Optimization and control of large-scale networked systems

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Neil K. Dhingra

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

December 2017

To Pallavi Kathuria, in loving memory.

Acknowledgements

No man is an island and in completing my graduate studies I have relied on the help and guidance of innumerable friends, relations, and mentors. I will try to give a partial accounting of their aid and support here.

I would first like to thank my advisor, Professor Mihailo R. Jovanović for being an amazing teacher, mentor, and friend. His relentless perfectionism and his insistence on detail has enhanced the quality and sophistication of my writing, research, and presentation. I am forever indebted to him for his guidance and mentorship.

Many other faculty members have contributed greatly to my time here. I learned much about convex optimization from Professor Tom Luo, both in his excellent classes and one-on-one meetings. I also benefitted tremendously from interactions with Tryphon Georgiou, Murti Salapaka, Peter J. Seiler, Andy Lamperski, Makan Fardad, and Anders Rantzer.

I am thankful for NASA and their amazing research scientists. Interacting with Marty Brenner was always stimulating; working with him broadened my perspectives and gave me insight into the practical applications of control theory.

I would also like to thank the professors who taught and mentored me at the University of Michigan and all my classmates and colleagues who inspired me to pursue graduate school. In particular, I would like to thank Sister Mary Elizabeth Merriam, under whose guidance I began my journey into academia.

The countless hours spent in our stuffy Keller Hall office were made bearable only by the incredible company I kept. I would like to thank Xiaofan Wu, Armin Zare, and Yongxin Chen, my compatriots in this endeavor. Whether watching the Timberwolves or running races, it has always been a pleasure. The senior statesmen of the lab – Rashad Moarref, Binh K. Lieu and Fu Lin – all offered invaluable guidance and counsel. I would like to thank "Mr. Fu" in particular for his help with optimization and sparsity-promoting optimal control. Much of the work in this thesis builds upon the foundation that he laid. Finally I would like to thank the junior graduate

students, Sepideh Hassan-Moghaddam, Wei Ran, and Dongsheng Ding – I am happy that I can leave knowing that the future of the lab is in good hands.

Beyond our lab group, interactions with Sei Zhen Khong at the IMA and Marcello Colombino at ETH were enjoyable socially and brought my research to a higher level. I also enjoyed my time with the other IMA postdocs, particularly Kaoru Yamamoto and Rohit Gupta.

Of course I must thank my incredible network of friends in Minneapolis and across the country. Katie Hartl, in particular, has always been a great friend with whom I could commiserate. Thank you to Kevin Nowland, for his friendship and mathematical advice and despite his dubious collegiate affiliation, and to Rohan Agarwal, for giving me a second home in Chicago. I am grateful for all my friends in Minneapolis, including my roommate Aarón as well as everyone I met through him, Brian and Julia and their wonderful children, Julius and Maria, Luke and Lindsey, Caitlin and Eric, Jane, Liz, Niko, Alex, Sarah, Stephen, Ben, Emily, Angela, Kyle, and the rest of the Scudder house crew, Taylor, Jordan, and everyone else at the Michigan Alumni Association, Megan, Sandy, Elizabeth, Azra, and many, many more. I would also like to thank Sisyphus, in whom I found a kindred spirit, and Albert Camus for providing an optimistic interpretation of this somewhat concerning correlation.

Thank you to my family for constantly providing love and support. Thanks especially to my parents, who have always been there for me, even when my car was stolen or broke down. Thank you to my sister Nina, for her support as well as for helping me start and accompanying me on different roadtrips from California to Minnesota. I am also indebted to my extended family, grandparents, cousins, aunts, and uncles for their support and love over the years. In particular, I would like to thank my little cousin – Pallavi, your wonderful smile and joyous laugh will stay with me forever.

Finally and most importantly I would like to thank Amy. Her constant support and love has given me strength and fortitude throughout my graduate studies. I loved our time in Minneapolis and our biking, running, and roadtrip adventures to all the breweries, lakes, waterfalls, ice caves, and stadia in the area. Thank you, Amy, I couldn't have done it without you.

Abstract

In this thesis, we design structured controllers for linear systems by solving regularized optimal control problems. We develop tractable methods for solving nonconvex regularized problems and then identify classes of problems for which regularized optimal control problems can be placed into a convex form.

We first develop novel methods based on reformulating the regularized optimization problem with an auxiliary variable. By exploiting the properties of proximal operators, we bring the associated augmented Lagrangian into a continuously differentiable form by constraining it to the manifold that corresponds to explicit minimization over the auxiliary variable. The new expression facilitates a method of multipliers algorithm that offers many advantages relative to existing methods, including guaranteed convergence for nonconvex problems and the ability to impose regularization in alternate coordinates. We then apply primal-descent dual-ascent Arrow-Hurwicz-Uzawa type gradient flow dynamics to solve regularized problems in a distributed manner. We prove global convergence for convex problems and use the theory of Integral Quadratic Constraints to establish conditions for exponential convergence for continuous- and discrete-time updates applied to strongly convex probems. Finally, we take advantage of generalizations of the Jacobian to develop a second-order algorithm which converges globally to the optimal solution for convex problems. Moreover, we prove local quadratic convergence for strongly convex problems.

We next study several classes of convex regularized optimal control problems. The problem of designing symmetric modifications to symmetric linear systems is convex in the underlying design variable and is thus appealing for the purpose of structured control. We show that even when the system and controller are not symmetric, their symmetric components can be used to perform structured design with stability and performance guarantees. We then examine the problem of designing structured diagonal modifications to positive systems. We prove convexity of the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control problems for this class of system and apply our results to leader selection in directed consensus networks and combination drug therapy design for HIV. We consider time-varying controllers and show that a constant controller is optimal for an induced-power performance index. Finally, we develop customized algorithms for large-scale actuator and sensor selection.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 A historical perspective

Feedback control theory is the science of designing a controller that connects a system's output to its input in order to achieve some desired behavior. A simple example is a thermostat: this controller achieves a target temperature by adjusting current through heating coils or strength of air conditioning to counteract unwanted deviations from a desired temperature. The concept of feedback control arguably traces back to irrigation canals in ancient Mesopotamia built to facilitate farming [1]. Modern mathematical study of control systems began in the 19th century with the study of centrifugal governors used to keep steam engines running at a constant speed [1]. During the Second World War, applications of control theory for weapons guidance systems drove the development of graphical single-input single-output control design techniques based on loop-shaping, Bode plots, and Nyquist diagrams [2]. After the war, the challenge of controlling rockets during the space race prompted the study of multi-input multi-output systems and eventually led to the formalism of optimal control [3].

The optimal control of linear systems with quadratic performance measures, such as the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms, is a cornerstone of systems theory. This framework provides a systematic way to balance closed-loop performance, robustness, and control effort. In the conventional formulation, an optimal controller is designed to minimize some measure of the amplification from exogenous sources of excitation to a regulated output which penalizes both the system state and the control effort. Typically, these optimal controllers do not have a particular structure and require a centralized implementation in which measurement and control must involve all

outputs and inputs and must be performed at a single, central location.

While they provide a powerful framework, most traditional optimal control tools are ill-suited to problems where controller structure is of tantamount importance. In many modern applications, constraints on communication, computation, and physics impose significant structural restrictions on the controller. Designing controllers subject to such restrictions is a challenging problem. In fact, even determining stabilizability was shown to be NP hard in general [4, 5].

Nevertheless, designing structured controllers is of growing importance for many modern problems, such as the distributed control of large-scale networks of dynamical systems. Systems of this type are found in applications ranging from distributed power generation, to deployment of teams of robotic agents, to control of segmented mirrors in extremely large telescopes, to control of fluid flows around wind turbines and vehicles. The development of analytical and computational methods for tractable analysis and design of such networks is a major challenge.

Recent technological advances have allowed the individual components of large-scale systems to be equipped with their own sensing, actuation, communication, computation, and decision making capabilities. Advances in Micro-Electro-Mechanical-Systems (MEMS) have enabled the development of arrays of sensors and actuators that can interact with one another. Strings of vehicular platoons, unmanned aerial vehicles (UAVs), and robotic agents constitute another set of examples of large-scale autonomous systems [6, 7].

In many of these applications, the scale of the problem, constraints on computing and communication resources, and wide-spread sensing and actuating capabilities pose additional requirements on controller complexity. Typically, these cannot be addressed using tools from standard optimal control theory. For example, a dense state-feedback controller resulting from the LQR framework would impose a prohibitive communication burden in large-scale networks. This is because forming every control input requires information from every subsystem in the distributed plant. The cost of creating and maintaining communication links makes such an all-to-all topology infeasible in most large-scale and distributed systems.

This has motivated the design of structured (both decentralized and distributed) controllers. Early efforts have centered on the design of decentralized strategies [8] and, during the last fifteen years, the emphasis has shifted to the design of distributed controllers [9–37]. Incorporating structure into systems theory has also enabled the analysis [38–41], control [42–45], and low complexity modeling [46, 47] of large-scale fluid flows. Two major issues have emerged: the identification of controller architectures of convex classes of structured control problems and optimal control design under *a priori* specified structural constraints.

In some cases, these questions can be resolved using standard tools under additional assumptions. Optimal control problems are often reformulated using the Youla parameterization [48,49]. The mapping from the controller to the Youla parameter is nonlinear which typically compromises convexity of the structural constraints in the distributed setup. It is thus important to identify subspaces which remain invariant under this nonlinear mapping for distributed systems. In [11, 15], the subspaces of *cone* and *funnel causal* systems have been introduced; these describe how information from every controller propagates through the distributed system. For spatially-invariant systems, the design of quadratically optimal controllers can be cast into a convex problem if the information in the controller propagates at least as fast as in the plant [11,15]. A similar but more general algebraic characterization of the constraint set was introduced and convexity was established under the condition of quadratic invariance in [16]. Other classes of convex distributed control problems include partially-nested systems [50–52], poset-causal systems [19, 25], and positive systems [53–59].

Since most of these convex formulations are expressed in terms of the impulse response parameters, they do not lend themselves easily to state-space characterization. Apart from very special instances, the optimal distributed design problem remains challenging. For poset-causal systems, explicit Riccati-based solutions for the optimal decentralized state-feedback problem were obtained in [19, 25]. For a two-player problem with block-triangular state-space matrices, the optimal decentralized output-feedback solution was recently provided in [23, 29]. Characterizing the structural properties of optimal distributed controllers is another important challenge. For spatially invariant systems, the quadratic optimal controllers are also spatially invariant and the information from other subsystems is exponentially discounted with the distance between the controller and the subsystems [9]. For systems on graphs, this spatially decaying property was studied in [60, 61] and it motivates the search for inherently localized controllers.

Optimal control under structural constraints remains a challenging problem [32]. Since these constraints are often combinatorial in nature (e.g., selecting k communication links between N nodes), structured optimal control has two main aspects: the identification of effective structures for the purpose of control (e.g., the communication topology between distributed controllers) and the design of controllers given a specific structure (e.g., the control law implemented over a given communication network).

1.2 Optimal control

The notion of controller structure can have different connotations and the design variable can impact the dynamics in different ways. We first define the dynamics we study and the performance metrics we use to quantify closed-loop performance. We consider the general dynamics,

$$\begin{aligned} \dot{\psi} &= (A + F(x))\,\psi + Bd \\ \zeta &= \begin{bmatrix} C \\ R(x) \end{bmatrix} \psi \end{aligned} \tag{1.1}$$

where $x \in \mathbb{R}^m$ is a design parameter, F: $\mathbb{R}^m \to \mathbb{R}^{n\times n}$ is a linear operator, $\psi(t) \in \mathbb{R}^n$ is the state vector, $d(t)$ represents an exogeneous input, and ζ is a regulated output. The matrix C represents a mapping from ψ to a state penalty and $R(x)$ is a mapping from the state to a measure of control effort. In Section 1.2.1, we illustrate how this formulation generalizes standard state feedback and encapsulates state feedback, output feedback, edge addition in directed and undirected consensus networks, leader selection, and other problems.

In the optimal control problems we consider, the objective is to improve the open-loop (i.e., no controller, or $x = 0$) performance of a system by implementing a static controller x. Solving optimal control problems amounts to designing controllers x that minimize a performance metric that quantifies the performance of the closed-loop transfer function from the exogenous input d to the regulated output ζ. This thesis mainly considers the $\mathcal{H}_2$- and $\mathcal{H}_\infty$-optimal control frameworks, where closed-loop performance is measured by variance amplification and the peak of the frequency response, respectively.

We first provide examples of problems that are included in formulation (1.1), provide the expressions for the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms, and then motivate the study of structured optimal control with a distributed systems example.

1.2.1 Applications

Static feedback

The standard static state feedback formulation,

$$\begin{aligned} \dot{\psi} &= A\psi + B_1 d + B_2 u \\ u &= -X\psi \\ \zeta &= \begin{bmatrix} C\psi \\ R^{1/2}u \end{bmatrix} \end{aligned} \tag{1.2}$$

with state-feedback matrix X is recovered from (1.1) by taking $x = \mathrm{vec}(X)$, $F(x) = -B_2 X$ and $R(x) = -R^{1/2}X$ for a positive semidefinite R. When the controller can only measure some output $\zeta_2 = C_2\psi$, taking $F(x) = -B_2 X C_2$ and $R(x) = -R^{1/2}XC_2$ yields the static output feedback formulation.

We describe and motivate this problem in the context of distributed systems in Section 1.3. In Section 2.5, we develop an algorithm for designing structured controllers for (1.2) following the formulation introduced by [62–64].

Edge addition in stochastically forced consensus networks

The problem of adding undirected 'controller' edges to an existing 'plant' consensus network also involves dynamics of the form (1.1),

$$\dot{\psi} = -\left(L + E\,\mathrm{diag}(x)E^T\right)\,\psi + d$$

where each element of the state vector, ψ_i, represents a node in the network, L is a directed graph Laplacian which contains information about how the nodes are connected in the 'plant' network, E contains information about the locations of potential added edges, x is a vector of added edges and $F(x) = -E\,\mathrm{diag}(x)E^T$ is the corresponding weighted graph Laplacian associated with the 'controller' network [65].

Consensus networks have garnered much interest for problems dealing with collective decision-making and collective sensing [66–69]. These networks have proven useful in applications as varied as modeling animal group dynamics, control for spacecraft flying in formation, and data fusion in sensor networks [70–72].

Adding edges to a consensus network improves its performance, but selecting a limited

number of edges to add is a combinatorial problem. Much recent work has focused on addressing the optimal edge addition problem using convex optimization techniques [73–75] and developing efficient algorithms that scale to very large networks [76]. However, most of these approaches consider undirected networks. We extend this problem and study edge addition in *directed* consensus networks in Sections 2.4 and 6.2.1.

Leader selection in consensus networks

This problem concerns the identification of influential nodes in consensus networks. These special nodes, so-called leaders, can be equipped with additional information in order to influence the network behavior in a beneficial way. One application is in vehicular formations, where the objective is for the vehicles to gather at a certain point. Here, the 'leaders' are equipped absolute measurements (e.g., from GPS units) and the other nodes must rely on relative measurements (e.g., their distance from certain neighbors). The dynamics of the problem are,

$$\dot{\psi} \;=\; -\left(L \;+\; \mathrm{diag}(x)\right)\,\psi \;+\; d$$

where L is a graph Laplacian and x is a nonnegative vector whose nonzero entries identify the leaders.

The question of how to optimally assign a predetermined number of nodes to act as leaders in a network of dynamical systems with a given topology has also recently emerged as a useful proxy for identifying important nodes in a network [77–84]. In [80], the authors develop a greedy algorithm for leader selection in undirected networks and use convex relaxations to quantify performance bounds. In [78, 79], the authors derive an explicit expression for the set of optimal leaders in terms of graph theoretic properties which is computationally tractable for very few for very many leaders. In [85], the authors characterize bounds on the convergence rate based on the distance between leaders and followers. Most of these results focus on undirected networks with the exception of [86] in which the authors derive the optimal leaders for one-dimensional directed path networks.

In contrast to earlier work, our framework can handle leader selection in *directed* consensus networks [87, 88]. We study this problem in detail in Section 7.3.

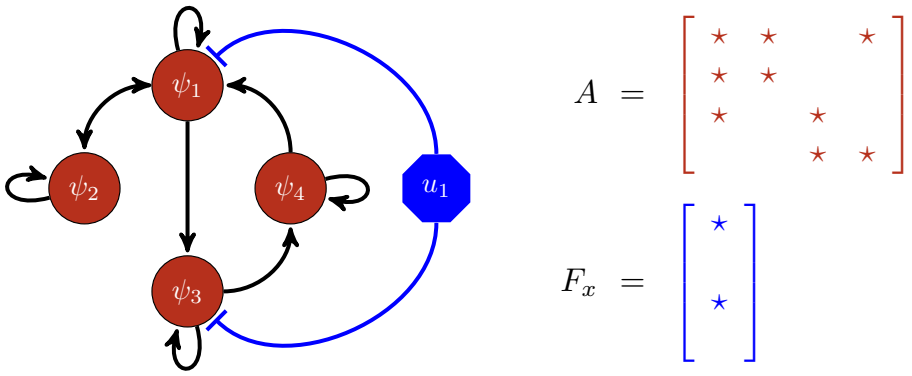


Figure 1.1: HIV replication-mutation pattern for a set of 4 mutants with a single drug that affects 2 mutants.

Combination drug therapy design for HIV treatment

The evolution-replication dynamics of HIV subject to treatment [89, 90] can be cast as,

$$\dot{\psi} \;=\; (A \;+\; \mathrm{diag}\,(F_x x))\;\psi \;+\; d.$$

Here, the elements of ψ represent populations of individual HIV mutants. The diagonal elements of A represent each mutant's replication rate and the off diagonal elements of A represent the probability of mutation from one mutant to another. The linear function $F(x) := \mathrm{diag}\,(F_x x)$ captures the effect of drug therapy. The components of the vector x are doses of different drugs and the kth column of matrix F_x is describes how efficiently drug k kills each HIV mutant. See Fig. 1.1 for an illustration.

Structure is particularly important for this problem since drug-drug interactions cannot always be captured by the linear model. These interactions can be vital for efficacy of the drug or may have fatal consequences, so it is important to apply constraints on dosages and to impose penalties that promote the satisfaction of combinatorial conditions. However, imposing constraints on drug dosages is challenging using existing positive systems tools for $\mathcal{H}_2$ and $\mathcal{H}_\infty$ design because the drug doses x do not appear explicitly as optimization variables [59, 89, 90], making regularization difficult.

We first study design of structured controllers by designing a conservative structured controller in Section 6.2.2. We then show convexity of the original problem in Section 7. Since our formulations use the drug doses x as an optimization variable, our results apply naturally

structured combination drug therapy design; we explore this application in Section 7.5.

1.2.2 Performance metrics

In what follows, we use the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms to quantify the closed-loop performance. In the centralized state-feedback problem, the optimal control law for LTI systems is given by static state-feedback. Even though it is not clear if the optimal *structured* controller is also memoryless, we design structured static controllers as a first iteration in design. In Chapter 7, we explore the use of a time-varying controller for a certain class of structured control problem.

The $\mathcal{H}_2$ performance metric is the steady-state amplification from white-in-time stochastic forcing at the disturbance input d to the performance output ζ of system (1.1),

$$f_2(x) \;:=\; \lim_{t\,\to\,\infty} \mathbb{E}\left(\zeta^T(t)\,\zeta(t)\right) \;=\; \lim_{t\,\to\,\infty} \mathbb{E}\left(\psi^T(t)\,C^T C\,\psi(t) \;+\; u^T(t) R^T(x) R(x)\,u(t)\right).$$

This quantity is determined by the square of the $\mathcal{H}_2$ norm of system (1.1) and it can be expressed as a function of the controller x as

$$f_2(x) \;=\; \begin{cases} \operatorname{trace}\big((C^T C \;+\; R^T(x)R(x))\,P\big), & x \text{ stabilizing} \\ +\infty, & \text{otherwise} \end{cases} \tag{1.3a}$$

where x must be such that the closed-loop dynamical generator, $A+F(x)$, corresponds to a stable system (the spectrum of $A+F(x)$ must lie completely in the open left-half plane). The matrix P is the closed-loop controllability gramian given by the solution to the Lyapunov equation,

$$(A \;+\; F(x))\,P \;+\; P\,(A \;+\; F(x))^T \;+\; B_1 B_1^T \;=\; 0.$$

When $F(x) = -B_2 X$, $R(x) = -R^{1/2} X$, and there are no structural constraints on the matrix X, the optimal $\mathcal{H}_2$ feedback gain is determined by the linear quadratic regulator (LQR) and it can be computed via the solution of an algebraic Riccati equation.

The $\mathcal{H}_\infty$ performance metric, which we denote by $f_\infty(\cdot)$, is the maximum induced $\mathcal{L}_2$ gain from d to ζ in system (1.1),

$$f_\infty(x) \;:=\; \sup_{\|d\|_{\mathcal{L}_2} \leq 1} \frac{\|\zeta\|_{\mathcal{L}_2}}{\|d\|_{\mathcal{L}_2}},$$

where the $\mathcal{L}_2$ norm of a signal v is defined as,

$$\|v\|_{\mathcal{L}_2}^2 \;:=\; \int_0^\infty v^T(t) v(t)\,\mathrm{d}t.$$

This performance metric corresponds to the peak of the frequency response,

$$f_\infty(x) \;=\; \sup_{\omega\in\mathbb{R}} \bar{\sigma}\left(C\left(j\omega I \,-\, (A \,+\, F(x))\right)^{-1} B\right). \tag{1.3b}$$

As with f_2, the optimal $\mathcal{H}_\infty$ control problem for unstructured state feedback can be cast in a convex form and readily solved. However, as we describe next, incorporating structural restrictions on x significantly complicates the design problem.

1.3 Distributed Control

To motivate the study of and explain the difficulty inherent to structured optimal control, we begin with a discussion of networks of dynamical systems. In this problem, the controller structure of interest is the communication topology induced by the control law. Traditional design techniques, such as LQR, yield controllers which require a centralized implementation where every subsystem communicates with a central node.

For large-scale systems, the computational and communication costs associated with such an all-to-all communication topology may be prohibitively high. It is thus of interest to design controllers with distributed structure and sparse communication topologies. We first draw a connection between sparsity of the feedback gain matrix and the induced communication topology, and then highlight challenges that arise in the design of structured state-feedback controllers.

Let us assume that (1.2) contains N individual subsystems, each with a local state and control inputs. We partition the state and control input vectors into subvectors corresponding to each subsystem, $\psi := [\,\psi_1^T \;\cdots\; \psi_N^T\,]^T$ and $u = [\,u_1^T \;\cdots\; u_N^T\,]^T$, and write the subsystem dynamics as,

$$\dot{\psi}_i \;=\; A_{ii}\,\psi_i \;+\; \sum_{j\neq i} A_{ij}\,\psi_j \;+\; B_{1i}\,d \;+\; B_{2,ii}\,u_i. \tag{1.4a}$$

The block-sparsity pattern of A determines the interaction topology between subsystems; when A_{ij} is zero, subsystem j has no direct effect on the evolution of the state of subsystem i.

With each subsystem we associate a controller that specifies the control input u_i. Standard optimal control techniques typically induce a communication topology which requires every local controller to have access to the state of every subsystem. In large-scale networks of dynamical systems, this may impose significant communication burden and implementation may be prohibitively expensive. It is thus of interest to explore the design of feedback laws that utilize *limited* information exchange within a large-scale network.

Under linear state-feedback $u = -X\psi$, the dynamics (1.4a) become,

$$\dot{\psi}_i \;=\; A_{ii}\,\psi_i \;+\; \sum_{j\neq i} A_{ij}\,\psi_j \;+\; B_{1i}\,d \;-\; B_{2,ii}\sum_j X_{ij}\,\psi_j. \tag{1.4b}$$

Thus, the block-sparsity pattern of the feedback gain matrix X determines the communication topology of the static controller: forming the control input u_i requires access to the states of each subsystem j for which X_{ij} is nonzero.

Figure 1.2a illustrates a network of coupled subsystems, associated controller topology, and the sparsity patterns of the corresponding matrices A and X. The subsystems in the physical layer are represented by blue octagons; their interaction topology is marked by the blue arrows which correspond to the sparsity pattern of the matrix A. Each local controller is represented by a yellow circle; the structure of the information exchange network between the two layers is marked by the red arrows which correspond to the sparsity pattern of the feedback gain matrix X.

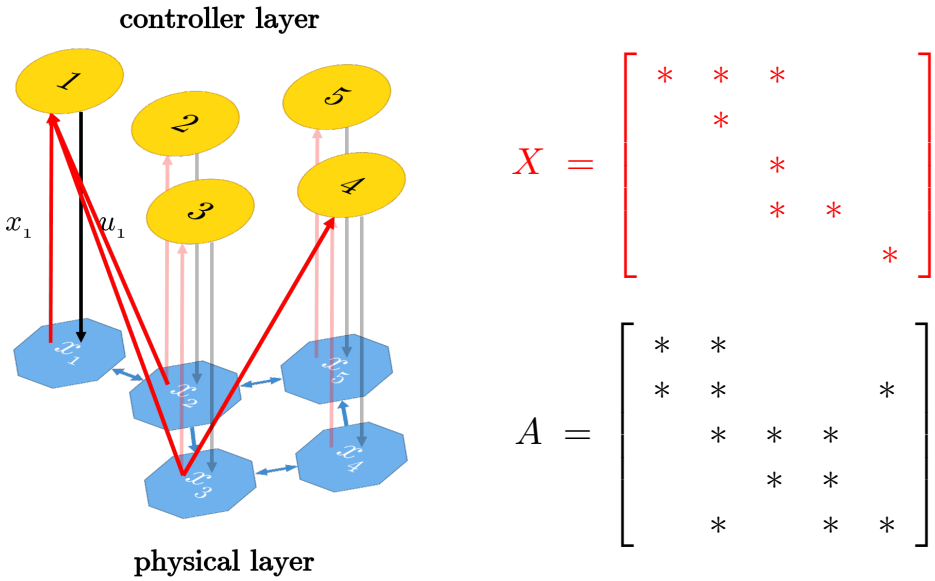
In the more general setup where the local controllers are dynamic (perhaps because they estimate the subsystem's state rather than directly measure it), it is important to determine the order of local controllers as well as the structure of the information exchange network in the controller layer; see Fig. 1.2b for an illustration. Recent advances have been made for particular classes of systems [23, 29], but addressing these questions in general remains an open challenge.

1.3.1 An example

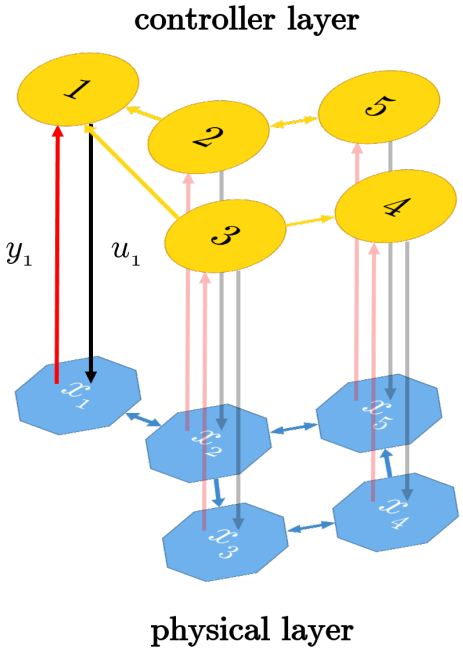
After permuting ψ relative to the definition in (1.4) for clarity of exposition, the state vector for the mass-spring system shown in Fig. 1.3 is determined by $\psi = \left[\, p^T \;\; v^T \,\right]^T$, where p and v are the vectors of positions and velocities of the N masses. We set all masses and spring constants to unity. Assuming that the control and disturbance inputs enter as forces and partitioning matrices in the state-space model (1.2) conformably with ψ yields

$$A \;=\; \begin{bmatrix} 0 & I \\ T & 0 \end{bmatrix}, \quad B_1 \;=\; B_2 \;=\; \begin{bmatrix} 0 \\ I \end{bmatrix},$$

where T is an $N \times N$ tridiagonal Toeplitz matrix with -2 on its main diagonal and 1 on its first sub- and super-diagonal. In the absence of the structural constraints, the solution to the Riccati equation yields the centralized $\mathcal{H}_2$-optimal controller, i.e., the linear quadratic regulator. In this case, the LQR, $X := [\, X_p \;\; X_v \,]$, has dense position and velocity feedback gain matrices X_p and



(a) The local controllers are memoryless.



(b) The local controllers are dynamic.

Figure 1.2: A network of 5 dynamical systems with associated local controllers.

Figure 1.3: Mass-spring system on a line.

X_v,

$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ u_3(t) \\ u_4(t) \end{bmatrix} = -\underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{X_p} \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ p_4(t) \end{bmatrix} - \underbrace{\begin{bmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}}_{X_v} \begin{bmatrix} v_1(t) \\ v_2(t) \\ v_3(t) \\ v_4(t) \end{bmatrix}.$$

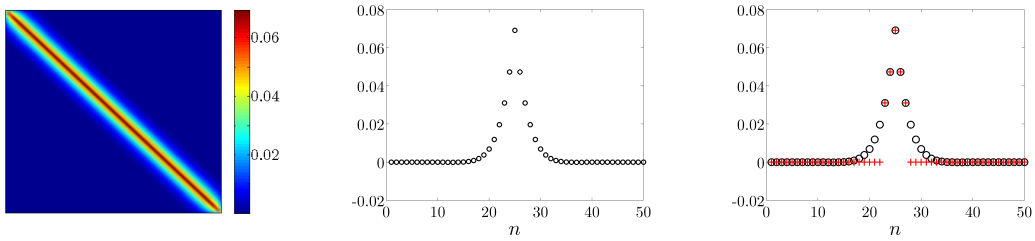
Even though these matrices are populated with non-zero elements, the gains that are used to form control actions for individual masses display interesting patterns. Figure 1.4 illustrates the optimal centralized position feedback gain matrix X_p in the system with 50 masses. Apart from the edges, both X_p and X_v (not shown) have almost constant diagonals and exponential off-diagonal decay. This suggests that good performance may be achievable if the smaller elements of the feedback gain matrix are set to 0. Since the small feedback gains represent interactions between masses that are spatially distant, such a strategy would allow the masses' controllers to interact in a distributed fashion.

More generally, for spatially invariant systems such as the mass-spring system on a circle, the optimal controllers with respect to quadratic performance indices (e.g., $\mathcal{H}_2$, $\mathcal{H}_\infty$) are also spatially invariant and they exponentially discount information with spatial distance [9]. Moreover, it has been suggested that optimal controllers for spatially-decaying systems over general graphs also possess spatially-decaying property [60, 61]. This motivates the search for inherently *localized* controllers and suggests that localized information exchange in the distributed controller may provide a viable strategy for controlling large-scale systems.

However, in general, low-magnitude elements in the feedback gain cannot reliably be assumed unimportant. It is difficult to provide error bounds on the deviation from optimality as a result of truncation (e.g., as in Fig. 1.4c) and it has been recognized that truncation of the centralized controller could significantly compromise the closed-loop performance and even yield a controller that does not guarantee closed-loop stability [60, 64].

1.3.2 Lack of convexity for structured control

We consider the state feedback case, i.e., $F(x) = B_2X$, for $\mathcal{H}_2$ optimal control. The design of the optimal state-feedback gain X, subject to constraints on its sparsity pattern (equivalently, on



(a) (b) (c)

Figure 1.4: (a) The optimal centralized position feedback gain matrix X_p in the system with 50 masses. Both X_p and X_v (not shown) have almost constant diagonals (modulo edges) and exponential off-diagonal decay. (b) Optimal centralized position gains for the middle mass $n = 25$. (c) Truncation of the optimal centralized position gains for the middle mass $n = 25$.

the communication topology in Fig. 1.2a) has a rich history and was recently revisited in [91,92]. Let the subspace $\mathcal{S}$ encapsulate these structural constraints and let us assume that there is a stabilizing $X \in \mathcal{S}$. The optimal control problem of determining stabilizing $X \in \mathcal{S}$ that minimizes the $\mathcal{H}_2$ norm of system (1.2) can be formulated as

$$\begin{aligned} &\text{minimize} && f(X) \\ &\text{subject to} && X \in \mathcal{S} \end{aligned} \tag{1.5}$$

and brought into the following form

$$\begin{aligned} \underset{P,\,X}{\text{minimize}} \quad & \text{trace}\left((Q + X^T R\, X)\, P\right) \\ \text{subject to} \quad & (A - B_2 X)\, P + P\,(A - B_2 X)^T + B_1 B_1^T = 0 \\ & P \succ 0, \quad X \in \mathcal{S} \end{aligned} \tag{1.6}$$

where $Q := C^T C$ and $R \succ 0$. In the absence of the structural constraint $X \in \mathcal{S}$, a standard change of variables [49]

$$Z := X\, P \tag{1.7}$$

can be used to express the square of the $\mathcal{H}_2$ norm as,

$$f(P, Y) = \text{trace}\,(Q\, P) + \text{trace}\,(R\, Z P^{-1}\, Z^T)$$

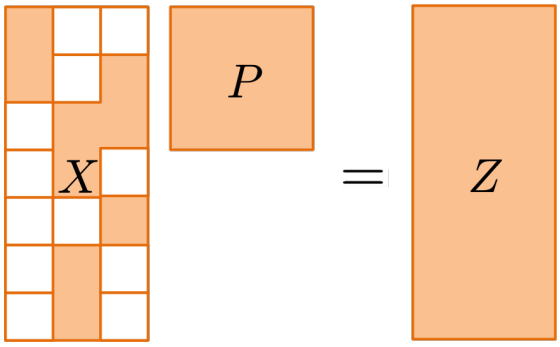


Figure 1.5: The change of variables that casts the unstructured state-feedback problem as an SDP, in general, does not preserve the structural properties of X.

and the Schur complement can be employed to cast the optimal state-feedback $\mathcal{H}_2$ control problem as an SDP,

$$\begin{array}{ll} \underset{P,\,Y,\,Z}{\text{minimize}} & \text{trace}\,(Q\,P) \;+\; \text{trace}\,(R\,\Theta) \\ \text{subject to} & (A\,P \;-\; B_2\,Z) \;+\; (A\,P \;-\; B_2\,Z)^T \;+\; B_1 B_1^T \;=\; 0 \\ & \begin{bmatrix} \Theta & Z \\ Z^T & P \end{bmatrix} \;\succeq\; 0. \end{array}$$

Since P is positive definite, it is invertible, and the optimal centralized (i.e., unstructured) X is determined by $X_c = ZP^{-1}$. This centralized solution coincides with the linear quadratic regulator, which can be explicitly determined by $X_c = R^{-1}B_2^T P$ where P is the unique positive definite solution of the algebraic Riccati equation, $A^T P + PA + Q - PB_2\,R^{-1}B_2^T P = 0$.

The above change of variables is, in general, not suitable for imposing structure on X. Although the constraint on the feedback gain matrix $X \in \mathcal{S}$ is linear and thus convex, the corresponding constraint on P and Z is bilinear, $ZP^{-1} \in \mathcal{S}$. This makes it difficult to translate the sparsity patterns of X to the sparsity patterns of P and Z (see Fig. 1.5 for an illustration), thereby limiting the use of these coordinates for structured design problems. In fact determining stabilizability, let alone achieving optimal closed-loop performance, is NP hard for general structured problems [4, 5].

If P is restricted to be diagonal, the sparsity structure of X coincides with the sparsity structure of Z. However, this may introduce considerable conservatism in the design and may not even lead to a feasible SDP characterization (even when the original nonconvex problem is feasible). One notable special case where this relaxation is tight appears in the $\mathcal{H}_\infty$ optimal

control of positive systems [54].

1.4 Regularized problems

In order to design structured controllers, we draw inspiration from recent advances in compressive sensing. Given some additional assumptions on data structure, it is possible to completely recover signals even when they are sampled at frequencies below Nyquist rate. While it may seem magical to a student fresh from an undergraduate signals and systems course, this result makes intuitive sense; using additional information about the signal's structure allows one to recover it with less data from elsewhere. The magic sauce is the concept of regularization, which allows one to incorporate additional knowledge about the signal structure into the recovery problem by augmenting the standard loss functions associated with signal sampling with nonsmooth structure-promoting penalty terms.

Regularized problems have found applications in diverse fields including compressive sensing [93], machine learning [94], statistics [95], image processing [96], and, as we develop in this thesis, control theory [32]. This is a class of composite optimization problems in which the objective function is a sum of a differentiable but possibly nonconvex component and a convex nondifferentiable component,

$$\text{minimize} \;\; f(x) \;+\; \gamma\, g(x).$$

The differentiable component of the objective function, $f(x)$, typically encodes some measure of performance. In a least-squares setting, $f(x) = \|Ax - b\|^2$ measures how well a candidate set of parameters x satisfies the linear relationship $Ax = b$. The nondifferentiable component of the objective function, $g(x)$, is a structure-promoting penalty function and γ specifies the emphasis on this structure relative to performance. When $\gamma = 0$, an unstructured x will be recovered and larger values of γ yield more structured optimal solutions. As an example, the ℓ_1 norm $g(x) = \sum |x_i|$ is a commonly used proxy for promoting sparsity of x.

In much of the work that utilizes regularization, γ can be chosen in a systematic way that arises from assumptions on the problem structure and problem data, e.g. number of nonzero entries and the statistics of noise affecting the measurements. In contrast, we use this framework somewhat artificially. Instead of using regularization to harness *a priori* knowledge of the structure of x in order to recover it, we use regularization to *artificially impose* structure on a design variable x to satisfy engineering requirements.

In feedback synthesis, we augment a traditional performance metric (such as the $\mathcal{H}_2$ or $\mathcal{H}_\infty$

norm) with a regularization function to promote certain structural properties in the optimal controller. For example, the ℓ_1 norm and the nuclear norm are commonly used nonsmooth convex regularizers that encourage sparse and low-rank optimal solutions, respectively.

Such regularized problems can be used to identify controller structure. This is particularly important because recently, it has been demonstrated that the *design of controller architectures* can have a more profound impact on the closed-loop performance than the optimal design under a given pre-specified architecture [97]. In [62, 64], tools and ideas from control theory, optimization, and compressive sensing have been combined to systematically address the challenge of designing controller architectures. The proposed approach introduces regularized versions of standard optimal control problems and aims to strike a balance between closed-loop performance and controller complexity. As discussed earlier, when the state vector and control inputs can be partitioned into subvectors that correspond to separate subsystems, promoting sparsity of the feedback gain matrix limits information exchange between the physical system and the controller. Sparse controller architectures can be designed by augmenting standard quadratic performance measures with sparsity-promoting penalty functions which serve as measures of controller complexity. Such an approach has received much recent attention [62, 64, 98–103].

Alongside sparse feedback synthesis, the critical question of sensor and actuator selection has been recently considered in [104,105]. Although, in general, finding the solution to this problem requires an intractable combinatorial search, by drawing upon recent developments in sparse representations, this problem can be cast as a semidefinite program (SDP). Moreover, it is also of interest to study problems where it is desired to impose structure on a linear function of the design variable [106,107]. This broader framework covers a wide variety of problems ranging from wide-area and distributed PI control of power networks [108–111], to combination drug therapy for HIV treatment [112], to edge addition [76,113,114] and leader selection [77,78,80,81,83,84,87] in consensus networks.

Several recent efforts have focused on establishing convexity for classes of these problems and on developing efficient algorithms for optimal controller design for both convex and nonconvex problems. Convex structured optimal control problems include symmetric modifications to symmetric linear systems [62, 115, 116], diagonal modifications to positive systems [87, 112], optimal sensor and actuator selection [104, 105], and edge addition to undirected consensus [76, 113, 114] and synchronization [117] networks. Algorithmic developments have employed alternating direction method of multipliers [64, 105], proximal gradient and Newton methods [114], as well as first- and second-order method of multipliers [107, 118–120] to efficiently perform identification of controller structure and structured feedback synthesis.

