic invariance is a property of P and S alone. Before we can prove these claims, we need two additional Lemmas.

Lemma 18. *Suppose $G \in \mathcal{R}^{m_2\times n_2}$. Let W_1 and W_2 be projectors.*

$$W_1 G W_2 = G \quad \iff \quad \begin{cases} \operatorname{range} G \subset \operatorname{range} W_1 \\ \operatorname{null} G \supset \operatorname{null} W_2 \end{cases}.$$

Proof. Suppose $\operatorname{range} G \subset \operatorname{range} W_1$. Let $x \in \mathcal{R}^{n_2}$. Now $Gx \in \operatorname{range} G \subset \operatorname{range} W_1$. So there exists some $y \in \mathcal{R}^{n_2}$ such that $Gx = W_1 y$. Since W_1 is a projector, $W_1 G x = W_1^2 y = W_1 y = Gx$. Therefore, $W_1 G = G$. Conversely, suppose $W_1 G = G$. Then

$\operatorname{range} G \subset \operatorname{range} W_1 G = \operatorname{range} W_1$. Similarly, we have $W_2^T G^T \iff \operatorname{range} G^T \subset \operatorname{range} W_2^T$. The result follows from taking orthogonal complements and using range-nullspace duality. ■

Lemma 19. *Suppose $P \in \mathcal{R}_{sp}^{m\times n}$ and $S \subset \mathcal{R}_p^{n_2 \times m_2}$ is an $\mathcal{R}_p$-module. Further suppose that S is internally quadratically invariant with respect to P, and let W_1, W_2 be projectors satisfying (7.1). Then,*

i) $\begin{bmatrix} I & 0 \\ 0 & W_1 \end{bmatrix} P \begin{bmatrix} I & 0 \\ 0 & W_2 \end{bmatrix} = P$

ii) $h(W_2 S W_1) = W_2 h(S) W_1$.

Proof. The range and nullspace requirement for W_1 and W_2 imply that

$$\operatorname{range} P \subset \operatorname{range} \begin{bmatrix} I & 0 \\ 0 & W_1 \end{bmatrix} \qquad \text{and} \qquad \operatorname{null} P \supset \operatorname{null} \begin{bmatrix} I & 0 \\ 0 & W_2 \end{bmatrix}.$$

Applying Lemma 18, we conclude that

$$\begin{bmatrix} I & 0 \\ 0 & W_1 \end{bmatrix} P \begin{bmatrix} I & 0 \\ 0 & W_2 \end{bmatrix} = P.$$

In particular, we have $W_1 P_{22} W_2 = P_{22}$. Using this identity, we have

$$\begin{aligned} h(W_2 K W_1) &= -(W_2 K W_1)\left[I - P_{22}(W_2 K W_1)\right]^{-1} \\ &= -W_2 K \left[I - W_1 P_{22} W_2 K\right]^{-1} W_1 \\ &= -W_2 K \left[I - P_{22} K\right]^{-1} W_2 \\ &= W_2 h(K) W_1. \end{aligned}$$

■

We will now show that internal quadratic invariance does not depend on the choice of projectors W_1 and W_2, it is a property of P and S alone. First, we show the required existence property, that it is always possible to construct projectors satisfying (7.1).

Lemma 20. *Suppose $A \in \mathcal{R}^{m \times n}$. There exists a proper projector $W \in \mathcal{R}_p^{m \times m}$ with the same range (or nullspace) as A.*

Proof. We may factor A as

$$A = \underbrace{\begin{bmatrix} U_1 & U_2 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} M_1 & 0 \\ 0 & 0 \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}}_{V},$$

where U and V are unimodular polynomial matrices and M is the Smith-McMillan form of A [38]. Then, it is straightforward to verify that $W_1 = U_1(U_1^T U_1)^{-1} U_1^T$ is a projector with $\operatorname{range} W_1 = \operatorname{range} A$. Similarly, $W_2 = V_1(V_1^T V_1)^{-1} V_1^T$ is a projector with $\operatorname{null} W_2 = \operatorname{null} A$.

One can show that the limit $\lim_{z \to \infty} W(z)$ is always a constant for projectors constructed in this fashion, and thus W is proper. ■

Next, we show that for fixed P and S, whether or not $W_2 S W_1$ is QI with respect to G does not depend on our choice of projectors.

Lemma 21. *Suppose $P \in \mathcal{R}^{m \times n}$, and $S \subset \mathcal{R}^{n_2 \times m_2}$ is an $\mathcal{R}$-module. Suppose further that W_i and Z_i are two sets of projectors satisfying (7.1). Then the two following statements are equivalent*

i) $W_2 S W_1$ is QI with respect to G

ii) $Z_2 S Z_1$ is QI with respect to G.

Proof. Since W_1 and Z_1 have the same range, each column of Z_1 is a linear combination of the columns of W_1. Therefore we can write $Z_1 = W_1 X$ for some $X \in \mathcal{R}^{m_2 \times m_2}$. Similar arguments imply that W_2^T and Z_2^T have the same range, so we may similarly conclude that $Z_2 = Y W_2$ for some $Y \in \mathcal{R}^{n_2 \times n_2}$.

Now suppose that $W_2 S W_1$ is QI with respect to G. Then for all $K \in S$,

$$(W_2 K W_1) G (W_2 K W_1) \in W_2 S W_1.$$

Multiply on the left by Y and on the right by X, and deduce that

$$(Z_2KW_1)G(W_2KZ_1) \in Z_2SZ_1.$$

From Lemma 19, we have

$$W_1GW_2 = Z_1GZ_2 = G$$

and so

$$(Z_2KZ_1)G(Z_2KZ_1) \in Z_2SZ_1,$$

and we conclude that Z_2SZ_1 is QI with respect to G. The same argument holds if we interchange W and Z, and this completes the proof. ■

Quadratic invariance is only a property of the information constraint S and the P_{22} block of the plant. However, internal quadratic invariance also depends on the other blocks P_{ij} of the plant. We now show that IQI is weaker than QI. That is, all QI systems are IQI.

Theorem 22. *If S is QI with respect to G. Then S is IQI with respect to P.*

Proof. Suppose S is QI with respect to G. Then for any $K \in S$, $KGK \in S$. Now choose W_1, W_2 as in Definition 17. It follows that $W_2KGKW_1 \in W_2SW_1$. We also have from Lemma 19 that $W_1GW_2 = G$. Thus, $(W_2KW_1)G(W_2GW_1) \in W_2SW_1$. ■

Most notably, if $\begin{bmatrix} P_{21} & P_{22} \end{bmatrix}$ and $\begin{bmatrix} P_{12} \\ P_{22} \end{bmatrix}$ have full normal rank, both projectors W_1 and W_2 as defined in (7.1) are the identity. In this case, the two notions are equivalent: S is IQI with respect to P if and only if S is QI with respect to G.

A simple consequence of Theorem 16 is that we may pre- and post-multiply by matrices W_1 and W_2, and the result will still hold.

Corollary 23. *Suppose $G \in \mathcal{R}_{sp}^{m\times n}$, $S \subset \mathcal{R}_p^{n\times m}$ is an $\mathcal{R}_p$-module, and $W_1 \in \mathcal{R}_p^{m\times m}$, $W_2 \in \mathcal{R}_p^{n\times n}$ are square matrices. Then,*

$$W_2SW_1 \text{ is QI with respect to } G \quad \Longleftrightarrow \quad h(W_2SW_1) = W_2SW_1.$$

Proof. It is straightforward to verify that if S is an $\mathcal{R}_p$-module, then so is $W_2 S W_1$. The proof follows immediately from Theorem 16. ■

We can now extend Theorem 16 to the IQI case. By choosing W_1 and W_2 in a particular way, we can find a sufficient condition under which the closed-loop map is convex. This condition is weaker than the QI condition, meaning that it is more general.

Theorem 24. *Let $P \in \mathcal{R}_{sp}^{m\times n}$, and suppose $S \subset \mathcal{R}_p^{n_2\times m_2}$ is an $\mathcal{R}_p$-module. If S is IQI with respect to P, then*

$$P_{12} h(S) P_{21} = P_{12} S P_{21}.$$

Proof. Let W_1 and W_2 be proper projectors satisfying (7.1). Using Corollary 23, we have $h(W_2 S W_1) = W_2 S W_1$. Applying Lemma 19, this is equivalent to

$$W_2 h(S) W_1 = W_2 S W_1.$$

Multiply on the left by P_{12} and on the right by P_{21}. Using Lemma 19 again, we have that $P_{12} W_2 = P_{12}$ and $W_1 P_{21} = P_{21}$. Therefore,

$$P_{12} h(S) P_{21} = P_{12} S P_{21}.$$

■

Theorem 24 is the main result of this chapter, and shows us that internal quadratic invariance leads to an affine set of achievable closed-loop maps. In Section 7.5, we will show how this theorem is applied to solve controller synthesis problems. First, however, will will present an alternative interpretation of IQI using reduction rather than projection.

7.4 Model Reduction Interpretation

As we saw in this chapter, the QI property depends on our choice of representation. If we have a plant P and two different information constraints S and $\tilde{S}$, it is possible that (P,S) and $(P,\tilde{S})$ have the same set of achievable closed-loop maps. Furthermore, it is possible for one of these pairs to be QI while the other pair is not. In this section, we show that if (P,S) is IQI but not QI, we can construct an equivalent system $(\tilde{P},\tilde{S})$ that is QI, and has fewer inputs and outputs than (P,S).

This idea of finding a *reduced* representation for a system also comes up in state-space theory. In state-space, a transfer function may be represented by many different choices of matrices (A,B,C,D). However, the *minimal* representation will have the smallest A possible.

We begin by observing that Lemma 20 still holds if we do not require W_1 and W_2 to be projectors. Indeed, if we define U_1 and V_1 as in Lemma 20, then we can instead set:

$$\begin{aligned}\operatorname{range} U_1 &= \operatorname{range} \begin{bmatrix} P_{21} & P_{22} \end{bmatrix} \\ \operatorname{null} V_1^T &= \operatorname{null} \begin{bmatrix} P_{12} \\ P_{22} \end{bmatrix}.\end{aligned} \tag{7.2}$$

The difference is that U_1 and V_1 are now skinny and have full normal rank. These definitions determine a matrix factorization similar to that of Lemma 19, but with a different P on the right:

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & U_1 \end{bmatrix} \begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \tilde{P}_{21} & \tilde{P}_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & V_1^T \end{bmatrix}.$$

It is clear that $\tilde{P}$ can be computed via the formula:

$$\begin{bmatrix} \tilde{P}_{11} & \tilde{P}_{12} \\ \tilde{P}_{21} & \tilde{P}_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & U_1^\dagger \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & V_1^\dagger \end{bmatrix}^T, \tag{7.3}$$

where we have defined $U_1^\dagger = (U_1^T U_1)^{-1} U_1^T$ and similarly for $V_1^\dagger$. Using these identities, we can compute the set of achievable closed-loop maps:

$$\begin{aligned}P_{11} + P_{12}K\left[I - P_{22}K\right]^{-1} P_{21} &= \tilde{P}_{11} + \tilde{P}_{12}V_1^TK\left[I - U_1\tilde{P}_{22}V_1^TK\right]^{-1}U_1\tilde{P}_{21}\\ &= \tilde{P}_{11} + \tilde{P}_{12}(V_1^TKU_1)\left[I - \tilde{P}_{22}(V_1^TKU_1)\right]^{-1}\tilde{P}_{21}.\end{aligned}$$

Thus, we may draw two equivalent block diagrams for our system; see Figure 7.1.

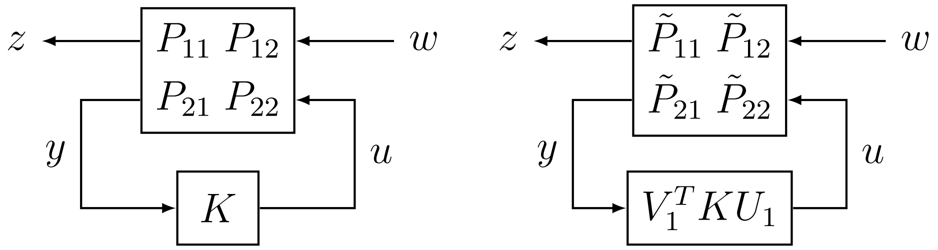

Figure 7.1: Two equivalent block diagrams

We can define a new information constraint $\tilde{S} = V_1^TSU_1$, which applies to the $\tilde{P}$ plant. Notice that because U_1 and V_1 are skinny matrices, $(\tilde{P}, \tilde{S})$ will have fewer inputs and outputs than (P, S). In this sense, we may think of it as a *reduced* system. The key result of this section is that testing internal quadratic invariance of (P, S) is equivalent to testing quadratic invariance of $(\tilde{P}, \tilde{S})$. The result is stated formally in Theorem 25.

Theorem 25. *Suppose $P \in \mathcal{R}_{sp}^{m\times n}$ and $S \subset \mathcal{R}_p^{n\times m}$ is an $\mathcal{R}_p$-module. Let U_1 and V_1 be skinny and full-normal-rank polynomial matrices that satisfy (7.2). Now compute $\tilde{P}$ using (7.3) and let $\tilde{S} = V_1^TSU_1$. Then S is IQI with respect to P if and only if $\tilde{S}$ is QI with respect to $\tilde{P}_{22}$.*

Proof. ($\Longrightarrow$) Suppose S is IQI with respect to P. Then for any choice of projectors W_1 and W_2 satisfying (7.1), we have that W_2SW_1 is QI with respect to P_{22}. In particular, we saw in Lemma 20 that one valid choice is $W_1 = U_1U_1^\dagger$ and $W_2 = (V_1V_1^\dagger)^T$. Applying this choice to the definition of QI, we find that $(W_2KW_1)P_{22}(W_2KW_1) \in W_2SW_1$. Substituting, we obtain

$$V_1^{\dagger^T}V_1^TKU_1U_1^\dagger P_{22}V_1^{\dagger^T}V_1^TKU_1U_1^\dagger \in V_1^{\dagger^T}V_1^TSU_1U_1^\dagger. \tag{7.4}$$

Now multiply on the left by V_1^T and on the right by U_1, and simplify.

$$V_1^T K U_1 U_1^\dagger P_{22} V_1^{\dagger T} V_1^T K U_1 \in V_1^T S U_1.$$

Recall from (7.3) that $\tilde{P}_{22} = U_1^\dagger P_{22} V_1^{\dagger T}$. So we have

$$V_1^T K U_1 \tilde{P}_{22} V_1^T K U_1 \in V_1^T S U_1. \tag{7.5}$$

In other words, $\tilde{S}$ is QI with respect to $\tilde{P}_{22}$, as required.