1.5 Design of structured controllers via regularization

The communication architecture $\mathcal{S}$ of the state-feedback controller in (1.6) is fixed and *a priori* specified which may impose limits on the achievable performance. For problems where the communication topology is *not* fixed, it is desirable to *design* a favorable communication topology while promoting sparsity of the communication links. To achieve this, an optimization framework which augments the $\mathcal{H}_2$ objective function with a penalty on the sparsity of the feedback gain matrix (i.e., the number of communication links) was introduced in [62–64].

1.5.1 Structure identification

Our objective is to design controller architecture that achieves a desired tradeoff between the quadratic performance of system (1.1) and the structure of the controller x. To address this challenge we consider a regularized optimal control problem

$$\begin{array}{lccc} \underset{x}{\text{minimize}} & f(x) & + & \gamma\, g(Tx) \\ & \downarrow & & \downarrow \\ & \textbf{closed-loop} & & \textbf{controller} \\ & \textbf{performance} & & \textbf{structure} \end{array} \tag{1.8}$$

where $f(x)$ is the $\mathcal{H}_2$ or $\mathcal{H}_\infty$ norm of system (1.1). In contrast to (1.6), no structural constraints are imposed on x in (1.8); instead, the objective is to manage structure of the controller by introducing a regularization term $g(Tx)$ into the optimal control problem. The matrix T allows regularization of structure in a different set of coordinates. The non-negative regularization parameter γ encodes the emphasis on controller structure relative to the closed-loop performance. For the state feedback problem $F(x) = B_2X$ regularized with a sparsity-promoting penalty function, $\gamma = 0$ yields the centralized LQR solution. As γ increases, larger emphasis is placed on obtaining a more structured feedback gain matrix X; see Fig. 1.6 for an illustration.

Problem (1.8) is difficult to solve directly because f is typically a nonconvex function of x and g is convex but not differentiable. While the nonlinear change of coordinates (1.7) yields a convex dependence of f on P and Y, in general, it introduces a nonconvex dependence of the regularization term g on these optimization variables.

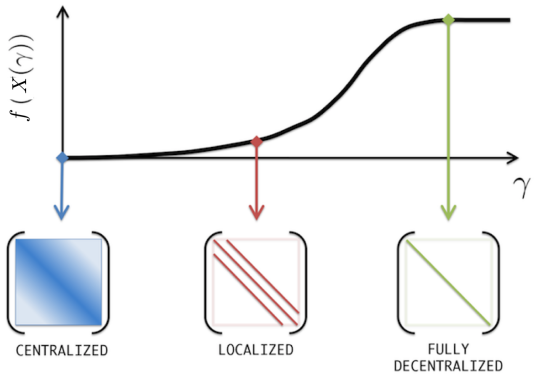


Figure 1.6: Increased emphasis on sparsity encourages sparser control architectures at the expense of deteriorating the closed-loop performance. For $\gamma = 0$ the optimal centralized controller X_c is obtained from the positive definite solution of the algebraic Riccati equation. Control architectures for $\gamma > 0$ are determined by $X(\gamma) := \text{argmin}_X \left(f(X) + \gamma\, g(X)\right)$ and they depend on interconnections in the distributed plant and the state and control performance weights Q and R.

Sparsity-promoting regularizers

Elementwise sparsity of x can be promoted by incorporating the cardinality function into the optimal control problem (1.8),

$$g_0(x) \;=\; \mathbf{card}\,(x)\,. \tag{1.9a}$$

This regularizer counts the number of nonzero elements in x and it yields a combinatorial optimization problem (1.8) whose solution typically requires an intractable combinatorial search. A weighted ℓ_1 penalty,

$$g_1(x) \;:=\; \|w \circ x\|_1 \;=\; \sum_i w_i\,|x_i| \tag{1.9b}$$

provides a convex proxy for promoting elementwise sparsity of x [121]. Here, w is the vector whose elements are non-negative weights w_i and $\circ$ is elementwise matrix multiplication. The weights w_i can be selected to place larger relative penalties on certain elements of x. Similarly, the sum of the ℓ_2 norms (*not* the ℓ_2 norm squared) of the subvectors $x_i \in \mathbb{R}^{n_i}$,

$$g_2(x) \;=\; \sum_i w_i\,\|x_i\|_2 \tag{1.9c}$$

enhances group sparsity; i.e., sparsity at the level of subvectors [122].

The ℓ_1 norm is the largest convex function that underestimates the cardinality function on the domain $[-1, 1]$ [123]; see Fig. 1.7 for an illustration in the scalar case. Both the ℓ_1

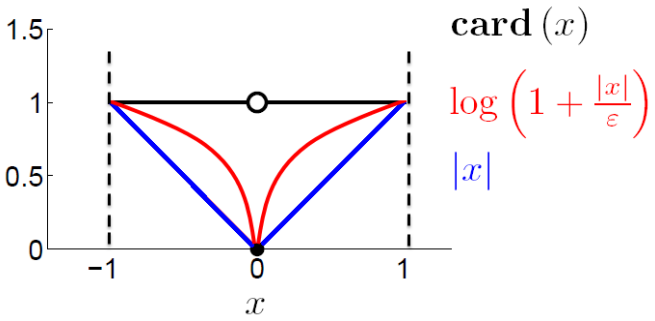


Figure 1.7: Cardinality function of a scalar variable x and the corresponding absolute value and logarithmic approximations on $x \in [-1,\, 1]$.

norm and its weighted version are convex relaxations of $\mathbf{card}(X)$. On the other hand, better approximation can be obtained with nonconvex functions, e.g., the sum-of-logs,

$$g_4(x) \;=\; \sum_i \log\left(1 \,+\, \frac{|x_i|}{\varepsilon}\right), \quad 0 \,<\, \varepsilon \,\ll\, 1. \tag{1.9d}$$

The weighted ℓ_1 norm attempts to bridge the difference between the ℓ_1 norm and the cardinality function. In contrast to the cardinality function that assigns the same cost to any nonzero element, the ℓ_1 norm penalizes more heavily the elements of larger magnitudes. The positive weights can be chosen to counteract this magnitude dependence of the ℓ_1 norm. For example, if the weights w_i are inversely proportional to the magnitude of x_i,

$$\begin{cases} w_i \;=\; 1/|x_{ij}|, & x_i \,\neq\, 0 \\ w_i \;=\; \infty, & x_i \,=\, 0 \end{cases}$$

then there is no difference between the weighted ℓ_1 norm of x and the cardinality function of x. This scheme, however, cannot be implemented, because the weights depend on the unknown feedback gain. A re-weighted algorithm that solves a sequence of weighted ℓ_1 optimization problems was proposed in [121]. In this, sequential linearization of the sum-of-logs function is used and the weights are determined by the solution of the optimization problem in the previous iteration. This algorithm has provided an effective heuristics for promoting sparsity in many emerging applications.

Additional intuition about the role of sparsity-promoting regularizers can be gained by considering a problem in which it is desired to find the sparsest feedback gain that provides a given

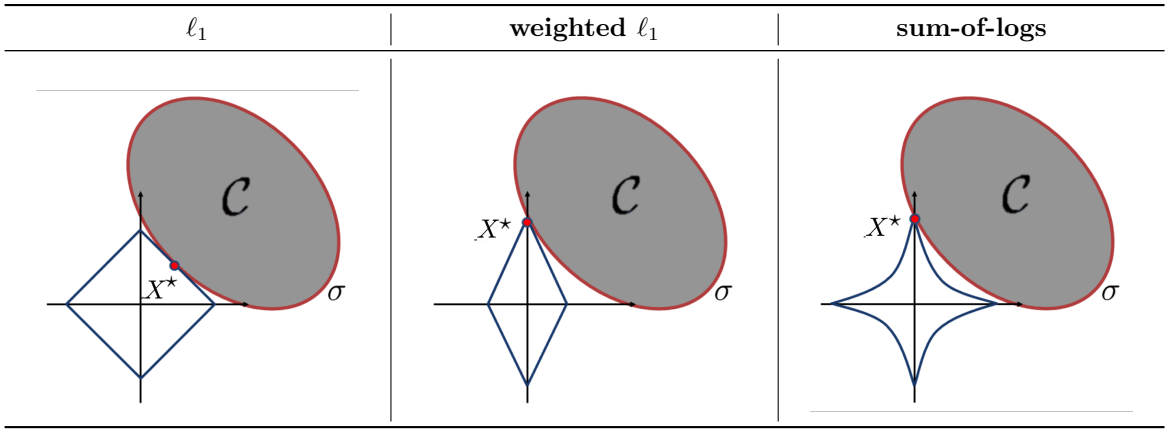


Figure 1.8: The solution $x^\star$ of the constrained problem (1.10) is the intersection of the constraint set $\mathcal{C} := \{x \,|\, f(x) \leq \sigma\}$ and the smallest sub-level set of g that touches $\mathcal{C}$. The penalty function g is the ℓ_1 norm (left); the weighted ℓ_1 norm with appropriate weights (middle); and the nonconvex sum-of-logs function (right).

level of $\mathcal{H}_2$ performance $\sigma > 0$,

$$\begin{array}{ll} \text{minimize} & \mathbf{card}\,(x) \\ \text{subject to} & f(x) \leq \sigma. \end{array}$$

Approximating $\mathbf{card}\,(x)$ with a penalty function $g(x)$ yields

$$\begin{array}{ll} \text{minimize} & g(x) \\ \text{subject to} & f(x) \leq \sigma. \end{array} \tag{1.10}$$

The solution to (1.10) is the intersection of the constraint set $\mathcal{C} := \{x \,|\, f(x) \leq \sigma\}$ and the smallest sub-level set of g that touches $\mathcal{C}$; see Fig. 1.8 for an illustration. In contrast to the ℓ_1 norm whose sub-level sets are determined by the convex ℓ_1 ball, the sub-level sets of the nonconvex sum-of-logs function have a star-like shape.

Other regularization functions

The nuclear norm of a matricial variable X,

$$g_*(X) \;:=\; \sum_i \sigma_i(X)$$

is the sum of the singular values of the matricial variable X and promotes a low-rank optimal solution. Enforcing low-rank solutions is often important for low-complexity modelling [46, 124–127]

The indicator function associated with a convex set $\mathcal{C}$,

$$g_{\mathcal{C}}(x) := \begin{cases} 0 & x \in \mathcal{C} \\ \infty & x \notin \mathcal{C} \end{cases}$$

can be used to enforce that x lie in the convex set $\mathcal{C}$. Such sets can encode structural properties as simple as box constraints and can even be used to promote more complicated combinatorial conditions such as mutual exclusivity or necessity between elements of x (i.e., either x_i or x_j may be nonzero or x_j may only be nonzero if x_i is nonzero) [128]. These sorts of constraints are very important in the context of the drug therapy example described in Section 1.2.1 since linear models often cannot capture the full complexity of drug-drug interactions.

Finally, we note that recent work has used the framework of atomic norms [129] to penalize communication between subsystems in large-scale distributed systems using more nuanced measures of controller complexity [100–103].

1.5.2 Polishing

After having identified the controller architecture, we optimize the closed-loop $\mathcal{H}_2$ performance over the identified structure $\mathcal{S}$. This is necessary because the presence of the regularizer in (1.8) often distorts the optimal solution. For example, in the ℓ_1 regularized case, the ℓ_1 norm imposes an additional penalty on the magnitude of the feedback gains, resulting in worse closed-loop performance. As a result, we obtain the final structured controller by solving a structured problem,

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \mathcal{S} \end{aligned}$$

which is equivalent to (1.8) where the regularizer is an indicator function corresponding to the identified structure $\mathcal{S}$.

For nuclear norm regularization, polishing amounts to optimizing over the singular values of the matrix obtained by solving (1.8). Regularization with an indicator function does not require a polishing step because it does not distort the value of the performance metric f in (1.8) as it enforces an exact condition on x and is not a proxy for a nonconvex or combinatorial constraint.

1.6 Outline of thesis

This thesis approaches the challenge of designing structured controllers for linear systems by solving regularized optimal control problems. In order to deal with the inherent difficulties of solving these optimization problems, we tackle two overarching problems. The first is that of developing tractable methods for solving regularized optimal control problems, which are in general nonconvex. The second is that of identifying classes of problems for which regularized optimal control can be placed into a convex form as well as developing methods to use convex design problems to provide suboptimal solutions for the harder, nonconvex problems.

In Part I, we study algorithms for solving regularized problems. In Chapter 2, we develop a novel splitting method based on a reformulation of the generalized regularized problem (1.8) with an auxiliary variable. By exploiting the properties of proximal operators, we bring the associated augmented Lagrangian into a continuously differentiable form by constraining it to a manifold that corresponds to explicit minimization over the auxiliary variable. The new expression, which we call the 'proximal augmented Lagrangian', facilitates a method of multipliers algorithm that offers many advantages relative to existing methods, including guaranteed convergence for nonconvex problems and the ability to regularize a linear function of the optimization variable. We then study primal-descent dual-ascent Arrow-Hurwicz-Uzawa type gradient flow dynamics to solve regularized problems in Chapter 3. We prove global convergence and establish conditions for exponential convergence for continuous- and discrete-time updates. In Chapter 4, we take advantage of generalizations of the gradient to apply second-order primal-dual updates to the proximal augmented Lagrangian and prove global and local quadratic convergence. Finally, in Chapter 5, we draw connections with other methods and provide additional insight.

In Part II, we identify classes of convex regularized optimal control problems. In Chapter 6, we study the problem of designing symmetric modifications to symmetric linear systems. This problem is convex in the underlying design variable and is thus appealing for the purpose of structured control. We show that even when the system and controller are not symmetric, their symmetric components can be used to perform structured design with stability and performance guarantees. In Chapter 7, we examine the problem of designing structured diagonal modifications to positive systems. We prove that the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control problems are convex, apply our results combination drug therapy design for HIV and leader selection in consensus networks, and show that a constant controller is optimal for an induced-power performance index. In Chapter 8, we examine the optimal actuator and sensor selection problems. Using a change of variables, this problem can be cast as a semidefinite program. Although convex,

the computational complexity required to solve problems of this class scales poorly with the problem dimension. We develop customized algorithms to solve this problem efficiently.

Finally, we provide concluding remarks and comment on future research in Chapter 9.

Notation

The set of real numbers is denoted by $\mathbb{R}$ and the set of nonnegative (positive) reals is denoted by $\mathbb{R}_+$ ($\mathbb{R}_{++}$). The integers are represented by $\mathbb{Z}$. The set of complex numbers are denoted by $\mathbb{C}$ and $\mathrm{j} = \sqrt{-1}$ is the imaginary unit. The operator $\Re(\cdot)$ ($\Im(\cdot)$) extracts the real (imaginary) component of a complex argument.

Given a matrix A, A^T denotes its transpose. The vector $\lambda(A)$ ($\sigma(A)$) indicates the eigenvalues (singular values) of A, $\lambda_i(A)$ denotes the ith largest eigenvalue of A and $\bar{\sigma}(A)$ denotes the principal (largest) singular value of A. We use $\mathrm{trace}(A)$ to denote its trace, and $\|A\|_F^2 := \mathrm{trace}\left(A^T A\right)$ to denote its Frobenius norm squared. The vector inner product is given by $\langle x,\, y\rangle := x^T y$ and the matricial inner product is given by $\langle X,\, Y\rangle := \mathrm{trace}(X^T Y)$. The Kronecker product is denoted by $\otimes$ and the Hadamard (entrywise) product is denoted by $\circ$.

We write $A \geq 0$ ($A > 0$) if A is entrywise nonnegative (positive) and $A \succeq 0$ ($A \succ 0$) to denote that A is symmetric and positive semidefinite (definite).

We use the $\mathrm{diag}(\cdot)$ operator to denote either the diagonal entries of a matrix or a diagonal matrix with elements of the vector $\cdot$ on its diagonal, depending on its argument. The symbol $\mathbb{E}$ denotes the expectation operator.

Given a set $\mathcal{C}$ we define the indicator function

$$I_{\mathcal{C}}(x) := \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{otherwise.} \end{cases}$$

We define the sparsity pattern $\mathrm{sp}\,(u)$ of a vector u to be the set of indices for which u_i is nonzero.

Definition 1. The adjoint of a linear operator F: $\mathbb{R}^m \to \mathbb{R}^{n\times n}$ is the linear operator $F^\dagger$: $\mathbb{R}^{n\times n} \to \mathbb{R}^m$ which satisfies

$$\langle X,\, F(x)\rangle \;=\; \left\langle F^\dagger(X),\, x\right\rangle.$$

Any additional notation specific to particular chapters will be introduced as required.

Part I

Proximal augmented Lagrangian algorithms

Chapter 2

Method of multipliers

We study a class of composite optimization problems in which the objective function is the sum of a differentiable but possibly nonconvex component and a convex nondifferentiable component. Problems of this form are encountered in diverse fields including compressive sensing [93], machine learning [94], statistics [95], image processing [96], and control [64]. In feedback synthesis, they typically arise when a traditional performance metric (such as the $\mathcal{H}_2$ or $\mathcal{H}_\infty$ norm) is augmented with a regularization function to promote certain structural properties in the optimal controller. For example, the ℓ_1 norm and the nuclear norm are commonly used nonsmooth convex regularizers that encourage sparse and low-rank optimal solutions, respectively.

In this chapter, we derive the proximal augmented Lagrangian, which allows us to apply the widely used method of multipliers (MM) to nondifferentiable composite optimization problems. We then illustrate the utility of this approach by applying it to edge addition in directed consensus networks and sparse feedback synthesis.

2.1 Background

The lack of a differentiability in objective function (1.8) due to the regularization function precludes the use of standard descent methods for smooth optimization. Proximal gradient methods [130] and their accelerated variants [131] generalize gradient descent, but typically require the nonsmooth term to be separable over the optimization variable. Furthermore, standard acceleration techniques are not well-suited for problems with constraint sets that do not admit an easy projection (e.g., closed-loop stability).

An alternative approach is to split the smooth and nonsmooth components in the objective

function over separate variables which are coupled via an equality constraint. Such a reformulation facilitates the use of the alternating direction method of multipliers (ADMM) [132]. This augmented-Lagrangian-based method splits the optimization problem into subproblems which are either smooth or easy to solve. It also allows for a broader class of regularizers than proximal gradient and it is convenient for distributed implementation. However, there are limited convergence guarantees for nonconvex problems and parameter tuning greatly affects its convergence rate.

The method of multipliers (MM) is the most widely used algorithm for solving constrained nonlinear programing problems [133–135]. In contrast to ADMM, it is guaranteed to converge for noncconvex problems and there are systematic ways to adjust algorithmic parameters. However, MM is not a splitting method and it requires *joint* minimization of the augmented Lagrangian with respect to *all* primal optimization variables. This subproblem is typically nonsmooth and as difficult to solve as the original optimization problem.

To treat this problem, we transform the augmented Lagrangian into a continuously differentiable form by exploiting the structure of proximal operators associated with nonsmooth regularizers. This new form is obtained by constraining the augmented Lagrangian to a manifold that corresponds to the explicit minimization over the variable in the nonsmooth term. The resulting expression, that we call the *proximal augmented Lagrangian*, is given in terms of the Moreau envelope of the nonsmooth regularizer and it is continuously differentiable. This allows us to take advantage of standard optimization tools, including gradient descent and quasi-Newton methods, and enjoy the convergence guarantees of the standard method of multipliers.

2.2 Problem formulation and background

2.2.1 Composite optimization problem

We consider a composite optimization problem,

$$\underset{x}{\text{minimize}} \quad f(x) \; + \; g\left(\mathcal{T}(x)\right) \tag{2.1}$$

where the optimization variable x belongs to a finite-dimensional Hilbert space (e.g., $\mathbb{R}^m$ or $\mathbb{R}^{m\times n}$) equipped with an inner product $\langle\cdot,\,\cdot\rangle$ and associated norm $\|\cdot\|$. The function f is continuously differentiable but possibly nonconvex, the function g is convex but potentially nondifferentiable, and $\mathcal{T}$ is a bounded linear operator. We further assume that g is proper and lower semicontinuous, that (2.1) is feasible, and that its minimum is finite.

Regularization of $\mathcal{T}(x)$ instead of x is important in the situations where the desired structure has a simple characterization in the co-domain of $\mathcal{T}$. Such problems arise in spatially-invariant systems, where it is convenient to perform standard control design in the spatial frequency domain [9] but necessary to promote structure in the physical space, and in consensus/synchronization networks, where the objective function is expressed in terms of the deviation of node values from the network average but it is desired to impose structure on the network edge weights [106].

2.2.2 Background on proximal operators

Problem (2.1) is difficult to solve directly because f is, in general, a nonconvex function of x and g is typically not differentiable. Since the existing approaches and our method utilize proximal operators, we first provide a brief overview; for additional information, see [130].

The proximal operator of the function g is given by,

$$\mathbf{prox}_{\mu g}(v) \;:=\; \underset{x}{\operatorname{argmin}}\; g(x) \;+\; \tfrac{1}{2\mu}\,\|x \,-\, v\|^2 \tag{2.2a}$$

and the associated optimal value specifies its Moreau envelope,

$$M_{\mu g}(v) \;:=\; \inf_{x}\; g(x) \;+\; \tfrac{1}{2\mu}\,\|x \,-\, v\|^2 \tag{2.2b}$$

where $\mu > 0$. The Moreau envelope is a continuously differentiable function, even when g is not, and its gradient is given by,

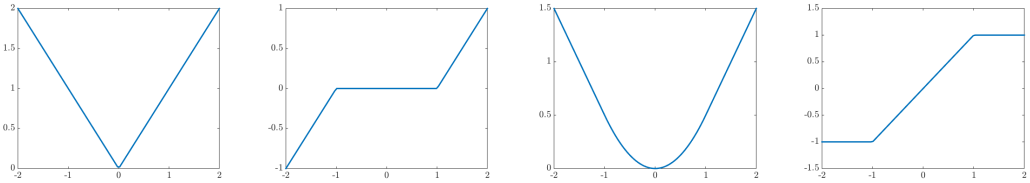
$$\nabla M_{\mu g}(v) \;=\; \tfrac{1}{\mu}\left(v \;-\; \mathbf{prox}_{\mu g}(v)\right). \tag{2.2c}$$

As a concrete example, when g is the ℓ_1 norm with unit weight $w = \mathbb{1}$, the proximal operator is determined by soft-thresholding,

$$\mathbf{prox}_{\mu g_1}(v_i) \;=\; \mathcal{S}_\mu(v_i) \;:=\; \operatorname{sign}(v_i)\max(|v_i| \,-\, \mu,\, 0) \tag{2.3a}$$

the associated Moreau envelope is given by the Huber function,

$$M_{\mu g_1}(v_i) \;=\; \begin{cases} v_i^2/(2\mu) & |v_i| \;\leq\; \mu \\ |v_i| \,-\, \mu/2 & |v_i| \;\geq\; \mu \end{cases} \tag{2.3b}$$



(a) $g_1(v)$ (b) $\mathbf{prox}_{\mu g_1}(v)$ (c) $M_{\mu g_1}(v)$ (d) $\nabla M_{\mu g_1}(v)$

Figure 2.1: The regularization function $g_1(v) = |v|$ for a scalar argument v with associated proximal operator, Moreau envelope, and Moreau envelope gradient when $\mu = 1$.

and the gradient of this Moreau envelope is the saturation function,

$$\nabla M_{\mu g_1}(v_i) \;=\; \operatorname{sign}(v_i)\min(|v_i|/\mu,\, 1). \tag{2.3c}$$

Figure 2.1 plots these functions for a scalar argument with $\mu = 1$.

2.2.3 Existing algorithms

Proximal gradient

The proximal gradient method generalizes standard gradient descent to certain classes of nonsmooth optimization problems. This method can be used to solve (2.1) when $g(\mathcal{T})$ has an easily computable proximal operator. The standard gradient descent update to find a minimizer $x^\star$ of a differentiable function $h(x)$ is given by

$$x^{l+1} \;=\; x^l \;-\; \alpha_l \nabla h(x^l)$$

where α_l is a step size. When $\mathcal{T} = I$, the proximal gradient method for problem (2.1) with step-size α_l is given by,

$$x^{l+1} \;=\; \mathbf{prox}_{\alpha_l g}(x^l \;-\; \alpha_l \nabla f(x^l)).$$

Clearly, this method is most effective when the proximal operator of g is easy to evaluate. When $g = 0$, the proximal gradient method simplifies to standard gradient descent, and when g is indicator function of a convex set, it simplifies to projected gradient descent.

In particular, the proximal gradient update for the ℓ_1-regularized least-squares problem (LASSO),

$$\underset{x}{\text{minimize}} \;\; \tfrac{1}{2}\,\|Ax \;-\; b\|^2 \;+\; \gamma\,\|x\|_1 \tag{2.4}$$

where γ is a positive regularization parameter, is given by the Iterative Soft-Thresholding Algorithm (ISTA),

$$x^{l+1} \;=\; \mathcal{S}_{\gamma\alpha_l}(x^l \;-\; \alpha_l A^T(Ax^l \;-\; b)).$$

When $g(x) = 0$, the proximal gradient method simplifies to standard gradient descent, and when $g(x)$ is the indicator function $I_{\mathcal{C}}(x)$ of the convex set $\mathcal{C}$, it simplifies to projected gradient descent.

Except in special cases, e.g, when $\mathcal{T} = I$ or is a diagonal operator, efficient computation of $\mathbf{prox}_{g(\mathcal{T})}$ does not necessarily follow from an efficiently computable $\mathbf{prox}_g$. This makes the use of proximal gradient method challenging for many applications and its convergence can be slow. Acceleration techniques [131, 136] improve the convergence rate, but they do not perform well in the face of constraints such as closed-loop stability.

Augmented Lagrangian methods

A common approach for dealing with a nondiagonal linear operator $\mathcal{T}$ in (2.1) is to introduce an additional optimization variable z

$$\begin{array}{ll} \text{minimize} & f(x) \;+\; g(z) \\ \text{subject to} & \mathcal{T}(x) \;-\; z \;=\; 0. \end{array} \tag{2.5}$$

The augmented Lagrangian is obtained by adding a quadratic penalty on the violation of the linear constraint to the regular Lagrangian associated with (2.5),

$$\mathcal{L}_\mu(x, z; y) \;=\; f(x) \;+\; g(z) \;+\; \langle y,\, \mathcal{T}(x) \;-\; z\rangle \;+\; \tfrac{1}{2\mu}\,\|\mathcal{T}(x) \;-\; z\|^2$$

where y is the Lagrange multiplier, μ is a positive parameter, and $\mathcal{L}_\mu$ is obtained by augmenting the regular Lagrangian with a quadratic penalty on the violation of the linear constraint in (2.5).

ADMM solves (2.5) via an iteration which involves minimization of $\mathcal{L}_\mu$ separately over x and z and a gradient ascent update (with step size $1/\mu$) of y [132],

$$x^{k+1} \;=\; \underset{x}{\operatorname{argmin}}\; \mathcal{L}_\mu(x,\, z^k;\, y^k) \tag{2.6a}$$

$$z^{k+1} \;=\; \underset{z}{\operatorname{argmin}}\; \mathcal{L}_\mu(x^{k+1},\, z;\, y^k) \tag{2.6b}$$

$$y^{k+1} \;=\; y^k \;+\; \tfrac{1}{\mu}\,(\mathcal{T}(x^{k+1}) \;-\; z^{k+1}). \tag{2.6c}$$

ADMM is appealing because, even when $\mathcal{T}$ is nondiagonal, the z-minimization step amounts

to evaluating $\mathbf{prox}_{\mu g}$, and the x-minimization step amounts to solving a smooth (but possibly nonconvex) optimization problem. Although it was recently shown that ADMM is guaranteed to converge to a stationary point of (2.5) for some classes of nonconvex problems [137], its rate of convergence is strongly influenced by the choice of μ.

The method of multipliers (MM) is the most widely used algorithm for solving constrained nonlinear programing problems and it guarantees convergence to a local minimum. In contrast to ADMM, each MM iteration requires *joint* minimization of the augmented Lagrangian with respect to the primal variables x and z,

$$(x^{k+1}, z^{k+1}) = \underset{x,\, z}{\operatorname{argmin}}\ \mathcal{L}_\mu(x,\, z;\, y^k) \tag{2.7a}$$

$$y^{k+1} = y^k + \tfrac{1}{\mu}\,(\mathcal{T}(x^{k+1}) - z^{k+1}). \tag{2.7b}$$

It is possible to refine MM to allow for inexact solutions to the (x, z)-minimization subproblem and adaptive updates of the penalty parameter μ. However, until now, MM has not been a feasible choice for solving (2.5) because the nonconvex and nondifferentiable (x, z)-minimization subproblem is as difficult as the original problem (2.1).

In what follows, we exploit the structure of the linear constraint in (2.5) and utilize the optimality conditions with respect to z to eliminate it from the augmented Lagrangian. This brings the (x, z)-minimization problem in (2.7) into a form that is continuously differentiable with respect to x.

2.3 MM with the proximal augmented Lagrangian

In this section, we derive the proximal augmented Lagrangian, a continuously differentiable function resulting from explicit minimization of the augmented Lagrangian over the auxiliary variable z. This brings the (x, z)-minimization problem in (2.7) into a form that is continuously differentiable with respect to x and facilitates the use of a wide suite of standard optimization tools, including the method of multipliers and the Arrow-Hurwicz-Uzawa method.

2.3.1 Derivation of the proximal augmented Lagrangian

The first main result of the paper is provided in Theorem 2.3.1. We use the proximal operator of the function g to eliminate the auxiliary optimization variable z from the augmented Lagrangian and transform (2.7a) into a tractable continuously differentiable problem.

Theorem 2.3.1. *For a proper, lower semicontinuous, and convex function function g, minimization of the augmented Lagrangian $\mathcal{L}_\mu(x,z;y)$ associated with problem* (2.5) *over (x,z) is equivalent to minimization of the* proximal augmented Lagrangian

$$\mathcal{L}_\mu(x;y) \;:=\; f(x) \;+\; M_{\mu g}(\mathcal{T}(x) \;+\; \mu y) \;-\; \tfrac{\mu}{2}\,\|y\|^2 \tag{2.8}$$

over x. Moreover, if f is continuously differentiable $\mathcal{L}_\mu(x;y)$ is continuously differentiable over x and y, and if f has a Lipschitz continuous gradient $\nabla\mathcal{L}_\mu(x;y)$ is Lipschitz continuous.

Proof. Through the completion of squares, the augmented Lagrangian $\mathcal{L}_\mu$ associated with (2.5) can be equivalently written as

$$\mathcal{L}_\mu(x,z;y) \;=\; f(x) \;+\; g(z) \;+\; \tfrac{1}{2\mu}\,\|z \;-\; (\mathcal{T}(x)+\mu y)\|^2 \;-\; \tfrac{\mu}{2}\,\|y\|^2.$$

Minimization with respect to z yields an explicit expression,

$$z^\star_\mu(x,y) \;=\; \mathbf{prox}_{\mu g}(\mathcal{T}(x) \;+\; \mu y) \tag{2.9}$$

and substitution of $z^\star_\mu$ into the augmented Lagrangian provides (2.8), i.e., $\mathcal{L}_\mu(x;y) = \mathcal{L}_\mu(x, z^\star_\mu(x,y);y)$. Continuous differentiability of $\mathcal{L}_\mu(x;y)$ follows from continuous differentiability of $M_{\mu g}$ and Lipschitz continuity of $\nabla\mathcal{L}_\mu(x;y)$ follows from Lipschitz continuity of $\mathbf{prox}_{\mu g}$ and boundedness of the linear operator $\mathcal{T}$; see (2.2c). □

Expression (2.8), that we refer to as the *proximal augmented Lagrangian*, characterizes $\mathcal{L}_\mu(x,z;y)$ on the manifold corresponding to explicit minimization over the auxilary variable z. Theorem 2.3.1 allows *joint* minimization of the augmented Lagrangian with respect to x and z, which is in general a nondifferentiable problem, to be achieved by minimizing differentiable function (2.8) over x. It thus facilitates the use of the method of multipliers as follows and the Arrow-Hurwicz-Uzawa gradient flow dynamics in Chapter 3.

2.3.2 MM based on the proximal augmented Lagrangian

Theorem 2.3.1 allows us to solve nondifferentiable subproblem (2.7a) by minimizing the continuously differentiable proximal augmented Lagrangian $\mathcal{L}_\mu(x;\,y^k)$ over x. Relative to ADMM, our customized MM algorithm guarantees convergence to a local minimum and offers systematic update rules for the parameter μ. Relative to proximal gradient, we can solve (2.1) with a general bounded linear operator $\mathcal{T}$ and can incorporate second order information about f.

Using reformulated expression (2.8) for the augmented Lagrangian, MM minimizes $\mathcal{L}_\mu(x; y^k)$ over the primal variable x and updates the dual variable y using gradient ascent with step-size $1/\mu$,

$$x^{k+1} \;=\; \underset{x}{\operatorname{argmin}}\; \mathcal{L}_\mu(x; y^k) \tag{2.10a}$$

$$y^{k+1} \;=\; y^k \;+\; \tfrac{1}{\mu}\, \nabla_y \mathcal{L}_\mu(x^{k+1}; y^k) \tag{2.10b}$$

where

$$\nabla_y \mathcal{L}_\mu(x^{k+1}; y^k) \;=\; \mathcal{T}(x^{k+1}) \;-\; \mathbf{prox}_{\mu g}(\mathcal{T}(x^{k+1}) + \mu y^k)$$

denotes the primal residual, i.e., the difference between $\mathcal{T}(x^{k+1})$ and $z^\star_\mu(x^{k+1}, y^k)$.

In contrast to ADMM, our approach does not attempt to avoid the lack of differentiability of g by fixing z to minimize over x. By constraining $\mathcal{L}_\mu(x, z; y)$ to the manifold resulting from explicit minimization over z, we guarantee continuous differentiability of the proximal augmented Lagrangian $\mathcal{L}_\mu(x; y)$. MM is a gradient ascent algorithm on the Lagrangian dual of problem (2.5) strengthened by a quadratic penalty on primal infeasibility. Since a closed-form expression of the dual is typically unavailable, MM uses the (x, z)-minimization subproblem (2.7a) to evaluate it computationally and then take a gradient ascent step (2.7b) in y. ADMM avoids this issue by solving simpler, separate subproblems over x and z. However, the x and z minimization steps (2.6a) and (2.6b) do not solve (2.7a) and thus unlike the y-update (2.7b) in MM, the y-update (2.6c) in ADMM is not a gradient ascent step on the strengthened dual. MM thus has stronger convergence results [132, 133] and may lead to fewer y-update steps.

Remark 2.3.2. *The proximal augmented Lagrangian enables MM because the x-minimization subproblem in MM 2.10a is not more difficult than in ADMM* (2.6a). *For LASSO problem* (2.4), *the z-update in ADMM* (2.6b) *is given by soft-thresholding, $z^{k+1} = \mathcal{S}_{\gamma\mu}(x^{k+1} + \mu y^k)$, and the x-update* (2.6a) *requires minimization of the quadratic function [132]. In contrast, the x-update* (2.10a) *in MM requires minimization of* $(1/2)\,\|Ax - b\|^2 + M_{\mu_k g}(x + \mu_k y^k)$, *where $M_{\mu_k g}(v)$ is the Moreau envelope associated with the ℓ_1 norm; i.e., the Huber function* (2.3b). *Although in this case the solution to* (2.6a) *can be characterized explicitly by a matrix inversion, this is not true in general. The computational cost associated with solving either* (2.6a) *or* (2.10a) *using first-order methods scales at the same rate.*

Algorithm

The procedure outlined in [135, Algorithm 17.4] allows minimization subproblem (2.10a) to be inexact, provides a method for adaptively adjusting μ_k, and describes a more refined update of the Lagrange multiplier y. We incorporate these refinements into our proximal augmented Lagrangian algorithm for solving (2.5). In Algorithm 1, η and ω are convergence tolerances, and $\mu_{\min}$ is a minimum value of the parameter μ. Because of the equivalence established in Theorem 2.3.1, convergence to a local minimum follows from the convergence results for the standard method of multipliers [135].

Algorithm 1 Proximal augmented Lagrangian algorithm for (2.5).

input: Initial point x^0 and Lagrange multiplier y^0
initialize: $\mu_0 = 10^{-1}$, $\mu_{\min} = 10^{-5}$, $\omega_0 = \mu_0$, and $\eta_0 = \mu_0^{0.1}$
for $k = 0, 1, 2, \ldots$
  Solve 2.10a such that
$$\|\nabla_x \mathcal{L}_\mu(x^{k+1}, y^k)\| \leq \omega_k$$
  if $\|\nabla_y \mathcal{L}_{\mu_k}(x^{k+1}; y^k)\| \leq \eta_k$
    if $\|\nabla_y \mathcal{L}_{\mu_k}(x^{k+1}; y^k)\| \leq \eta$ and $\|\nabla_x \mathcal{L}_\mu(x^{k+1}, y^k)\| \leq \omega$
      stop with approximate solution x^{k+1}
    else
      $y^{k+1} = y^k + \frac{1}{\mu_k} \nabla_y \mathcal{L}_{\mu_k}(x^{k+1}; y^k)$, $\quad \mu_{k+1} = \mu_k$
      $\eta_{k+1} = \eta_k \, \mu_{k+1}^{0.9}$, $\quad \omega_{k+1} = \omega_k \, \mu_{k+1}$
    endif
  else
    $y^{k+1} = y^k$, $\quad \mu_{k+1} = \max(\mu_k/5, \mu_{\min})$
    $\eta_{k+1} = \mu_{k+1}^{0.1}$, $\quad \omega_{k+1} = \mu_{k+1}$
  endif
endfor

2.3.3 Minimization of the proximal augmented Lagrangian over x

MM based on the proximal augmented Lagrangian alternates between minimization of $\mathcal{L}_\mu(x; y)$ with respect to x (for fixed values of μ and y) and the update of μ and y. Since $\mathcal{L}_\mu(x; y)$ is once continuously differentiable, many techniques can be used to find a solution to subproblem (2.10a). We next summarize three of them.

Gradient descent

The gradient of the proximal augmented Lagrangian with respect to x is given by,

$$\nabla_x \mathcal{L}_\mu(x;y) \;=\; \nabla f(x) \;+\; \tfrac{1}{\mu}\mathcal{T}^\dagger(\mathcal{T}(x) \;+\; \mu y \;-\; \mathbf{prox}_{\mu g}(\mathcal{T}(x) \;+\; \mu y)).$$

Gradient descent with backtracking rules, such as the Armijo rule, can be used to find a solution to (2.10a).