($\Longleftarrow$) Suppose that $\tilde{S}$ is QI with respect to $\tilde{P}_{22}$. Begin with (7.5), multiply on the left by $V_1^{\dagger T}$ and on the right by $U_1^\dagger$, and obtain (7.4). ■

In other words, if we want to use quadratic invariance as a test for tractability, we should always perform the test on the reduced system $(\tilde{P}, \tilde{S})$. A reduced system exists whenever the matrices U_1 or V_1 are strictly skinny. And this happens when either $\begin{bmatrix} P_{21} & P_{22} \end{bmatrix}$ or $\begin{bmatrix} P_{12} \\ P_{22} \end{bmatrix}$ is not full-normal-rank, respectively. In Section 7.6, we show a numerical example that illustrates reduction.

7.5 IQI and Optimization

In this section, we show how to solve optimal control synthesis problems when the information constraint S is internally quadratically invariant with respect to the plant P. Return to the optimization formulation (5.2), but now suppose that $P \in \mathcal{R}_{sp}$:

$$\begin{aligned} \text{minimize} \quad & \left\| P_{11} - P_{12} Q P_{21} \right\| \\ \text{subject to} \quad & Q \in h(S). \end{aligned}$$

We saw that in the QI case, Theorem 16 guarantees us that $h(S) = S$. So the problem may be transformed into:

$$\begin{aligned} \text{minimize} \quad & \left\| P_{11} - P_{12} Q P_{21} \right\| \\ \text{subject to} \quad & Q \in S. \end{aligned} \tag{7.6}$$

Once we have found the optimal Q_{opt}, we recover K via $K_{\text{opt}} = h(Q_{\text{opt}})$. In the IQI case, Theorem 24 guarantees us that $P_{12}h(S)P_{21} = P_{12}SP_{21}$, even though $h(S) \neq S$. So the optimization problem in the IQI case reduces to (7.6), just as in the QI case.

The main difference is that in the IQI case, $h(S) \neq S$, so we must do something different to obtain K_{opt}. The procedure is explained in the following Theorem.

Theorem 26. *Suppose $P \in \mathcal{R}_{sp}^{m\times n}$ and $S \subset \mathcal{R}_p^{n\times m}$ is an $\mathcal{R}_p$-module. Further suppose that S is internally quadratically invariant with respect to P, and let W_1, W_2 be projectors satisfying (7.1). Let K_{opt} be the solution to the optimization problem*

$$\begin{aligned} &\textit{minimize} && \left\|P_{11} + P_{12}K(I-GK)^{-1}P_{21}\right\| \\ &\textit{subject to} && K \in S. \end{aligned} \tag{7.7}$$

Then K_{opt} satisfies the constrained linear equations

$$W_2 K_{opt} W_1 = h(W_2 Q_{opt} W_1), \tag{7.8}$$

where Q_{opt} is the solution to the optimization problem

$$\begin{aligned} &\textit{minimize} && \left\|P_{11} - P_{12}QP_{21}\right\| \\ &\textit{subject to} && Q \in S. \end{aligned} \tag{7.9}$$

Proof. Once we have found $Q_{\text{opt}} \in S$ that solves (7.9), then we know from Corollary 23 that there exists some $K_{\text{opt}} \in S$ that satisfies

$$h(W_2 K_{\text{opt}} W_1) = W_2 Q_{\text{opt}} W_1. \tag{7.10}$$

If we apply h to this equation, we find that K_{opt} must satisfy (7.8), as required. We can verify that the K_{opt} obtained from this equation is optimal by computing:

$$\begin{aligned}
P_{11} - P_{12}Q_{\text{opt}}P_{21} &= P_{11} - P_{12}W_2Q_{\text{opt}}W_1P_{21} && \text{(Lemma 19, part } i\text{)}\\
&= P_{11} - P_{12}h(W_2K_{\text{opt}}W_1)P_{21} && \text{(Equation (7.10))}\\
&= P_{11} - P_{12}W_2h(K_{\text{opt}})W_1P_{21} && \text{(Lemma 19, part } ii\text{)}\\
&= P_{11} - P_{12}h(K_{\text{opt}})P_{21} && \text{(Lemma 19, part } i\text{).}
\end{aligned}$$

Thus the cost we optimized to find Q_{opt} is the same as the cost we were trying to optimize in (7.7). ■

In other words, we can solve IQI optimization problems by first solving the simpler optimization (7.9) for Q_{opt}, and then solving the linear equations (7.8) for K_{opt}. We new present some examples of internally quadratically invariant systems.

7.6 Examples

7.6.1 Simple Example

Consider the following plant and information constraint:

$$P = \left[\begin{array}{c:ccc} a & b_1 & b_2 & b_2 \\ \hdashline c_1 & g_1 & 0 & 0 \\ c_1 & g_1 & 0 & 0 \\ c_2 & g_2 & g_3 & g_3 \end{array}\right], \qquad S = \left\{ \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix} \;\middle|\; k_i \in \mathbb{R} \right\},$$

where a, b_i, c_i, g_i are real numbers. Note that the dotted lines in P are simply there to show how P is partitioned into its four blocks. It does not denote a state-space notation. Since S is diagonal, it is clear that $KP_{22}K$ will never be diagonal, and thus S is not QI with respect to P_{22}. Now compute W_1, W_2 using Definition 17 and U_1, V_1

using (7.2):

$$\operatorname{range}\begin{bmatrix} c_1 & g_1 & 0 & 0 \\ c_1 & g_1 & 0 & 0 \\ c_2 & g_2 & g_3 & g_3 \end{bmatrix} = \operatorname{range}\underbrace{\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{W_1} = \operatorname{range}\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{U_1}$$

$$, \operatorname{null}\begin{bmatrix} b_1 & b_2 & b_2 \\ g_1 & 0 & 0 \\ g_1 & 0 & 0 \\ g_2 & g_3 & g_3 \end{bmatrix} = \operatorname{null}\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}}_{W_2} = \operatorname{null}\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}}_{V_1^T}.$$

The set W_2SW_1 is given by

$$W_2SW_1 = \left\{ \begin{bmatrix} k_1 & k_1 & 0 \\ k_2 & k_2 & k_3 \\ k_2 & k_2 & k_3 \end{bmatrix} \;\middle|\; k_i \in \mathbb{R} \right\},$$

and notice that

$$(W_2KW_1)P_{22}(W_2KW_1) = \begin{bmatrix} v_1 & v_1 & 0 \\ v_2 & v_2 & v_3 \\ v_2 & v_2 & v_3 \end{bmatrix}, \in W_2SW_1$$

where $v_1 = \frac{1}{2}g_1k_1^2$, $v_2 = \frac{1}{4}(g_1k_1k_2 + g_2k_1k_3 + g_3k_2k_3)$, and $v_3 = \frac{1}{2}g_3k_3^2$. Therefore, W_2SW_1 is QI with respect to G, and so S is internally quadratically invariant with respect to P.