Proximal gradient

Gradient descent does not exploit the structure of the Moreau envelope of the function g; in some cases, it may be advantageous to use proximal operator associated with the Moreau envelope to solve 2.10a. In particular, when $\mathcal{T} = I$, this requires evaluation of the proximal operator of the Moreau envelope,

$$\mathbf{prox}_{\alpha M_{\mu g}}(v) \;:=\; \underset{x}{\operatorname{argmin}}\; M_{\mu g}(x) \;+\; \tfrac{1}{2\alpha}\,\|x \;-\; v\|^2.$$

Since $M_{\mu g}$ is continuously differentiable, the optimality conditions are given by

$$\begin{aligned} 0 \;&=\; \nabla M_{\mu g}(x) \;+\; \tfrac{1}{\alpha}\,(x \;-\; v) \\ &=\; \tfrac{1}{\mu}\,(x \;-\; \mathbf{prox}_{\mu g}(x)) \;+\; \tfrac{1}{\alpha}\,(x \;-\; v). \end{aligned}$$

Thus, $\mathbf{prox}_{\alpha M_{\mu g}}(v) = x^*$ where x^* satisfies,

$$x^* \;=\; \tfrac{1}{\mu+\alpha}\left(\alpha\,\mathbf{prox}_{\mu g}(x^*) \;+\; \mu\, v\right).$$

For separable g, the proximal operators associated with the function g and its Moreau envelope are easily computable. For example, the proximal operator of the ℓ_1 norm is determined by soft-thresholding (2.3a) and the ith element of $\mathbf{prox}_{\alpha M_{\mu g}}(v)$ is given by

$$\mathbf{prox}_{\alpha M_{\mu g}}(v_i) \;=\; \begin{cases} \frac{\mu}{\mu+\alpha}\, v_i, & |v_i| \;\leq\; \mu \;+\; \alpha \\ v_i \;-\; \alpha\,\mathrm{sign}(v_i), & |v_i| \;\geq\; \mu \;+\; \alpha. \end{cases}$$

In [118], proximal gradient methods were used for subproblem (2.10a) to solve the sparse feedback synthesis problem described in Section 2.5. For associated problem (2.1), computational savings were shown relative to standard proximal gradient and ADMM.

Quasi-Newton method

When the proximal operator associated with g is continuously differentiable, $x \in \mathbb{R}^m$, and $\mathcal{T}(x) = Tx$, the Hessian of the proximal augmented Lagrangian is given by,

$$\nabla_{xx}\mathcal{L}_\mu(x;y) \;=\; \nabla^2 f(x) \;+\; \tfrac{1}{\mu}\,T^T(I \,-\, B_p)\,T$$

where $(1/\mu)(I-B_p)$ is the Hessian of the Moreau envelope and B_p is the Jacobian of $\mathbf{prox}_{\mu g}(x+\mu y)$. Although $\mathbf{prox}_{\mu g}$ is not differentiable in general, it is Lipschitz continuous and therefore differentiable almost everywhere [138]. When $\mathbf{prox}_{\mu g}$ is not differentiable, the Dini derivative or the Clarke subgradient [138, 139] can be used to obtain a generalized Jacobian B_p. For the soft-thresholding operator (2.3a), $B_p = \mathrm{diag}\{\beta\}$, where the ith component of the vector β is $\beta_i \in \{0,\, |v_i| < \mu;\; 1,\, |v_i| > \mu;\; \{0,1\},\, |v_i| = \mu\}$. To improve computational efficiency, we employ the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method [135, Algorithm 7.4] to estimate the Hessian $\nabla_{xx}\mathcal{L}_\mu(x;y^k)$.

L-BFGS estimates the Hessian inverse, H_l, to compute the search direction via $r = -H_l(\nabla\mathcal{L}_\mu)$. Instead of explictly forming H_l, it computes r via a low-rank operation $q = -S_l(\nabla\mathcal{L}_\mu)$ followed by an easily-computable full-rank operation and another low-rank operation, $r = S_l^\dagger(H_l^0(q))$. The operations S_l and $S_l^\dagger$ encode updates to an initial Hessian-inverse approximation H_l^0 based on first-order information at previous iterates. The initial estimate of the Hessian inverse can be adaptively updated and is commonly taken to be the scaling operation, $H_l^0(v) = \kappa v$ where [135]

$$\kappa \;=\; \frac{\left\langle x^l \,-\, x^{l-1},\, \nabla\mathcal{L}_\mu(x^l,y) \,-\, \nabla\mathcal{L}_\mu(x^{l-1},y)\right\rangle}{\|\nabla\mathcal{L}_\mu(x^l,y) \,-\, \nabla\mathcal{L}_\mu(x^{l-1},y)\|^2}. \tag{2.11}$$

Remark 2.3.3. *For regularization functions that do not admit simply computable proximal operators,* $\mathbf{prox}_{\mu g}$ *has to be evaluated numerically by solving* (2.2a)*. If this is expensive, the primal-descent dual-ascent algorithm of Chapter 3 offers an appealing alternative because it requires one evaluation of* $\mathbf{prox}_{\mu g}$ *per iteration. When the regularization function* g *is nonconvex, the proximal operator may not be single-valued and the Moreau envelope may not be continuously differentiable. In spite of this, the convergence of proximal algorithms has been established for nonconvex, proper, lower semicontinuous regularizers that obey the Kurdyka-Łojasiewicz inequality [140].*

Algorithm 2 Limited memory BFGS (L-BFGS)

input: Gradient $\nabla\mathcal{L}_\mu(x^l;y)$, estimate of Hessian inverse H_l^0, and history $s^l = x^{l+1} - x^l$, gradients $t^l = \nabla\mathcal{L}_\mu(x^{l+1};y) - \nabla\mathcal{L}_\mu(x^l;y)$ and $\rho^l = 1/\left\langle t^l,\, s^l\right\rangle$ for last n_l iterations

output: Search direction $r = H_l(\nabla\mathcal{L}_\mu(x^l;y))$

$q = \nabla\mathcal{L}_\mu(x^l;y)$

for $i = l-1,\ldots,l-n_l$

$\alpha^i = \rho^i\left\langle s^i,\, q\right\rangle$

$q = q - \alpha^i t^i$

endfor

$r = H_l^0(q)$

for $i = l-n_l,\ldots,l-1$

$\tau = \rho^i\left\langle t^i,\, r\right\rangle$

$r = r + s^i(\alpha^i - \tau)$

endfor

2.4 Example: Edge addition in directed consensus networks

To illustrate the utility of MM with the proximal augmented Lagrangian, we consider the problem of edge addition described in Section 1.2.1. The x minimization subproblem is solved using L-BFGS.

It is desired to minimize the $\mathcal{H}_2$ norm of the closed-loop system,

$$\dot{\psi} \;=\; -(L_p + L_x)\,\psi + d, \qquad \zeta \;=\; \begin{bmatrix} Q^{1/2} \\ -R^{1/2}L_x \end{bmatrix}\psi$$

where L_p and L_x are the Laplacians of the plant and controller networks, respectively, $R \succ 0$, and $Q := I - (1/N)\mathbb{1}\mathbb{1}^T$.

To ensure convergence of ψ to the average of the initial node values, we require that the closed-loop graph Laplacian, $L = L_p + L_x$, is balanced. This condition amounts to the linear constraint, $\mathbb{1}^T L = 0$. We express the directed graph Laplacian of the controller network as,

$$L_x \;=\; \sum_{i\neq j} L_{ij}\, w_{ij} \;=:\; \sum_{l=1}^{M} L_l\, w_l$$

where $w_{ij} \geq 0$ is the added edge weight that connects node j to node i, $L_{ij} := \mathrm{e}_i\mathrm{e}_i^T - \mathrm{e}_i\mathrm{e}_j^T$, e_i is the ith basis vector in $\mathbb{R}^n$, and the integer l indexes the edges such that $w_l = w_{ij}$ and $L_l = L_{ij}$.

For simplicity, we assume that the plant network L_p is balanced and connected. Thus, enforcing that L is balanced amounts to enforcing

$$\mathbb{1}^T L_x \;=\; \mathbb{1}^T \left(\sum_l L_l \, w_l \right) \;=\; 0.$$

This imposes a linear constraint on the vector of edge weights $w \in \mathbb{R}^M$, $Ew = 0$, where E is the incidence matrix of the directed controller network [7]. When any edge may be added to a network with N nodes, the number of potential added edges is $M = N^2 - N$.

Any vector of edge weights w that corresponds to a balanced graph can be written as $w = Tx$ where the columns of T span the nullspace of the matrix E. The matrix T can be obtained via the singular value decomposition of E and it provides a basis for the space of balanced graphs, i.e., the cycle space [7] of the controller network. Each feasible controller network can be expressed in terms of this basis,

$$L_x \;=\; \sum_l L_l \, [Tx]_l \;=\; \sum_l L_l \left[\sum_k (T\mathrm{e}_k) \, x_k \right]_l \;=:\; \sum_k \hat{L}_k \, x_k \tag{2.12a}$$

where the matrices $\hat{L}_k$ are given by,

$$\hat{L}_k \;:=\; \sum_l L_l \, [T\mathrm{e}_k]_l. \tag{2.12b}$$

Since the controller and plant networks each have an eigenvalue at zero corresponding to $\mathbb{1}$, we introduce the change of variables,

$$\begin{bmatrix} \phi \\ \bar{\psi} \end{bmatrix} \;=\; \begin{bmatrix} V^T \\ \mathbb{1}^T \end{bmatrix} \psi \quad \Leftrightarrow \quad \psi \;=\; V\phi \;+\; \bar{\psi}\,\mathbb{1} \tag{2.12c}$$

where the columns of V form an orthonormal basis of the orthogonal complement to the subspace spanned by the vector of all ones. The average mode $\bar{\psi}$ is marginally stable and decoupled from the dynamics that govern the evolution of ϕ. Since $Q = I - (1/N)\mathbb{1}\mathbb{1}^T$, $\bar{\psi}$ is not detectable from ζ and the dynamics that govern the evolution of deviations from average are,

$$\dot{\phi} \;=\; \hat{A}\,\phi \;+\; \hat{B}\,d, \qquad \zeta \;=\; \begin{bmatrix} Q^{1/2}\,V \\ -R^{1/2}\,L_x\,V \end{bmatrix} \phi \tag{2.12d}$$

where $\hat{A} := -V^T(L_p + L_x)V$ and $\hat{B} := V^T$. The square of the $\mathcal{H}_2$ norm of this system is determined by,

$$f_2(x) \;=\; \left\langle V^T(Q \,+\, L_x^T R L_x)\, V,\, P_c \right\rangle \tag{2.12e}$$

where P_c is the controllability gramian of system (2.12d),

$$0 \;=\; \hat{A}\, P_c \;+\; P_c \hat{A}^T \;+\; \hat{B}\hat{B}^T. \tag{2.12f}$$

To balance the closed-loop $\mathcal{H}_2$ performance with the number of added edges, we introduce a regularized optimization problem

$$x_\gamma \;=\; \operatorname{argmin} \;\; f_2(x) \;+\; \gamma\, \mathbb{1}^T T x \;+\; I_+(Tx). \tag{2.13}$$

Here, the regularization parameter $\gamma > 0$ specifies the emphasis on sparsity relative to the closed-loop performance f, and I_+ is the indicator function associated with the nonnegative orthant $\mathbb{R}_+^M$. When the desired level of sparsity for the vector of the added edge weights $w_\gamma = T x_\gamma$ has been attained, optimal weights for the identified set of edges are obtained by solving,

$$\underset{x}{\text{minimize}} \quad f_2(x) \;+\; I_{\mathcal{W}_\gamma}(Tx) \;+\; I_+(Tx) \tag{2.14}$$

where $\mathcal{W}_\gamma$ is the set of vectors with the same sparsity pattern as w_γ and $I_{\mathcal{W}_\gamma}$ is the indicator function associated with this set.

2.4.1 Implementation

We next derive the gradient of the square of the $\mathcal{H}_2$ norm of system (2.12d) and provide implementation details for MM with the proximal augmented Lagrangian with the regularization functions in (2.13) and (2.14).

Lemma 2.4.1. *Let a graph Laplacian of a directed plant network L_p be balanced and connected and let $\hat{A}$, $\hat{B}$, L_x, and V be as defined in* (2.12a)–(2.12d)*. The gradient of $f_2(x)$ defined in* (2.12e) *is given by,*

$$\nabla f_2(x) \;=\; 2\,\mathrm{vec}\left(\left\langle (R\, L_x V \,-\, V P_o)\, P_c V^T,\, \hat{L}_k \right\rangle\right)$$

where P_c and P_o are the controllability and observability gramians of system (2.12d)*,*

$$\begin{aligned} \hat{A}\,P_c \;+\; P_c\hat{A}^T \;+\; \hat{B}\hat{B}^T \;&=\; 0 \\ \hat{A}^T P_o \;+\; P_o\hat{A} \;+\; V^T(Q \;+\; L_x^T R L_x)\,V \;&=\; 0. \end{aligned}$$

Proof. The Lagrangian associated with the minimization of the function f_2 given by (2.12e) subject to constraint (2.12f) is

$$\left\langle V^T(Q \;+\; L_x^T R L_x)\,V,\, P_c\right\rangle \;+\; \left\langle \hat{A}\,P_c \;+\; P_c\hat{A}^T \;+\; \hat{B}\hat{B}^T,\, P_o\right\rangle$$

where P_o is the Lagrange multiplier. Variations with respect to P_o and P_c yield the Lyapunov equations for the controllability and observability gramians P_c and P_o, respectively. Using the definition of $\hat{A}$, the Lagrangian can be rewritten as,

$$\left\langle (RL_xV \;-\; 2VP_o)P_cV^T,\, L_x\right\rangle \;+\; \left\langle \hat{B}\hat{B}^T \;-\; V^TL_pVP_c \;-\; P_cV^TL_p^TV,\, P_o\right\rangle \;+\; \left\langle V^TQV,\, P_c\right\rangle.$$

Using definition (2.12a) of L_x and taking variation with respect to each x_k yields the expression for $\nabla f_2(x)$. □

The proximal operator associated with the regularization function $g_s(z) := \gamma \mathbb{1}^T z + I_+(z)$ in (2.13) is given by

$$\mathbf{prox}_{\mu g_s}(v_i) \;=\; \max\left(0,\, v_i \;-\; \gamma\mu\right).$$

The associated Moreau envelope for each element of v is

$$M_{\mu g_s}(v_i) \;=\; \begin{cases} \frac{1}{2\mu}v_i^2 & v_i \;\leq\; \gamma\mu \\ \gamma\left(v_i \;-\; \gamma\frac{\mu}{2}\right) & v_i \;>\; \gamma\mu \end{cases}$$

and its gradient is

$$\nabla M_{\mu g_s}(v_i) \;=\; \max\left(\tfrac{1}{\mu}v_i,\, \gamma\right).$$

The proximal operator associated with the regularization function in (2.14), $g_p(z) := I_{\mathcal{W}_\gamma}(z) + I_+(z)$, is a projection onto the intersection of the set $\mathcal{W}_\gamma$ and the nonnegative orthant,

$$\mathbf{prox}_{\mu g_p}(v) \;=\; \mathcal{P}_{\mathcal{E}}(v)$$

where $\mathcal{E} := \mathcal{W}_\gamma \cap \mathbb{R}^M_+$. The associated Moreau envelope is

$$M_{\mu g_p}(v) \;=\; \tfrac{1}{2\mu}\,\|v \;-\; \mathcal{P}_{\mathcal{E}}(v)\|^2$$

and its gradient is determined by

$$\nabla M_{\mu g_p}(v) \;=\; \tfrac{1}{\mu}\,(v \;-\; \mathcal{P}_{\mathcal{E}}(v)).$$

We solve (2.13) and (2.14) using Algorithm 1, where L-BFGS (i.e., Algorithm 2) is employed in the x-minimization subproblem (2.10a).

2.4.2 Computational experiments

For the plant network shown in Fig. 2.2, Fig. 2.3 illustrates the tradeoff between the number of added edges and the closed-loop $\mathcal{H}_2$ norm. The added edges are identified by computing the γ-parameterized homotopy path for problem (2.13), and the optimal edge weights are obtained by solving (2.14). The red dashed lines in Fig. 2.2 show the optimal set of 2 added edges. These are obtained for $\gamma = 3.5$ and they yield 23.91% performance loss relative to the setup in which all edges in the controller graph are used. We note that the same set of edges is obtained by conducting an exhaustive search. This suggests that the proposed convex regularizers may offer a good proxy for solving difficult combinatorial optimization problems.

We also consider simple directed cycle graphs with $N = 5$ to 50 nodes and $M = N^2 - N$ potential added edges. We solve (2.13) for $\gamma = 0.01, 0.1, 0.2$, and $R = I$ using MM with the proximal augmented Lagrangian (PAL), ADMM, and ADMM with an adaptive heuristic for updating μ [132] (ADMM μ). The x-minimization subproblems in each algorithm are solved using L-BFGS. Since $g_s(Tx)$ and $g_p(Tx)$ are not separable in x, proximal gradient cannot be used here.

Figure 2.4a shows the time required to solve problem (2.13) in terms of the total number of potential added edges; Fig. 2.4b demonstrates that PAL requires fewer outer iterations; and Fig. 2.4c illustrates that the average computation time per outer iteration is roughly equivalent for all three methods. Even with an adaptive heuristic update of μ [132], ADMM requires more outer iterations which increases overall solve time relative to the proximal augmented Lagrangian method. Thus, compared to ADMM, PAL provides computational advantage by reducing the number of outer iterations (indexed by k in Algorithm 1 and in (2.6)).

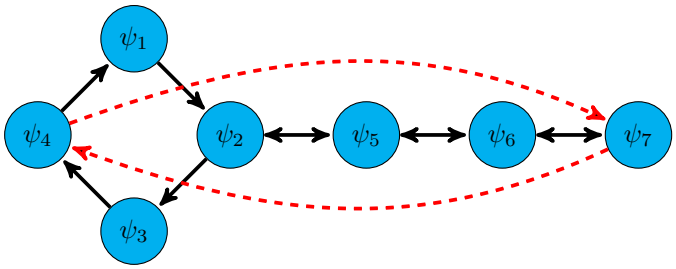


Figure 2.2: A balanced plant graph with 7 nodes and 10 directed edges (solid black lines). A sparse set of 2 added edges (dashed red lines) is identified by solving (2.13) with $\gamma = 3.5$ and $R = I$.

2.5 Example: Sparsity-promoting optimal control

Here, we apply the method of multipliers algorithm to apply a sparsity-promoting penalty on the optimal state-feedback design problem described in Section 1.2.1. As described in Section 1.3, sparsity of the feedback gain associated with a distributed system corresponds to a controller with a distributed architecture. We use the proximal gradient method with BB step-size selection to solve the $\mathcal{L}_\mu$ minimization subproblem.

The design of state-feedback controllers which balance performance with sparsity has been the subject of considerable attention in recent years [62, 64, 98–102, 104, 105, 141]. We consider the sparsity-promoting optimal control problem applied to the linear system,

$$\dot{\psi} \;=\; (A \,-\, B_2 X)\,\psi \,+\, B_1\, d, \qquad\qquad \zeta \;=\; \begin{bmatrix} Q^{1/2} \\ -\,R^{1/2} X \end{bmatrix} \psi \tag{2.15}$$

where d is an exogenous disturbance, ζ is the performance output, and $Q = Q^T \succeq 0$ and $R = R^T \succ 0$ are the state and control performance weights. System (2.15) describes closed-loop dynamics under the state-feedback control law,

$$u \;=\; -X\psi,$$

We make the standard assumptions that (A, B_2) is stabilizable and $(A, Q^{1/2})$ is detectable.

The state-feedback gain X which minimizes the closed-loop $\mathcal{H}_2$ norm is, in general, a dense

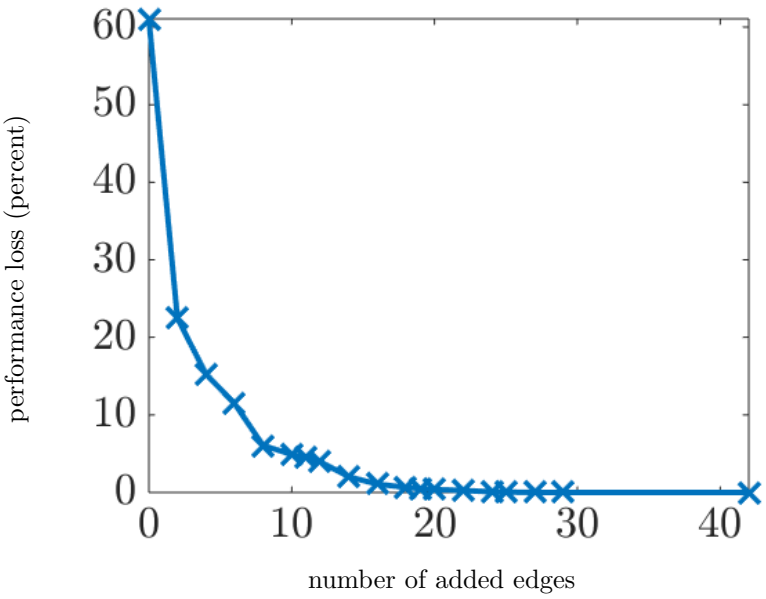
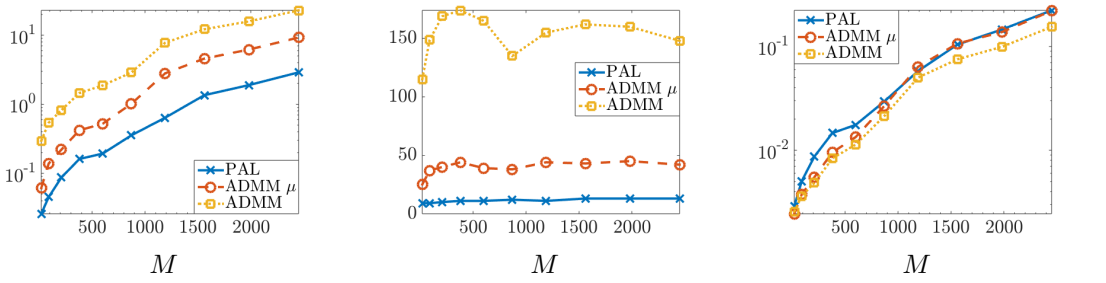


Figure 2.3: Tradeoff between performance and sparsity resulting from the solution to (2.13)-(2.14) for the network shown in Fig. 2.2. Performance loss is measured relative to the optimal centralized controller (i.e., the setup in which all edges in the controller network are used).



(a) Total solve time (s) (b) Number of outer iterations (c) Solve time (s) per outer iteration

Figure 2.4: (a) Total time; (b) number of outer iterations; and (c) average time per outer iteration required to solve (2.13) with $\gamma = 0.01, 0.1, 0.2$ for a cycle graph with $N = 5$ to 50 nodes and $m = 20$ to 2450 potential added edges using PAL (**solid blue** $-\times-$), ADMM (**dashed red** - -○- -), and ADMM with the adaptive μ-update heuristic (**dotted yellow** $\cdots\square\cdots$). PAL requires fewer outer iterations and thus a smaller total solve time.

matrix. In [62, 64], the authors studied the problem of designing feedback gain matrices which balance $\mathcal{H}_2$ performance with the sparsity of X. This was achieved by considering a regularized optimal control problem (2.1) where $f = f_2$ is the closed-loop $\mathcal{H}_2$ norm, $g(X)$ encodes some structural constraint or penalty on X, and $\gamma > 0$ encodes the emphasis on this penalty relative to the $\mathcal{H}_2$ performance. In particular we consider regularization by the ℓ_1 norm,

$$\text{minimize} \quad f_2(X) \; + \; \gamma \|X\|_1$$

where f_2 is the $\mathcal{H}_2$ norm (1.3a) of system (2.15),

$$f_2(X) \; = \; \text{trace}(P_c(Q + X^T R X))$$

and its gradient with respect to X is given by

$$\nabla f_2(X) \; = \; 2(RX \; - \; B_2^T P_o) P_c$$

where P_o and P_c are observability and controllability gramians of the closed-loop system,

$$\begin{aligned} A_{\text{cl}}^T P_o \; + \; P_o A_{\text{cl}} \quad &= \quad -(Q \; + \; X^T R X) \\ A_{\text{cl}} P_c \; + \; P_c A_{\text{cl}}^T \quad &= \quad -B_1 B_1^T \end{aligned}$$

and $A_{\text{cl}} := A \; - \; B_2 X$. Furthermore, $\nabla f_2(X)$ is a Lipschitz continuous function on the set of stabilizing feedback gains [142].

For $\gamma = 0$, the problem simplifies to the $\mathcal{H}_2$ state-feedback problem whose solution is given by the standard linear quadratic regulator. A typical approach is to solve (2.1) for a series of different γ and to generate a set of feedback gains with different levels of sparsity. From this set, a sparse feedback gain can be selected or γ can be refined to yield sparser or denser controllers.

The proximal augmented Lagrangian associated with this regularized problem is

$$f_2(X) \; + \; \gamma h_\kappa(X \; + \; \mu Y) \; - \; \tfrac{\mu}{2} \|Y\|_F^2$$

where $\kappa := \mu\gamma$ and the Huber function

$$h_\kappa(V) \; = \; \sum_{i,j} \begin{cases} \frac{1}{2} V_{ij}^2, & |V_{ij}| \; \leq \; \kappa \\ \kappa \, (|V_{ij}| \; - \; \frac{1}{2} \kappa), & |V_{ij}| \; \geq \; \kappa \end{cases}$$

is the Moreau envelope of the ℓ_1 norm.

2.5.1 Minimization of the proximal augmented Lagrangian

The main computational burden in the method of multipliers lies in finding a solution to the optimization problem (2.10a),

$$\underset{X}{\text{minimize}} \quad f_2(X) \; + \; \gamma h_\kappa(X \; + \; \mu Y^k).$$

Although the differentiability of $\mathcal{L}_\mu$ implies that gradient descent may be employed to update X, we utilize the proximal gradient method to exploit the structure of the Moreau envelope h_κ. We use the notation $\{X^l\}$ to denote the sequence of inner iterates that converge to a solution of (2.10a).

Proximal gradient step for minimizing $\mathcal{L}_\mu$

Proximal gradient descent, described in Section 2.2.2, provides a generalization of standard gradient descent which can be applied to nonsmooth optimization problems. Here, we apply it to solve subproblm (2.10a) in MM with the proximal augmented Lagrangian. The proximal gradient update is given by $X^{l+1} = X^l + \widetilde{X}$ where $\widetilde{X}$ minimizes

$$\tfrac{1}{2\alpha}\|\widetilde{X}\|_F^2 \; + \; \left\langle \nabla f_2(X^l), \, \widetilde{X} \right\rangle \; + \; \gamma \, h_\kappa(X^l \; + \; \widetilde{X} \; + \; \mu Y)$$

over $\widetilde{X}$ and α is a step size [130]. Note that this problem is separable over the elements of $\widetilde{X}$. By defining $a := \widetilde{X}_{ij}$, $b := (\nabla f(X^l))_{ij}$, and $c := (X^l + \mu Y^k)_{ij}$, optimization over each element of $\widetilde{X}$ can be expressed as,

$$\underset{a}{\text{minimize}} \quad \tfrac{1}{2\alpha} \, a^2 \; + \; b \, a \; + \; \gamma \, h_\kappa(a + c).$$

Setting the gradient to zero yields, $a + \alpha(b + \gamma \, \text{sat}_\kappa(a + c)) = 0$ and considering the separate cases when $\text{sat}_\kappa(a + c) = \kappa$, $a + c$, and $-\kappa$ yields the optimal a,

$$a^\star \; = \; \begin{cases} -\alpha(b \; + \; \gamma\kappa), & \alpha b - c \; \geq \; \kappa(\alpha\gamma \; + \; 1) \\ -\dfrac{\alpha}{1 \; + \; \alpha\gamma}(b \; + \; \gamma c), & |\alpha b - c| \leq \; \kappa(\alpha\gamma \; + \; 1) \\ -\alpha(b \; - \; \gamma\kappa), & \alpha b - c \; \leq -\kappa(\alpha\gamma \; + \; 1). \end{cases}$$

Substituting values for a, b, and c yields the proximal gradient update.

Step-size selection

Since the objective function is not smooth, an Armijo backtracking rule cannot be used. Instead, we backtrack from $\alpha_{l,0}$ by selecting the smallest nonnegative integer r such that $\alpha_l = c^r \alpha_{l,0}$ with $c \in (0,1)$ such that X^{l+1} is stabilizing and satisfies a descent condition based on the quadratic approximation

$$f_2(X^{l+1}) \;\leq\; f_2(X^l) \;+\; \left\langle \nabla f_2(X^l),\, X^{l+1} \,-\, X^l \right\rangle \;+\; \tfrac{1}{2\alpha_l} \left\| X^{l+1} \,-\, X^l \right\|_F^2.$$

When ∇f_2 is Lipschitz continuous and $\alpha < \frac{1}{L_f}$ where L_f is a Lipschitz constant of ∇f_2, this condition is always satisfied. This backtracking rule adaptively estimates the Lipschitz constant of $\nabla f(X)$ to ensure convergence [131].

To improve the speed of the proximal gradient algorithm, we initialize the step-size using the Barzilai-Borwein (BB) method [143],

$$\alpha_{l,0} \;=\; \frac{\|X^l \,-\, X^{l-1}\|_F^2}{\langle X^{l-1} \,-\, X^l,\, \nabla f_2(X^{l-1}) \,-\, \nabla f(X^l) \rangle}.$$

Figure 2.5 illustrates the utility of the proximal gradient method over standard gradient descent and the advantage of BB step-size initialization.

2.5.2 Proximal gradient applied directly to (2.1)

For this problem, it is also possible to solve (2.1) directly using proximal gradient descent. This algorithm is guaranteed to converge to a local optimal point [144], but we find that in practice it takes longer to find a solution than the method of multipliers. The proximal operator for the weighted ℓ_1-norm is the elementwise softhresholding operator defined in (2.3a) and the update for solving (2.1) directly is given by

$$X^{k+1} \;=\; \mathcal{S}_\beta\big(X^k \,-\, \alpha_k \nabla f_2(X^k)\big)$$

where $\beta := \gamma \alpha_k$ and α_k is the step-size. The backtracking and BB step-size initialization rules described in Section 2.5.1 are also used here.

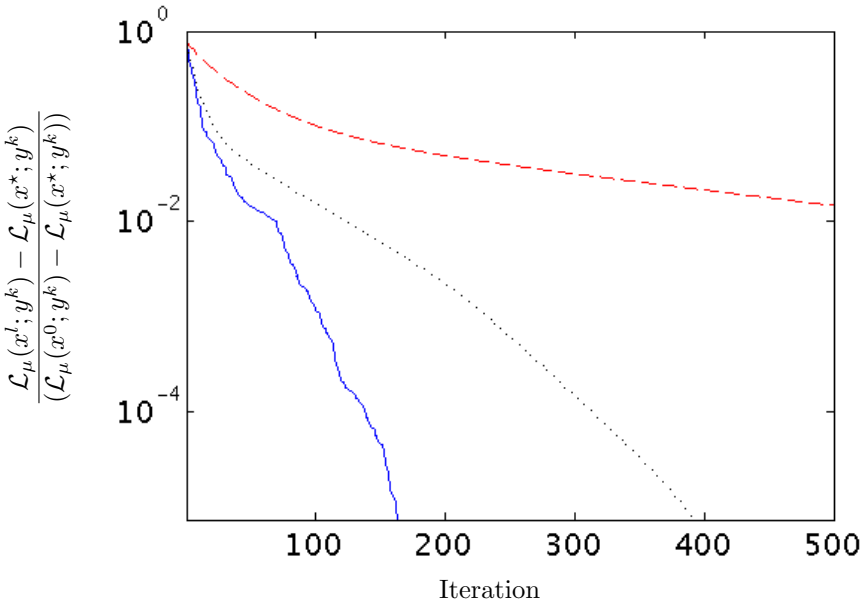


Figure 2.5: Comparison of proximal gradient with BB step-size selection (**solid blue** —), proximal gradient without BB step-size selection (**dotted black** $\cdots$) and gradient descent with BB step-size selection (**dashed red - - -**) for the x-minimization step (2.10a) for an unstable network with 20 subsystems, $\gamma = 0.0844$, and $\mu = 0.10$. The y-axis shows the distance from the optimal objective value relative to initial distance.

2.5.3 Numerical experiments

We next illustrate the utility of our approach using two examples. We compare our method of multipliers algorithm with the ADMM algorithm from [64] and a direct application of the proximal gradient method.

Mass-spring system

Consider a series of N masses connected by linear springs. The dynamics of each mass are described by

$$\ddot{p}_i \;=\; -\left(p_i \;-\; p_{i+1}\right) \;-\; \left(p_i \;-\; p_{i-1}\right) \;+\; d_i \;+\; u_i$$

where p_i is the position of the ith mass. When the first and last masses are affixed to rigid bodies, the aggregate dynamics are given by

$$\begin{bmatrix} \dot{p} \\ \dot{v} \end{bmatrix} \;=\; \begin{bmatrix} 0 & I \\ -T & 0 \end{bmatrix} \begin{bmatrix} p \\ v \end{bmatrix} \;+\; \begin{bmatrix} 0 \\ I \end{bmatrix} d \;+\; \begin{bmatrix} 0 \\ I \end{bmatrix} u$$

where p, v, and d are the position, velocity and disturbance vectors, and T is a Toeplitz matrix with 2 on the main diagonal and -1 on the first super- and sub-diagonals.

In Figure 2.6, we compare the time required to compute a series of sparse feedback gains for 10 values of γ, linearly spaced between 0.001 and 1.0. Taking $\gamma = 1.0$ corresponds to roughly 6% nonzero elements in the feedback gain matrix.

Among the three algorithms, ADMM is the fastest; however, the method of multipliers is comparable and scales at the same rate. Direct proximal gradient was the slowest and exhibited the worst scaling. Since the mass-spring system has benign dynamics, we next consider an unstable network.

Unstable network

Let N nodes be uniformly randomly distributed in a box. Each node is an unstable second order system coupled with nearby nodes via an exponentially decaying function of the Euclidean distance $\delta(i,j)$ between them [60],

$$\begin{bmatrix} \dot{\psi}_{1i} \\ \dot{\psi}_{2i} \end{bmatrix} \;=\; \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \psi_{1i} \\ \psi_{2i} \end{bmatrix} \;+\; \sum_{j \neq i} \mathrm{e}^{-\delta(i,j)} \begin{bmatrix} \psi_{1j} \\ \psi_{2j} \end{bmatrix} \;+\; \begin{bmatrix} 0 \\ 1 \end{bmatrix} \left(d_i \;+\; u_i\right)$$

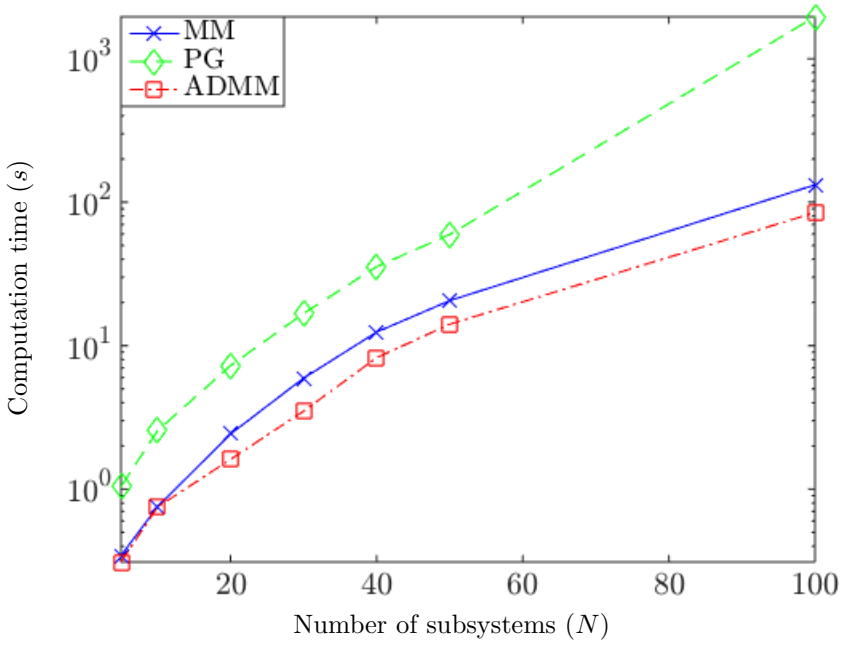


Figure 2.6: Computation time required to solve (2.1) for 10 evenly spaced values of γ from 0.001 to 1.0 for a mass-spring example with $N = 5, 10, 20, 30, 40, 50, 100$ masses. Performance of direct proximal gradient (**dashed green** $- - \diamond - -$), the method of multipliers (**solid blue** $—\times—$) and ADMM (**dash-dot red** $- \cdot \square - \cdot$) is displayed. All algorithms use BB step-size initialization.

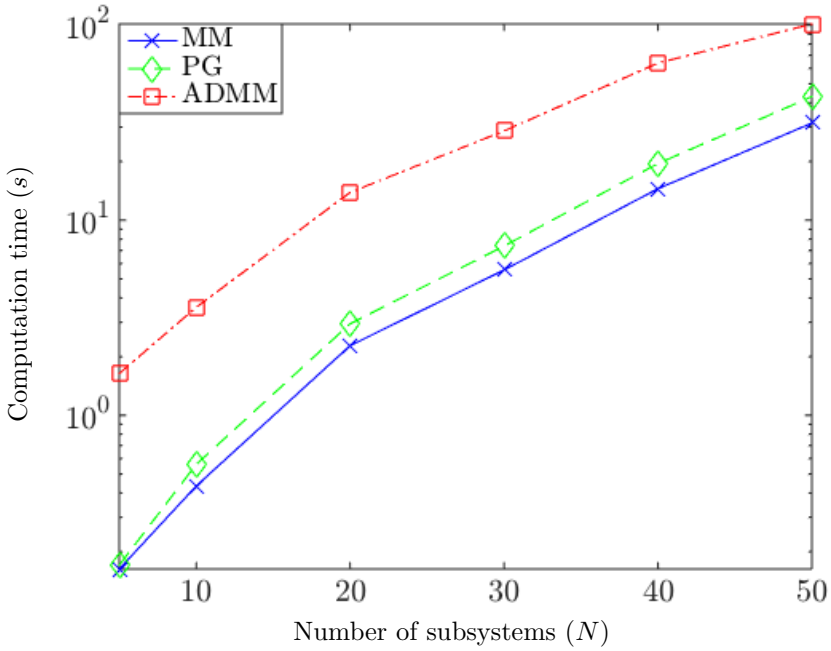


Figure 2.7: Computation time required to solve (2.1) for 20 evenly spaced values of γ from 0.001 to 0.05 for unstable network examples with $N = 5, 10, 20, 30, 40, 50$ nodes. Performance of direct proximal gradient (**dashed green** $- - \diamond - -$), the method of multipliers (**solid blue** $—\times—$) and ADMM (**dash-dot red** $- \cdot \square - \cdot$) is displayed. All algorithms use BB step-size initialization.

where Q and R are taken to be the identity. Note that simple truncation of the centralized controller could result in a non-stabilizing feedback matrix [60]. We solve (2.5) for γ varying from 0.001 to 0.05 in 20 linearly spaced increments. On average, $\gamma = 0.05$ corresponds to approximately 25% nonzero entries in the feedback gain matrix.

Computation times for N varying from 5 to 50, are shown in Figure 2.7. Since networks are randomly generated, we average the computation time for 5 networks of each size. For this more complicated example, the method of multipliers algorithm is the fastest and ADMM is the slowest.

Chapter 3

First-order primal-dual algorithm

We examine Arrow-Hurwicz-Uzawa gradient flow dynamics for the proximal augmented Lagrangian. Such dynamics can be used to identify saddle points of the Lagrangian [145] and have enjoyed recent renewed interest in the context of networked optimization because, in many cases, the gradient can be computed in a distributed manner [146]. Our approach yields a dynamical system with a continuous right-hand side for a broad class of nonsmooth optimization problems. This is in contrast to existing techniques which employ subgradient methods [147] or use discontinuous projected dynamics [148–150] for problems with inequality constraints. Furthermore, since the proximal augmented Lagrangian is neither strictly concave nor linear in the dual variable, we require additional developments relative to recent advances [151] to establish convergence.

3.1 Arrow-Hurwicz-Uzawa gradient flow

In this section, we apply the continuous-time Arrow-Hurwicz-Uzawa gradient flow [145] to the proximal augmented Lagrangian (2.8). Simultaneous update of the primal and dual variables in

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}_\mu(x;y) \\ \nabla_y \mathcal{L}_\mu(x;y) \end{bmatrix} \tag{3.1}$$

should be compared and contrasted with the approach presented in Section 2.3 where we alternate between minimization of $\mathcal{L}_\mu(x;y)$ over x and gradient ascent over y. For convex, differentiable f and convex g, we show that the gradient flow dynamics (3.1) globally converge to the

set of saddle points of the proximal augmented Lagrangian.

We first characterize the optimal primal-dual pairs of optimization problem (2.5) with the Lagrangian,

$$f(x) \; + \; g(z) \; + \; \langle y, \mathcal{T}(x) \; - \; z \rangle \, .$$

The associated first-order optimality conditions are given by,

$$\begin{aligned} 0 \quad &= \quad \nabla f(x^\star) \; + \; \mathcal{T}^\dagger(y^\star) && \text{(3.2a)} \\ 0 \quad &\in \quad \partial g(z^\star) \; - \; y^\star && \text{(3.2b)} \\ 0 \quad &= \quad \mathcal{T}(x^\star) \; - \; z^\star && \text{(3.2c)} \end{aligned}$$

where ∂g is the subgradient of g. Clearly, these are equivalent to the optimality condition for (2.1), $0 \in \nabla f(x^\star) + \mathcal{T}^\dagger(\partial g(\mathcal{T}(x^\star)))$.

Remark 3.1.1. *Since the proximal augmented Lagrangian* (2.8) *is neither linear nor strictly concave in* y*, we cannot use [151] to show global convergence. For LASSO problem* (2.4)*,* $\mathbf{prox}_{\mu g}(x+\mu y)$ *is differentiable at points where* $|x_i+\mu y_i| \neq \gamma\mu$ *for all* i*. When it is differentiable,* $\mathcal{L}_\mu(x;y)$ *is locally quadratic in* y *and* $\nabla_{yy}\mathcal{L}_\mu(x;y) = -\mu \, \mathrm{diag}(\beta)$ *where the* i*th element* $\beta_i = 1$ *if* $|x_i + \mu y_i| > \gamma\mu$ *and* 0 *if* $|x_i + \mu y_i| < \gamma\mu$*. Since* $\nabla_{yy}\mathcal{L}_\mu(x;y) \neq 0$*, the proximal augmented Lagrangian* (2.8) *is not linear in* y*. Furthermore, since* $\nabla_{yy}\mathcal{L}_\mu(x;y)$ *is not negative definite,* $\mathcal{L}_\mu(x;y)$ *is not strictly concave in* y*.*

To show convergence, we introduce a simple lemma about proximal operators which follows almost directly from their definition. Even though we state the result for $x \in \mathbb{R}^m$, g: $\mathbb{R}^M \to \mathbb{R}$ and $\mathcal{T}(x) = Tx$ where $T \in \mathbb{R}^{M \times m}$ is a given matrix, the proof for x in a Hilbert space and a bounded linear operator $\mathcal{T}$ follows from similar arguments.

Lemma 3.1.2. *Let* g*:* $\mathbb{R}^M \to \mathbb{R}$ *be a proper, lower semicontinuous, convex function and let* $\mathbf{prox}_{\mu g}$*:* $\mathbb{R}^M \to \mathbb{R}^M$ *be the corresponding proximal operator. Then, for any* $a, b \in \mathbb{R}^M$*, we can write*

$$\mathbf{prox}_{\mu g}(a) \; - \; \mathbf{prox}_{\mu g}(b) \; = \; D\,(a \; - \; b) \tag{3.3}$$

where D *is a symmetric matrix satisfying* $I \succeq D \succeq 0$*.*

Proof. Let $\tilde{c} := a - b$, $\tilde{p} := \mathbf{prox}_{\mu g}(a) - \mathbf{prox}_{\mu g}(b)$, and $D := \{\tilde{p}\,\tilde{p}^T/(\tilde{p}^T\tilde{c}),\ \tilde{p} \neq 0;\ 0,\ \text{otherwise}\}$. Then, by construction, $D = D^T \succeq 0$ and (3.3) can be written as $\tilde{p} = D\tilde{c}$. Since $\mathbf{prox}_{\mu g}$ is firmly nonexpansive [130], $\tilde{p}^T\tilde{c} \geq \|\tilde{p}\|^2$ or, equivalently, $\tilde{c}^T D(I - D)\tilde{c} \geq 0$ for every $\tilde{c} \in \mathbb{R}^M$. Positive semi-definiteness of $I - D$ thus follows from $D \succeq 0$ and commutativity of D and $I - D$. $\square$

Theorem 3.1.3. *Let f be a continuously differentiable convex function, and let g be a proper, lower semicontinuous, convex function. Then, for the primal-descent dual-ascent gradient flow dynamics* (3.1),

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\left(\nabla f(x) + T^T \nabla M_{\mu g}(Tx + \mu y)\right) \\ \mu \left(\nabla M_{\mu g}(Tx + \mu y) - y\right) \end{bmatrix} \tag{3.4}$$

the set of optimal primal-dual pairs $(x^\star, y^\star)$ of (2.5) *is globally asymptotically stable, and $x^\star$ is an optimal solution of* (2.1).

Proof. We introduce a change of coordinates $\tilde{x} := x - x^\star$, $\tilde{y} := y - y^\star$ and a Lyapunov function candidate,

$$V(\tilde{x}, \tilde{y}) = \tfrac{1}{2} \langle \tilde{x}, \tilde{x} \rangle + \tfrac{1}{2} \langle \tilde{y}, \tilde{y} \rangle$$

where $(x^\star, z^\star; y^\star)$ is an optimal solution to (2.5) that satisfies (3.2). The dynamics in the $(\tilde{x}, \tilde{y})$-coordinates are given by,

$$\begin{bmatrix} \dot{\tilde{x}} \\ \dot{\tilde{y}} \end{bmatrix} = \begin{bmatrix} -(\nabla f(x) - \nabla f(x^\star) + \frac{1}{\mu} T^T \tilde{m}) \\ \tilde{m} - \mu \tilde{y} \end{bmatrix} \tag{3.5}$$

where

$$\tilde{m} := \mu \left(\nabla M_{\mu g}(Tx + \mu y) - \nabla M_{\mu g}(Tx^\star + \mu y^\star)\right). \tag{3.6a}$$

Using expression (3.3) to construct D such that,

$$D(T\tilde{x} + \mu\tilde{y}) = \mathbf{prox}_{\mu g}(Tx + \mu y) - \mathbf{prox}_{\mu g}(Tx^\star + \mu y^\star) \tag{3.6b}$$

and definition (2.2c) of $\nabla M_{\mu g}$, we can write

$$\tilde{m} = (I - D)(T\tilde{x} + \mu\tilde{y}). \tag{3.6c}$$

Thus, the derivative of V along the solutions of (3.5) is given by

$$\dot{V}(\tilde{x}, \tilde{y}) = -\langle \tilde{x}, \nabla f(x) - \nabla f(x^\star) \rangle - \tfrac{1}{\mu} \langle T\tilde{x}, (I - D) T\tilde{x} \rangle - \mu \langle \tilde{y}, D\tilde{y} \rangle .$$

Since f is convex, $\langle \tilde{x}, \nabla f(x) - \nabla f(x^\star) \rangle \geq 0$ and Lemma 3.1.2 imply $\dot{V} \leq 0$.

When $\nabla f(x) = \nabla f(x^\star)$ and when matrices D and $T^T(I - D)T$ have nontrivial kernels, it is possible that $\dot{V} = 0$ for a nonzero $\tilde{y} \in \ker\{D\}$ and $\tilde{x}$ such that $T\tilde{x} \in \ker\{(I - D)\}$. From (3.6c),

these conditions imply $\tilde{m} = \mu\tilde{y}$ and (3.5) simplifies to,

$$\dot{\tilde{x}} \;=\; -T^T\tilde{y}, \quad \dot{\tilde{y}} \;=\; 0.$$

Thus, the only invariant set for dynamics (3.5) is $\nabla f(x^\star) = \nabla f(x)$, $\tilde{y} \in \ker\{T^T\} \cap \ker\{D\}$, and $T\tilde{x} \in \ker\{I - D\}$. Global asymptotic stability of these points follows from LaSalle's invariance principle. To complete the proof, we show that any x and y that lie in this invariant set also satisfy the optimality conditions (3.2) of problem (2.5) with $z = z^\star_\mu(x, y)$ and thus that x solves problem (2.1).

Since $x^\star$ and $y^\star$ are optimal, (3.2a) implies $\nabla f(x^\star) + T^T y^\star = 0$. For any x and y that lie in the invariant set of (3.5), we can replace $\nabla f(x^\star)$ with $\nabla f(x)$ and add $T^T\tilde{y} = 0$ to obtain

$$\nabla f(x) \,+\, T^T(y^\star \,+\, \tilde{y}) \;=\; \nabla f(x) \,+\, T^T y \;=\; 0$$

which implies that x and y also satisfy (3.2a). Furthermore, $(I - D)T\tilde{x} = 0$ and $D\tilde{y} = 0$ can be combined to yield $T\tilde{x} - D(T\tilde{x} + \mu\tilde{y}) = 0$, which, by (3.6b), leads to

$$T\tilde{x} \,-\, (\mathbf{prox}_{\mu g}(Tx \,+\, \mu y) \,-\, \mathbf{prox}_{\mu g}(Tx^\star \,+\, \mu y^\star)) \;=\; 0.$$

By optimality condition (3.2b), $Tx^\star = z^\star$, and by definition (2.9), $z^\star = z^\star_\mu(x^\star, y^\star) = \mathbf{prox}_{\mu g}(Tx^\star + \mu y^\star)$. It thereby follows that $Tx - \mathbf{prox}_{\mu g}(Tx + \mu y) = 0$ and thus

$$Tx \;=\; \mathbf{prox}_{\mu g}(Tx \,+\, \mu y) \;=\; z^\star_\mu(x, y) \;=\; z$$

which implies that x and z satisfy (3.2c).

Finally, the optimality condition of the minimization problem (2.2a) that defines $\mathbf{prox}_{\mu g}(v)$ is $\partial g(z) + \frac{1}{\mu}(z - v) \ni 0$. Setting $v = Tx + \mu y$ from the expression (2.9) that characterizes the $z^\star_\mu$-manifold and $Tx = z$ by (3.2c) yields the final optimality condition (3.2b). □

We note that gradient flow dynamics (3.1) are convenient for distributed implementation. If the state vector x corresponds to the concatenated states of individual agents, x_i, the sparsity pattern of T and the structure of the gradient map $\nabla f\colon \mathbb{R}^m \to \mathbb{R}^m$ dictate the communication topology required to form $\nabla\mathcal{L}_\mu$ in (3.1). For example, if $f(x) = \sum f_i(x_i)$ is separable over the agents, then $\nabla f_i(x_i)$ can be formed locally. If in addition T^T is an incidence matrix of an undirected network with the graph Laplacian $T^T T$, each agent must share its state x_i with its neighbors and maintain dual variables y_i that correspond to its edges.

Our approach provides several advantages over existing distributed optimization algorithms. Even for problems (2.1) with non-differentiable regularizers g, a formulation based on the proximal augmented Lagrangian yields gradient flow dynamics (3.1) with a continuous right-hand side. This is in contrast to existing approaches which employ subgradient methods [147] or use discontinuous projected dynamics [148–151]. Furthemore, for problems where T is not diagonal, a distributed proximal gradient scheme cannot be implemented because the proximal operator of $g(Tx)$ may not be separable. Finally, relative to a distributed ADMM scheme, our method does not require solving an x-minimization subproblem in each iteration.

3.2 Exponential convergence for strongly convex f

In this section, we will show that for a strongly convex f with a Lipschitz continuous gradient, the dynamics described by (3.1) converge exponentially for a sufficiently large μ. Establishing exponential convergence is particularly interesting since $\mathcal{L}_\mu$ is neither strictly concave nor linear in the dual variable. We then turn our attention to discrete-time dynamics, which are directly related to an algorithmic implementation of the continuous dynamics (3.1) with a fixed step-size. For this setup, we also show exponential convergence (i.e., linear convergence in standard optimization terminology) for a suitably selected penalty-parameter μ and step-size.

To show these results, we employ the integral quadratic constraint (IQC) framework recently formulated by [152] for studying optimization algorithms as linear systems G connected via feedback with a nonlinear block corresponding to the gradient of f. To study (3.1), we introduce an additional nonlinear block corresponding to the Moreau envelope; see Fig. 3.1.

We express 3.1, or equivalently 3.4, as a linear system G connected in feedback with nonlinearities that correspond to the gradients of f and of the Moreau envelope of g; see Fig. 3.1. These nonlinearities can be conservatively characterized by IQCs. Exponential stability of G connected in feedback with *any* nonlinearity that satisfies these IQCs implies exponential convergence of 3.1 to the primal-dual optimal solution of (2.5). In what follows, we adjust the tools of [152–155] to our setup and establish exponential convergence by evaluating the feasibility of an LMI. We assume that the function f is m_f-strongly convex with an L_f-Lipschitz continuous gradient. Characterizing additional structural restrictions on f and g with IQCs may lead to tighter bounds on the rate of convergence.

In this work, we use a static filter $\eta = \Psi(u, \zeta) = [u^T \quad \zeta^T]^T$ and IQCs which correspond to Lipschitz continuity. In general, filters with nontrivial dynamics and stricter IQCs may yield tighter bounds on the exponential rate of convergence.

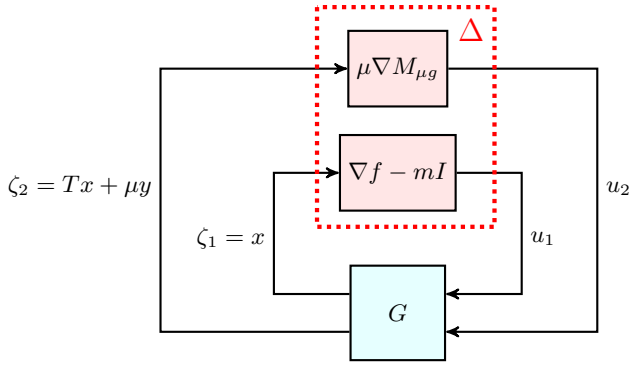


Figure 3.1: Block diagram of primal-descent dual-ascent dynamics where G is a linear system connected via feedback with nonlinearities.

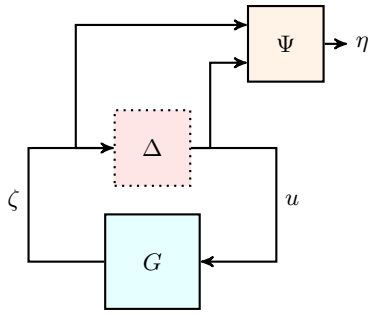


Figure 3.2: Block diagram of primal-descent dual-ascent dynamics where the nonlinearities are replaced by an IQC imposed on their inputs and outputs.

3.2.1 Continuous-time dynamics

As illustrated in Fig. 3.1, 3.4 can be expressed as a linear system G connected via feedback to a nonlinear block Δ,

$$\dot{w} \;=\; A\,w \;+\; B\,u, \quad \xi \;=\; C\,w, \quad u \;=\; \Delta(\xi)$$

$$A = \begin{bmatrix} -m_f I & \\ & -\mu I \end{bmatrix},\; B = \begin{bmatrix} -I & -\frac{1}{\mu}T^T \\ 0 & I \end{bmatrix},\; C = \begin{bmatrix} I & 0 \\ T & \mu I \end{bmatrix}$$

where $w := [x^T \; y^T]^T$, $\xi := [\xi_1^T \; \xi_2^T]^T$, and $u := [u_1^T \; u_2^T]^T$. Nonlinearity Δ maps the system outputs $\xi_1 = x$ and $\xi_2 = Tx + \mu y$ to the inputs u_1 and u_2 via $u_1 = \Delta_1(\xi_1) := \nabla f(\xi_1) - m_f \xi_1$ and $u_2 = \Delta_2(\xi_2) := \mu \nabla M_{\mu g}(\xi_2) = \xi_2 - \mathbf{prox}_{\mu g}(\xi_2)$.

When the mapping $u_i = \Delta_i(\xi_i)$ is the $\hat{L}$-Lipschitz continuous gradient of a convex function,

it satisfies the IQC [152, Lemma 6]

$$\begin{bmatrix} \xi_i - \xi_0 \\ u_i - u_0 \end{bmatrix}^T \begin{bmatrix} 0 & \hat{L}I \\ \hat{L}I & -2I \end{bmatrix} \begin{bmatrix} \xi_i - \xi_0 \\ u_i - u_0 \end{bmatrix} \geq 0 \tag{3.7}$$

where ξ_0 is some reference point and $u_0 = \Delta_i(\xi_0)$. Since f is strongly convex with parameter m_f, the mapping $\Delta_1(\xi_1)$ is the gradient of the convex function $f(\xi_1) - (m_f/2)\|\xi_1\|^2$. Lipschitz continuity of ∇f with parameter L_f implies Lipschitz continuity of $\Delta_1(\xi_1)$ with parameter $L := L_f - m_f$; thus, Δ_1 satisfies (3.7) with $\hat{L} = L$. Similarly, $\Delta_2(\xi_2)$ is the scaled gradient of the convex Moreau envelope and is Lipschitz continuous with parameter 1; thus, Δ_2 also satisfies (3.7) with $\hat{L} = 1$. These two IQCs can be combined into

$$(\eta - \eta_0)^T \Pi (\eta - \eta_0) \geq 0, \quad \eta := [\xi^T \; u^T]^T. \tag{3.8}$$

For a linear system G connected in feedback with nonlinearities that satisfy IQC (3.8), [153, Theorem 3] establishes ρ-exponential convergence, i.e., $\|w(t) - w^\star\| \leq \tau \mathrm{e}^{-\rho t}\|w(0) - w^\star\|$ for some $\tau, \rho > 0$, by verifying the existence of a matrix $P \succ 0$ such that,

$$\begin{bmatrix} A_\rho^T P + P A_\rho & PB \\ B^T P & 0 \end{bmatrix} + \begin{bmatrix} C^T & 0 \\ 0 & I \end{bmatrix} \Pi \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \preceq 0, \tag{3.9}$$

where $A_\rho := A + \rho I$. In Theorem 3.2.1, we determine a scalar condition that ensures exponential stability when TT^T is a full rank matrix.

Theorem 3.2.1. *Let f be strongly convex with parameter m_f, let its gradient be Lipschitz continuous with parameter L_f, let g be proper, lower semicontinuous, and convex, and let TT^T be full rank. Then, if $\mu \geq L_f - m_f$, there is a $\rho > 0$ such that the dynamics 3.1 converge ρ-exponentially to the optimal point of* (2.5).

Proof. Since any function that is Lipschitz continuous with parameter L is also Lipschitz continuous with parameter $\hat{L} > L$, we establish the result for $\mu = L := L_f - m_f$. We utilize [153, Theorem 3] to show ρ-exponential convergence by verifying matrix inequality (3.9) through a series of equivalent expressions (3.10). We first apply the KYP Lemma [156, Theorem 1] to (3.9) to obtain an equivalent frequency domain characterization

$$\begin{bmatrix} G_\rho(\mathrm{j}\omega) \\ I \end{bmatrix}^* \Pi \begin{bmatrix} G_\rho(\mathrm{j}\omega) \\ I \end{bmatrix} \preceq 0, \quad \forall\, \omega \in \mathbb{R} \tag{3.10a}$$

where $G_\rho(\mathrm{j}\omega) = C(\mathrm{j}\omega I - A_\rho)^{-1}B$. Evaluating the left-hand side of (3.10a) for $L = \mu$ and dividing by -2 yields the matrix inequality

$$\begin{bmatrix} \dfrac{\mu\hat{m} + \hat{m}^2 + \omega^2}{\hat{m}^2 + \omega^2}\, I & \dfrac{\hat{m}}{\hat{m}^2 + \omega^2}\, T^T \\ * & \dfrac{\hat{m}/\mu}{\hat{m}^2 + \omega^2}\, TT^T + \dfrac{\omega^2 - \rho\hat{\mu}}{\hat{\mu}^2 + \omega^2}\, I \end{bmatrix} \succ 0 \tag{3.10b}$$

where $\hat{m} := m_f - \rho > 0$ and $\hat{\mu} := \mu - \rho > 0$ so that A_ρ is Hurwitz, i.e., the system G_ρ is stable. Since the $(1,1)$ block in (3.10b) is positive definite for all ω, the matrix in (3.10b) is positive definite if and only if the corresponding Schur complement is positive definite,

$$\frac{\hat{m}/\mu}{\mu\hat{m} + \hat{m}^2 + \omega^2}\, TT^T \;+\; \frac{\omega^2 - \rho\hat{\mu}}{\hat{\mu}^2 + \omega^2}\, I \;\succ\; 0. \tag{3.10c}$$

We exploit the symmetry of TT^T to diagonalize (3.10c) via a unitary coordinate transformation. This yields m scalar inequalities parametrized by the eigenvalues λ_i of TT^T. Multiplying the left-hand side of these inequalities by the positive quantity $(\hat{\mu}^2 + \omega^2)(\mu\hat{m} + \hat{m}^2 + \omega^2)$ yields a set of equivalent, quadratic in ω^2, conditions,

$$\omega^4 \;+\; (\tfrac{\hat{m}\lambda_i}{\mu} + \hat{m}^2 + \mu\hat{m} - \rho\hat{\mu})\omega^2 \;+\; \tfrac{\hat{m}\hat{\mu}^2\lambda_i}{\mu} \;-\; \rho\hat{\mu}(\mu\hat{m} + \hat{m}^2) \;>\; 0. \tag{3.10d}$$

Condition (3.10d) is satisfied for all $\omega \in \mathbb{R}$ if there are no $\omega^2 \geq 0$ for which the left-hand side is nonpositive. When $\rho = 0$, both the constant term and the coefficient of ω^2 are strictly positive, which implies that the roots of (3.10d) as a function of ω^2 are either not real or lie in the domain $\omega^2 < 0$, which cannot occur for $\omega \in \mathbb{R}$. Finally, continuity of (3.10d) with respect to ρ implies the existence a positive ρ that satisfies (3.10d) for all $\omega \in \mathbb{R}$. □

Remark 3.2.2. *Each eigenvalue λ_i of a full rank matrix TT^T is positive and hence to estimate the exponential convergence rate it suffices to check* (3.10d) *only for the smallest λ_i. A sufficient condition for* (3.10d) *to hold for each $\omega \in \mathbb{R}$ is positivity of the constant term and the coefficient multiplying ω^2. These can be expressed as quadratic inequalities in ρ that admit explicit solutions, thereby providing an estimate of the rate of exponential convergence.*

3.2.2 Discrete time

The discrete time dynamics can be expressed as a discrete time linear system G connected via feedback to a nonlinear block Δ,

$$\begin{gathered} w^{k+1} \;=\; A\,w^k \;+\; B\,u^k, \quad \xi \;=\; C\,w^k, \quad u \;=\; \Delta(\xi) \\ A = \begin{bmatrix} \alpha_m I & \\ & \alpha_\mu I \end{bmatrix},\; B = \alpha \begin{bmatrix} -I & -\frac{1}{\mu}T^T \\ 0 & I \end{bmatrix},\; C = \begin{bmatrix} I & 0 \\ T & \mu I \end{bmatrix} \end{gathered} \tag{3.11}$$

where $\alpha_m := 1 - \alpha m_f$, $\alpha_\mu := 1 - \alpha\mu$, $w := [\,x^T\; y^T]^T$, $\xi := [\,\xi_1^T\; \xi_2^T]^T$, and $u := [\,u_1^T\; u_2^T]^T$. The nonlinearity Δ is the same as in the continuous-time case and it still satisfies IQC (3.8).

For a discrete time linear system G connected in feedback with nonlinearities that satisfy IQC (3.8), [152, Theorem 4] establishes ρ-exponential convergence, i.e., $\|w^k - w^\star\| \leq \tau\rho^k\|w(0) - w^\star\|$ for some $\tau \geq 0, \rho \in [0,1)$ is guaranteed by verifying the existence of a matrix $P \succ 0, \lambda \geq 0$ such that,

$$\begin{bmatrix} A_\rho^T P A_\rho - I & A_\rho^T P B_\rho \\ B_\rho^T P A_\rho & B_\rho^T P B_\rho \end{bmatrix} + \lambda \begin{bmatrix} C^T & 0 \\ 0 & I \end{bmatrix} \Pi \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \preceq 0, \tag{3.12}$$

where $A_\rho := A/\rho$ and $B_\rho := B/\rho$. In Theorem 3.2.3, we determine a scalar condition that ensures exponential stability when $TT^T = I$. The extension to the general case where TT^T is full rank is straightforward and the layout of the proof is meant to facilitate such development; see Remark 3.2.4.

Theorem 3.2.3. *Let f be strongly convex with parameter m_f, let its gradient be Lipschitz continuous with parameter L_f, let g be proper, lower semicontinuous, and convex, and let $TT^T = I$. Then, if $\mu \geq L_f - m_f$ and $\alpha \in (0, \min(\frac{1}{2m}, \frac{1}{2\mu}, \mu))$, there is a $\rho \in (0,1)$ such that the dynamics 3.11 converge ρ-exponentially to the optimal solution.*

Proof. Since any function that is Lipschitz continuous with parameter L is also Lipschitz continuous with parameter $\hat{L} > L$, we establish the result for $\mu = L := L_f - m_f$. We utilize [152, Theorem 4] to show ρ-exponential convergence by verifying matrix inequality (3.12) through a series of equivalent expressions (3.13). We first apply the KYP Lemma [156, Theorem 2] to (3.12) to obtain an equivalent frequency domain characterization

$$\begin{bmatrix} G_\rho(\mathrm{e}^{\mathrm{j}\omega}) \\ I \end{bmatrix}^* \Pi \begin{bmatrix} G_\rho(\mathrm{e}^{\mathrm{j}\omega}) \\ I \end{bmatrix} \preceq 0, \quad \forall\, \omega \in \mathbb{R} \tag{3.13a}$$

where $G_\rho(e^{j\omega}) = C(e^{j\omega}I - A_\rho)^{-1}B_\rho = C(\rho\, e^{j\omega}I - A)^{-1}B$. Evaluating the left-hand side of (3.13a) for $L = \mu$ and dividing by -2 yields the matrix inequality

$$\begin{bmatrix} a(\omega)I & b(\omega)T \\ * & c(\omega)I + d(\omega)TT^T \end{bmatrix} \succeq 0, \quad \forall \omega \in \mathbb{R} \tag{3.13b}$$

where

$$\begin{aligned} a(\omega) &:= 1 + \alpha\mu \frac{\alpha m - 1 + \rho\cos\omega}{\rho^2 + (\alpha m - 1)^2 + 2\rho(\alpha m - 1)\cos\omega} \\ b(\omega) &:= \alpha \frac{\alpha m - 1 + \rho\cos\omega}{\rho^2 + (\alpha m - 1)^2 + 2\rho(\alpha m - 1)\cos\omega} \\ c(\omega) &:= 1 - \alpha\mu \frac{\alpha\mu - 1 + \rho\cos\omega}{\rho^2 + (\alpha\mu - 1)^2 + 2\rho(\alpha\mu - 1)\cos\omega} \\ d(\omega) &:= \tfrac{\alpha}{\mu} \frac{\alpha m - 1\rho\cos\omega}{\rho^2 + (\alpha m - 1)^2 2\rho(\alpha m - 1)\cos\omega}. \end{aligned}$$

Since the range of $\cos\omega$ over $\omega \in \mathbb{R}$ is $[-1, 1]$, we can replace $\cos\omega$ by ζ and check the matrix inequality

$$\begin{bmatrix} a(\zeta)I & b(\zeta)T \\ * & c(\zeta)I + d(\zeta)TT^T \end{bmatrix} \succ 0, \quad \forall \zeta \in [-1, 1]$$

where

$$\begin{aligned} a(\zeta) &:= 1 + \alpha\mu \frac{\alpha m - 1 + \rho\zeta}{\rho^2 + (\alpha m - 1)^2 + 2\rho(\alpha m - 1)\zeta} &&=: 1 + \alpha\mu\frac{n}{d} \\ b(\zeta) &:= \alpha \frac{\alpha m - 1 + \rho\zeta}{\rho^2 + (\alpha m - 1)^2 + 2\rho(\alpha m - 1)\zeta} &&=: \alpha\frac{n}{d} \\ c(\zeta) &:= 1 - \alpha\mu \frac{\alpha\mu - 1 + \rho\zeta}{\rho^2 + (\alpha\mu - 1)^2 + 2\rho(\alpha\mu - 1)\zeta} \\ d(\zeta) &:= \tfrac{\alpha}{\mu} \frac{\alpha m - 1 + \rho\zeta}{\rho^2 + (\alpha m - 1)^2 + 2\rho(\alpha m - 1)\zeta} &&=: \frac{\alpha}{\mu}\frac{n}{d}. \end{aligned} \tag{3.13c}$$

This is equivalent to checking positive definiteness of the $(1, 1)$ block via,

$$a(\zeta) > 0 \tag{3.14a}$$

and the Schur complement. After diagonalizing TT^T with a unitary transformation, this amounts to checking

$$c(\zeta) + d(\zeta)\lambda_i - \frac{(b(\zeta))^* b(\zeta)}{a(\zeta)}\lambda_i > 0 \tag{3.14b}$$

for each λ_i where λ_i are the eigenvalues of TT^T. To show exponential convergence at some rate ρ, we show (3.12) via (3.14) for $\rho = 1$. By continuity of (3.14) in ρ, this establishes (3.12) for some $\rho < 1$. For the remainder of the proof, we take $\rho = 1$.

Since $a(\zeta)$ is a linear fractional transformation of ζ, it is quasiconvex and thus condition (3.14a) may be established by checking $a(1) > 0$ and $a(-1) > 0$ [123],

$$1 \;+\; \alpha\mu \frac{\alpha m}{1 + (\alpha m - 1)^2 + 2(\alpha m - 1)} \quad = \quad 1 \;+\; \frac{\mu}{m} \;>\; 0 \tag{3.15a}$$

$$1 \;+\; \alpha\mu \frac{\alpha m - 2}{1 + (\alpha m - 1)^2 - 2(\alpha m - 1)} \quad = \quad 1 \;+\; \frac{\alpha\mu}{\alpha m - 2} \;>\; 0 \tag{3.15b}$$

Condition (3.15a) is clearly satisfied. By assumption, $\alpha\mu \in (0, \frac{1}{2})$ and $\alpha m \in (0, \frac{1}{2})$ which implies that $\alpha m - 2 \in (-2, -\frac{3}{2})$ and $\frac{\alpha\mu}{\alpha m - 2} \in (0, -\frac{1}{3})$, thereby establishing condition (3.15b).

Expressing the Schur complement (3.14b) in terms of the n and d from (3.13c) yields,

$$\begin{aligned}
c(\zeta) \;+\; \frac{\alpha}{\mu}\frac{n}{d}\lambda_i \;-\; \left(\alpha\frac{n}{d}\right)\left(1 \;+\; \alpha\mu\frac{n}{d}\right)^{-1}\left(\alpha\frac{n}{d}\right)\lambda_i \;&=\; c(\zeta) \;+\; \frac{\alpha}{\mu}\frac{n}{d}\lambda_i \;-\; \alpha^2\frac{n^2}{d^2}\frac{d}{\alpha\mu n + d}\lambda_i \\
&=\; c(\zeta) \;+\; \frac{\alpha}{\mu}\frac{n}{d}\lambda_i \;-\; \frac{\alpha^2 n^2}{d(\alpha\mu n + d)}\lambda_i \\
&=\; c(\zeta) \;+\; \frac{(\alpha/\mu)n(\alpha\mu n + d) \;-\; \alpha^2 n^2}{d(\alpha\mu n + d)}\lambda_i \\
&=\; c(\zeta) \;+\; \frac{(\alpha/\mu)n}{\alpha\mu n + d}\lambda_i
\end{aligned}$$

Substituting to remove n and d yields

$$s(\zeta) \;:=\; c(\zeta) \;+\; \frac{(\alpha/\mu)n}{\alpha\mu n + d}\lambda_i \;=:\; 1 \;+\; s_1(\zeta) \;+\; s_2(\zeta)\lambda_i \;>\; 0. \tag{3.16}$$

where

$$\begin{aligned}
s_1(\zeta) \;&:=\; -\frac{\alpha\mu(\alpha\mu - 1 + \zeta)}{1 + (\alpha\mu - 1)^2 + 2(\alpha\mu - 1)\zeta} \\
s_2(\zeta) \;&:=\; \frac{(\alpha/\mu)(\alpha m - 1 + \zeta)}{\alpha\mu(\alpha m - 1 + \zeta) + 1 + (\alpha m - 1)^2 + 2(\alpha m - 1)\zeta}
\end{aligned}$$

Since both s_1 and s_2 are linear fractional transformations of ζ, they are quasiconvex which implies $s_1(\zeta) \in [s_1(-1), s_1(1)]$ and $s_2(\zeta) \in [s_2(-1), s_2(1)]$ for all $\zeta \in [-1, 1]$. Evaluating these functions at $\zeta = \pm 1$,

$$s_1(-1) \;\in\; (0, \tfrac{1}{3}\alpha\mu), \quad s_1(1) \;=\; -1, \quad s_2(-1) \;=\; \tfrac{(\alpha/\mu)}{\alpha m + \alpha\mu - 2} > -1, \quad s_2(1) \;=\; \tfrac{(\alpha/\mu)}{\alpha\mu + \alpha m} > 1 \tag{3.17}$$

yields

$$s_1(\zeta) \;\in\; (-1, \tfrac{1}{3}\alpha\mu], \qquad s_2(\zeta) > -1, \quad \forall \;\; \zeta \;\in\; [-1, 1] \tag{3.18}$$

Since $TT^T = I$, $\lambda_i = 1$ for all i. However, although $s(1) > 0$ and $s(-1) > 0$, the set of quasiconvex functions is not closed under addition so inequality (3.16) does not follow for all $\zeta \in [-1, 1]$. From (3.18), we can only claim that $s(\zeta) > -1$ for all $\zeta \in [-1, 1]$, which is clearly insufficient. To obtain the result, we consider segments of $[-1, 1]$. These segments vary based on two cases: $\mu \geq m$ and $\mu \leq m$.

First, let us consider the case where $\mu \leq m$. We consider two segments $[-1, \eta]$ and $[\eta, 1]$ where $\eta = 1 - \alpha m$. Note that the denominator of $s_1(\zeta)$ is always positive so

$$s_1(\eta) \;=\; \frac{-\alpha^2\mu(\mu - m)}{1 + (\alpha\mu - 1)^2 + 2(\alpha\mu - 1)(1 - \alpha m)}.$$

is positive because $\mu - m \leq 0$. Together with (3.17), this implies

$$s_1(\zeta) \;\geq\; 0 \;\; \forall\zeta \;\in [-1, \eta], \qquad\qquad s_1(\zeta) \;\geq\; -1 \;\; \forall\zeta \;\in [\eta, 1] \tag{3.19a}$$

Combining (3.17) with $s_2(\eta) = 0$ yields,

$$s_2(\zeta) \;\geq\; -1 \;\; \forall\zeta \;\in [-1, \eta], \qquad\qquad s_2(\zeta) \;\geq\; 0 \;\; \forall\zeta \;\in [\eta, 1]. \tag{3.19b}$$

Together, equations (3.19) imply that (3.14b) holds when $\mu \leq m$.

Now consider the case $\mu \geq m$. We consider three segments $[-1, \delta]$, $[\delta, \eta]$, and $[\eta, 1]$ where $\delta = 1 - \alpha\mu$. Combining $s_1(\delta) = 0$ and $s_2(\eta) = 0$ with (3.17) yields

$$s_1(\zeta) \;\geq\; 0 \;\; \forall\zeta \;\in [-1, \delta], \qquad\qquad s_2(\zeta) \;\geq\; 0 \;\; \forall\zeta \;\in [\eta, 1]. \tag{3.20a}$$

Together with (3.18), equation (3.20a) implies that (3.14b) holds for $\zeta \in [-1, \delta]$ and $\zeta \in [\eta, 1]$. To consider $\zeta \in [\delta, \eta]$, we must consider $s_1(\eta)$ and $s_2(\delta)$. For this case, where $\mu \geq m$,

$$s_1(\eta) \;=\; \frac{(-\alpha\mu)\alpha(\mu - m)}{1 + (\alpha\mu - 1)((\alpha\mu - 1) + 2(1 - \alpha m))}$$

is negative. However, since $|\alpha\mu| \leq \frac{1}{2}$, $|\alpha\mu - 1| \leq \frac{1}{2}$, $|\alpha(m - \mu)| \leq \frac{1}{2}$ and $|1 - \alpha m| \leq \frac{1}{2}$, it is readily shown that $s_1(\eta) > -\frac{1}{3}$. Now consider,

$$s_2(\delta) \;=\; \frac{(\alpha/\mu)(\alpha(m - \mu))}{\alpha\mu(\alpha(m - \mu)) \;+\; 1 + (\alpha m - 1)^2 + 2(\alpha m - 1)(1 - \alpha\mu)}.$$

which is also necessarily negative. Using the same bounds, $|\alpha\mu| \leq 1$, $|\alpha\mu - 1| \leq 1$, $|\alpha(m - \mu)| \leq 1$

and $|1-\alpha m| \leq 1$, it is readily shown that $s_2(\delta) \geq -\frac{1}{2}$. This implies,

$$s_1(\zeta) \in [-\tfrac{1}{3}, 0] \ \ \forall \zeta \in [\delta, \eta], \qquad\qquad s_2(\zeta) \in [-\tfrac{1}{2}, 0] \ \ \forall \zeta \in [\delta, \eta]. \tag{3.20b}$$

from which it follows that $s(\zeta) > 0$ for all $\zeta \in [\delta, \eta]$, establishing condition (3.14b) and completing the proof. □

Remark 3.2.4. *The conditions on the step size α and the spectrum of TT^T in Theorem 3.2.3 are conservative and lead to the simple verification of inequalities* (3.15b) *and* (3.16). *These inequalities can be posed as quadratic and quartic equations in α, respectively. Since these orders of polynomial have explicit solutions, more rigorous analysis of these conditions would lead to explicit, tighter bounds on α and an overt dependence on λ_i.*

3.3 Distributed implementation

Dynamics (3.1) and (3.11) are convenient for distributed implementation. If the state vector x corresponds to the concatenated states of individual agents, x_i, the sparsity pattern of T and the structure of the gradient map $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$ dictate the communication topology required to form $\nabla \mathcal{L}_\mu$ in (3.1). For example, if $f(x) = \sum f_i(x_i)$ is separable over the agents, then $\nabla f_i(x_i)$ can be formed locally. If in addition T^T is an incidence matrix of an undirected network with M edges and the graph Laplacian $T^T T$, each agent must share its state x_i with its neighbors and maintain dual variables y_i that correspond to the edges to which it is connected.

Distributed implementation of a primal-descent dual-ascent dynamics on the proximal augmented Lagrangian provides several advantages over existing distributed optimization algorithms. For problems where T is not diagonal, a distributed proximal gradient scheme cannot be implemented because the proximal operator of $g(Tx)$ may not be separable. It has a continuous right-hand side even for problems (2.1) with non-differentiable regularizers g, unlike existing approaches which employ subgradient methods [147] or use discontinuous projected dynamics [148–151]. Finally, relative to a distributed ADMM scheme, our method does not require solving an x-minimization subproblem in each iteration.