Using the model reduction ideas of Section 7.4, we can establish the IQI property by computing the reduced system $(\tilde{P}, \tilde{S})$, and we find that:

$$\tilde{P} = \left[\begin{array}{c:cc} a & b_1 & b_2 \\ \hdashline c_1 & g_1 & 0 \\ c_2 & g_2 & g_3 \end{array}\right], \qquad \tilde{S} = \left\{ \begin{bmatrix} k_1 & 0 \\ k_2 & k_3 \end{bmatrix} \;\middle|\; k_i \in \mathbb{R} \right\}.$$

According to Theorem 25, the reduced system $(\tilde{P}, \tilde{S})$ should be QI. This is clear, as both $\tilde{P}_{22}$ and $\tilde{S}$ are lower-triangular.

Applying Theorem 24, we have that $P_{12}h(S)P_{21}$ is an affine set, and so we may use convex programming to solve an associated control synthesis problem. Note that $h(S)$ is *not* affine in this case, because S being QI with respect to P_{22} is both necessary and sufficient for $h(S)$ to be affine.

7.6.2 Networked System with Delays

Suppose we have two discrete-time systems G_1 and G_2, controlled by K_1 and K_2 respectively. Controller K_1 receives a measurement from G_1, and a one-timestep-delayed measurement from G_2. Similarly, K_2 receives a measurement from G_2, and a one-timestep-delayed measurement from G_1. Further suppose that G_1 and G_2 are coupled, so that G_1 has an additional input that depends on the state of G_2 and vice-versa. The coupling has a delay of one timestep, as shown in Figure 7.2.

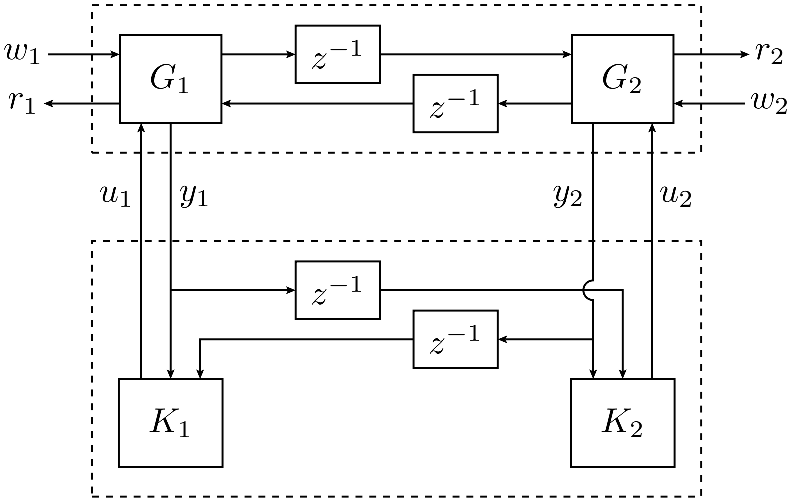

Figure 7.2: Two coupled systems with controllers that receive delayed measurements

Let G_1 be the stable second-order plant with discrete-time state-space equations:

$$\begin{aligned} x_1(t+1) &= \begin{bmatrix} 0.9 & 0.3 \\ -0.6 & 0.8 \end{bmatrix} x_1(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u_1(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v_2(t-1) \\ r_1(t) &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x_1(t) + \begin{bmatrix} 0 \\ \mu \end{bmatrix} u_1(t) \\ \begin{bmatrix} y_1(t) \\ v_1(t) \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ 0.1 & 0.2 \end{bmatrix} x_1(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_1(t), \end{aligned}$$

where r_i are the regulated outputs we wish to keep small, u_i are the inputs provided by the controllers, and v_i is the coupling between the two systems. The equations are the same for G_2, except the subscripts 1 and 2 are interchanged. Taking z-transforms and eliminating the state x, we obtain the plant

$$\begin{bmatrix} r \\ y \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix},$$

where the various transfer functions are:

$$P_{11} = 0, \qquad P_{21} = I, \qquad P_{12} = \begin{bmatrix} P_{22} \\ \mu I \end{bmatrix},$$

$$P_{22} = \frac{6z}{\Delta} \begin{bmatrix} 2z(10z^2 - 17z + 9) & 4z - 3 \\ 4z - 3 & 2z(10z^2 - 17z + 9) \end{bmatrix},$$

and $\Delta = (20z^3 - 34z^2 + 14z + 3)(20z^3 - 34z^2 + 22z - 3)$. The controller has a special structure, because of the delays associated with the measurements:

$$S = \left\{ \begin{bmatrix} K_{11} & z^{-1}K_{12} \\ z^{-1}K_{21} & K_{22} \end{bmatrix} \middle| \; K_{ij} \in \mathcal{R}_p \right\}.$$

It is straightforward to verify that $KP_{22}K \in S$ for all $K \in S$. So we conclude that (P, S) is QI. A more general version of this problem is analyzed in [50].

Note that there are multiple ways of generating a pair (P, S) that represents the

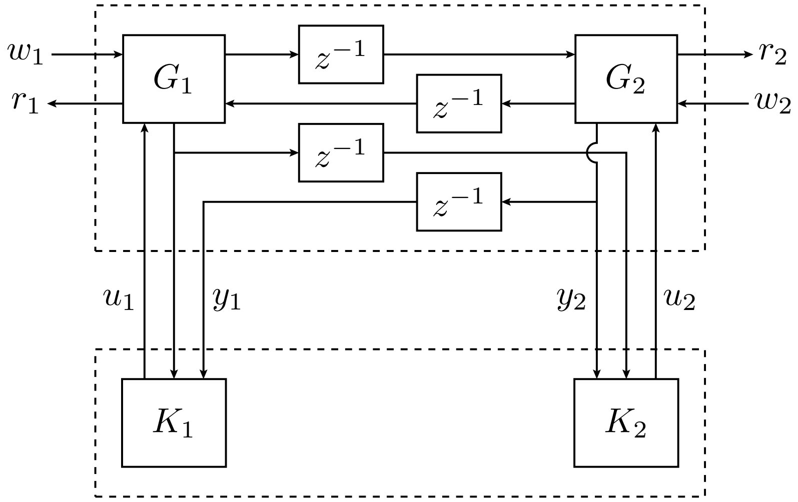

Figure 7.3: Alternate diagram for the system of Figure 7.2

system in Figure 7.2. For example, we could absorb the measurement delays into P, as shown in Figure 7.3, so that $\hat{P}_{11} = P_{11}$, $\hat{P}_{12} = P_{12}$,

$$\hat{P}_{21} = \begin{bmatrix} 1 & 0 \\ z^{-1} & 0 \\ 0 & 1 \\ 0 & z^{-1} \end{bmatrix}, \quad \hat{P}_{22} = \frac{6z}{\Delta} \begin{bmatrix} 2z(10z^2 - 17z + 9) & 4z - 3 \\ 2(10z^2 - 17z + 9) & (4z - 3)/z \\ 4z - 3 & 2z(10z^2 - 17z + 9) \\ (4z - 3)/z & 2(10z^2 - 17z + 9) \end{bmatrix}.$$

The new information constraint is sparse:

$$\hat{S} = \left\{ \begin{bmatrix} K_1 & 0 & 0 & K_2 \\ 0 & K_3 & K_4 & 0 \end{bmatrix} \;\middle|\; K_i \in \mathcal{R}_p \right\}.$$

It is straightforward to verify that $\hat{S}$ is not QI with respect to $\hat{P}_{22}$. However, this problem is IQI. We can transform $(\hat{P}, \hat{S})$ to (P, S) by first computing the skinny and full-normal-rank matrices U_1 and V_1.