3.3.1 Example: Optimal placement

In this example [123], a number of agents try to minimize some weighted measure of distances from their neighbors. The set of neighbors for each node is determined by an *a priori* specified

network. The optimization variable x contains the locations of the 'mobile' agents The locations of fixed agents b enter as a problem parameter. The optimization problem can be written as,

$$\text{minimize} \quad \sum_{(i,j)\,\in\,\mathcal{E}_{xx}} f_{ij}(x_i \;-\; x_j) \;+\; \sum_{(i,j)\,\in\,\mathcal{E}_{xb}} f_{ij}(x_i \;-\; b_j)$$

where $x_i, b_i \in \mathbb{R}^p$ are position vectors of mobile and fixed agents in a p-dimensional space, $\mathcal{E}_{xx}$ and $\mathcal{E}_{xb}$ denote the sets of edges from mobile agents to each other and to the fixed agents respectively, and f_{ij} is some measure of distance. In this example, we take $f_{ij} = (1/2)\|\cdot\|^2$ and impose an additional constraint on the distances between neighboring mobile agents,

$$\|x_i \;-\; x_j\| \;\leq\; c, \quad \forall (i,j) \;\in\; \mathcal{E}_{xx}.$$

We can reexpress the problem as,

$$\text{minimize} \quad \underbrace{\frac{1}{2}\left\| \begin{bmatrix} A \\ T \end{bmatrix} x \;-\; \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|^2}_{f(x)} \;+\; \underbrace{I_{[-c,c]}(Tx)}_{g(\mathcal{T}(x))} \tag{3.21}$$

where x contains the positions of the mobile agents, $Ax - b$ forms vectors between mobile agents and neighboring fixed agents, Tx forms vectors connecting neighboring mobile agents, and $I_{[-c,c]}$ is an indicator function associated with the set of vectors v whose subvectors $v_i \in \mathbb{R}^p$ have norm less than c.

Distributed implementation

We form the proximal augmented Lagrangian and implement the gradient dynamics described in (3.1). The proximal operator assoicated with $I_{[-c,c]}$ is a projection which scales each subvector with norm larger than c such that it has norm c. Each subvector x_i of the state corresponds to the position of the ith agent and each subvector y_i of the Lagrange multiplier is associated with the difference in position between two agents connected by an edge in $\mathcal{E}_{xx}$.

The continuous time dynamics described in (3.1) and the expression for the gradient of f,

$$\nabla f(x) \;=\; (A^T A \;+\; T^T T)x \;+\; A^T b$$

reveal the distributed nature of the dynamics. The matrix $A^T A = \text{blkdiag}(\ldots, d_i I_p, \ldots)$ where

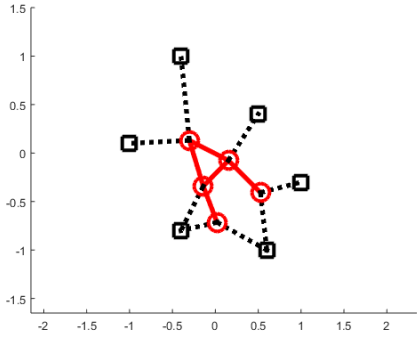
d_i is the number of fixed agents to which the ith mobile agent is connected and I_p is a p-dimensional identity matrix. The matrix T is of the form $E \otimes I_p$ where E is the incidence matrix corresponding to $\mathcal{E}_{xx}$ so $T^T T$ is of the form $L \otimes I_p$ where L is the graph laplacian associated with $\mathcal{E}_{xx}$.

Each agent can form its dynamics only the positions of itself and its neighbors and the Lagrange multipliers associated with the edges to which it is connected. The dynamics of y_i may be independently determined by agents.

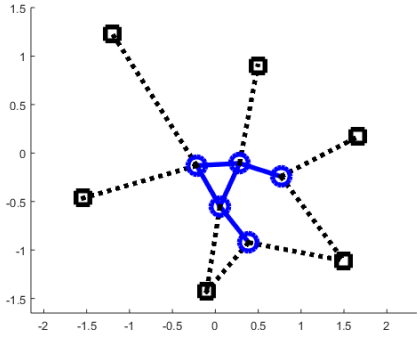
Simulation results

We simulated the gradient dynamics for the constrained optimal placement problem for a small system with 5 mobile agents and 6 fixed agents. Figs. 3.3a and 3.3b show the optimal mobile agent positions (solution to (3.21)) with **red** $\circ$ or **blue** $\circ$ circles given two configurations of fixed agents (**black** $\square$) with $\mathcal{E}_{xx}$ shown by **solid red** — or **solid blue** — edges, $\mathcal{E}_{xb}$ is shown by **dotted black** $\cdots$ edges, and $c = 0.5$. The second configuration corresponds to rotating the fixed agents by $\pi/8$ and scaling their positions by 1.6. In Fig. 3.3a, 3 out of the 5 inequality constraints are active and in Fig. 3.3b, all 5 inequality constraints are active.

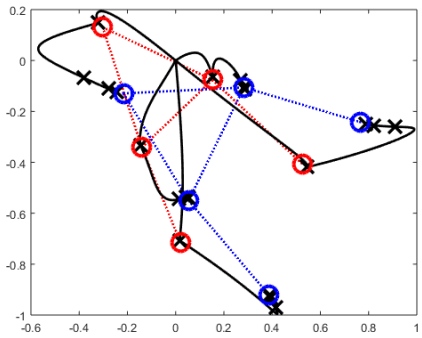
To illustrate the dynamics, we simulated (3.1) from $t = 0$ to $t = 10$ with a discrete jump in fixed agent positions. The mobile agents were initially placed at the origin. The fixed agents were placed as in Fig. 3.3a until time $t = 2.5$, at which they were placed as in Fig. 3.3b. The trajectories of each mobile agent are shown in Fig. 3.3c and the Euclidean distance from their optimal position (i.e., as in Figs. 3.3a and 3.3b) with respect to time is plotted in Fig. 3.3d.



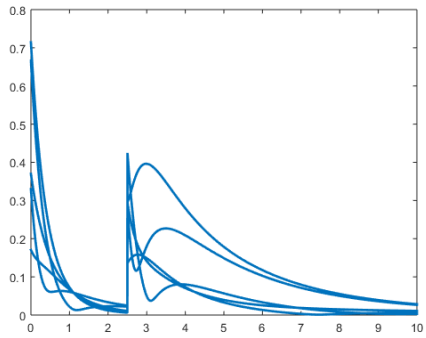
(a) First configuration. 3 out of 5 inequality constraints active.



(b) Second configuration. 5 out of 5 inequality constraints active.



(c) Trajectories with black × at position every 2.5 time steps.



(d) Euclidean distance from optimal with respect to time.

Figure 3.3: Subfigures (a) and (b) show two optimal configurations of mobile agents (○) with corresponding fixed agents (**black** □), $\mathcal{E}_{xx}$ is (**solid red** — and **solid blue** — edges) and $\mathcal{E}_{xb}$ (**dotted black** ⋯ edges). Subfigures (c) and (d) show the agent trajectories and distance from optimal configuration (a) until $t = 2.5$ and from optimal configuration (b) until the final $t = 10$.

Chapter 4

Second-order primal-dual method

Although the MM and Arrow-Hurwicz-Uzawa primal-descent dual-ascent gradient flow approaches are typically simple and computationally efficient, they tend to converge slowly to high-accuracy solutions. As with smooth problems, second order information can enhance the speed of convergence. A generalization of Newton's method to non-smooth problems has been developed in [157, 158], but it requires solving a regularized quadratic subproblem to determine a search direction. Except in cases where the Hessian of the smooth component has a special structure, this problem may be difficult to solve. More recently, proximal Newton method for non-smooth composite minimization was developed in [159]. It was shown that, even in the situations when the search direction is computed approximately, the developed algorithm shares convergence properties with traditional Newton's method. Related ideas have been successfully utilized in a number of applications, including sparse inverse covariance estimation in graphical models [160, 161].

Recent work has extended the method of multipliers to incorporate second-order updates of the primal and dual variables [162–164]. Since the standard method of multipliers seeks the saddle point of the augmented Lagrangian, it is challenging to assess joint progress of primal and dual iterates. In [162], Gill and Robinson have introduced the notion of primal-dual augmented Lagrangian that serves as the merit function for measuring progress of second-order updates. The proposed approach is applicable to non-convex problems but it requires that the objective function is twice continuously differentiable. In [165], a forward-backward envelope was utilized to cast the original non-smooth optimization problem as the unconstrained minimization of a continuously differentiable function and generalized sub-differential calculus was employed to derive second-order updates. Generalized second order methods have also been applied to saddle

point problems directly [166].

We draw on these recent advances and develop a new algorithm for non-smooth composite optimization which efficiently forms second-order updates of both primal and dual variables. To motivate our approach, we show global exponential stability of the corresponding continuous-time differential inclusion when f is strongly convex. Using the merit function employed in [163, 164], we refine the algorithm by adding adaptive updates for the penalty parameter. We show global convergence and local quadratic convergence of this algorithm for strongly convex f.

4.1 Problem formulation and background

We consider problem (2.1) with additional assumptions on the function f and linear operator T. Although our strongest results require strong convexity of f, our theory and techniques are applicable as long as the Hessian of f is positive definite.

Assumption 1. The function f is twice continuously differentiable, has an L_f Lipschitz continuous gradient ∇f, and is strictly convex with $\nabla^2 f \succ 0$; the function g is proper, lower semicontinuous, and convex; and the matrix T has full row rank.

We now provide background on generalizations of the gradient for nondifferentiable functions and briefly overview existing second order methods for solving (2.1).

4.1.1 Generalization of the gradient and Jacobian

Although $\mathbf{prox}_{\mu g}$ is typically not differentiable, it is Lipschitz continuous and therefore differentiable almost everywhere [138]. One generalization of the gradient for such functions is given by the B-subdifferential set [167], which applies to locally Lipschitz continuous functions h: $\mathbb{R}^m \to \mathbb{R}$. Let C_h be a set at which h is differentiable. Each element in the set $\partial_B h(\bar{z})$ is the limit point of a sequence of gradients $\{\nabla h(z_k)\}$ evaluated at a sequence of points $\{z_k\} \subset C_h$ whose limit is $\bar{z}$,

$$\partial_B h(\bar{z}) := \{J \mid \exists \{z_k\} \subset C_h,\ z_k \to \bar{z},\ \nabla h(z_k) \to J\}. \tag{4.1}$$

If h is continuously differentiable in the neighborhood of a point z, the B-subdifferential set becomes single valued and it is given by the gradient, $\partial_B h(z) = \nabla h(z)$. In general, $\partial_B h(\bar{z})$ is not a convex set; e.g., if $h(z) = |z|$, $\partial_B h(0) = \{-1, 1\}$.

The Clarke subdifferential set of h: $\mathbb{R}^m \to \mathbb{R}$ at $\bar{z}$ is the convex hull of the B-subdifferential set [139],

$$\partial_C h(\bar{z}) \;:=\; \mathrm{conv}\,(\partial_B h(\bar{z})).$$

When h is a convex function, the Clarke subdifferential set is equal to the subdifferential set $\partial h(\bar{z})$ which defines the supporting hyperplanes of h at $\bar{z}$ [168, Chapter VI]. For a function G: $\mathbb{R}^m \to \mathbb{R}^n$, the B-generalization of the Jacobian at a point $\bar{z}$ is given by

$$\partial_B G(\bar{z}) \;:=\; \begin{bmatrix} J_1^T & \dots & J_n^T \end{bmatrix}^T$$

where each $J_i \in \partial_B G_i(\bar{z})$ is a member of the B-subdifferential set of the ith component of G evaluated at $\bar{z}$. The Clarke generalization of the Jacobian at a point $\bar{z}$, $\partial_C G(\bar{z})$, has the same structure where each $J_i \in \partial_C G_i(\bar{z})$ is a member of the Clarke subdifferential of $G_i(\bar{z})$.

4.1.2 Semismoothness

The mapping G: $\mathbb{R}^m \to \mathbb{R}^n$ is semismooth at $\bar{z}$ if for any sequence $z_k \to \bar{z}$, the sequence of Clarke generalized Jacobians $J_{G_k} \in \partial_C G(z_k)$ provides a first order approximation of G,

$$\|G(z_k) \,-\, G(\bar{z}) \,+\, J_{G_k}(\bar{z} \,-\, z_k)\| \;=\; o(\|z_k \,-\, \bar{z}\|). \tag{4.2}$$

where $\phi(k) = o(\psi(k))$ denotes that $\phi(k)/\psi(k) \to 0$ as k tends to infinity [169]. The function G is strongly semismooth if this approximation satisfies the stronger condition,

$$\|G(z_k) \,-\, G(\bar{z}) \,+\, J_{G_k}(\bar{z} \,-\, z_k)\| \;=\; O(\|z_k \,-\, \bar{z}\|^2),$$

where $\phi(k) = O(\psi(k))$ signifies that $|\phi(k)| \leq L\psi(k)$ for some positive constant L and positive $\psi(k)$ [169].

Remark 4.1.1. *(Strong) semismoothness of the proximal operator leads to fast asymptotic convergence of the differential inclusion (see Section 4.2) and the efficient algorithm (see Section 4.3). Proximal operators associated with many typical regularization functions (e.g., the ℓ_1 and nuclear norms [170], piecewise quadratic functions [171], and indicator functions of affine convex sets [171]) are strongly semismooth. In general, semismoothness of $\mathbf{prox}_{\mu g}$ follows from semismoothness of the projection onto the epigraph of g [171]. However, there are convex sets onto which projection is not directionally differentiable [172]. The indicator functions associated with such sets or functions whose epigraphs are described by such sets may induce proximal*

operators which are not semismooth.

4.1.3 Existing second order methods

We discussed existing first order methods in Section 2.2.3. While simple to implement, their slow convergence to high-accuracy solutions motivates the development of second order methods for solving (2.1). A generalization of Newton's method for nonsmooth problems (2.1) with $T = I$ was developed in [157, 158, 165, 173]. A sequential quadratic approximation of the smooth part of the objective function is utilized and a search direction $\tilde{x}$ is obtained as the solution of a regularized quadratic subproblem,

$$\underset{\tilde{x}}{\text{minimize}} \quad \tfrac{1}{2}\,\tilde{x}^T H\tilde{x} \;+\; \nabla f(x^k)^T\tilde{x} \;+\; g(x^k + \tilde{x}) \tag{4.3}$$

where x^k is the current iterate and H is the Hessian of f. This method generalizes the projected Newton method [174] to a broader class of regularizers. For example, when g is the ℓ_1 norm, this amounts to solving a LASSO problem [175], which can be a challenging task. Coordinate descent is often used to solve this subproblem [173] and it has been observed to perform well in practice [76, 160, 161].

The Forward-Backward Envelope (FBE) was introduced in [176–178]. The FBE is a once-continuously differentiable nonconvex function of x and its minimum corresponds to the solution of (2.1) with $T = I$. As demonstrated in Chapter 5, the FBE can be obtained from the proximal augmented Lagrangian that we introduce in Section 2.3. Since the generalized Hessian of FBE involves third-order derivatives of f (which may be expensive to compute), references [176–178] employ either truncated or quasi-Newton methods to obtain a second order update to x.

4.1.4 Second order updates

Even though Newton's method is primarily used for solving minimization problems in modern optimization, it was originally formulated as a root-finding technique and it has long been employed for finding stationary points [179]. In [166], a generalized Jacobian was used to extend Newton's method to semismooth problems. We employ this generalization of Newton's method to $\nabla\mathcal{L}_\mu(x;y)$ in order to compute the saddle point of the proximal augmented Lagrangian. The unique saddle point of $\mathcal{L}_\mu(x;y)$ is given by the optimal primal-dual pair $(x^\star, y^\star)$ and it thus provides the solution to (2.1).

Generalized Newton updates

Let $H := \nabla^2 f(x)$. We use the B-generalized Jacobian of the proximal operator $\mathbf{prox}_{\mu g}$, $\mathbb{P}_B := \partial_B \mathbf{prox}_{\mu g}(Tx + \mu y)$, to define the set of B-generalized Hessians of the proximal augmented Lagrangian,

$$\partial_B^2 \mathcal{L}_\mu \;:=\; \left\{ \begin{bmatrix} H + \frac{1}{\mu} T^T(I \,-\, P)T & T^T(I \,-\, P) \\ (I \,-\, P)T & -\mu P \end{bmatrix}, P \in \mathbb{P}_B \right\} \tag{4.4a}$$

and the Clarke generalized Jacobian $\mathbb{P}_C := \partial_C \mathbf{prox}_{\mu g}(Tx + \mu y)$ to define the set of Clarke generalized Hessians of the proximal augmented Lagrangian,

$$\partial_C^2 \mathcal{L}_\mu \;:=\; \left\{ \begin{bmatrix} H + \frac{1}{\mu} T^T(I \,-\, P)T & T^T(I \,-\, P) \\ (I \,-\, P)T & -\mu P \end{bmatrix}, P \in \mathbb{P}_C \right\}. \tag{4.4b}$$

Note that $\partial_B^2 \mathcal{L}_\mu(x; y) \subset \partial_C^2 \mathcal{L}_\mu(x; y)$ because $\mathbb{P}_B \subset \mathbb{P}_C$.

We introduce the composite variable, $w := [\, x^T \;\; y^T \,]^T$, use $\mathcal{L}_\mu(w)$ interchangeably with $\mathcal{L}_\mu(x; y)$, and suppress the dependance of H and P on w to reduce notational clutter. For simplicity of exposition, we assume that $\mathbf{prox}_{\mu g}$ is semismooth and state the results for the Clarke generalized Hessian (4.4b), i.e., $\partial^2 \mathcal{L}_\mu(w) = \partial_C^2 \mathcal{L}_\mu(w)$. As described in Remark 4.2.6, analogous convergence results for non-semismooth $\mathbf{prox}_{\mu g}$ can be obtained for the B-generalized Hessian (4.4a), i.e., $\partial^2 \mathcal{L}_\mu(w) = \partial_B^2 \mathcal{L}_\mu(w)$.

We use the Clarke generalized Hessian (4.4b) to obtain a second order update $\tilde{w}$ by linearizing the stationarity condition $\nabla \mathcal{L}_\mu(w) = 0$ around the current iterate w^k,

$$\partial_C^2 \mathcal{L}_\mu(w^k)\, \tilde{w}^k \;=\; -\nabla \mathcal{L}_\mu(w^k). \tag{4.5}$$

Since $\mathbf{prox}_{\mu g}$ is firmly nonexpansive, $0 \preceq P \preceq I$. In Lemma 4.1.2 we use this fact to prove that the second order update $\tilde{w}$ is well-defined for any generalized Hessian (4.4) of the proximal augmented Lagrangian $\mathcal{L}_\mu(x; y)$ as long as f is strictly convex with $\nabla^2 f(x) \succ 0$ for all $x \in \mathbb{R}^m$.

Lemma 4.1.2. *Let $H \in \mathbb{R}^{n \times n}$ be symmetric positive definite, $H \succ 0$, let $P \in \mathbb{R}^{M \times M}$ be symmetric positive semidefinite with eigenvalues less than one, $0 \preceq P \preceq I$, let $T \in \mathbb{R}^{M \times m}$ be full row rank, and let $\mu > 0$. Then, the matrix*

$$\begin{bmatrix} H + \frac{1}{\mu} T^T(I - P)T & T^T(I - P) \\ (I - P)T & -\mu P \end{bmatrix}$$

is invertible and it has n positive and m negative eigenvalues.

Proof. The Haynsworth inertia additivity formula [180] implies that the inertia of matrix (4.4) is determined by the sum of the inertias of matrices,

$$H \ + \ \tfrac{1}{\mu}\, T^T(I \ - \ P)T \tag{4.6a}$$

and

$$-\mu P \ - \ (I-P)\, T \left(H + \tfrac{1}{\mu}\, T^T(I-P)T\right)^{-1} T^T(I-P). \tag{4.6b}$$

Matrix (4.6a) is positive definite because $H \succ 0$ and both P and $I - P$ are positive semidefinite. Matrix (4.6b) is negative definite because the kernels of P and $I - P$ have no nontrivial intersection and T has full row rank. □

Fast local convergence

The use of generalized Newton updates for solving the nonlinear equation $G(x) = 0$ for nondifferentiable G was studied in [166]. We apply this framework to the stationarity condition $\nabla \mathcal{L}_\mu(w) = 0$ when $\mathbf{prox}_{\mu g}$ is (strongly) semismooth and show that second order updates (4.5) converge (quadratically) superlinearly within a neighborhood of the optimal primal-dual pair.

Proposition 4.1.1. *Let $\mathbf{prox}_{\mu g}$ be (strongly) semismooth, and let $\tilde{w}^k$ be defined by* (4.5)*. Then, there is a neighborhood of the optimal solution $w^\star$ in which the second order iterates $w^{k+1} = w^k + \tilde{w}^k$ converge (quadratically) superlinearly to $w^\star$.*

Proof. Lemma 4.1.2 establishes that $\partial_C^2 \mathcal{L}_\mu(w)$ is nonsingular for any $P \in \mathbb{P}_C$. Since the gradient $\nabla \mathcal{L}_\mu(w)$ of the proximal augmented Lagrangian is Lipschitz continuous by Theorem 2.3.1, nonsingularity of $\partial_C^2 \mathcal{L}_\mu(w)$ and (strong) semismoothness of the proximal operator guarantee (quadratic) superlinear convergence of the iterates by [166, Theorem 3.2]. □

4.2 A globally convergent differential inclusion

Since we apply a generalization of Newton's method to a saddle point problem and the second order updates are set valued, convergence to the optimal point is not immediate. Although we showed local convergence rates in Proposition 4.1.1 by leveraging the results of [166], proof of the global convergence is more subtle and it is established next.

To justify the development of a discrete-time algorithm based on the search direction resulting from (4.5), we first examine the corresponding differential inclusion,

$$\dot{w} \;\in\; -(\partial_C^2 \mathcal{L}_\mu(w))^{-1}\, \nabla \mathcal{L}_\mu(w) \tag{4.7}$$

where $\partial_C^2 \mathcal{L}_\mu$ is the Clarke generalized Hessian (4.4b) of $\mathcal{L}_\mu$. We assume existence of a solution and prove asymptotic stability of (4.7) under Assumption 1 and global exponential stability under an additional assumption that f is strongly convex.

Assumption 2. Differential inclusion (4.7) has a solution.

4.2.1 Asymptotic stability

We first establish asymptotic stability of differential inclusion (4.7).

Theorem 4.2.1. *Let Assumptions 1 and 2 hold and let* $\mathbf{prox}_{\mu g}$ *be semismooth. Then, differential inclusion* (4.7) *is asymptotically stable. Moreover,*

$$V(w) \;:=\; \tfrac{1}{2}\, \|\nabla \mathcal{L}_\mu(w)\|^2 \tag{4.8}$$

provides a Lyapunov function and

$$\dot{V}(t) \;=\; -\,2V(t). \tag{4.9}$$

Proof. Lyapunov function candidate (4.8) is a positive function of w everywhere apart from the optimal primal-dual pair $w^\star$ at which it is zero. It remains to show that V is decreasing along the solutions $w(t)$ of (4.7), i.e., that $\dot{V}$ is strictly negative for all $w(t) \neq w^\star$,

$$\dot{V}(t) \;:=\; \tfrac{\mathrm{d}}{\mathrm{d}t} V(w(t)) \;=\; -\,2V(w(t))$$

For Lyapunov function candidates $\hat{V}(w)$ which are differentiable with respect to w, $\dot{\hat{V}} = \dot{w}^T \nabla \hat{V}$ by the chain rule. Although (4.8) is *not* differentiable with respect to w, we show that $V(w(t))$ *is* differentiable *along the solutions* of (4.7). Instead of employing the chain rule, we use the limit that defines the derivative,

$$\dot{V}(t) \;:=\; \tfrac{\mathrm{d}}{\mathrm{d}t} V(w(t)) \;=\; \lim_{s \to 0} \frac{V(w(t) \,+\, s\tilde{w}(t)) \,-\, V(w(t))}{s} \tag{4.10}$$

to show that $\dot{V}$ exists and is negative along the solutions of (4.7). Here, $\tilde{w}$ is determined by the dynamics (4.7),

$$\tilde{w}(t) \;=\; -\,H_C^{-1}\nabla\mathcal{L}_\mu(w(t)), \tag{4.11}$$

for some $H_C \in \partial_C^2\mathcal{L}_\mu(w(t))$ and (4.10) is equivalent to the directional derivative of $V(w)$ in the direction $\tilde{w}$. We first introduce

$$h_s(t) \;:=\; \frac{V(w(t) \,+\, s\tilde{w}(t)) \,-\, V(w(t))}{s}$$

which yields $\dot{V}$ in the limit $s \to 0$. We then rewrite $h_s(t)$ as the limit point of a sequence of functions $\{h_{s,k}(t)\}$ so that

$$\dot{V}(t) \;=\; \lim_{s\,\to\,0}\; h_s(t) \;=\; \lim_{s\,\to\,0}\,\lim_{k\,\to\,\infty}\; h_{s,k}(t) \tag{4.12}$$

and use the Moore-Osgood theorem [181, Theorem 7.11] to exchange the order of the limits and establish that $\dot{V}(t) = -2V(t)$.

Let C_g denote a subset of $\mathbb{R}^{m+M}$ over which $\mathbf{prox}_{\mu g}(Tx+\mu y)$ is differentiable (and therefore V is differentiable with respect to w) and let $\{w_k\}$ be a sequence of points in C_g that converges to w. We define the sequence of functions $\{h_{s,k}(t)\}$,

$$h_{s,k}(t) \;:=\; \frac{V(w_k(t) \,+\, s\tilde{w}(t)) \;-\; V(w_k(t))}{s}.$$

To employ the Moore-Osgood theorem, it remains to show that $h_{s,k}(t)$ converges pointwise (for any k) as $s \to 0$ and that $h_{s,k}(t)$ converges uniformly on some interval $s \in [0, \bar{s}]$ as $k \to \infty$.

Since $\{w_k\} \subset C_g$, $V(w_k)$ is differentiable for every $k \in \mathbb{Z}_+$ and $\nabla V(w_k) = \nabla^2\mathcal{L}_\mu(w_k) = \partial_C^2\mathcal{L}_\mu(w_k)$. It thus follows that

$$\lim_{s\,\to\,0}\; h_{s,k}(t) \;=\; \tilde{w}_k^T(t)\,\nabla^2\mathcal{L}_\mu(w_k)\nabla\mathcal{L}_\mu(w_k(t)) \tag{4.13}$$

pointwise (for any k). Local Lipschitz continuity of V with respect to w_k implies uniform convergence of $h_{s,k}(t)$ to $h_s(t)$ on $s \in (0, \bar{s}\,]$ where $\bar{s} > 0$. Therefore, the Moore-Osgood theorem

on exchanging limits [181, Theorem 7.11] in conjunction with (4.13) implies

$$\begin{aligned}
\dot{V}(t) &= \lim_{s \to 0} h_s(t) \\
&= \lim_{s \to 0} \lim_{k \to \infty} h_{s,k}(t) = \lim_{k \to \infty} \lim_{s \to 0} h_{s,k}(t) \\
&= \lim_{k \to \infty} \tilde{w}^T(t)\, \nabla^2 \mathcal{L}_\mu(w_k) \nabla \mathcal{L}_\mu(w_k(t)) \qquad (4.14a) \\
&= \tilde{w}^T(t)\, H_B \nabla \mathcal{L}_\mu(w(t)) \\
&= -(\nabla \mathcal{L}_\mu(w(t)))^T H_C^{-1}\, H_B \nabla \mathcal{L}_\mu(w(t))
\end{aligned}$$

for some $H_B \in \partial_B^2 \mathcal{L}_\mu(w(t))$ by the definition of the B-subdifferential (4.1). This immediately establishes (4.9) for any $\tilde{w}$ (4.11) such that $C \in \partial_B^2 \mathcal{L}_\mu(w(t))$ by choosing C_g such that $H_B = H_C$.

Semismoothness implies directional differentiability of V and thereby equivalence of (4.10) and (4.14a) by [182, Proposition 2.3]. It follows that

$$\dot{V}(t) \;=\; (\nabla \mathcal{L}_\mu(w(t)))^T H_C^{-1}\, H_{B_i} \nabla \mathcal{L}_\mu(w(t)) \;=\; (\nabla \mathcal{L}_\mu(w(t)))^T H_C^{-1}\, H_{B_j} \nabla \mathcal{L}_\mu(w(t)) \qquad (4.14b)$$

for any $H_{B_i}, H_{B_j} \in \partial_B^2 \mathcal{L}_\mu(w(t))$. By definition of the Clarke subgradient, any $H_C \notin \partial_B^2 \mathcal{L}_\mu(w(t))$ can be expressed as

$$H_C \;=\; \sum_i \alpha_i H_{B_i} \qquad (4.14c)$$

where $H_{B_i} \in \partial_B^2 \mathcal{L}_\mu(w(t))$ and $\alpha \in \mathcal{P}$ where $\mathcal{P}$ is the simplex $\mathcal{P} := \{\alpha \in \mathbb{R}_+^r \mid \sum_i \alpha_i = 1\}$ and r is the cardinality of the set $\partial_B^2 \mathcal{L}_\mu(w)$. By substituting (4.14c) into the final line of (4.14a) and noting the equivalence (4.14b), the directional derivative (4.10) can be expressed as

$$\dot{V}(t) \;=\; -(\nabla \mathcal{L}_\mu(w(t)))^T \left(\sum_i \alpha_i H_{B_i} \right)^{-1} \left(\sum_i \beta_i H_{B_i} \right) \nabla \mathcal{L}_\mu(w(t))$$

for any $\beta \in \mathcal{P}$. Taking $\beta = \alpha$ yields $\dot{V}(t) = -\|\mathcal{L}_\mu(w(t))\|^2 = -2V(t)$, completing the proof. □

4.2.2 Global exponential stability

To establish global asymptotic stability, we show that the Lyapunov function (4.8) is radially unbounded, and to prove exponential stability we bound it with quadratic functions. We first provide two lemmas that characterize the mappings $\mathbf{prox}_{\mu g}$ and ∇f in terms of the spectral properties of matrices that describe the corresponding input-output relations at given points.

Lemma 4.2.2. *Let f be strongly convex with parameter m_f and let its gradient ∇f be Lipschitz continuous with parameter L_f. Then, for any $a, b \in \mathbb{R}^n$ there exists a symmetric matrix $G_{a,b}$ satisfying $m_f I \preceq G_{a,b} \preceq L_f I$ such that*

$$\nabla f(a) \; - \; \nabla f(b) \; = \; G_{a,b}\,(a \; - \; b).$$

Proof. Let $c := a - b$, $d := \nabla f(a) - \nabla f(b)$, $e := d - m_f c$, and

$$\hat{G}_{a,b} \quad := \quad \{ee^T/(e^T c),\, e \neq 0;\; 0,\, \text{otherwise}\} \tag{4.15a}$$

$$G_{a,b} \quad := \quad \hat{G}_{a,b} \; + \; m_f I. \tag{4.15b}$$

Clearly, by construction, $\hat{G}_{a,b} = \hat{G}_{a,b}^T \succeq 0$. It is also readily verified that $G_{a,b}\,c = d$ when $e^T c \neq 0$. It thus remains to show that (i) $G_{a,b}\,c = d$ when $e^T c = 0$; and (ii) $\hat{G}_{a,b} \preceq (L_f - m_f) I$.

(i) Since f is m_f strongly convex and ∇f is L_f Lipschitz continuous, $h(x) := f(x) - \frac{m_f}{2}\|x\|^2$ is convex and $\nabla h(x) = \nabla f(x) - m_f x$ is $L_f - m_f$ Lipschitz continuous. Furthermore, we have $e = \nabla h(a) - \nabla h(b)$, and [152, Proposition 5] implies

$$e^T c \; \geq \; \tfrac{1}{L_f - m_f}\,\|e\|^2, \quad \text{for all } c \in \mathbb{R}^n. \tag{4.16}$$

This shows that $e^T c = 0$ only if $e := d - m_f c = 0$ and, thus, $d = m_f c = G_{a,b}\,c$ when $e^T c = 0$. Therefore, there always exist a symmetric matrix $G_{a,b}$ such that $G_{a,b}\,c = d$.

(ii) When $e^T c \neq 0$, $\hat{G}_{a,b}$ is a rank one matrix and its only nonzero eigenvalue is $\|e\|^2/(e^T c)$; this follows from $\hat{G}_{a,b}\,e \; = \; (\|e\|^2/(e^T c))\,e$. In this case, inequality (4.16) implies $e^T c > 0$ and (4.16) is equivalent to $1/(e^T c) \leq (L_f - m_f)/\|e\|^2$. Thus, $\|e\|^2/(e^T c) \leq L_f - m_f$ and $\hat{G}_{a,b} \preceq (L_f - m_f) I$ when $e^T c \neq 0$. Since $\hat{G}_{a,b} = 0$ when $e^T c = 0$, $\hat{G}_{a,b} \preceq (L_f - m_f) I$ for all a and b. Finally, $\hat{G}_{a,b} \succeq 0$ and (4.15b) imply $m_f I \preceq G_{a,b} \preceq L_f I$. □

Remark 4.2.3. *Although matrices $D_{a,b}$ and $G_{a,b}$ in Lemmas 3.1.2 and 4.2.2 depend on the operating point, their spectral properties, $0 \preceq D_{a,b} \preceq I$ and $m_f I \preceq G_{a,b} \preceq L_f I$, hold for all a and b. These lemmas can be interpreted as a combination between a generalization of the mean value theorem [181, Theorem 5.9] to vector-valued functions and spectral bounds on the operators $\mathbf{prox}_{\mu g}\colon \mathbb{R}^m \to \mathbb{R}^m$ and $\nabla f\colon \mathbb{R}^n \to \mathbb{R}^n$ arising from firm nonexpansiveness of $\mathbf{prox}_{\mu g}$, strong convexity of f, and Lipschitz continuity of ∇f.*

We now combine Lemmas 3.1.2 and 4.2.2 to establish quadratic upper and lower bounds

for Lyapunov function (4.8) and thereby prove global exponential stability of differential inclusion (4.7) for strongly convex f.

Theorem 4.2.4. *Let Assumptions 1 and 2 hold, let* $\mathbf{prox}_{\mu g}$ *be semismooth, and let* f *be* m_f *strongly convex. Then, differential inclusion* (4.7) *is globally exponentially stable, i.e., there exists* $\kappa > 0$ *such that* $\|w(t) - w^\star\| \leq \kappa\, \mathrm{e}^{-t}\, \|w(0) - w^\star\|$.

Proof. Given the assumptions, Theorem 4.2.1 establishes asymptotic stability of (4.7) with the dissipation rate $\dot{V}(w) \;=\; -2V(w)$. It remains to show the existence of positive constants κ_1 and κ_2 such that Lyapunov function (4.8) satisfies

$$\tfrac{\kappa_1}{2}\, \|\tilde{w}\|^2 \;\leq\; V(w) \;\leq\; \tfrac{\kappa_2}{2}\, \|\tilde{w}\|^2 \tag{4.17}$$

where $\tilde{w} := w - w^\star$ and $w^\star := (x^\star, y^\star)$ is the optimal primal-dual pair. The upper bound in (4.17) follows from Lipschitz continuity of $\nabla \mathcal{L}_\mu(w)$ (see Theorem 2.3.1), with κ_2 determined by the Lipschitz constant of $\nabla \mathcal{L}_\mu(w)$.

To show the lower bound in (4.17), and thus establish radial unboundedness of $V(w)$, we construct matrices that relate $V(w)$ to $\tilde{w}$. Lemmas 3.1.2 and 4.2.2 imply the existence of symmetric matrices $D_{\tilde{w}}$ and $G_{\tilde{w}}$ such that $0 \preceq D_{\tilde{w}} \preceq I$, $m_f I \preceq G_{\tilde{w}} \preceq L_f I$, and

$$\begin{aligned} \mathbf{prox}_{\mu g}(Tx + \mu y) - \mathbf{prox}_{\mu g}(Tx^\star + \mu y^\star) \;&=\; D_{\tilde{w}}\, (T\tilde{x} + \mu \tilde{y}) \\ f(x) - f(x^\star) \;&=\; G_{\tilde{w}}\, \tilde{x}. \end{aligned}$$

As noted in Remark 4.2.3, although $D_{\tilde{w}}$ and $G_{\tilde{w}}$ depend on the operating point, their spectral properties hold for all $\tilde{w}$.

Since $\nabla \mathcal{L}_\mu(w^\star) = 0$, we can write

$$\nabla \mathcal{L}_\mu(w) \;=\; \nabla \mathcal{L}_\mu(w) \;-\; \nabla \mathcal{L}_\mu(w^\star) \;=\; Q_{\tilde{w}}\, \tilde{w}$$

and express Lyapunov function (4.8) as

$$V(w) \;=\; \tfrac{1}{2}\, \tilde{w}^T Q_{\tilde{w}}^T\, Q_{\tilde{w}}\, \tilde{w}$$

where

$$Q_{\tilde{w}} \;:=\; \begin{bmatrix} G_{\tilde{w}} + \frac{1}{\mu}\, T^T (I - D_{\tilde{w}}) T & T^T (I - D_{\tilde{w}}) \\ (I - D_{\tilde{w}}) T & -\mu D_{\tilde{w}} \end{bmatrix}$$

for some $(D_{\tilde{w}}, G_{\tilde{w}}) \in \Omega_{\tilde{w}}$,

$$\Omega_{\tilde{w}} := \{(D_{\tilde{w}}, G_{\tilde{w}}) \mid 0 \preceq D_{\tilde{w}} \preceq I,\ m_f I \preceq G_{\tilde{w}} \preceq L_f I\}.$$

The set $\Omega_{\tilde{w}}$ is closed and bounded and the minimum eigenvalue of $Q_{\tilde{w}}^T Q_{\tilde{w}}$ is a continuous function of $G_{\tilde{w}}$ and $D_{\tilde{w}}$. Thus, the extreme value theorem [181, Theorem 4.14] implies that its infimum over $\Omega_{\tilde{w}}$,

$$\kappa_1 = \inf_{(D_{\tilde{w}}, G_{\tilde{w}}) \in \Omega_{\tilde{w}}} \lambda_{\min}\left(Q_{\tilde{w}}^T Q_{\tilde{w}}\right)$$

is achieved. By Lemma 4.1.2, $Q_{\tilde{w}}$ is a full rank matrix, which implies that $Q_{\tilde{w}}^T Q_{\tilde{w}} \succ 0$ for all $\tilde{w}$ and therefore that κ_1 is positive. Thus, $V(w) \geq \frac{\kappa_1}{2} \|\tilde{w}\|^2$, establishing condition (4.17).

Condition (4.9) and [183, Lemma 2.5] imply $V(w(t)) = \mathrm{e}^{-2t} V(w(0))$. It then follows from (4.17) that

$$\|w(t) - w^\star\|^2 \leq (\kappa_2/\kappa_1)\, \mathrm{e}^{-2t} \|w(0) - w^\star\|^2.$$

Taking the square root completes the proof and provides an upper bound for the constant κ, $\kappa \leq \sqrt{\kappa_2/\kappa_1}$. □

Remark 4.2.5. *The* rate *of exponential convergence established by Theorem 4.2.4 is independent of m_f, L_f, and μ. This is a consequence of insensitivity of Newton-like methods to poor conditioning. In contrast, the first order primal-dual method considered in [107] requires a sufficiently large μ for exponential convergence. In our second order primal-dual method, problem conditioning and parameter selection affect the multiplicative constant κ but not the rate of convergence.*

Remark 4.2.6. *When differential inclusion* (4.7) *is defined with the B-generalized Hessian* (4.4a)*, Theorems 4.2.1 and 4.2.4 hold even for proximal operators which are not semismooth. In this case, Theorem 4.2.1 is complete at equation* (4.14a) *and Theorem 4.2.4 applies without modification.*

4.3 A second order primal-dual algorithm

An algorithm based on the second order updates (4.5) requires step size selection to ensure global convergence. This is challenging for saddle point problems because standard notions, such as sufficient descent, cannot be applied to assess the progress of the iterates. Instead, it is necessary to identify a merit function whose minimum lies at the stationary point and whose

sufficient descent can be used to evaluate progress towards the saddle point.