Had we used this representation for our problem, we would still be able to transform it into a QI representation. Since $\begin{bmatrix} \hat{P}_{21} & \hat{P}_{22} \end{bmatrix}$ has a normal rank of 2 (not full-normal-rank), our system is reducible. Indeed,

$$\begin{bmatrix} \hat{P}_{21} & \hat{P}_{22} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ z^{-1} & 0 \\ 0 & 1 \\ 0 & z^{-1} \end{bmatrix}}_{U_1} \begin{bmatrix} P_{21} & P_{22} \end{bmatrix}, \qquad \text{and } V_1 = I.$$

We can also verify that $S = \hat{S}U_1$. Therefore, the reduction procedure of Section 7.4 transforms $(\hat{P}, \hat{S})$ to (P, S).

Chapter 8

Future Directions

In this chapter, we discuss three potential directions for future research. Each idea presented here is inspired by a result from Chapters 5–7.

8.1 More General Reduction

In Section 7.4, we show that to test for QI, it is sufficient to test the system representation with the fewest inputs and outputs. We also show how to transform a system into this *minimal* representation by suitably choosing U_1 and V_1. However, we only describe one way of performing reduction; by using the transformation depicted in Figure 7.1. Other ways of finding a representation with a reduced number of inputs and outputs exist, but it is not always advantageous to find such reductions. An avenue for future research would be to investigate the link between model reduction and convexity.

In this section, we will show an example that illustrates that reduction is not always beneficial from the viewpoint of quadratic invariance. The example is inspired by the well-known Kalman decomposition for state-space systems. Namely, if a state-space system is put into Kalman form, it can be reduced:

$$\left[\begin{array}{cccc|c} A_{11} & 0 & A_{13} & 0 & B_1 \\ A_{21} & A_{22} & A_{23} & A_{24} & B_2 \\ 0 & 0 & A_{33} & 0 & 0 \\ 0 & 0 & A_{43} & A_{44} & 0 \\ \hline C_1 & 0 & C_3 & 0 & D \end{array}\right] = \left[\begin{array}{c|c} A_{11} & B_1 \\ \hline C_1 & D \end{array}\right] = D + C_1(sI - A_{11})^{-1}B_1.$$

The state-space notation is actually a special case of an LFT. In Figure 8.1, we show a block diagram that represents a general state-space system. It is identical in structure to the block diagram used for general feedback loops, as shown in Figure 5.1.

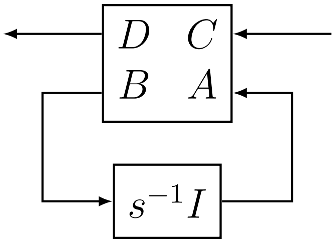

Figure 8.1: Block-diagram representing a state-space system

The reduction afforded by the Kalman decomposition can be generalized to the multidimensional case if we use a diagonal controller. In other words,

$$P = \left[\begin{array}{c:cccc} d & c_1 & 0 & c_3 & 0 \\ \hdashline b_1 & a_{11} & 0 & a_{13} & 0 \\ b_2 & a_{21} & a_{22} & a_{23} & a_{24} \\ 0 & 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{43} & a_{44} \end{array}\right], \qquad K = \begin{bmatrix} k_1 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 \\ 0 & 0 & k_3 & 0 \\ 0 & 0 & 0 & k_4 \end{bmatrix}, \tag{8.1}$$

has the same closed-loop map as:

$$P = \left[\begin{array}{c:c} d & c_1 \\ \hdashline b_1 & a_{11} \end{array}\right], \qquad K = \begin{bmatrix} k_1 \end{bmatrix}.$$

We saw in Section 7.4 that it is always in our best interest to reduce the number of inputs and outputs. However, that observation only holds for reductions described by Figure 7.1. It does not hold with Kalman-type reductions. Consider (8.1), for example. It is neither QI nor IQI, but we know that the second, third, and fourth inputs and outputs can be eliminated without changing the closed-loop map.

If we remove the second and third inputs and outputs, we obtain

$$P = \left[\begin{array}{c:cc} d & c_1 & 0 \\ \hdashline b_1 & a_{11} & 0 \\ 0 & 0 & a_{44} \end{array}\right], \qquad K = \begin{bmatrix} k_1 & 0 \\ 0 & k_4 \end{bmatrix}, \tag{8.2}$$

which is quadratically invariant. If we remove the third and fourth instead,

$$P = \left[\begin{array}{c:cc} d & c_1 & 0 \\ \hdashline b_1 & a_{11} & 0 \\ b_2 & a_{21} & a_{22} \end{array}\right], \qquad K = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}, \tag{8.3}$$

which is not quadratically invariant. Even though (8.2) and (8.3) have the same number of inputs and outputs, only one of the two is QI. So, it is unclear how convexity or quadratic invariance fits in when we consider more general types of reduction.

While the reduction above was clear thanks to the special forms of the matrices, finding a reduction in the general case is much more difficult; it amounts to solving a pair of coupled linear matrix inequalities subject to a rank constraint [33].

8.2 More General Convex Sets

We saw in Corollary 7 that QI is necessary and sufficient for the convexity of $h(S)$. However, there is no known converse result for the convexity of the set of achievable closed-loop maps $P_{11} + P_{12}h(S)P_{21}$. In other words, we do not have a complete characterization of when the set of achievable closed loop maps is convex.