An approach based on discretization of differential inclusion (4.7) and Lyapunov function (4.8) as a merit function leads to Algorithm 4 in Section 4.5.1. However, such a merit function is nonconvex and nondifferentiable in general which makes the utility of backtracking (e.g., the Armijo rule) unclear. Moreover, Algorithm 4 employs a fixed penalty parameter μ. *A priori* selection of this parameter is difficult and it has a large effect on the convergence speed.

Instead, we employ the primal-dual augmented Lagrangian introduced in [162] as a merit function and incorporate an adaptive μ update. This merit function is convex in both x and y and it facilitates an implementation with outstanding practical performance. Drawing upon recent advancements for constrained optimization of twice differentiable functions [163, 164], we show that our algorithm converges to the solution of (2.5). Finally, our algorithm exhibits local (quadratic) superlinear convergence for (strongly) semismooth $\mathbf{prox}_{\mu g}$.

4.3.1 Merit function

The primal-dual augmented Lagrangian,

$$\mathcal{M}_\mu(x, z; y, \lambda) := \mathcal{L}_\mu(x, z; \lambda) + \tfrac{1}{2\mu} \|Tx - z + \mu(\lambda - y)\|^2$$

was introduced in [162], where λ is an estimate of the optimal Lagrange multiplier $y^\star$. Following [162, Theorem 3.1], it can be shown that the optimal primal-dual pair $(x^\star, z^\star; y^\star)$ of optimization problem (2.5) is a stationary point of $\mathcal{M}_\mu(x, z; y, y^\star)$. Furthermore, for any fixed λ, $\mathcal{M}_\mu$ is a convex function of x, z, and y and it has a unique global minimizer.

In contrast to [162], we study problems in which a component of the objective function is not differentiable. As in Theorem 2.3.1, the Moreau envelope associated with the nondifferentiable component g allows us to eliminate the dependence of the primal-dual augmented Lagrangian $\mathcal{M}_\mu$ on z,

$$\begin{aligned} \hat{z}^\star_\mu(x; y, \lambda) &= \operatorname*{argmin}_z \mathcal{M}_\mu(x, z; y, \lambda) \\ &= \mathbf{prox}_{\frac{\mu}{2} g}(Tx + \tfrac{\mu}{2}(2\lambda - y)) \end{aligned}$$

and to express $\mathcal{M}_\mu$ as a continuously differentiable function,

$$\mathcal{M}_\mu(x; y, \lambda) := \mathcal{M}_\mu(x, \hat{z}^\star_\mu(x; y, \lambda); y, \lambda) = f(x) + M_{\frac{\mu}{2} g}\big(Tx + \tfrac{\mu}{2}(2\lambda - y)\big) + \tfrac{\mu}{4}\|y\|^2 - \tfrac{\mu}{2}\|\lambda\|^2.$$

For notational compactness, we suppress the dependence on λ and write $\mathcal{M}_\mu(w)$ when λ is fixed.

Remark 4.3.1. *The primal-dual augmented Lagrangian is not a Lyapunov function unless* $\lambda = y^\star$. *We establish convergence by minimizing* $\mathcal{M}_\mu(x; y, \lambda)$ *over* $(x; y)$ *– a convex problem – while adaptively updating the Lagrange multiplier estimate* λ.

In [162, 164], the authors obtain a search direction using the Hessian of the merit function, $\nabla^2 \mathcal{M}_\mu$. Instead of implementing an analogous update using generalized Hessian $\partial^2 \mathcal{M}_\mu$, we take advantage of the efficient inversion of $\partial^2 \mathcal{L}_\mu$ (see Section 4.3.2) to define the update

$$\partial^2 \mathcal{L}_\mu(w^k)\, \tilde{w} \;=\; -\, \text{blkdiag}(I, -I)\, \nabla \mathcal{M}_{2\mu}(w^k) \tag{4.18}$$

where the identity matrices are sized conformably with the dimensions of x and y, and

$$\nabla \mathcal{M}_{2\mu}(w) = \begin{bmatrix} \nabla f(x) + T^T \nabla M_{\mu g}(Tx + \mu(2\lambda - y)) \\ \mu(y - \nabla M_{\mu g}(Tx + \mu(2\lambda - y))) \end{bmatrix}. \tag{4.19}$$

Multiplication by $\text{blkdiag}(I, -I)$ is used to ensure descent in the dual direction and $\mathcal{M}_{2\mu}$ is employed because $\hat{z}^\star_\mu(x; y, \lambda)$ is determined by the proximal operator associated with $(\mu/2)g$. When $\lambda = y$, $\nabla_x \mathcal{M}_{2\mu} = \nabla_x \mathcal{L}_\mu$, $\nabla_y \mathcal{M}_{2\mu} = -\nabla_y \mathcal{L}_\mu$, and (4.18) becomes equivalent to the second order update (4.5).

Lemma 4.3.2. *Let* $\tilde{w}$ *solve* (4.18). *Then, for the fixed value of the Lagrange multiplier estimate* λ *and any* $\sigma \in (0, 1]$,

$$d \;:=\; (1 \,-\, \sigma)\, \tilde{w} \;-\; \sigma\, \nabla \mathcal{M}_{2\mu}(w, \lambda) \tag{4.20}$$

is a descent direction of the merit function $\mathcal{M}_{2\mu}(w, \lambda)$.

Proof. By multiplying (4.18) with the nonsingular matrix

$$\Pi \;:=\; \begin{bmatrix} I & -\frac{1}{\mu} T^T \\ 0 & I \end{bmatrix} \tag{4.21}$$

we can express it as

$$\begin{bmatrix} H & T^T \\ (I \,-\, P)T & -\mu P \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} -(\nabla f(x) + T^T y) \\ \nabla_y \mathcal{M}_{2\mu}(x; y, \lambda) \end{bmatrix} \tag{4.22}$$

where $H := \nabla^2 f(x) \succ 0$. Using (4.19) and (4.22), $\nabla \mathcal{M}_{2\mu}(w)$ can be expressed as,

$$\nabla \mathcal{M}_{2\mu}(w) = \begin{bmatrix} -(H + \frac{1}{\mu} T^T (I - P)T)\tilde{x} \; - \; T^T (I - P)\tilde{y} \\ (I \; - \; P)T\tilde{x} \; - \; \mu P \tilde{y} \end{bmatrix}.$$

Thus,

$$\tilde{w}^T \nabla \mathcal{M}_{2\mu}(w) \; = \; -\tilde{x}^T (H + \tfrac{1}{\mu} T^T (I \; - \; P)T)\tilde{x} \; - \; \mu \tilde{y}^T P \tilde{y}$$

is negative semidefinite, and the inner product

$$d^T \nabla \mathcal{M}_{2\mu}(w) \; = \; (1 - \sigma)\, \tilde{w}^T \nabla \mathcal{M}_{2\mu}(w) \; - \; \sigma \|\nabla \mathcal{M}_{2\mu}(w)\|^2$$

is negative definite when $\nabla \mathcal{M}_{2\mu}$ is nonzero. □

4.3.2 Second order primal-dual algorithm

We now develop a customized algorithm that alternates between minimizing the merit function $\mathcal{M}_\mu(x; y, \lambda)$ over $(x; y)$ and updating λ. Near the optimal solution, the algorithm approaches second order updates (4.5) with unit step size, leading to local (quadratic) superlinear convergence for (strongly) semismooth $\mathbf{prox}_{\mu g}$.

Our approach builds on the sequential quadratic programming method described in [162–164] and it uses the primal-dual augmented Lagrangian as a merit function to assess progress of iterates to the optimal solution. Inspired by [184], we ensure sufficient progress with damped second order updates.

The following two quantities

$$\begin{aligned} r \; &:= \; Tx \; - \; \mathbf{prox}_{\mu g}(Tx \; + \; \mu y) \\ s \; &:= \; Tx \; - \; \mathbf{prox}_{\mu g}(Tx \; + \; \mu\,(2\lambda \; - \; y)) \end{aligned}$$

appear in the proof of global convergence. Note that r is the primal residual of optimization problem (2.5) and that $\nabla \mathcal{M}_{2\mu}$ can be equivalently expressed as

$$\nabla \mathcal{M}_{2\mu}(w) \; = \; \begin{bmatrix} \nabla f(x) \; + \; \frac{1}{\mu}\, T^T (s + \mu(2\lambda - y)) \\ -(s \; + \; 2\mu(\lambda - y)) \end{bmatrix}. \tag{4.23}$$

Global convergence

We now establish global convergence of Algorithm 3 under an assumption that the sequence of gradients generated by the algorithm, $\nabla f(x^k)$, is bounded. This assumption is standard for augmented Lagrangian based methods [164,185] and it does not lead to a loss of generality when f is strongly convex.

Theorem 4.3.3. *Let Assumption 1 hold and let the sequence $\{\nabla f(x^k)\}$ resulting from Algorithm 3 be bounded. Then, the sequence of iterates $\left\{w^k\right\}$ converges to the optimal primal-dual point of problem* (2.5) *and the Lagrange multiplier estimates $\{\lambda^k\}$ converge to the optimal Lagrange multiplier.*

Proof. Since $\mathcal{V}_{2\mu}(w,\lambda)$ is convex in w for any fixed λ, condition (4.26) in Algorithm 3 will be satisfied after finite number of iterations. Combining (4.26) and (4.23) shows that $s^k + 2\mu^k(\lambda^k - y^k) \to 0$ and $\nabla f(x^k) + \frac{1}{\mu^k} T^T(s^k + \mu(2\lambda^k - y^k)) \to 0$. Together, these statements imply that the dual residual $\nabla f(x^k) + T^T y^k$ of (2.5) converges to zero.

To show that the primal residual r^k converges to zero, we first show that $s^k \to 0$. If Step 2*a* in Algorithm 3 is executed infinitely often, $s^k \to 0$ since it satisfies (4.25) at every iteration and $\eta \in (0,1)$. If Step 2*a* is executed finitely often, there is k_0 after which $\lambda^k = \lambda^{k_0}$. By adding and subtracting $2\mu^k \nabla f(x^k) + T^T s^k + 4\mu^k T^T(\lambda^{k_0} - y^k)$ and rearranging terms, we can write

$$\begin{aligned} T^T s^k \;=\; & 2\mu^k(\nabla f(x^k) \;+\; \tfrac{1}{\mu^k} T^T(s^k + \mu^k(2\lambda^{k_0} - y^k))) \\ & - \;2\mu^k \nabla f(x^k) \;-\; T^T(s^k + 2\mu^k(\lambda^{k_0} - y^k)) \;-\; 2\mu^k T^T \lambda^{k_0}. \end{aligned}$$

Taking the norm of each side and applying the triangle inequality, (4.26) and (4.23) yields

$$\|T^T s^k\| \;\leq\; 2\mu^k \epsilon^k + 2\mu^k \|\nabla f(x^k)\| + \|T^T\|\epsilon_k + 2\mu^k \|T^T \lambda^{k_0}\|. \tag{4.24}$$

This inequality implies that $T^T s^k \to 0$ because $\nabla f(x^k)$ is bounded, $\epsilon^k \to 0$, and $\mu^k \to 0$. Since T has full row rank, T^T has full column rank and it follows that $s^k \to 0$.

Substituting $s^k \to 0$ and $\nabla f(x^k) + T^T y^k \to 0$ into the first row of (4.23) and applying (4.26) implies $\lambda^k \to y^k$. Thus, $s^k \to r^k$, implying that the iterates asymptotically drive the primal residual r^k to zero, thereby completing the proof. □

Remark 4.3.4. *Despite the assumption that $\{\nabla f(x^k)\}$ is bounded, Theorem 4.3.3 can be used to ensure global convergence whenever f is strongly convex. We show in Lemma 4.5.1 in Section 4.5.2 that a bounded set $\mathcal{C}_f$ containing the optimal point can be identified* a priori. *One*

can thus artificially bound $\nabla f(x)$ for all $x \notin \mathcal{C}_f$ to satisfy the conditions of Theorem 4.3.3 and guarantee global convergence to the solution of (2.1).

Algorithm 3 Second order primal-dual algorithm for nonsmooth composite optimization.

input: Initial point $w^0 = (x^0, y^0)$, and parameters $\eta \in (0,1)$, $\beta \in (0,1)$, $\tau_a, \tau_b \in (0,1)$, $\epsilon^k \geq 0$ such that $\epsilon^k \to 0$.
initialize: Set $\lambda^0 = y^0$.

Step 1**:** If

$$\|s^k\| \;\leq\; \eta \|s^{k-1}\| \tag{4.25}$$

go to Step 2a. If not, go to Step 2b.

Step 2a**:** Set

$$\mu^{k+1} \;=\; \tau_a \mu^k, \qquad \lambda^{k+1} \;=\; y^k$$

Step 2b**:** Set

$$\mu^{k+1} \;=\; \tau_b \mu^k, \qquad \lambda^{k+1} \;=\; \lambda^k$$

Step 3**:** Using a backtracking line search, perform a sequence of inner iterations to choose w^{k+1} until

$$\|\nabla \mathcal{M}_{2\mu^{k+1}}(w^{k+1}, \lambda^{k+1})\| \;\leq\; \epsilon^k \tag{4.26}$$

where the search direction d is obtained using (4.20)–(4.22) with

$$\sigma \;=\; 0, \qquad \frac{(\tilde{w}^k)^T \nabla \mathcal{M}_{2\mu^{k+1}}(w^k)}{\|\nabla \mathcal{M}_{2\mu^{k+1}}(w^k)\|^2} \;\leq\; -\beta, \tag{4.27a}$$

$$\sigma \;\in\; (0,1], \qquad \text{otherwise.} \tag{4.27b}$$

Asymptotic convergence rate

The invertibility of the generalized Hessian $\partial^2 \mathcal{L}_\mu(w)$ allows us to establish local convergence rates for the second order updates (4.5) when $\mathbf{prox}_{\mu g}$ is (strongly) semismooth.

We now show that the updates in Algorithm 3 are equivalent to the second order updates (4.5) as $k \to \infty$. Thus, if $\mathbf{prox}_{\mu g}$ is (strongly) semismooth, the sequence of iterates generated by Algorithm 3 converges (quadratically) superlinearly to the optimal point in some neighborhood of it.

Theorem 4.3.5. *Let the conditions of Theorem 4.3.3 hold, let* $\mathbf{prox}_{\mu g}$ *be (strongly) semismooth, and let ϵ^k be such that $\|w^k - w^\star\| = O(\epsilon^k)$. Then, in a neighborhood of the optimal point $w^\star$, the iterates w^k converge (quadratically) superlinearly to $w^\star$.*

Proof. From Theorem 4.3.3, $\lambda^k \to y^k$ and thus $\nabla \mathcal{M}_{2\mu}(w^k) \to \nabla \mathcal{L}_\mu(w^k)$. Descent of the Lyapunov function in Theorem 4.2.1 therefore implies that the update in Step 3 of Algorithm 3 is

given by (4.27a), which is equivalent to (4.5) because $\lambda = y$. The assumption on $\{\epsilon^k\}$ in conjunction with Proposition 4.1.1, and [182, Theorem 3.2] imply that this update asymptotically satisfies (4.26) in one iteration with a unit step size. Therefore, Step 3 reduces to (4.5) in some neighborhood of the optimal solution and Proposition 4.1.1 implies that w^k converges to $w^\star$ (quadratically) superlinearly. □

Efficient computation of the Newton direction

When g is (block) separable, the matrix P in (4.4) is (block) diagonal. We next demonstrate that the solution to (4.22) can be efficiently computed when $T = I$ and P is a sparse diagonal matrix whose entries are either 0 or 1. The extensions to a low rank P, to a P with entries between 0 and 1, or to a general diagonal T follow from similar arguments.

These conditions occur, for example, when $g(z) = \gamma\|z\|_1$. The matrix P is sparse when $\mathbf{prox}_{\mu g}(x + \mu y) = \mathcal{S}_{\gamma\mu}(x + \mu y)$ is sparse. Larger values of γ are more likely to produce a sequence of iterates w^k for which P is sparse and thus the second order search directions (4.22) are cheaper to compute.

We can write (4.22) as

$$\begin{bmatrix} H & I \\ I - P & -\mu P \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} \vartheta \\ \theta \end{bmatrix}, \tag{4.28a}$$

permute it according to the entries of P which are 1 and 0, respectively, and partition the matrices H, P, and $I - P$ conformably such that

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12}^T & H_{22} \end{bmatrix}, \quad P = \begin{bmatrix} I & \\ & 0 \end{bmatrix}.$$

Let v denote either the primal variable x or the dual variable y. We use v_1 to denote the subvector of v corresponding to the entries of P which are equal to 1 and v_2 to denote the subvector corresponding to the zero diagonal entries of P.

Note that $(I - P)v = 0$ when $v_2 = 0$ and $Pv = 0$ when $v_1 = 0$. As a result, $\tilde{x}_2$ and $\tilde{y}_1$ are explicitly determined by the bottom row of the system of equations (4.28a),

$$\begin{bmatrix} 0 & -\mu I \\ I & 0 \end{bmatrix} \begin{bmatrix} \tilde{x}_2 \\ \tilde{y}_1 \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{4.28b}$$

Substitution of the subvectors $\tilde{x}_2$ and $\tilde{y}_1$ into (4.28a) yields,

$$H_{11}\tilde{x}_1 \;=\; \vartheta_1 \,+\, H_{12}\tilde{x}_2 \,+\, \tilde{y}_1 \tag{4.28c}$$

which must be solved for $\tilde{x}_1$. Finally, the computation of $\tilde{y}_2$ requires only matrix-vector products,

$$\tilde{y}_2 \;=\; -\left(\vartheta_2 \,+\, H_{21}\tilde{x}_1 \,+\, H_{22}\tilde{x}_2\right). \tag{4.28d}$$

Thus, the major computational burden in solving (4.22) lies in performing a Cholesky factorization to solve (4.28c), where H_{11} is a matrix of a much smaller size than H.

4.4 Computational experiments

In this section, we illustrate the merits and the effectiveness of our approach. We first apply our algorithm to the ℓ_1-regularized least squares problem and then study a system theoretic problem of controlling a spatially-invariant system. All computations were performed in Matlab R2014*b* on a 2012 Macbook Air with a 2GHz Intel Core i7 processor and 8GB of RAM. All random computational experiments are averaged over 20 trials.

4.4.1 Example: ℓ_1-regularized least squares

The LASSO problem (2.4) regularizes a least squares objective with a γ-weighted ℓ_1 penalty, As described in Section 2.2.2, the associated proximal operator is given by soft-thresholding $\mathcal{S}_{\gamma\mu}$, the Moreau envelope is the Huber function, and its gradient is the saturation function. Thus, $P \in \mathbb{P}$ is diagonal and P_{ii} is 0 when $|x_i + \mu y_i| < \gamma\mu$, 1 outside this interval, and between 0 and 1 on the boundary. Larger values of the regularization parameter γ induce sparser solutions for which one can expect a sparser sequence of iterates. Note that we require strong convexity of the least squares penalty; i.e., that A^TA is positive definite.

In Fig. 4.1, we show the distance of the iterates from the optimal for the standard proximal gradient algorithm ISTA, its accelerated version FISTA, and our customized second order primal-dual algorithm for a problem where A^TA has condition number 3.26×10^4. We plot distance from the optimal point as a function of both iteration number and solve time. Although our method always requires much fewer iterations, it is most effective when γ is large. In this case the most computationally demanding step (4.28c) required to determine the second order search direction (4.22) involves a smaller matrix inversion; see Section 4.3.2 for details. In Fig. 4.2,

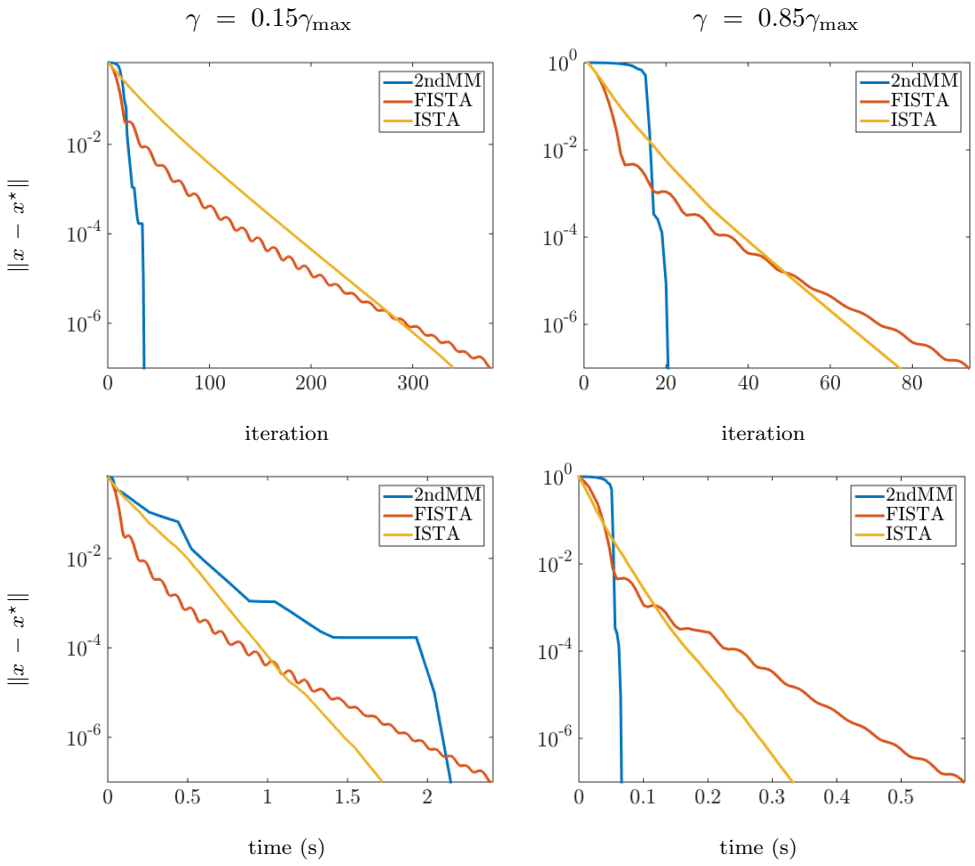


Figure 4.1: Distance from optimal solution as a function of the iteration number and solve time when solving LASSO for two values of γ using ISTA, FISTA, and our algorithm (2ndMM).

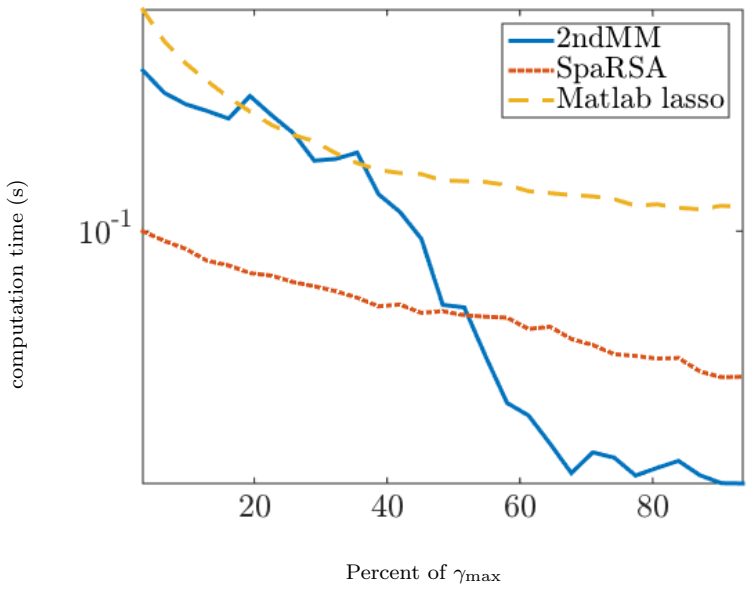


Figure 4.2: Solve times for LASSO with $n = 1000$ obtained using ISTA, FISTA, and our algorithm (2ndMM) as a function of the sparsity-promoting parameter γ.

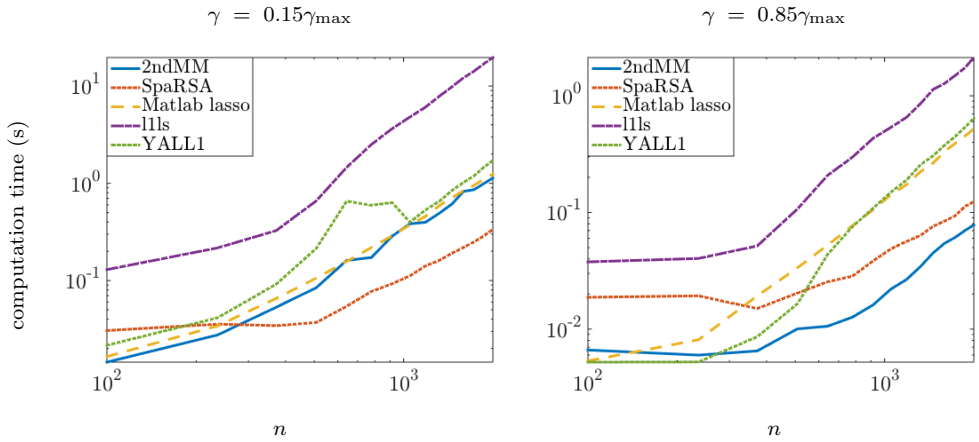


Figure 4.3: Comparison of our algorithm (2ndMM) with state-of-the-art methods for LASSO with problem dimension varying from $n = 100$ to 2000.

we show the solve times for $n = 1000$ as the sparsity-promoting parameter γ ranges from 0 to $\gamma_{\max} = \|A^T b\|$, where $\gamma_{\max}$ yields a zero solution. All numerical experiments consist of 20 averaged trials.

In Fig. 4.3, we compare the performance of our algorithm with the LASSO function in Matlab (a coordinate descent method [186]), SpaRSA [187], an interior point method [188], and YALL1 [189]. Problem instances were randomly generated with $A \in \mathbb{R}^{m \times n}$, n ranging from 100 to 2000, $m = 3n$, and $\gamma = 0.15\gamma_{\max}$ or $0.85\gamma_{\max}$. The solve times and scaling of our algorithm is competitive with these state-of-the-art methods. For larger values of γ, the second order search direction (4.22) is cheaper to compute and our algorithm is the fastest.

4.4.2 Example: Distributed control of a spatially-invariant system

We now apply our algorithm to a structured control design problem aimed at balancing closed-loop $\mathcal{H}_2$ performance with spatial support of a state-feedback controller. Following the problem formulation of [64], ADMM was used in [106, 190] to design sparse feedback gains for spatially-invariant systems. Herein, we demonstrate that our algorithm provides significant computational advantage over both the ADMM algorithm and a proximal Newton scheme.

Spatially-invariant systems

Let us consider

$$\begin{aligned} \dot{\psi} &= A\,\psi + u + d \\ \zeta &= \begin{bmatrix} Q^{1/2}\,\psi \\ R^{1/2}\,u \end{bmatrix} \end{aligned} \tag{4.29}$$

where ψ, u, d, and ζ are the system state, control input, white stochastic disturbance, and performance output and A, $Q \succeq 0$, and $R \succ 0$ are $n \times n$ circulant matrices. Such systems evolve over a discrete spatially-periodic domain; they can be used to model spatially-invariant vehicular platoons [97] and can result from a spatial discretization of fluid flows [38].

Any circulant matrix can be diagonalized via the discrete Fourier transform (DFT). Thus, the coordinate transformation $\psi := T\hat{\psi}$, $u := T\hat{u}$, $d := T\hat{d}$, where T^{-1} is the DFT matrix, brings the state equation in (4.29) into,

$$\dot{\hat{\psi}} = \hat{A}\,\hat{\psi} + \hat{u} + \hat{d}. \tag{4.30}$$

Here, $\hat{A} := T^{-1}AT$ is a diagonal matrix whose main diagonal $\hat{a}$ is determined by the DFT of

the first row a of the matrix A,

$$\hat{a}_k \;:=\; \sum_{i\,=\,0}^{n-1} a_i \, \mathrm{e}^{-\mathrm{j}\frac{2\pi i k}{n}}, \quad k \,\in\, \{0,\ldots,n-1\}.$$

We are interested in designing a structured state-feedback controller, $u = -Z\psi$, that minimizes the closed-loop $\mathcal{H}_2$ norm, i.e., the variance amplification from the disturbance d to the regulated output ζ. Since the optimal unstructured Z for spatially-invariant system (4.29) is a circulant matrix [9], we restrict our attention to circulant feedback gains Z. Thus, Z can also be diagonalized via a DFT and we equivalently take $x := \hat{z}$ as our optimization variable where $T^{-1}ZT = \mathrm{diag}\,(\hat{z})$.

For simplicity, we assume that A and Z are symmetric. In this case, $\hat{a}$ and $\hat{z}$ are real vectors and the closed-loop $\mathcal{H}_2$ norm of system (4.29) takes separable form $f_2(x) = \sum_{k\,=\,0}^{n-1} f^k(x_k)$,

$$f^k(x_k) \;=\; \begin{cases} \dfrac{\hat{q}_k \,+\, \hat{r}_k x_k^2}{2\,(x_k \,-\, \hat{a}_k)}, & x_k \,>\, \hat{a}_k \\ \qquad\qquad \infty, & \text{otherwise} \end{cases}$$

where $x_k > \hat{a}_k$ guarantees closed-loop stability. To promote sparsity of Z, we consider a regularized optimization problem (2.5),

$$\begin{array}{ll} \underset{x,\,z}{\text{minimize}} & f_2(x) \,+\, \gamma\,\|z\|_1 \\ \text{subject to} & Tx \,-\, z \,=\, 0 \end{array} \tag{4.31a}$$

where γ is a positive regularization parameter, T is the inverse DFT matrix, and $z \in \mathbb{R}^n$ denotes the first row of the symmetric circulant matrix Z. Formulation (4.31a) signifies that while it is convenient to quantify the $\mathcal{H}_2$ norm in the spatial frequency domain, sparsity has to be promoted in the physical space.

By solving (4.31a) over a range of γ, we identify distributed controller structures which are specified by the sparsity pattern of the solutions $z_\gamma^\star$ to (4.31a) at different values of γ. After selecting a controller structure associated with a particular value of γ, we solve a 'polishing' or 'debiasing' problem,

$$\begin{array}{ll} \underset{x,z}{\text{minimize}} & f_2(x) \,+\, I_{\mathrm{sp}(z_\gamma^\star)}(z) \\ \text{subject to} & Tx \,-\, z \,=\, 0 \end{array} \tag{4.31b}$$

where $I_{\mathrm{sp}(z_\gamma^\star)}(z)$ is the indicator function associated with the sparsity pattern of $z_\gamma^\star$. The solution to this problem is the optimal controller for system (4.29) with the desired structure, i.e., the

same sparsity pattern as $z_\gamma^\star$. This step is necessary because the ℓ_1 norm in (4.31a) imposes an additional penalty on z that compromises closed-loop performance.

Implementation

The elements of the gradient of f are

$$\frac{\mathrm{d}f^k(x_k)}{\mathrm{d}x_k} \;=\; \frac{\hat{r}_k x_k^2 \,-\, 2\hat{a}_k\hat{r}_k x_k \,-\, \hat{q}_k}{2\,(x_k \,-\, \hat{a}_k)^2}$$

the Hessian $\nabla^2 f_2$ is a diagonal matrix with non-zero entries,

$$\frac{\mathrm{d}^2 f^k(x_k)}{\mathrm{d}x_k^2} \;=\; \frac{\hat{q}_k \,+\, \hat{r}_k\hat{a}_k^2}{(x_k \,-\, \hat{a}_k)^3}$$

and the proximal operator associated with the nonsmooth regularizer in (4.31a) is given by soft-thresholding $\mathcal{S}_{\gamma\mu}$.

While the optimal unstructured controller can be obtained by solving n uncoupled scalar quadratic equations for x_k, sparsity-promoting problem (4.31a) is not in a separable form (because of the linear constraint) and computing the second order update (4.5) requires solving a system of equations

$$\begin{bmatrix} H & T^* \\ (I-P)T & -\mu P \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \;=\; \begin{bmatrix} \vartheta \\ \theta \end{bmatrix} \tag{4.32}$$

Pre-multiplying by the matrix $\mathrm{blkdiag}(T, I)$ and changing variables to solve for $\tilde{z} := T\tilde{x}$ brings (4.5) into

$$\begin{bmatrix} T\,H\,T^{-1} & \frac{1}{n}\,I \\ I-P & -\mu P \end{bmatrix} \begin{bmatrix} \tilde{z} \\ \tilde{y} \end{bmatrix} \;=\; \begin{bmatrix} T\vartheta \\ \theta \end{bmatrix}$$

which is of the same form as (4.28a). This equation can be solved efficiently when P is sparse (i.e., $\gamma\mu$ is large; cf. Section 4.3.2) and $\tilde{x}$ can be recovered from $\tilde{z}$ via FFT.

Since the Hessian H of a separable function f is a diagonal matrix, the search direction can also be efficiently computed when $I - P$ is sparse (i.e., $\gamma\mu$ is small). As in Section 4.3.2, the component of $\tilde{y}$ in the support of P is determined from the bottom row of (4.32). The top row of (4.32) implies $\tilde{x} = H^{-1}(\vartheta - T^*\tilde{y})$ and substitution into the bottom row yields

$$(I-P)\,T\,H^{-1}\,T^*\,(I-P)\,\tilde{y} \;=\; \tilde{\theta}$$

where $\tilde{\theta} := (I-P)(TH^{-1}(\vartheta - T^*P\tilde{y}) - \theta)$ is a known vector. Thus, the component of $\tilde{y}$ in

the support of $I - P$ can be determined by inverting a matrix whose size is determined by the support of $I - P$ and $\tilde{x}$ is readily obtained from $\tilde{y}$ and ϑ. The operations involving T and T^* can be performed via FFT; since H is diagonal, multiplication by these matrices is cheap and the computational burden in solving (4.5) again arises from a limited matrix inversion. In contrast to Section 4.3.2, the computation of the search direction using this approach is efficient when $I - P$ is sparse, i.e., $\gamma\mu$ is small.

Swift-Hohenberg equation

We consider the linearized Swift-Hohenberg equation [191],

$$\partial_t \psi(t,\xi) \;=\; (cI \;-\; (I \;+\; \partial_{\xi\xi})^2)\,\psi(t,\xi) \;+\; u(t,\xi) \;+\; d(t,\xi)$$

with periodic boundary conditions on a spatial domain $\xi \in [-\pi, \pi]$. Finite-dimensional approximation and diagonalization via the DFT (with an even number of Fourier modes n) yields (4.30) with $\hat{a}_k = c - (1 - k^2)^2$ where $k = \{-n/2 + 1, \ldots, n/2\}$ is a spatial wavenumber.

Figure 4.4a shows the optimal centralized controller and solutions to (4.31a) for $c = -0.01$, $n = 64$, $Q = R = I$, and $\gamma = 4 \times 10^{-4}$, 4×10^{-3}, and 4. As further illustrated in Fig. 4.4b, the optimal solutions to (4.31a) become sparser as γ is increased.

In Fig. 4.5, we demonstrate the utility of using regularized problems to navigate the tradeoff between controller performance and structure, an approach pioneered by [64]. The polished optimal structured controllers (**solid blue** —○—) were designed by first solving (4.31a) to identify an optimal structure and then solving (4.31b) to further improve the closed-loop performance. To illustrate the importance of polishing step (4.31b), we also show the closed-loop performance of unpolished optimal structured controllers (**dashed red - -×- -**) resulting from (4.31a). Finally, to evaluate the controller *structures* identified by (4.31a), we show the closed-loop performance of polished 'reference' structured controllers (**dotted yellow** $\cdots+\cdots$). Instead of solving (4.31a), the reference structures are *a priori* specified as nearest neighbor symmetric controllers of the same cardinality as controllers resulting from (4.31a). Among controllers with the same number of nonzero entries, the polished optimal structured controller consistently achieves the best closed-loop performance.

We compare the computational efficiency of our approach with the proximal Newton method [173] and ADMM [106, 190]. The proximal Newton method requires solving a LASSO subproblem (4.3), for which we employ SpaRSA [187]. Since A is circulant, the x-minimization step in ADMM (2.6) requires solving n uncoupled cubic scalar equations. In general, when A

is block circulant, the DFT only block diagonalizes the dynamics and thus the x-minimization step has to be solved via an iterative procedure [106, 190].

Figure 4.6 shows the time to solve (4.31a) with $\gamma = 0.004$ using our method, proximal Newton, and ADMM. Our algorithm and ADMM were stopped when the primal and dual residuals were below 1×10^{-8}. The proximal Newton method was stopped when the norm of the difference between two consecutive iterates was smaller than 1×10^{-8}. In Fig. 4.7, we show the per iteration cost and the total number of iterations required to find the optimal solution using each method. Our algorithm clearly outperforms proximal Newton and ADMM.

Although proximal Newton requires a similar number of iterations, the LASSO subproblem (4.3) that determines its search direction is much more expensive; this increases the computation cost of each iteration and slows the overall algorithm. Moreover, for larger problem sizes, the proximal Newton method struggles with finding a stabilizing search direction because $\nabla^2 f_2$ seems to bring it away from the set of stabilizing feedback gains. It appears that our method circumvents this issue because its iterates lie in a larger lifted space in which stability is easier to enforce via backtracking.

On the other hand, while the x- and z- minimization steps in ADMM are quite efficient, as a first order method ADMM requires a large number of iterations to reach high-accuracy solutions. Our algorithm achieves better performance because its use of second order information leads to relatively few iterations and the structured matrix inversion leads to efficient computation of the search direction.

4.5 Additional considerations

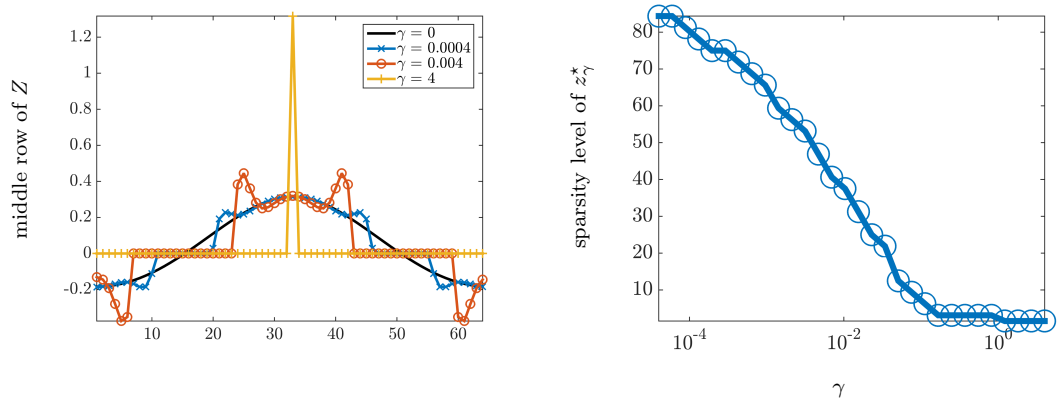
4.5.1 Algorithm based on $V(w)$ as a merit function

Since Theorem 4.2.4 establishes global convergence of the differential inclusion, one algorithmic approach is to implement a Forward Euler discretization of differential inclusion (4.7). A natural choice of merit function is the Lyapunov function V defined in (4.8). A simple corollary of Theorem 4.2.1 shows that the second order update (4.5) is a descent direction for V.