In both the QI and IQI cases, this set turns out to be affine. However, it is possible to have a convex set of achievable closed-loop maps that is not affine. For example, consider the following plant partitioned into its four blocks:

$$P = \left[\begin{array}{cccccccc:c}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & -1 & 0 & -1 \\
\hdashline
0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 0 & -1 & 0 & -1
\end{array}\right],$$

and define the information constraint set S to be the following diagonal set:

$$S = \left\{ \left[\begin{array}{cccccccc}
t & & & & & & & \\
& t & & & & & & \\
& & t & & & & & \\
& & & t & & & & \\
& & & & s & & & \\
& & & & & s & & \\
& & & & & & s & \\
& & & & & & & s
\end{array}\right] \;\middle|\; s, t \in \mathbb{R} \right\}.$$

One can verify that (P, S) is neither QI nor IQI. However, if we compute the set of achievable closed-loop maps, we find:

$$\begin{aligned}
f(P,S) &= \left\{ P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21} \;\middle|\; K \in S \right\} \\
&= \left\{ \left[\begin{array}{c} \frac{2\,t}{(s^2+1)\,(t^2+1)} \\ \frac{t^2-1}{(s^2+1)\,(t^2+1)} \end{array}\right] \;\middle|\; s, t \in \mathbb{R} \right\}.
\end{aligned}$$

The closure of $f(P,S)$ is the unit disk in $\mathbb{R}^2$. That is,

$$\operatorname{clos} f(P,S) = \left\{ x \in \mathbb{R}^2 \;\middle|\; x_1^2 + x_2^2 \le 1 \right\}.$$

This set is clearly convex, so we have a guarantee that the solution to the original problem can be reduced to a convex optimization problem. We can solve for x_1 and x_2 first, and then change coordinates to find the corresponding s and t.

So while QI and IQI provide nice sufficient conditions under which the set of achievable closed-loop maps is affine, it can also happen that the set of achievable closed-loop maps is a more general convex set. Some research has been done in addressing the more general case. For example, Shin et. al. [52] show how to compute the set of achievable closed-loop maps using elimination theory. The solution is expressed as the projection of a semialgebraic set.

While this is the only general characterization currently known for the set of achievable closed-loop maps, there are still some drawbacks. First, computing the solutions involves finding a Groebner basis. There are systematic algorithms for doing so which are guaranteed to converge after a finite number of steps, but this number may be very large. Second, once we have found a representation for the set of achievable closed-loop maps, determining convexity remains an open question. Unless the equations happen to be linear or some other simple form such as the unit disc in the example above, there is no known efficient way to verify convexity.

8.3 Stabilization

Suppose that in addition to solving (7.7), we would like K to be a *stabilizing* controller. That is, we would like the closed-loop interconnection to be internally stable. In the QI case, the stabilizing controllers were parametrized when P is stable [49]. More recently, a structured coprime factorization was found that yields a parametrization of stabilizing controllers even when P is unstable [51]. In both cases, the best stabilizing K is found by solving a related optimization problem where the variable to be optimized is a Youla-type parameter Q which is constrained to be stable. Unfortunately, there is no simple answer to the stabilization question for IQI problems. We will now demonstrate this fact with a counterexample.

Define $\mathcal{C}$ to be the set of stable matrices. If $P \in \mathcal{C}$, then K is stabilizing if and only if $h(K) \in \mathcal{C}$. Therefore, the set of achievable stabilized closed-loop maps is

$$f(S) = P_{11} - P_{12}\left(h(S) \cap \mathcal{C}\right) P_{21}.$$

If S is QI with respect to P_{22}, then $h(S) = S$ and the set becomes

$$f_{\mathrm{QI}}(S) = P_{11} - P_{12}\left(S \cap \mathcal{C}\right) P_{21}.$$

This set is affine, and thus amenable to a convex optimization approach. Unfortunately, we cannot make the same simplification in the IQI case. Even though $P_{12}h(S)P_{21} = P_{12}SP_{21}$,

$$\begin{aligned} f_{\mathrm{IQI}}(S) &= P_{11} - P_{12}\left(h(S) \cap \mathcal{C}\right) P_{21} \\ &\neq P_{11} - P_{12}\left(S \cap \mathcal{C}\right) P_{21}. \end{aligned}$$

Indeed, it turns out that $P_{12}\left(h(S) \cap \mathcal{C}\right) P_{21}$ may not even be an $\mathcal{R}_p$-module. To see why, consider the system defined by the matrices

$$P_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \quad P_{21} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad P_{22} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix},$$

and define the set of admissible controllers

$$S = \left\{ \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix} \;\middle|\; k_i \in \mathcal{R}_p \right\}.$$

Note that S is IQI with respect to P, because if we define

$$W_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad W_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

the three following properties hold:

i) $W_i^2 = W_i$, so W_i are projectors,

ii) $\begin{bmatrix} I & 0 \\ 0 & W_1 \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & W_2 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$,

iii) $W_2 S W_1$ is QI with respect to P_{22}.

Therefore, we have from Theorem 24 that

$$P_{12} h(S) P_{21} = P_{12} S P_{21}.$$

Since P_{22} is stable, a stabilizing K corresponds to a stable Q. If compute Q for this particular problem, we find:

$$\begin{aligned} Q &= - \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} k_1 & 0 & 0 \\ 0 & k_2 & 0 \\ 0 & 0 & k_3 \end{bmatrix} \right)^{-1} \\ &= - \begin{bmatrix} k_1(1-k_1)^{-1} & 0 & 0 \\ k_1 k_2 (1-k_1)^{-1} & k_2 & 0 \\ (k_2+1) k_1 k_3 (1-k_1)^{-1} (1-k_3)^{-1} & k_2 k_3 (1-k_3)^{-1} & k_3 (1-k_3)^{-1} \end{bmatrix} \\ &= - \begin{bmatrix} q_1 & 0 & 0 \\ q_1 q_2 & q_2 & 0 \\ q_1 q_3 (q_2+1) & q_2 q_3 & q_3 \end{bmatrix}, \end{aligned}$$

where we have defined $q_1 = k_1(1-k_1)^{-1}$, $q_2 = k_2$, and $q_3 = k_3(1-k_3)^{-1}$. Note that every element of this matrix is stable if and only if the q_i are stable. Therefore,

$$\begin{aligned}
&\{-P_{12}h(K)P_{21} \mid K \text{ stabilizing}\} \\
&\quad = \{-P_{12}QP_{21} \mid Q \in \mathcal{C}\} \\
&\quad = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} q_1 & 0 & 0 \\ q_1q_2 & q_2 & 0 \\ q_1q_3(q_2+1) & q_2q_3 & q_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \;\middle|\; q_i \in \mathcal{C} \right\} \\
&\quad = \left\{ \begin{bmatrix} q_1 & 0 \\ q_1q_3 + q_2(1+q_1)(1+q_3) & q_3 \end{bmatrix} \;\middle|\; q_i \in \mathcal{C} \right\}.
\end{aligned}$$

Note that the matrix $X = \begin{bmatrix} \frac{-2}{s+1} & 0 \\ 0 & \frac{-2}{s+1} \end{bmatrix}$ does not belong to this set. Indeed, it would constrain $q_1 = q_3 = \frac{-2}{s+1}$, and would force $q_2 = \frac{-q_1q_3}{(1+q_1)(1+q_3)} = \frac{-4}{(s-1)^2}$, which is not stable. However, it is clear that both $X_1 = \begin{bmatrix} \frac{-2}{s+1} & 0 \\ 0 & 0 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 0 & 0 \\ 0 & \frac{-2}{s+1} \end{bmatrix}$ belong to the set ($q_1 = \frac{-2}{s+1}, q_2 = q_3 = 0$) and ($q_1 = q_2 = 0, q_3 = \frac{-2}{s+1}$), respectively. So we conclude that the set $P_{12}(h(S) \cap \mathcal{C})P_{21}$, and therefore the set of achievable stabilized closed-loop maps, does ***not*** form an $\mathcal{R}_p$-module.