Corollary 4.5.0.1. *The second order update* (4.5) *is a descent direction for the merit function* V *defined in* (4.8).

Proof. Follows from (4.9) in Theorem 4.2.1. □

Corollary 4.5.0.1 enables the use of a backtracking Armijo rule for step size selection. A



(a) (b)

Figure 4.4: (a) The middle row of the circulant feedback gain matrix Z; and (b) the sparsity level of $z_\gamma^\star$ (relative to the sparsity level of the optimal centralized controller $z_0^\star$) resulting from the solutions to (4.31a) for the linearized Swift-Hohenberg equation with $n = 64$ Fourier modes and $c = -0.01$.

natural choice of stopping criterion for such an algorithm is a condition on the size of the primal and dual residuals. Moreover, proposition 4.1.1 suggests fast asymptotic convergence when $\mathbf{prox}_{\mu g}$ is semismooth. The LASSO example in Fig. 4.8 verifies this intuition when solving LASSO to a threshold of 1×10^{-8} for the primal and dual residuals.

Algorithm 4 Second order primal-dual algorithm for nonsmooth composite optimization based on discretizing (4.7).

input: Initial point x_0, y_0, backtracking constant $\alpha \in (0,1)$, Armijo parameter $\sigma \in (0,1)$, and stopping tolerances ε_1, ε_2.

While: $\|Tx^k - \mathbf{prox}_{\mu g}(Tx^k + \mu y^k)\| > \varepsilon_1$ or
$\|\nabla f(x^k) - T^T y^k\| > \varepsilon_2$

Step 1: Compute $\tilde{w}^k$ as defined in (4.5)

Step 2: Choose the smallest $j \in \mathbb{Z}_+$ such that

$$V(w^k + \alpha^j \tilde{w}^k) \;\leq\; V(w^k) \;-\; \sigma\, \alpha^j \, \|\nabla \mathcal{L}_\mu(w^k)\|^2$$

Step 3: Update the primal and dual variables

$$w^{k+1} \;=\; w^k \;+\; \alpha^j \tilde{w}^k$$

However, such an implementation would require a fixed penalty parameter μ, which typically has a large effect on the convergence speed of augmented Lagrangian algorithms and is difficult to select *a priori*. Moreover, stability of the solution to a differential equation does not always

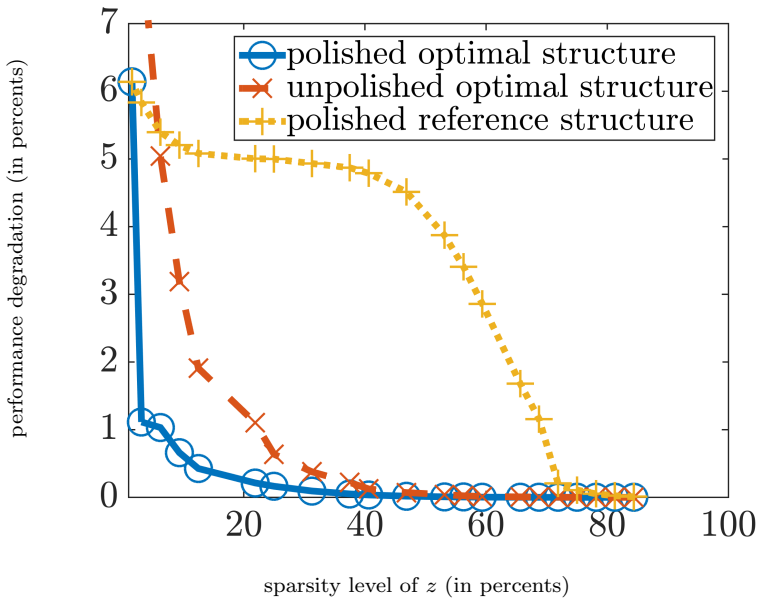


Figure 4.5: Performance degradation (in percents) of structured controllers relative to the optimal centralized controller: polished optimal structured controller obtained by solving (4.31a) and (4.31b) (**solid blue** —○—); unpolished optimal structured controller obtained by solving only (4.31a) (**dashed red - -×- -**); and optimal structured controller obtained by solving (4.31b) for an *a priori* specified nearest neighbor reference structure (**dotted yellow** $\cdots + \cdots$).

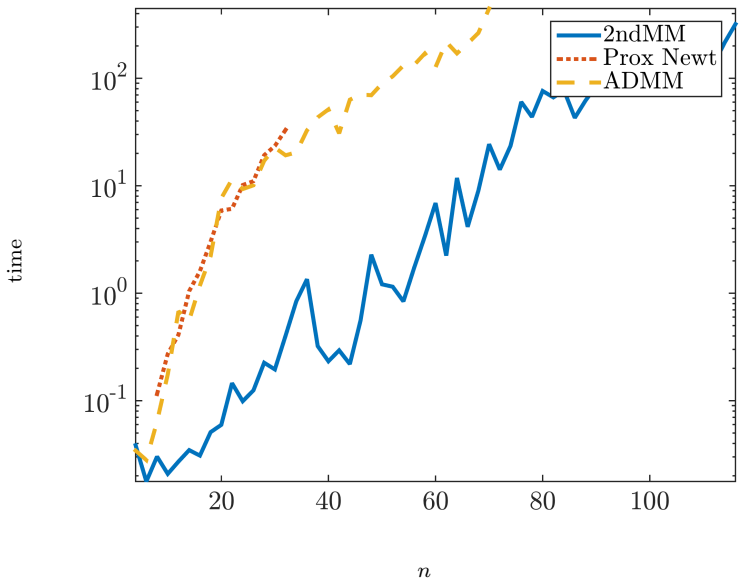


Figure 4.6: Total time to compute the solution to (4.31a) with $\gamma = 0.004$ using our algorithm (2ndMM), proximal Newton, and ADMM.

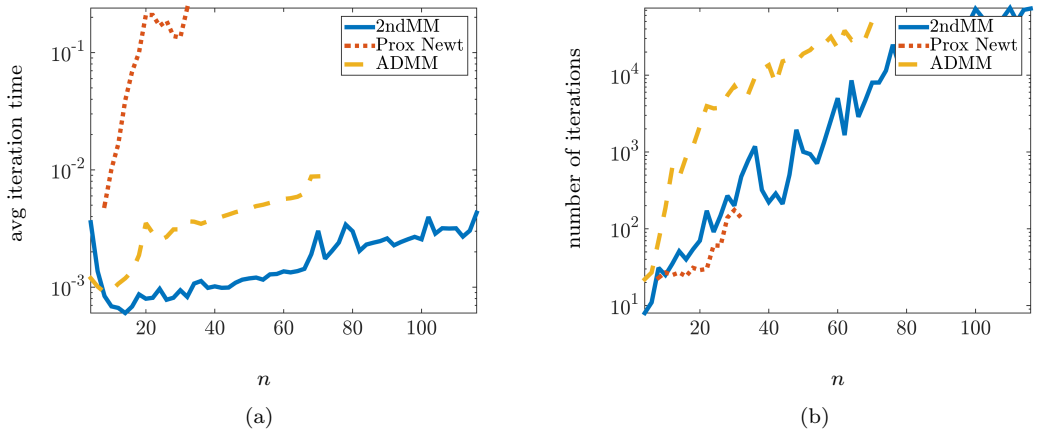


Figure 4.7: Comparison of (a) times to compute an iteration (averaged over all iterations); and (b) numbers of iterations required to solve (4.31a) with $\gamma = 0.004$.

imply stability of its discretization.

An example

We implement Algorithm 4 to solve the LASSO problem (2.4) studied in Section 4.4. The LASSO problem was randomly generated with $n = 500$, $m = 1000$, and $\gamma = 0.85\gamma_{\max}$. Figure 4.8 illustrates the quadratic asymptotic convergence of Algorithm 4 and a strong influence of μ.

4.5.2 Bounding $\nabla f(x)$ for Algorithm 3

A bounded set $\mathcal{C}_f$ containing the solution to (2.1) can always be identified from any point $\bar{x}$, $\nabla f(\bar{x})$, any element of the subgradient $\partial g(\bar{x})$, and a lower bound on the parameter of strong convexity.

Lemma 4.5.1. *Let Assumption 1 hold and let f be m_f-strongly convex. Then, for any $\bar{x}$, the optimal solution to* (2.1) *lies within the ball of radius $\frac{2}{m_f}\|\nabla f(\bar{x}) + T^T \partial g(T\bar{x})\|$ centered at $\bar{x}$.*

Proof. Given any point $\bar{x}$, strong convexity of f and convexity of g imply that,

$$f(x) + g(Tx) - f(\bar{x}) - g(T\bar{x}) \geq \left\langle \nabla f(\bar{x}) + T^T \partial g(T\bar{x}),\, x - \bar{x} \right\rangle + \tfrac{m_f}{2}\|x - \bar{x}\|^2 \qquad (4.33)$$

for all x where $\partial g(T\bar{x})$ is any member of the subgradient of $g(T\bar{x})$. For any x with $\|x - \bar{x}\| \geq \frac{2}{m_f}\|\nabla f(\bar{x}) + T^T \partial g(T\bar{x})\|$, the right-hand side of (4.33) must be nonnegative which implies that $f(x) + g(Tx) \geq f(\bar{x}) + g(T\bar{x})$ and thus that x cannot solve (2.1). □

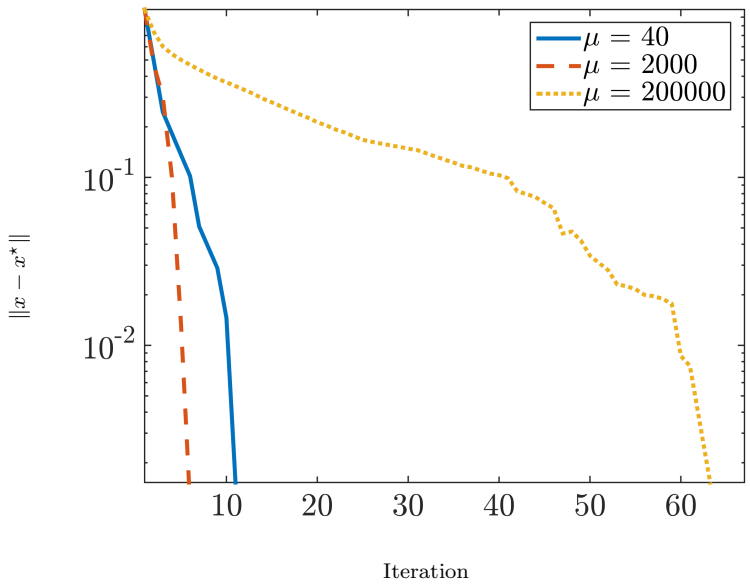


Figure 4.8: Distance from the optimal solution as a function of iteration number when solving LASSO using Algorithm 4 for different values of μ.

For any strongly convex function f and convex function g, a function $\hat{f}$ can be selected such that

$$\text{argmin } f(x) + g(Tx) = \text{argmin } \hat{f}(x) + g(Tx).$$

Here, $\hat{f}$ is identical to f in some closed set $\mathcal{C}_f$ containing $x^\star$,

$$\hat{f}(x) := \begin{cases} f(x), & x \in \mathcal{C}_f \\ \tilde{f}(x), & x \notin \mathcal{C}_f \end{cases}$$

and $\tilde{f}(x)$ is chosen such that $\hat{f}(x)$ is convex and twice differentiable, $\nabla \hat{f}(x)$ is uniformly bounded, and $\nabla^2 \hat{f}(x) \succ 0$. Lemma 4.5.1 implies that a set $\mathcal{C}_f$ that contains the optimal solution $x^\star$ of (2.1) can be identified *a priori* from any given point $\bar{x}$. Since $\hat{f}(x)$ satisfies all the conditions of Theorem 4.3.3, Algorithm 3 can be used to solve (2.1) and, since $\mathcal{C}_f$ contains $x^\star$, its optimal solution will also solve (2.1).

An example

When $x \in \mathbb{R}$, $f(x) = \frac{1}{2}x^2$, and $\mathcal{C}_f = [-1, 1]$, a potential choice of $\tilde{f}(x)$ is given by,

$$\tilde{f}(x) \;=\; \begin{cases} -2x \,+\, \mathrm{e}^{x+1} \,-\, 2.5, & x \,\leq\, -1 \\ 2x \,+\, \mathrm{e}^{-x+1} \,-\, 2.5, & x \,\geq\quad 1. \end{cases}$$

For this choice of $\tilde{f}$, the gradient of $\hat{f}$ is continuous and bounded,

$$\nabla \hat{f}(x) \;=\; \begin{cases} -2 + \mathrm{e}^{x+1}, & x \,\leq\, -1 \\ \qquad\quad x, & x \,\in\, [-1, 1] \\ 2 - \mathrm{e}^{-x+1}, & x \,\geq\, 1 \end{cases}$$

and the Hessian of $\hat{f}$ is determined by

$$\nabla^2 \hat{f}(x) \;=\; \begin{cases} \mathrm{e}^{x+1}, & x \,\leq\, -1 \\ \qquad 1, & x \,\in\, [-1, 1] \\ \mathrm{e}^{-x+1}, & x \,\geq\, 1. \end{cases}$$

Chapter 5

Connections with other methods and discussion

The proximal augmented Lagrangian $\mathcal{L}_\mu(x;y)$ is obtained by constraining $\mathcal{L}_\mu(x,z;y)$ to the manifold

$$\begin{aligned}\mathcal{Z} &:= \{(x, z_\mu^\star; y) \mid z_\mu^\star = \operatorname*{argmin}_z \mathcal{L}_\mu(x,z;y)\} \\ &= \{(x, z_\mu^\star; y) \mid Tx + \mu y \in z_\mu^\star + \mu\, \partial_C g(z_\mu^\star)\}\end{aligned}$$

which results from the explicit minimization over the auxiliary variable z. Herein, we interpret the second order search direction as a linearized update to the KKT conditions for problem (2.5) and discuss connections to the alternative algorithms.

5.1 Second order updates as linearized KKT corrections

The second order update (4.5) can be viewed as a first order correction to the KKT conditions for optimization problem (2.5),

$$\begin{aligned}0 &= \nabla f(x) + T^T y \\ 0 &= Tx - z \\ 0 &\in \partial_C g(z) - y.\end{aligned} \tag{5.1}$$

Substitution of $z_\mu^\star$ into (5.1) makes the last two conditions redundant; when combined with the definition of the manifold $\mathcal{Z}$, $Tx = z$ implies $y \in \partial_C g(z)$ and $y \in \partial_C g(z)$ implies $Tx = z$.

Multiplying equation (4.5) for the second order update with the nonsingular matrix (4.21) recovers the equivalent expression

$$\begin{bmatrix} H & T^T \\ (I - P)T & -\mu P \end{bmatrix} \begin{bmatrix} \tilde{x}^k \\ \tilde{y}^k \end{bmatrix} = - \begin{bmatrix} \nabla f(x^k) + T^T y^k \\ r^k \end{bmatrix} \tag{5.2}$$

where $r^k := Tx^k - z^\star_\mu(x^k; y^k) = Tx^k - \mathbf{prox}_{\mu g}(Tx^k + \mu y^k)$ is the primal residual of (2.5). Thus, (5.2) describes a first order correction to the first and second KKT conditions in (5.1).

5.2 Connections with other methods

We now discuss broader implications of our framework and draw connections to the existing methods for solving versions of (2.1). Many techniques for solving composite minimization problems of the form (2.1) can be expressed in terms of functions embedded in the augmented Lagrangian; see Table 5.1. Trivially, the objective function in (2.1) corresponds to $\mathcal{L}_\mu(x, z; y)$ over the manifold $z = Tx$. The Lagrange dual of a problem equivalent to (2.5),

$$\begin{array}{ll} \underset{x,\, z}{\text{minimize}} & f(x) \,+\, g(z) \,+\, \frac{1}{2\mu} \|Tx - z\|^2 \\ \text{subject to} & Tx - z \,=\, 0 \end{array} \tag{5.3}$$

is recovered by collapsing $\mathcal{L}_\mu(x, z; y)$ onto the intersection of the z-minimization manifold $\mathcal{Z}$ and the x-minimization manifold,

$$\begin{aligned} \mathcal{X} \;&:=\; \{(x^\star, z; y) \mid x^\star \,=\, \underset{x}{\text{argmin}}\; \mathcal{L}_\mu(x, z; y)\} \\ &=\; \{(x^\star, z; y) \mid \mu\,(\nabla f(x^\star) \,+\, T^T y) \,=\, T^T(z \,-\, Tx^\star)\}. \end{aligned}$$

The Lagrange dual of (2.5) is recovered from the Lagrange dual of (5.3) in the limit $\mu \to \infty$.

5.2.1 MM and ADMM

MM implements gradient ascent on the dual of (5.3) by collapsing $\mathcal{L}_\mu(x, z; y)$ computationally onto $\mathcal{X} \cap \mathcal{Z}$. The joint (x, z)-minimization step in (2.7) evaluates the Lagrange dual at discrete iterates y^k by finding the corresponding (x, z)-pair on $\mathcal{X} \cap \mathcal{Z}$; i.e., the iterate $(x^{k+1}, z^{k+1}; y^k)$ generated by (2.7) lies on the manifold $\mathcal{X} \cap \mathcal{Z}$. Note that, in this form, joint (x, z)-minimization is a challenging nondifferentiable optimization problem.

ADMM avoids this challenge by collapsing $\mathcal{L}_\mu(x, z; y)$ onto $\mathcal{X}$ and $\mathcal{Z}$ separately. While

the underlying x- and z-minimization subproblems in ADMM are relatively simple, the iterate $(x^{k+1}, z^{k+1}; y^k)$ generated by (2.6) does not typically lie on the $\mathcal{X} \cap \mathcal{Z}$ manifold. Thus ADMM does not represent gradient ascent on the dual of (5.3), causing looser theoretical guarantees and often worse practical performance.

By collapsing the augmented Lagrangian onto $\mathcal{Z}$, the proximal augmented Lagrangian (2.8) allows us to express the (x, z)-minimization step in MM as a tractable continuously differentiable problem (cf. Theorem 2.3.1). This avoids challenges associated with ADMM and it does not increase computational complexity in the x-minimization step in (2.7) relative to ADMM when using first order methods. We finally note that unlike the Rockafellar's *proximal method of multipliers* [192] which applies the proximal point algorithm to the primal-dual optimality conditions, our framework *reformulates* the *standard method of multipliers* and develops second order algorithm to solve nonsmooth composite optimization problems.

5.2.2 First order methods

A special instance of our framework has strong connections with the existing methods for distributed optimization on graphs; e.g., [146, 147, 193]. The networked optimization problem of minimizing $f(\bar{x}) = \sum f^i(\bar{x})$ over a single variable $\bar{x}$ can be reformulated as $\sum f^i(x_i) + g(Tx)$ where the components f^i of the objective function are distributed over independent agents x_i, x is the aggregate state, T^T is the incidence matrix of a strongly connected and balanced graph, and g is the indicator function associated with the set $Tx = 0$. The $g(Tx)$ term ensures that at feasible points, $x_i = x_j = \bar{x}$ for all i and j. It is easy to show that $\nabla M_{\mu g}(v) = \frac{1}{\mu}\, v$ and that the primal-descent dual-ascent dynamics 3.1 are given by,

$$\begin{aligned} \dot{x} &= -\nabla f(x) - \tfrac{1}{\mu} L\, x - \tilde{y} \\ \dot{\tilde{y}} &= L\, x \end{aligned} \tag{5.4}$$

where $\tilde{y} := T^T y$ and $L := T^T T$ is the graph Laplacian. The only difference relative to [193, Equation 11] is that $-\tilde{y}$ appears instead of $-L\tilde{y}$ in the equation for the dynamics of the primal variable x.

We also include a comparison with the EXTRA algorithm [194]. The discretized version of the first order primal-dual dynamics (5.4) is equivalent to

$$\begin{aligned} x^{k+1} &= x^k - \alpha_x(\nabla f(x^k) + \tfrac{1}{\mu} L x^k + \bar{y}^k) \\ \bar{y}^{k+1} &= \bar{y}^k + \alpha_y L x^k \end{aligned} \tag{5.5}$$

x	z	y	**function**	**methods**
$z = Tx$		-	objective function of (2.1)	subgradient iteration [134]
-	-	-	augmented Lagrangian	ADMM [64, 132]
$\mathcal{X}$	$\mathcal{Z}$	-	Lagrange dual of (5.3)	Dual ascent (if explicit expression available)
-	$\mathcal{Z}$	-	proximal augmented Lagrangian	MM [107], Arrow-Hurwicz-Uzawa [107], **Second order primal-dual**
-	$\mathcal{Z}$	$\mathcal{Y}$	Forward-Backward Envelope	Forward-Backward Truncated Newton [176], related to proximal gradient [176, Section 2.1]

Table 5.1: Summary of different functions embedded in the augmented Lagrangian of (2.5) and methods for solving (2.1) based on these functions.

where $\bar{y} := T^T y$, $L := T^T T$ is the graph Laplacian, and α_x and α_y are the primal and dual step sizes, respectively. By noting that that $y^k = \sum_{t=0}^{k-1}(W - \tilde{W})x^t$, the EXTRA algorithm [194, Equation (2.13)],

$$x^{k+1} \;=\; Wx^k \;-\; \alpha\nabla f(x^k) \;+\; \tfrac{1}{\mu}Lx^k \;+\; \sum_{t=0}^{k-1}(W - \tilde{W})x^t$$

is recovered exactly from (5.5) by setting $\alpha_x = \alpha$ and $\alpha_y = \frac{\alpha}{2\mu}$ and taking $W = I - \frac{\alpha}{\mu}L$, $\tilde{W} = \frac{1}{2}(I + W)$. Of course, EXTRA does not require that W be stated in the form of a Laplacian, but we can take $T = (I - W)^{1/2}$ and still recover EXTRA.

5.2.3 Second order methods

We identify the saddle point of $\mathcal{L}_\mu(x, z; y)$ by forming second order updates to x and y along the $\mathcal{Z}$ manifold. Just as Newton's method approximates an objective function with a convex quadratic function, we approximate $\mathcal{L}_\mu(x; y)$ with a quadratic saddle function.

Constraining the dual variable itself yields connections with other methods. When $T = I$, (5.1) implies that the optimal dual variable is given by $y^\star = -\nabla f(x^\star)$, so it is natural to collapse the augmented Lagrangian onto the manifold

$$\mathcal{Y} \;:=\; \{(x, z; y^\star) \mid y^\star \;=\; -\nabla f(x)\}.$$

The augmented Lagrangian over the manifold $\mathcal{Z} \cap \mathcal{Y}$ corresponds to the Forward-Backward Envelope (FBE) introduced in [176]. The proximal gradient algorithm with step size μ can be

recovered from a variable metric gradient iteration on the FBE [176]. In [176–178], the approximate line search and quasi-Newton methods based on the FBE were developed to solve (2.1) with $T = I$. Since the Hessian of the FBE involves third order derivatives of f, these techniques employ either truncated- or quasi-Newton methods.

The approach advanced in the current paper applies a second order method to the augmented Lagrangian that is constrained over the larger manifold $\mathcal{Z}$. Relative to alternatives, our framework offers several advantages. First, while the FBE is in general a nonconvex function of x, $\mathcal{L}_\mu(x; y)$ is always convex in x and concave in y. Furthermore, we can compute the Hessian exactly using only second order derivatives of f and its structure allows for efficient computation of the search direction. Finally, our formulation allows us to leverage recent advances in second order methods for augmented Lagrangian methods, e.g., [162–164].

Part II

Structured optimal control

Chapter 6

Symmetry and spatial invariance

In this chapter, we propose a principled convex approach for structured $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal controller design. We consider problems where the dynamical generator is an affine function of the design variable. When the dynamical generator is symmetric, we show that the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of the closed-loop system are convex functions of the design variable. This allows us to impose regularization penalties directly on the design variable in order to promote structure.

We then show that even when the dynamical generator is not symmetric, its symmetric component can be used to provide an upper bound on its $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norm. Thus, we can design structured controllers by solving the convex design problem for the symmetric component of the system and implement the resulting controllers on the original system. We show that this procedure guarantees stability and a level of $\mathcal{H}_2/\mathcal{H}_\infty$ performance.

Although convex, the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms for symmetric systems are expressed using linear matrix inequalities (LMIs) and the semidefinite programs (SDPs) that result from the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control problems do not scale favorably with the problem dimension. To address this challenge, we develop a customized ADMM algorithm and provide a procedure to gain computational advantage when the system and its controller are jointly block-diagonalizable. Such a structure arises, for example, in spatially-invariant systems [9]. In [190], we took advantage of this property to develop an efficient and scalable algorithm for sparsity-promoting $\mathcal{H}_2$ optimal feedback design.

6.1 Symmetric systems

We consider the class of systems (1.1) with $R(x) = 0$ and $B_1 = I$,

$$\begin{aligned} \dot{\psi} &= (A + F(x))\,\psi + d \\ \zeta &= \psi. \end{aligned} \tag{6.1}$$

Instead of limiting the standard measure of control effort, we regularize x directly by imposing a quadratic penalty, $x^T R x$ with $R \succ 0$, to limit the magnitude of x, and an ℓ_1 penalty, $\|x\|_1 := \sum_i |x_i|$, to promote sparsity in x. This is because in many motivating examples for this section, it is of interest to regulate the size of the controller x itself and not its effect on the system state, $R(x)\psi$. For example, in the combination drug therapy example first described in Section 1.2.1, it makes sense to penalize the size of the drug dose but not the quantity of virus each drug is killing.

Under the above assumptions, we can design controllers for (6.1) by designing controllers for

$$\begin{aligned} \dot{\psi} &= (A_s + F_s(x))\,\psi + d \\ \zeta &= \psi \end{aligned} \tag{6.2}$$

where $A_s := (A + A^T)/2$ is the symmetric part of A and $F_s(x) := (F(x) + F^T(x))/2$. Since any matrix A can be decomposed into its symmetric A_s and antisymmetric components, any system (6.1) can be placed into the form of (6.2)

We first show that the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control problems for (6.2) are convex in x and thus that we can impose regularization direclty on the controller. We then show that stability of the symmetric system (6.2) implies stability of the corresponding original system (6.1) and that the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of the symmetric system (6.2) are an upper bound on the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of the original system (6.1). We then show that when the difference between A and A_s of the order ϵ, the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of systems (6.1) and (6.2) differ only by $O(\epsilon^2)$.

We note that this method is more appropriate when $F(x)$ is symmetric. If $F_s(x) \neq F(x)$, the neglected effect of the asymmetric component of $F(x)$ makes the degree of conservatism unpredictable.

6.1.1 Convex formulation

The $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of (6.2) can be expressed in a convex form. Although more general symmetric systems can be cast as convex problems, here we assume $B = C = I$ and $R(x) = 0$

to facilitate the transition to the discussion of spectral properties and performance bounds.

$\mathcal{H}_2$-optimal control

The $\mathcal{H}_2$ norm of system (6.2) is given by trace (P_s) where,

$$\big(A_s \ + \ F_s(x)\big)P_s \ + \ P_s\big(A_s \ + \ F_s(x)\big) \ + \ I \ = \ 0.$$

Since $A_s \ = \ A_s^T$ is symmetric, the controllability gramian of system (6.2) can be explicitly expressed as,

$$P_s \ = \ -\tfrac{1}{2}\ (A_s \ + \ F_s(x))^{-1}$$

and, by taking a Schur complement, the regularized optimal $\mathcal{H}_2$ control problem can be cast in a convex function of x and an auxiliary variable Θ,

$$\begin{array}{ll} \underset{x,\Theta}{\text{minimize}} & \tfrac{1}{2}\text{trace}(\Theta) \ + \ g(x) \\ \text{subject to} & \begin{bmatrix} \Theta & I \\ I & -(A_s \ + \ F_s(x)) \end{bmatrix} \succ 0. \end{array} \tag{6.3}$$

The matrix $A_s + F_s(x)$ is always invertible when it is Hurwitz. We note the structured LQR problem (i.e., $R(x) = R^{1/2}F_s(x)$) for symmetric systems can also be expressed as an SDP by taking the Schur complement of $F_s(x)RF_s(x)$. Finally, we drop the constant factor $1/2$ for compactness.

$\mathcal{H}_\infty$-optimal control

To formulate the $\mathcal{H}_\infty$ optimal control problem as an SDP, we first state a simple lemma about system (6.2).

Proposition 6.1.1. *The disturbance that achieves the maximum induced $\mathcal{L}_2$ amplification for system* (6.2) *corresponds to the constant signal $d(t) = v$ where v is the right principal singular vector of A_s^{-1}. I.e., the peak of the frequency response of system* (6.2) *occurs at the temporal frequency $\omega = 0$.*

Proof. A symmetric matrix can be diagonalized as $A_s = U\Lambda U^T$ where Λ is a diagonal matrix with the eigenvalues of A_s on the main diagonal and the columns of U contain the corresponding

eigenvectors. For such a matrix,

$$(\mathrm{j}\omega I \ - \ A_s)^{-1} \ = \ U \ \mathrm{diag}\left\{\tfrac{1}{\mathrm{j}\omega - \lambda_i}\right\} U^T.$$

It is clear that $\omega = 0$ maximizes the singular values of the above matrix. Thus, the $\mathcal{H}_\infty$ norm of (6.2) can be characterized by $\sigma_{\max}\left(-(A_s \ + \ F_s(x))^{-1}\right)$. □

The regularized $\mathcal{H}_\infty$-optimal control problem for symmetric systems can therefore be expressed as,

$$\begin{aligned} \underset{x,\Theta}{\text{minimize}} \quad & \sigma_{\max}(\Theta) \ + \ g(x) \\ \text{subject to} \quad & \begin{bmatrix} \Theta & I \\ I & -(A_s + F_s(x)) \end{bmatrix} \succeq 0. \end{aligned} \tag{6.4}$$

As we show in the next section, this convex problem can be used for structured $\mathcal{H}_\infty$ control design. This is particularly advantageous because many of the existing algorithms for general structured $\mathcal{H}_2$ control cannot be extended to the structured $\mathcal{H}_\infty$ problem because the $\mathcal{H}_\infty$ norm is not differentiable in general.

In Section 6.1.2, we show that stability of the symmetric system (6.2) implies stability of the corresponding original system (6.1). Furthermore, the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of the symmetric system are upper bounds on the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of the original system. Finally, when the difference between A and A_s is small (of the order ε, $\mathcal{O}(\varepsilon)$), the $\mathcal{H}_2/\mathcal{H}_\infty$ norms of systems (6.1) and (6.2) differ only by $\mathcal{O}(\varepsilon^2)$.

This formulation can be applied to the design of edges in consensus networks, leader selection, and combination drug therapy design for HIV.

6.1.2 Stability and performance

We now justify the use of the symmetric component (6.2) to design controllers for the original system (6.1). First, we show that the stability of the symmetric system implies stability of the full system.

Lemma 6.1.1. *Let the symmetric part of A be Hurwitz. Then, A is Hurwitz.*

Proof. We show this by contradiction. Since the symmetric part of A, $A_s := (A + A^T)/2$ is symmetric and Hurwitz, it is negative definite,

$$v^* A_s v \ < \ 0 \quad \text{for all } v \neq 0.$$

Assume that A is not Hurwitz. Then there is a v such that $Av = \lambda v$ with $\Re(\lambda) \geq 0$. Furthermore, $v^*Av = \lambda v^*v$. However,

$$\begin{aligned} v^*Av &= v^*A_s v \ + \ \tfrac{1}{2}\, v^* \left(A \ - \ A^T\right) v \\ &= v^*A_s v \ + \ \mathrm{j}\,\mathrm{Im}(\lambda)\ \|v\|_2^2. \end{aligned}$$

Since $A_s \prec 0$, $v^*A_s v$ cannot have a nonnegative real part. □

Remark 6.1.2. *Lemma 6.1.1 is not a necessary condition; A may be Hurwitz even if A_s is not.*

6.1.3 Performance bounds

We next show that the $\mathcal{H}_2$ norm of the symmetric system (6.2) is an upper bound on the $\mathcal{H}_2$ norm of the system (6.1). First, we present a useful theorem from linear algebra [195].

Theorem 6.1.3 (Theorem IX.3.1 in [195])**.** *Let A be any matrix and let $A_s = \frac{1}{2}(A + A^T)$ be the symmetric component of A. Then*

$$\|\mathrm{e}^{A}\| \ \leq \ \|\mathrm{e}^{A_s}\|$$

for every unitarily invariant norm.

The statement about the $\mathcal{H}_2$ norms of systems (6.1) and (6.2) is a simple corollary of Theorem 6.1.3.

Corollary 6.1.3.1. *When the systems* (6.1) *and* (6.2) *are stable, the $\mathcal{H}_2$ norm of* (6.1) *is bounded from above by the $\mathcal{H}_2$ norm of system* (6.2)*.*

Proof. Recall that the $\mathcal{H}_2$ norm of a system,

$$\dot{\psi} \ = \ A\psi \ + \ d$$

with A Hurwitz is given by $\mathrm{trace}(P_c)$ where

$$AP_c \ + \ P_cA^T \ + \ I \ = \ 0$$

and that X can be expressed as

$$P_c \ = \ \int_0^\infty \mathrm{e}^{At}\mathrm{e}^{A^T t}\mathrm{d}t.$$

Using the linearity of the trace and of integration, we can rewrite the expression for the $\mathcal{H}_2$ norm as,

$$\text{trace}\left(\int_0^\infty \mathrm{e}^{At}\mathrm{e}^{A^T t}\mathrm{d}t\right) \;=\; \int_0^\infty \|\mathrm{e}^{At}\|_F^2 \mathrm{d}t.$$

Since the Frobenius norm is unitarily invariant, by Theorem 6.1.3 $\|\mathrm{e}^{At}\|_F^2 \leq \|\mathrm{e}^{A_s t}\|_F^2$ for any t and therefore,

$$\int_0^\infty \|\mathrm{e}^{At}\|_F^2 \;\mathrm{d}t \;\leq\; \int_0^\infty \|\mathrm{e}^{A_s t}\|_F^2 \;\mathrm{d}t.$$

Since the right hand side is the $\mathcal{H}_2$ norm of system (6.2), this completes the proof. □

Note: Theorem 6.1.3 relies on the fact that the sum of the k largest eigenvalues of P_s are larger than the sum of the k largest eigenvalues of P for any integer k. As a result, Corollary 6.1.3.1 does not extend to a general state penalty matrix Q where the $\mathcal{H}_2$ norm would be trace(QX). For the same reason, the result requires $\mathbb{E}(dd^T) = I$.

We now show that an analogous bound holds for the $\mathcal{H}_\infty$ norm.

Proposition 6.1.2. *When the systems* (6.1) *and* (6.2) *are stable, the $\mathcal{H}_\infty$ norm of the general system* (6.1) *is bounded from above by the $\mathcal{H}_\infty$ norm of the symmetric system* (6.2).

Proof. From the bounded real lemma [49], the $\mathcal{H}_\infty$ norm of the general system (6.1) is less than ζ if there exists a $P \succ 0$ such that,

$$A^T P \;+\; PA \;+\; I \;+\; \zeta^{-2} P^2 \;\prec\; 0. \tag{6.5}$$

From Proposition 6.1.1, the $\mathcal{H}_\infty$ norm of symmetric system (6.2) is $\zeta = \sigma_{\max}(A_s^{-1})$. Taking $P = \zeta I$ for any $\zeta > \sigma_{\max}(A_s^{-1})$ satisfies the above LMI; substituting P and A_s into (6.5) yields,

$$2\zeta A_s \;+\; 2I \;\prec\; 0.$$

Since A_s is Hurwitz, $A_s \prec 0$. Since $\zeta > -\lambda_{\max}(A_s^{-1})$, $\zeta^{-1} < -\lambda_{\min}(A_s)$, so $A_s \prec -\zeta^{-1} I$. Therefore the LMI is satisfied. Since $A_a = -A_a^T$, setting $P = \zeta I$ implies,

$$A^T P \;+\; PA \;=\; 2\zeta A_s.$$

Substituting A and $P = \zeta I$ into (6.5) yields,

$$A^T P \;+\; PA \;+\; I \;+\; \zeta^{-2} P^2 \;=\; 2\zeta A_s \;+\; 2I \;\prec\; 0.$$

□

6.1.4 Small asymmetric perturbations

We next show that in addition to being an upper bound, the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms of the symmetric and full systems are close when A is nearly normal. In what follows, we show that when a normal system is subject to an antinormal perturbation of $\mathcal{O}(\varepsilon)$, the first order correction to the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms is zero. A similar but more restrictive result appeared in [117] for the design of an interconnection graph for synchronizing oscillator networks. We present a result for systems with normal dynamical generators A_n. Since a normal matrix A_n is a matrix that commutes with A_n^T, this set includes symmetric matrices A_s.

Proposition 6.1.3. *Let A_n be a normal matrix. The $\mathcal{O}(\varepsilon)$ correction to the $\mathcal{H}_2$ norm of the system*

$$\dot{x} \;=\; A_n x \;+\; d$$

from an $\mathcal{O}(\varepsilon)$ antisymmetric perturbation A_a is zero.

Proof. The $\mathcal{H}_2$ norm of the above system is given by $\text{trace}(P_n)$ where,

$$A_n P_n \;+\; P_n A_n^T \;+\; I \;=\; 0. \tag{6.6}$$

From Lemma 1 in [196], $P_n = -(A_n + A_n)^{-1}$. Perturbing A_n by a small antisymmetric matrix εA_a yields a small correction term $\varepsilon\tilde{P}$ in the controllability gramian. Collecting the $\mathcal{O}(\varepsilon)$ terms from the Lyapunov equation,

$$(A_n \;+\; \varepsilon A_a)(P_n \;+\; \varepsilon\tilde{P}) \;+\; (P_n \;+\; \varepsilon\tilde{P})(A_n \;+\; \varepsilon A_a)^T \;+\; I \;=\; 0.$$

recovers the linear equation,

$$A_n\tilde{P} \;+\; \tilde{P}A_n \;+\; A_a P_n \;+\; P_n A_a^T \;=\; 0.$$

The $\mathcal{O}(\varepsilon)$ correction to the $\mathcal{H}_2$ norm is given by,

$$\begin{aligned}
\text{trace}(\tilde{P}) &= \text{trace}\left(P_n\left(A_a P_n \;+\; P_n A_a^T\right)\right) \\
&= \text{trace}\left(\left(A_a \;-\; A_a\right)P_n^2\right) \\
&= 0
\end{aligned}$$

since $A_a = -A_a^T$. □

To illustrate this result for a small system, we provide an explicit expression for the $\mathcal{H}_2$ norm of a 2-dimensional system in terms of the corresponding symmetric system.

2-dimensional system

To build intuition, we give the analytical expression for the $\mathcal{H}_2$ norm of a 2-dimensional system in terms of the norm of its symmetric component. The A and A_s matrices are,

$$A \;=\; \begin{bmatrix} a & e+h \\ e-h & d \end{bmatrix}, \quad A_s \;=\; \begin{bmatrix} a & e \\ e & d \end{bmatrix}.$$

Solving for $P_s = -0.5A_s^{-1}$ and taking the trace gives the $\mathcal{H}_2$ norm of the symmetric system,

$$\text{trace}(P_s) = -\tfrac{a+d}{2(ad-e^2)}$$

Explicitly solving the Lyapunov equation $AP_c + P_cA^T + I = 0$ for P and taking its trace yields the $\mathcal{H}_2$ norm of the original system,

$$\text{trace}(P_c) = -\tfrac{(b-c)^2+(a+d)^2}{2(a+d)(ad-bc)}.$$

With some algebra, the above simplifies to, $\text{trace}(P_s) = -(a+d)(2(ad-e^2))^{-1}$ and,

$$\text{trace}(P_c) = \text{trace}(P_s)\left(\frac{(a+d)^2(ad-e^2)+4(ad-e^2)h^2}{(a+d)^2(ad-e)+(a+d)^2h^2}\right).$$

We show that a similar property holds for the $\mathcal{H}_\infty$ norm.