Now suppose that $P_{11} = -X = \begin{bmatrix} \frac{2}{s+1} & 0 \\ 0 & \frac{2}{s+1} \end{bmatrix}$. If we naively attempt to solve the modified problem (7.9), we have:

$$\begin{aligned}
&\text{minimize} && \left\| \frac{2}{s+1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \right\| \\
&\text{subject to} && q_i \in \mathcal{C}.
\end{aligned} \tag{8.4}$$

This simplifies to:

$$\begin{array}{ll} \text{minimize} & \left\| \dfrac{2}{s+1} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} q_1 & 0 \\ q_2 & q_3 \end{bmatrix} \right\| \\ \text{subject to} & \hat{q}_i \in \mathcal{C}. \end{array} \tag{8.5}$$

We can achieve a norm of zero, by setting $q_1 = q_3 = \frac{-2}{s+1}$ and $q_2 = 0$. This solution is unique, and clearly optimal. However, we already showed that $X \notin P_{12}(h(S) \cap \mathcal{C})P_{21}$. So achieving a norm of zero in (7.7) is impossible, even though we did it in (7.9).

The problem is that in modifying (7.9) by adding the stability constraint to Q, we implicitly assumed that because we had $P_{12}h(S)P_{21} = P_{12}SP_{21}$, we would also have

$$P_{12}\left(h(S) \cap \mathcal{C}\right)P_{21} = P_{12}\left(S \cap \mathcal{C}\right)P_{21},$$

which is false. In general, the equals sign should be replaced by $\subseteq$. Thus, if we are seeking a stabilizing solution, the method of Theorem 26 can modified and used to find a K_{opt}, but in general it will only be a *lower bound* on the optimal cost. There is no guarantee that the K_{opt} obtained will be feasible.

Chapter 9

Conclusions

In this work, we investigated two open research questions. In Chapters 2–4, we developed an efficient and scalable algorithm to perform wavefront reconstruction on adaptive optics hardware. Using numerical simulations, we found that our algorithm performed comparably to other proposed methods, but with a significant computational speedup. Finally, we tested our method at the 3.1 m main telescope at Palomar Observatory. We found that our algorithm produced indistinguishable images from those found using the conventional optimal reconstructor, but at a fraction of the cost. The computational speedup will be even greater for future large adaptive optics systems, since our algorithm's computational complexity is $\mathcal{O}(n)$ rather than the conventional $\mathcal{O}(n^2)$.

In Chapters 5–7, we characterized a new class of tractable decentralized problems, which we called *internally quadratically invariant*. We prove that such problems have a convex set of achievable closed-loop maps, making them amenable to a convex optimization approach. This is the largest class of tractable decentralized control problems known to date, and represents a significant step toward a complete characterization of convex decentralized control problems.

Taking a step back, both parts of this thesis fit into the broad category of complex systems. However, the nature of each part is very different. In adaptive optics, the problem is solvable, but computationally intensive. Our approach was to find an approximate solution method that was fast and scalable. We found that with a very slight approximation that went unnoticed on a real telescope, we achieved a very large computational speedup.

In decentralized control, some problems are known to be hopelessly complicated. The limitation is not with our computers, but rather in the nature of the problem. Rather than finding ways to cut corners and save on computational effort, our goal became to discover which types of decentralized problems have any hope of being solved at all. We found a simple algebraic condition that ensures the convexity of the set of achievable closed loop maps, thus paving the way for efficient numerical approaches.

While this thesis provides significant contributions in two areas related to complex systems, much work remains to be done. Future large-scale systems will require further innovation on both fronts: designing network architectures that are provably tractable, and ensuring that the numerical methods used therein are efficient and scalable.

Adaptive Optics References

[1] M. C. Britton. Arroyo C++ library: object oriented class libraries for the simulation of electromagnetic wave propagation through turbulence. http://eraserhead.caltech.edu/arroyo/arroyo.html.

[2] M. C. Britton. Arroyo. *Proc. SPIE*, 5497:290–300, September 2004.

[3] R. Dekany, A. Bouchez, M. Britton, V. Velur, M. Troy, J. C. Shelton, and J. Roberts. PALM-3000: visible light AO on the 5.1-meter Telescope. *Proc. SPIE*, 6272:62720G, July 2006.

[4] B. L. Ellerbroek. Efficient computation of minimum-variance wave-front reconstructors with sparse matrix techniques. *J. Opt. Soc. Am. A*, 19(9):1803–1816, September 2002.

[5] B. L. Ellerbroek and F. J. Rigaut. Scaling multiconjugate adaptive optics performance estimates to extremely large telescopes. *Proc. SPIE*, 4007:1088–1099, July 2000.

[6] B. L. Ellerbroek and C. R. Vogel. Simulations of closed-loop wavefront reconstruction for multiconjugate adaptive optics on giant telescopes. *Proc. SPIE*, 5169:206–217, December 2003.

[7] D. T. Gavel and D. Wiberg. Toward Strehl-optimizing adaptive optics controllers. *Proc. SPIE*, 4839:890–901, February 2003.

[8] L. Gilles. Order-N sparse minimum-variance open-loop reconstructor for extreme adaptive optics. *Opt. Lett.*, 28(20):1927–1929, October 2003.

[9] L. Gilles. Closed-loop stability and performance analysis of least-squares and minimum-variance control algorithms for multiconjugate adaptive optics. *Appl. Opt.*, 44(6):993–1002, February 2005.

[10] L. Gilles, B. L. Ellerbroek, and C. Vogel. A comparison of multigrid V-Cycle versus Fourier domain preconditioning for laser guide star atmospheric tomography. In *Adaptive Optics: Analysis and Methods/Computational Optical Sensing and Imaging/Information Photonics/Signal Recovery and Synthesis Topical Meetings on CD-ROM*. Optical Society of America, 2007. paper JTuA1.

[11] L. Gilles, C. R. Vogel, and B. L. Ellerbroek. Multigrid preconditioned conjugate-gradient method for large-scale wave-front reconstruction. *J. Opt. Soc. Am. A*, 19(9):1817–1822, September 2002.

[12] J. W. Hardy. *Adaptive Optics for Astronomical Telescopes*. Oxford University Press US, 1998.

[13] T. L. Hayward, B. Brandl, B. Pirger, C. Blacken, G. E. Gull, J. Schoenwald, and J. R. Houck. PHARO: a near-infrared camera for the Palomar adaptive optics system. *Publications of the Astronomical Society of the Pacific*, 113:105–118, January 2001.

[14] V. Kornilov, A. Tokovinin, N. Shatsky, O. Voziakova, S. Potanin, and B. Safonov. Combined MASS-DIMM instruments for atmospheric turbulence studies. *Mon. Not. R. Astron. Soc.*, 382:1268–1278, December 2007.

[15] B. Le Roux, J.-M. Conan, C. Kulcsár, H.-F. Raynaud, L. M. Mugnier, and T. Fusco. Optimal control law for classical and multiconjugate adaptive optics. *J. Opt. Soc. Am. A*, 21(7):1261–1276, July 2004.

[16] L. Lessard, D. MacMynowski, M. West, A. Bouchez, and S. Lall. Experimental validation of single-iteration multigrid wavefront reconstruction at the Palomar Observatory. *Opt. Lett.*, 33(18):2047–2049, 2008.

[17] L. Lessard, M. West, D. MacMynowski, and S. Lall. Warm-started wavefront reconstruction for adaptive optics. *J. Opt. Soc. Am. A*, 25(5):1147–1155, 2008.

[18] D. G. MacMartin. Local, hierarchic, and iterative reconstructors for adaptive optics. *J. Opt. Soc. Am. A*, 20(6):1084–1093, June 2003.

[19] R. N. Paschall and D. J. Anderson. Linear quadratic Gaussian control of a deformable mirror adaptive optics system with time-delayed measurements. *Appl. Opt.*, 32(31):6347–6358, November 1993.

[20] P. Piatrou and M. Roggemann. Performance analysis of Kalman filter and minimum variance controllers for multi conjugate adaptive optics. *Proc. SPIE*, 5894:288–296, August 2005.

[21] L. A. Poyneer, D. T. Gavel, and J. M. Brase. Fast wave-front reconstruction in large adaptive optics systems with use of the Fourier transform. *J. Opt. Soc. Am. A*, 19(10):2100–2111, October 2002.

[22] L. A. Poyneer, B. A. Macintosh, and J.-P. Véran. Fourier transform wavefront control with adaptive prediction of the atmosphere. *J. Opt. Soc. Am. A*, 24(9):2645–2660, September 2007.

[23] L. A. Poyneer, M. Troy, B. Macintosh, and D.T. Gavel. Experimental validation of Fourier-transform wave-front reconstruction at the Palomar Observatory. *Opt. Lett.*, 28(10):798–800, 2003.

[24] H. Ren, R. Dekany, and M. Britton. Large-scale wave-front reconstruction for adaptive optics systems by use of a recursive filtering algorithm. *Appl. Opt.*, 44(13):2626–2637, May 2005.

[25] F. Shi, D. G.MacMartin, M. Troy, G. L. Brack, R. S. Burruss, and R. G. Dekany. Sparse-matrix wavefront reconstruction: simulations and experiments. *Proc. SPIE*, 4839:1035–1044, 2003.

[26] V. I. Tatarskii. *Wave Propagation in a Turbulent Medium.* McGraw-Hill New York, 1961.

[27] U. Trottenberg, A. Schüller, and C. W. Oosterlee. *Multigrid Methods.* Academic Press, 2000.

[28] M. Troy, R. G. Dekany, G. Brack, B. R. Oppenheimer, E. E. Bloemhof, T. Trinh, F. G. Dekens, F. Shi, T. L. Hayward, and B. Brandl. Palomar adaptive optics project: status and performance. *Proc. SPIE*, 4007:31–40, July 2000.

[29] C. R. Vogel and Q. Yang. Multigrid algorithm for least-squares wavefront reconstruction. *Appl. Opt.*, 45(4):705–715, February 2006.

[30] E. P. Wallner. Optimal wave-front correction using slope measurements. *J. Opt. Soc. Am.*, 73(12):1771–1776, December 1983.

[31] Q. Yang, C. R. Vogel, and B. L. Ellerbroek. Fourier domain preconditioned conjugate gradient algorithm for atmospheric tomography. *Appl. Opt.*, 45(21):5281–5293, July 2006.

Convexity References

[32] R. Bansal and T. Basar. Stochastic teams with nonclassical information revisited: When is an affine law optimal? *IEEE Transactions on Automatic Control*, 32(6):554–559, 1987.

[33] C. L. Beck, J. Doyle, and K. Glover. Model-reduction of multidimensional and uncertain systems. *IEEE Transactions on Automatic Control*, 41(10):1466–1477, 1996.

[34] V. D. Blondel and J. N. Tsitsiklis. A survey of computational complexity results in systems and control. *Automatica*, 36(9):1249–1274, 2000.

[35] R. D'Andrea and G. E. Dullerud. Distributed control design for spatially interconnected systems. *IEEE Transactions on Automatic Control*, 48(9):1478–1495, 2003.

[36] E. Fornasini and G. Marchesini. Doubly-indexed dynamical systems: State-space models and structural properties. *Theory of Computing Systems*, 12(1):59–72, 1978.

[37] Y-C. Ho and K-C. Chu. Team decision theory and information structures in optimal control problems—Part I. *IEEE Transactions on Automatic Control*, 17(1):15–22, 1972.

[38] T. Kailath. *Linear systems.* Prentice-Hall Englewood Cliffs, NJ, 1980.

[39] R. E. Kalman. Algebraic structure of linear dynamical systems, i. the module of σ. *Proceedings of the National Academy of Sciences of the United States of America*, 54(6):1503, 1965.

[40] L. Lessard and S. Lall. Reduction of decentralized control problems to tractable representations. In *IEEE Conference on Decision and Control*, pages 1621–1626, 2009.

[41] L. Lessard and S. Lall. An algebraic framework for quadratic invariance. In *IEEE Conference on Decision and Control*, pages 2698–2703, 2010.

[42] L. Lessard and S. Lall. Internal quadratic invariance and decentralized control. In *American Control Conference*, pages 5596–5601, 2010.

[43] L. Lessard and S. Lall. Quadratic invariance is necessary and sufficient for convexity. In *American Control Conference*, pages 5360–5362, 2011.

[44] H. Matsumura and M. Reid. *Commutative ring theory.* Cambridge University Press, 1989.

[45] R. Radner. Team decision problems. *The Annals of Mathematical Statistics*, 33(3):857–881, 1962.

[46] R. Roesser. A discrete state-space model for linear image processing. *IEEE Transactions on Automatic Control*, 20(1):1–10, 1975.

[47] M. Rotkowitz. On information structures, convexity, and linear optimality. In *IEEE Conference on Decision and Control*, pages 1642–1647, 2008.

[48] M. Rotkowitz, R. Cogill, and S. Lall. Convexity of optimal control over networks with delays and arbitrary topology. *International Journal of Systems, Control and Communication*, 2(1):30–54, 2010.

[49] M. Rotkowitz and S. Lall. Decentralized control information structures preserved under feedback. In *IEEE Conference on Decision and Control*, pages 569–575, 2002.

[50] M. Rotkowitz and S. Lall. A characterization of convex problems in decentralized control. *IEEE Transactions on Automatic Control*, 51(2):274–286, 2006.

[51] S. Sabău and N.C. Martins. On the stabilization of LTI decentralized configurations under quadratically invariant sparsity constraints. In *Allerton Conference on Communication, Control, and Computing*, pages 1004–1010. IEEE, 2010.

[52] H. S. Shin and S. Lall. Optimal decentralized control of linear systems via Groebner bases and variable elimination. In *American Control Conference*, pages 5608–5613, 2010.

[53] M. Vidyasagar, H. Schneider, and B. Francis. Algebraic and topological aspects of feedback stabilization. *IEEE Transactions on Automatic Control*, 27(4):880–894, 1982.

[54] H. S. Witsenhausen. A counterexample in stochastic optimum control. *SIAM Journal on Control*, 6:131, 1968.