Proposition 6.1.4. *Let A_s be a symmetric matrix. The $\mathcal{O}(\varepsilon)$ correction to the $\mathcal{H}_\infty$ norm of the system*

$$\dot{x} \;=\; A_s x \;+\; d$$

from an $\mathcal{O}(\varepsilon)$ antisymmetric perturbation A_a is zero.

Proof. From Proposition 6.1.1, the $\mathcal{H}_\infty$ norm of the symmetric system is given by $\sigma_{\max}(-A_s^{-1})$. The maximum singular value of a matrix is equivalent to,

$$\sigma_{\max}(X) \;=\; \sup_{\|v\|_2 \le 1, \|w\|_2 \le 1} v^T X w.$$

Since A_s is symmetric, $w = v$. Taking an $\mathcal{O}(\varepsilon)$ antisymmetric perturbation A_a to the above expression,

$$\sigma_{\max}(-(A_s + \varepsilon A_a)^{-1}) = -v^T A_s^{-1} v + \varepsilon v^T A_s^{-1} A_a A_s^{-1} v. + \mathcal{O}(\varepsilon^2)$$

Since A_a is antisymmetric, $\left\langle A_s^{-1} v v^T A_s^{-1}, A_a \right\rangle = 0$ and thus the first-order correction is 0. □

6.2 Computational experiments

6.2.1 Example: Directed Consensus Network

In this example, we illustrate the utility of the approach described in Section 6.1.2. Consider the network dynamics given by a directed network as described in Section 1.2.1,

$$\dot{\psi} = -(L + E \operatorname{diag}(x) E^T)\psi$$

where L is a directed graph Laplacian, $F(x) = E \operatorname{diag}(x) E^T$ represents the addition of undirected links, v is a vector that contains weights of these added links, and the incidence matrix E describes which edges may be added or altered. The regularization on x is given by,

$$g(x) = \|x\|_2^2 + \gamma \sum_i |x_i|$$

where the quadratic term limits the size of the edge weights, the ℓ_1 norm promotes sparsity of added links, and $\gamma > 0$ parametrizes the importance of sparsity.

For this concrete example, the network topology is given by Figure 6.1. The potential added edges can connect the following pairs of nodes: $(1)-(2)$, $(1)-(3)$, $(1)-(5)$, $(1)-(6)$, $(2)-(5)$, $(2)-(6)$, $(3)-(6)$, and $(4)-(5)$.

Controllers were designed by solving problems (6.3) and (6.4) for the symmetric version of the network over 50 log-distributed values of $\gamma \in [10^{-4}, 1]$. The closed-loop $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms obtained by applying these controllers to the symmetric and original systems are shown in Fig. 6.2. Fig. 6.1 also shows which edges were added for $\gamma = 1$.

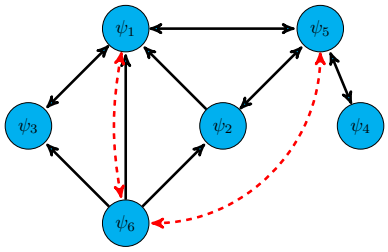


Figure 6.1: Directed network (**solid black** — arrows) with added undirected edges (**dashed red - - -** arrows). Both the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal structured control problems yielded the same set of added edges. In addition to these edges, the controllers tuned the weights of the edges $(1)-(3)$ and $(1)-(5)$.

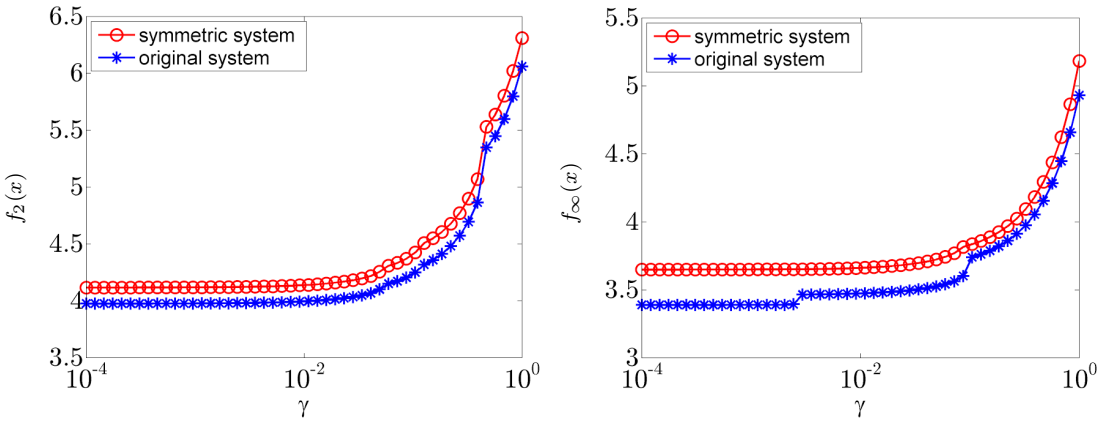


Figure 6.2: $\mathcal{H}_2$ and $\mathcal{H}_\infty$ performance of the closed-loop symmetric system and the original system subject to a controller designed at various values of γ.

6.2.2 Example: Combination drug therapy design via symmetric systems

In this section, we consider the combination drug therapy problem introduced in Section 1.2.1,

$$\dot{\psi} \;=\; (A \;+\; \mathrm{diag}\,(F_x x))\,\psi \;+\; d.$$

where ψ represents a vector of HIV mutant populations, A contains their evolutionary dynamics, and F_x contains information about drug treatment. The entry A_{ii} represents how fast mutant i replicates and the entry A_{ij} represents the probability of mutation from mutant j into mutant i. Each element x_k represents the amount of drug k administered to the patient. The ith element of the kth column of of F_x specifies how quickly drug k destroys mutant i.

Combination drug therapy is desirable because using only one drug often leads to the mutant population adapting to the weaknesses of that particular drug [197]. However, because of potential side effects and drug-drug interactions, it is not desirable to use a large number of different drugs. Furthermore, large doses can have additional side effects [198].

Since the probability of mutation is often orders of magnitude less than the rate of replication [199], Proposition 6.1.3 justifies design using the symmetric component of the system. We therefore minimize the $\mathcal{H}_2$ norm of the system,

$$\dot{\psi} \;=\; (A_s \;+\; \mathrm{diag}(F_x x))\,\psi \;+\; d.$$

where $A_s = \frac{1}{2}(A \;+\; A^T)$. Since this system is symmetric, it can be posed as problem (6.3).

Primal and dual optimization problems

We first state the problem in a form which is convenient for implementation of the alternating direction method of multipliers (ADMM), a technique well suited to large scale problems that has been recently successfully applied to sparse control synthesis problems [64, 105]. Using the auxiliary variable $G := -(A_s + \mathrm{diag}(F_x x))$, problem (6.3) becomes,

$$\begin{array}{rl} \text{minimize} & \mathrm{trace}(G^{-1}) \;+\; \frac{1}{2}x^T R x \;+\; \gamma \mathbb{1}^T u \\ \text{subject to} & G \;+\; A_s \;+\; \mathrm{diag}(F_x x) \;=\; 0 \\ & G \;\succ\; 0, \;\; x \;\geq\; 0. \end{array} \tag{6.7}$$

Since a negative drug dosage is not possible, the ℓ_1 norm of x can be written as $\mathbb{1}^T x$. The Lagrangian is,

$$\mathcal{L}(G, x, Y, \lambda) \;=\; \mathrm{trace}(G^{-1}) \;+\; \tfrac{1}{2}x^T R x \;+\; \gamma \mathbb{1}^T x \;-\; \lambda^T x \;+\; \langle Y,\, G \;+\; A \;+\; \mathrm{diag}(F_x x)\rangle$$

where $Y \;=\; Y^T$ and $\lambda \;\geq\; 0$. We omit the Lagrange multiplier associated with the positive definiteness of G for compactness; it will become a slack variable and will result in a requirement on the positive definiteness of the dual variable Y. We substitute the equivalent expression,

$$\mathrm{trace}\,(Y \;\mathrm{diag}(F_x x)) \;\;=\;\; -\; y^T F_x x$$

where $y = \mathrm{diag}(Y)$ into the Lagrangian, differentiate it with respect to G and x and set the

gradients equal to zero,

$$\begin{aligned} 0 &= -G^{-2} + Y \\ 0 &= Rx + \gamma\mathbb{1} - \lambda + F_x^T y. \end{aligned}$$

The optimal G and x are therefore,

$$\begin{aligned} G &= Y^{-1/2} \\ x &= -R^{-1}(F_x^T y + \gamma\mathbb{1} - \lambda). \end{aligned}$$

Substituting these expressions into the Lagrangian yields

$$2\,\mathrm{trace}(Y^{\frac{1}{2}}) + \mathrm{trace}(AY) - \tfrac{1}{2}\,(F_x^T y + \gamma\mathbb{1} - \lambda)^T R^{-1}(F_x^T y + \gamma\mathbb{1} - \lambda).$$

The dual problem is to maximize the above expression over $\lambda \geq 0$ and $Y \succ 0$. Since the above is concave and quadratic in λ, its maximum value is achieved at $\lambda = F_x^T y + \gamma\mathbb{1}$. However, λ must be nonnegative and thus

$$\lambda^\star = \begin{cases} F_x^T y + \gamma\mathbb{1}, & F_x^T y + \gamma\mathbb{1} > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, λ can be eliminated by substituting $\lambda^\star$ and the dual problem can be written as,

$$\begin{array}{ll} \text{maximize} & 2\,\mathrm{trace}(Y^{1/2}) + \mathrm{trace}(AY) - \tfrac{1}{2}\,\max(-F_x^T y - \gamma\mathbb{1}, 0)^T R^{-1}(-F_x^T y - \gamma\mathbb{1}) \\ \text{subject to} & Y \succeq 0. \end{array} \tag{6.8}$$

Alternating direction method of multipliers

To apply ADMM to (6.7), we first form the corresponding augmented Lagrangian,

$$\mathcal{L}_\mu(G, x, Y) := \mathrm{trace}(G^{-1}) + x^T R x + \gamma\,\mathbb{1}^T x + \langle Y,\, G + A + \mathrm{diag}(F_x x)\rangle + \tfrac{1}{2\mu}\|G + A + \mathrm{diag}(F_x x)\|_F^2.$$

Relative to the standard Lagrangian, $\mathcal{L}_\mu$ contains an additional quadratic penalty on the violation of the linear constraint. The positive parameter μ specifies the magnitude on the constraint violation penalty at each iteration.

The ADMM iteration uses the update sequence [132]

$$\begin{aligned} G^{k+1} &= \underset{G}{\text{argmin}} \quad \mathcal{L}_\rho(G,\, x^k,\, Y^k) \\ x^{k+1} &= \underset{x}{\text{argmin}} \quad \mathcal{L}_\rho(G^{k+1},\, x,\, Y^k) \\ Y^{k+1} &= Y^k \;+\; \tfrac{1}{\mu}\,\left(G^{k+1} + A + \text{diag}(F_x x^{k+1})\right) \end{aligned}$$

to find the optimal solution to the original problem. The stopping criteria depend on the primal residual, which quantifies how well G^k and x^k satisfy the linear constraint, and the dual residual, which quantifies the difference between x^k and x^{k-1}. We refer the reader to [132] for details.

This algorithm is advantageous because the subproblems are much simpler than the original problem. The G-minimization step has an explicit solution, the x-minimization step takes a standard problem form for which there are efficient algorithms, and the Y-update step is algebraic.

G-minimization

The G-minimization step amounts to solving,

$$\begin{aligned} \underset{G}{\text{minimize}} \quad & \text{trace}(G^{-1}) + \tfrac{1}{2\mu}\|G - V^k\|_F^2 \\ \text{subject to} \quad & G \succ 0 \end{aligned}$$

where $V_k := -(A + \text{diag}(F_x x^k)) - \frac{1}{2\mu} Y^k$. Seting the gradient,

$$-G^{-2} \;+\; \tfrac{1}{\mu} G \;-\; V^k$$

to zero is an explicit exercise with the positive definiteness constraint. Since the powers of G appear with no coefficients, the optimal G has the same eigenstructure as V^k. Its eigenvalues are determined by the positive real solution to the cubic equation

$$\tfrac{1}{\mu}\,\lambda_i^3 \;+\; \sigma_i\,\lambda_i^2 \;-\; 1$$

where σ_i is the corresponding eigenvalue of V_k. By the convexity of the G-minimization problem, there can be only one positive real solution to the above equation.

x-minimization

The x-minimization step is equivalent to solving,

$$\underset{x}{\text{minimize}} \quad \gamma \mathbb{1}^T x \ + \ x^T R x \ + \ \tfrac{1}{2\mu}\|F_x x \ - \ w^k\|_2^2.$$

where $w^k := g^{k+1} + a + \mu y^k$, $g^k = \text{diag}(G^k)$, $a = \text{diag}(A)$, and $y^k = \text{diag}(Y^k)$. The objective function is the sum of a quadratic term and an ℓ_1 norm: a problem form is commonly referred to as LASSO. This problem has attracted lots of attention in recent years and there are many efficient methods for computing its solution; see Section 4.4.1 for a comparison of many standard techniques with the second order method we develop in Chapter 4. We solve it using Iterative Soft-Thresholding (ISTA) [131].

Computational complexity

Generic SDP solvers scale with n^6, where n is the dimension of the positive definiteness constraint. The G-minimization requires $\mathcal{O}(n^3)$ operations because it requires an eigenvalue decomposition, the x-minimization step requires $\mathcal{O}(nr)$, operations and the Y-update step requires $\mathcal{O}(n^2)$ operations.

Numerical example

Following the example given in [89, 90] based on [200], we examine a system with 35 mutants and 5 potential types of drugs. The diagonal entries are 0.5 and the off diagonal elements range from the order of 10^{-8} to 10^{-6}. The structure of A is shown in Fig. 6.3. It is evident that the matrix is not symmetric.

We next use our algorithm to design control inputs u for the symmetric system with varying levels of the sparsity promoting parameter γ. As γ is increased, sparsity is prioritized over $\mathcal{H}_2$ performance of the system and therefore the performance degrades. In Fig. 6.4, we show the difference in $\mathcal{H}_2$ performance between the symmetric and original systems as a function of γ. In this problem, the symmetric model is a very good approximation of the original system, even up to extremely large levels of γ.

Since the off-diagonal entries of the matrix A are small, we artificially increase them by a constant factor to study our approach for problems with larger degrees of asymmetry. We take,

$$A_c \ = \ c(A - I \circ A) \ + \ 0.5(I \circ A)$$

where c is a constant factor and $\circ$ is the Hadamard (element wise) product. This modification means that A_c does not have physically relevant implications for the drug synthesis problem, but it illustrates the utility of our approach with a realistic problem structure. Figure 6.5 compares the $\mathcal{H}_2$ performance for $c = 10^5, 1.4 \times 10^6$, and 1.9×10^7. Compared to the diagonal entries of A_c, the maximum off-diagonal element is of the same order, one order of magnitude higher, and two orders of magnitude higher respectively.

When the off-diagonal elements are of the same order of magnitude or smaller than the diagonal elements, there is almost no difference between the symmetric and full models. As the off-diagonal elements get larger, the fidelity of the approximation suffers. Unsurprisingly, as the system becomes more asymmetric, the symmetric approximation becomes more conservative [201].

We note that for a realistic synthesis problem, γ would be varied to find sparsity structures for x. Once a desired sparsity structure is identified, (6.3) would be solved with $\gamma = 0$ but x constrained to have that particular sparsity structure. This process, known as polishing in compressive sensing literature, would yield a x which is sparse but has better performance than simply obtaining x from (6.3) where g represents ℓ_1 regularization.

6.3 Computational advantages for structured problems

Structured control is often of interest for large-scale systems. As such, the computational scaling of algorithms used to compute optimal controllers is very important. In this section, we identify a class of systems which are amenable to scalable distributed algorithms.

When A and $F(x)$ are always jointly block-diagonalizable, the dynamics of the system can be expressed as the sum of independent subsystems. Define $\hat{\psi} := \Pi\psi$ and let Π be a unitary matrix such that,

$$\dot{\hat{\psi}} \;=\; (\hat{A} \;+\; \hat{F}(x))\hat{\psi}$$

where

$$\hat{A} := \Pi A \Pi^T, \qquad \hat{F}(x) := \Pi F(x) \Pi^T,$$

and, for any choice of x, $\hat{A} + \hat{F}(x) = \text{blkdiag}(\hat{A}_{11} + \hat{F}_{11}, \cdots, \hat{A}_{NN} + \hat{F}_{NN})$ is block-diagonal with N blocks of size $n \times n$ each.

For problems of this form, computing optimal control strategies is much more efficient in the $\hat{\psi}$ coordinates because the majority of the computational burden in solving problems (6.3) and (6.4) comes from the $nN \times nN$ LMI constraint involved in minimizing the performance

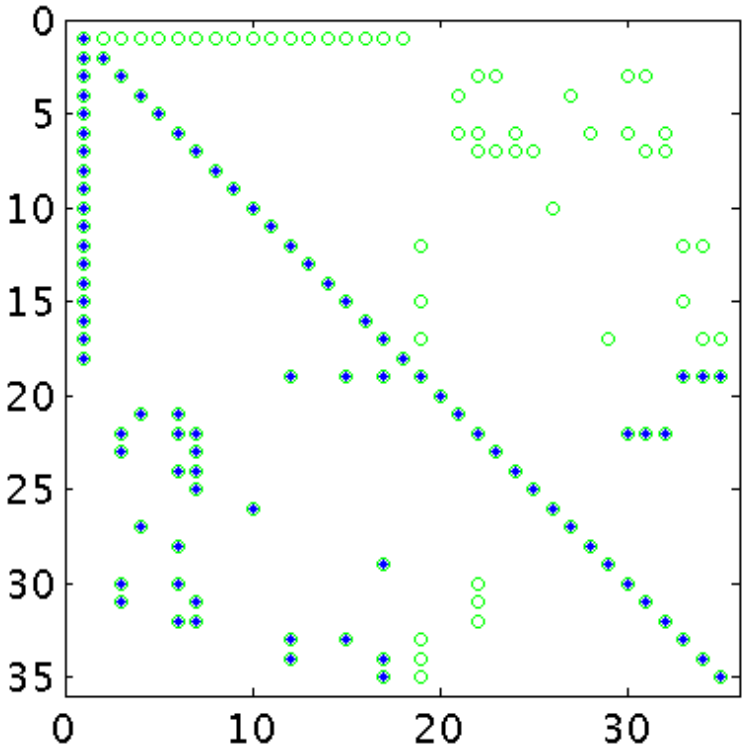


Figure 6.3: Sparsity structure of the matrix A and its symmetric counterpart. The elements of A are shown with blue dots, and the elements in its symmetric component A_s are overlaid in green circles.

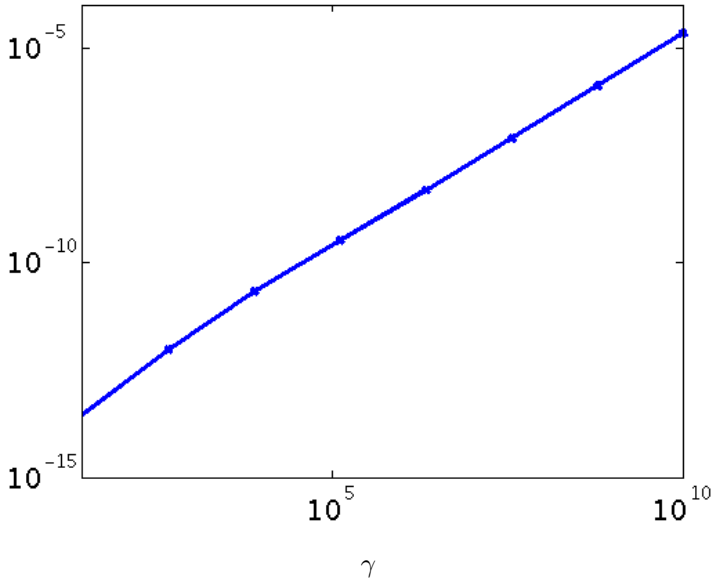


Figure 6.4: Difference in $\mathcal{H}_2$ norm between the symmetric and original systems with different controllers designed as a function of γ, normalized by the $\mathcal{H}_2$ norm of original system.

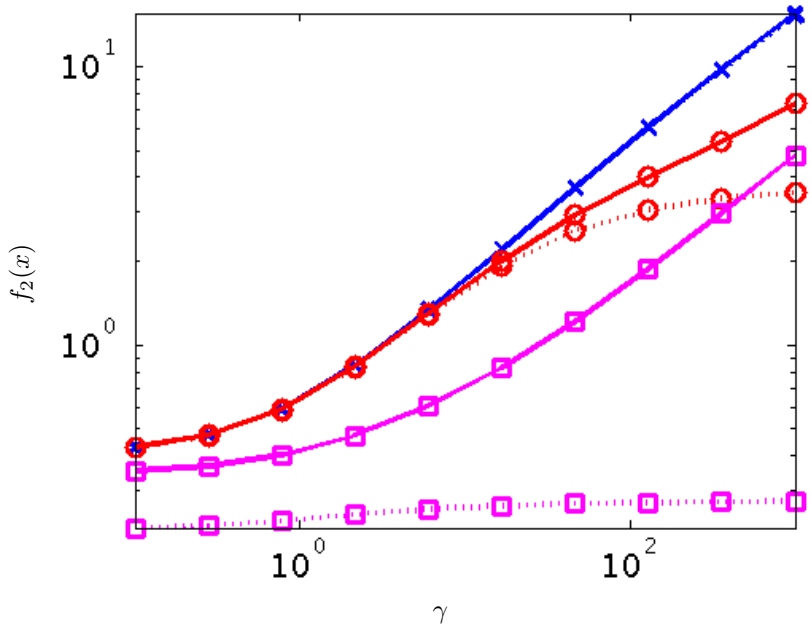


Figure 6.5: The **solid** — lines are the $\mathcal{H}_2$ norms of the symmetric systems and the **dotted - - -** lines are the $\mathcal{H}_2$ norms of the original systems. The **blue** $\times$, **red** $\circ$, and **magenta** $\square$ designate $c = 10^5, 1.4 \times 10^6$, and 1.9×10^7 respectively.

metrics $f_2(x)$ or $f_\infty(x)$.

For this class of system, the $\mathcal{H}_2$-optimal control problem (6.3) can be expressed as,

$$\begin{aligned} \underset{v,\Theta_i}{\text{minimize}} \quad & \frac{1}{2}\sum_i \text{trace}(\Theta_i) \;+\; g(x) \\ \text{subject to} \quad & \begin{bmatrix} \Theta_i & I \\ I & -((\hat{A}_s)_{ii} + (\hat{F}_s(x))_{ii}) \end{bmatrix} \succeq 0. \end{aligned} \tag{6.9}$$

which is an SDP with N separate $n \times n$ LMI blocks. Since SDPs scale with the sixth power of the LMI blocks, solving this reformulation scales with n^6 as opposed to n^6N^6.

Analogously, the structured $\mathcal{H}_\infty$-optimal control problem (6.4) can be cast as,

$$\begin{aligned} \underset{v,\Theta_i}{\text{minimize}} \quad & \max_i(\sigma_{\max}(\Theta_i)) \;+\; g(x) \\ \text{subject to} \quad & \begin{bmatrix} \Theta_i & I \\ I & (\hat{A}_s)_{ii} + (\hat{F}_s(x))_{ii} \end{bmatrix} \succeq 0. \end{aligned} \tag{6.10}$$

One important class of system which satisfies these assumptions is spatially-invariant systems. This structure was used in [190] to develop efficient techniques for sparse feedback synthesis.

6.3.1 Spatially-invariant systems

Spatially-invariant systems have a block-circulant structure which is block-diagonalizable by a Discrete Fourier Transform (DFT). A spatially-invariant system can be represented by N subsystems with n states each. The state vector $\psi \in \mathbb{R}^{nN}$ is composed of N subvectors $\psi_i \in \mathbb{R}^n$ which denotes the state of the subsystem. The matrix $A \in \mathbb{R}^{nN \times nN}$ is block-circulant with blocks of the size $n \times n$. For example, when $N = 3$,

$$A \;=\; \begin{bmatrix} A_0 & A_1 & A_{-1} \\ A_{-1} & A_0 & A_1 \\ A_1 & A_{-1} & A_0 \end{bmatrix}$$

where the blocks A_0, A_{-1}, $A_1 \in \mathbb{R}^{n \times n}$.

It was shown in [9] that the optimal feedback controller for a spatially-invariant system is

itself spatially-invariant. Assuming that the optimal *sparse* feedback controller is also spatially-invariant is equivalent to assuming that $F(x)$ is block-circulant. Block circulant matrices are block-diagonalizable by the appropriate DFT. Let the block Fourier matrix be

$$\Phi \;:=\; \Phi_N \otimes I_n,$$

where I_n is the $n \times n$ identity matrix, Φ_N is the $N \times N$ discrete Fourier transform matrix, and $\otimes$ represents the Kronecker product. By introducing the change of variables $\hat{\psi} := \Phi\psi$, where

$$\hat{\psi} = \left[\begin{array}{ccc} \hat{\psi}_1^T & \cdots & \hat{\psi}_N^T \end{array} \right]^T,$$

and $\hat{\psi}_i \in \mathbb{R}^n$, the original system's dynamics can be expressed as N independent $n \times n$ subsystems,

$$\hat{A} \;=\; \text{blkdiag}(\hat{A}_{11}, \hat{A}_{22}, \hat{A}_{33})$$

Consequently, the optimal structured control problems (6.3) and (6.4) can be cast as (6.9) and (6.10), which are more amenable to efficient computation.

6.3.2 Swift-Hohenberg Equation

Here we illustrate the utility of the block-diagonalization. Consider a particular realization of the Swift-Hohenberg equation [191],

$$\partial_t \psi(t,\phi) \quad = \quad \beta\, \psi(t,\phi) \;-\; (1 + \partial_{\phi\phi})^2\, \psi(t,\phi) \;+\; x(\phi)\psi(t,\phi) \;+\; d(t,\phi)$$

where $\beta \in \mathbb{R}$, and $\psi(t,\cdot), v(\cdot) \in L_2(-\infty,\infty)$, $d(t,\phi)$ is a white-in-time-and-space stochastic disturbance and $x(\phi)$ is a spatially-invariant feedback controller which is to be designed. A finite dimensional approximation of this system can be obtained by using the differentiation suite from [202] to discretize the problem into N points and approximating the infinite domain with periodic boundary conditions over the domain $L_2[0, 2\pi]$. A sparse $\mathcal{H}_2$ feedback controller $x(\phi)$ can then be identified by solving problem (6.3).

We contrast this method with the approach we advocate in Section 6.3, where we use the DFT to decompose the system into N first-order systems corresponding to eigenfunctions of the Swift-Hohenberg equation and solve problem (6.9).

The state vector takes the form of $\psi(\phi)$ evaluated at grid points in ϕ where the dynamics

are given by

$$\dot{\psi} \;=\; (A \;+\; X)\psi \;+\; d.$$

and $A = \beta I - (I + D^2)^2$. Here D is a discrete differentiation matrix from [202], and X is the circulant state feedback matrix.

Using the DFT over ϕ, the Swift-Hohenberg equation can be expressed as a set of independent first-order systems,

$$\dot{\hat{\psi}}_\phi \;=\; (a_\phi \;+\; \hat{x}_\phi)\hat{\psi}_\phi \;+\; \hat{d}$$

where $a_\phi := \beta - (1 - \kappa_\phi^2)^2$, and the new coordinates are $\hat{\psi} := \Pi\psi$, Π is the DFT matrix, κ_ϕ is the wavenumber (spatial frequency), and $\hat{x}$ represents X in the Fourier space; i.e., $X = \Pi^T \operatorname{diag}(\hat{x})\Pi$.

We take the regularization term to be

$$g(x) \;=\; \|X\|_F^2 \;+\; \gamma\|X\|_1$$

where $\|X\|_1 := \sum_{ij} |X_{ij}|$ is the elementwise ℓ_1 norm and γ is a parameter which specifies the emphasis on sparsity relative to performance.

For the $\mathcal{H}_2$ problem, the regularized optimal control problem is of the form of (6.3) with $F_s(x) = X$ and X is circulant. In that formulation, the problem is an SDP with one $N \times N$ LMI block. In the Fourier space, the problem can be expressed as (6.9), which takes the particular form,

$$\begin{aligned} \underset{\hat{x}}{\text{minimize}} \quad & \frac{1}{2}\sum \frac{1}{-(a_\phi \;+\; \hat{x}_\phi)} \;+\; g\left(\Pi^T \operatorname{diag}(\hat{x})\Pi\right) \\ \text{subject to} \quad & -\left(a_\phi \;+\; \hat{x}_\phi\right) \;\geq\; 0 \end{aligned}$$

which does not require the large SDP constraints in (6.3).

We solved the regularized $\mathcal{H}_2$ optimal control problem by solving the general formulation (6.3) and the more efficient formulation (6.9) for $\beta = 0.1$, $\gamma = 1$ and N varying from 5 to 51 using CVX, a general purpose convex optimization solver [203].

Taking advantage of spatial invariance yields a significant computational advantage, as can be seen in Figure 6.6. Although both expressions of the problem yield the same solution, solving the realization in (6.9) is much faster and allows us to examine much larger problem dimensions. In Figure 6.7, we show the spatially-invariant feedback controller for one point in the domain, i.e., one row of X, computed for $N = 101$ at $\gamma = 0$, $\gamma = 0.1$, and $\gamma = 10$..

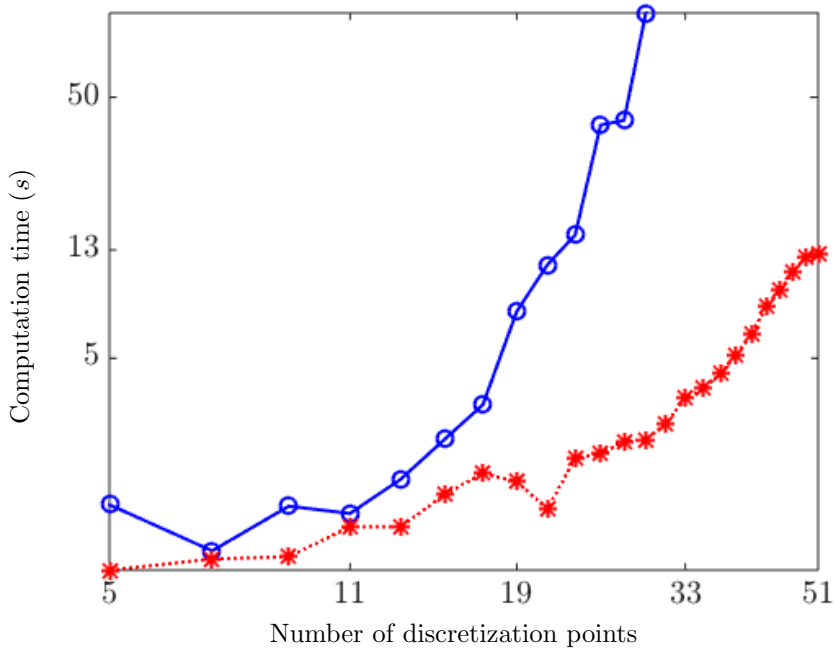


Figure 6.6: Computation time for the general formulation (6.3) (**solid blue** —○—) and that which takes advantage of spatial invariance (6.9) (**dotted red** ···∗···).

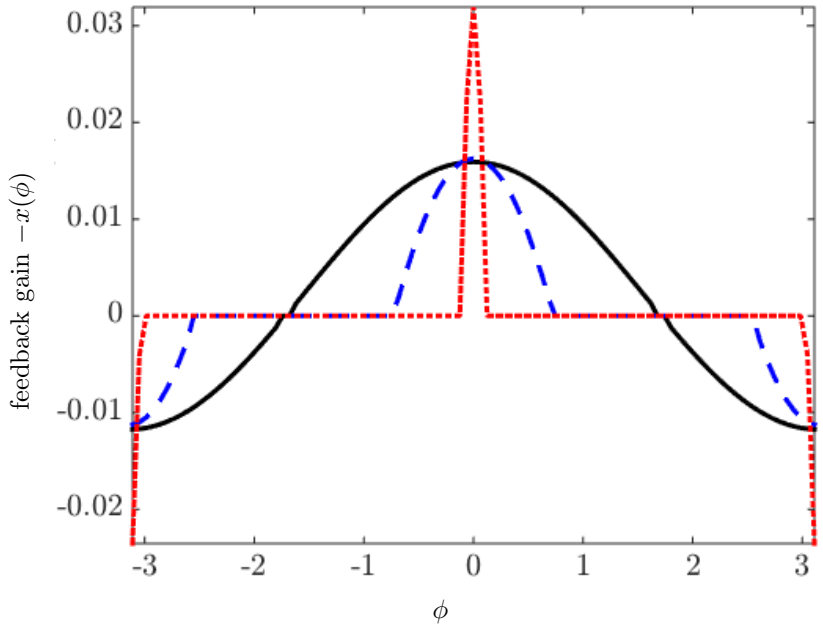


Figure 6.7: Feedback gain $-x(\phi)$ for the node at position $\phi = 0$, computed with $N = 51$ and $\gamma = 0$ (**solid black** —), $\gamma = 0.1$ (**dashed blue - - -**), and $\gamma = 10$ (**dotted red** ···).

Chapter 7

Structured decentralized control of positive systems

Positive systems have received much attention in recent years because of convenient properties that arise from their structure. A system is called positive if, for every nonnegative initial condition and input signal, its state and output remain nonnegative [53]. Such systems appear in the models of heat transfer, chemical networks, and probabilistic networks. In [54], the authors show that the KYP lemma greatly simplifies for positive systems, thereby allowing for decentralized $\mathcal{H}_\infty$ synthesis via Semidefinite Programming (SDP). In [57], it is shown that static output-feedback can be solved via Linear Programming (LP) for a class of positive systems. In [55,56], the authors develop necessary and sufficient conditions for robust stability of positive systems with respect to induced $\mathcal{L}_1$–$\mathcal{L}_\infty$ norm-bounded perturbations. In [204–206], it is shown that the structured singular value is equal to its convex upper bound for positive systems. Thus, assessing robust stability with respect to induced $\mathcal{L}_2$ norm-bounded perturbation is also tractable.

Most of the recent literature focuses on control design for positive systems with respect to induced $\mathcal{L}_1$–$\mathcal{L}_\infty$ norms or induced $\mathcal{L}_2$ norms [54, 56, 58]. In this chapter, we show that a class of $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal control problems which are not tractable for general systems are convex for positive systems. Moreover, they are convex in the original controller variables which (i) allows us to formulate convex optimization problems where the control parameter is kept as an optimization variable; and (ii) facilitates a straightforward implementation of constraints or regularizers on the control parameter in the optimal control formulation.

7.1 Problem formulation and background

We begin with a review of positive systems and then pose the problem.

7.1.1 Background on positive systems

We first provide relevant notation and definitions to facilitate discussion.

Notation

The set of $n \times n$ Metzler matrices (matrices with nonnegative off diagonal elements) is denoted by $\mathbb{M}^{n\times n}$. The set of $n \times n$ nonnegative (positive) diagonal matrices is denoted by $\mathbb{D}^n_+$ ($\mathbb{D}^n_{++}$).

Definition 2 (Graph associated to a matrix)**.** Given $A \in \mathbb{R}^{n\times n}$ we denote the graph associated to A as $\mathcal{G}(A) = (\mathcal{V}, \mathcal{E})$, with the set of vertices $\mathcal{V} = \{1, ..., n\}$ and the set of edges $\mathcal{E} := \{(i, j) \,|A_{ji} \neq 0 \,\}$.

Definition 3 (Strongly connected graph)**.** A graph $(\mathcal{V}, \mathcal{E})$ is strongly connected if there is a directed path linking every two distinct nodes in $\mathcal{V}$.

Definition 4 (Weakly connected graph)**.** A graph $(\mathcal{V}, \mathcal{E})$ is weakly connected if replacing its edges with undirected edges results in a strongly connected graph.

Positive systems

We begin with a definition of a linear positive system.

Definition 5. A linear system described by the state-space representation,

$$\begin{aligned} \dot{\psi} &= A\psi + Bd \\ \zeta &= C\psi + Dd, \end{aligned}$$

is *positive* if and only if A is Metzler and B, C, and D are nonnegative matrices.

We now state two lemmas that are useful for the analysis of positive LTI systems.

Lemma 7.1.1. *Let $A \in \mathbb{M}^{n\times n}$ and let $Q \in \mathbb{R}^{n\times n}$ be a positive definite matrix with nonnegative entries. Then*

(a) $\mathrm{e}^A \geq 0$*;*

(b) For Hurwitz A, the solution P to the algebraic Lyapunov equation,

$$AP \;+\; PA^T \;+\; Q \;=\; 0$$

is elementwise nonnegative.

Proof. (a) The Metzler matrix A can be written as $\tilde{A} - \alpha I$ with $\tilde{A} \geq 0$ and $\alpha > 0$. Since $\mathrm{e}^{\tilde{A}} = \sum_{k=0}^{\infty} \tilde{A}^k/k!$ and $\tilde{A}^k \geq 0$ for all k, $e^{\tilde{A}} \geq 0$. Thus, $\mathrm{e}^{-\alpha} > 0$ implies $\mathrm{e}^{A} = \mathrm{e}^{-\alpha}\mathrm{e}^{\tilde{A}} \geq 0$.

(b) Apply (a) to $P = \int_0^\infty \mathrm{e}^{At}\, Q\, \mathrm{e}^{A^T t}\, \mathrm{d}t$. □

Lemma 7.1.2. *The left and right principal singular vectors, w and v, of $A \in \mathbb{R}_+^{n\times n}$ are nonnegative. If $A \in \mathbb{R}_{++}^{n\times n}$, w and v are positive and unique.*

Proof. Follows from the application of the Perron theorem [180, Theorem 8.2.11] to AA^T and A^TA. □

7.1.2 Problem formulation

We consider a class of positive systems (1.1) subject to structured decentralized control,

$$\begin{aligned} \dot{\psi} \;&=\; (A \;+\; F(x))\psi \;+\; Bd \\ \zeta \;&=\; C\psi \end{aligned} \tag{7.1}$$

where A is a Metzler matrix, $F(x)$: $\mathbb{R}^m \to \mathbb{D}^n$, and $B, C \geq 0$. Many applications fit into this form, including leader selection and combination drug therapy design for HIV. Our objective is to design a stabilizing *diagonal* matrix $F(x)$ that minimizes amplification from d to z. To study this class of problem, we introduce the following assumption.

Assumption 3. The matrix A in (7.1) is Metzler, the matrices B and C are nonnegative, and the diagonal matrix $F(x) := \mathrm{diag}\,(F_x x)$ is a linear function of x with $F_x \in \mathbb{R}^{n\times m}$.

We now review some recent results. Under Assumption 3, the matrix $A + F(x)$ is Metzler and its largest eigenvalue is real and a convex function of x [207]. Recently, it has been shown that the weighted $\mathcal{L}_1$ norm of the response of system (7.1) from a nonnegative initial condition $x_0 \geq 0$,

$$\int_0^T c^T \psi(t)\, \mathrm{d}t$$

is a convex function of u for every $c \in \mathbb{R}_+^n$ [58, 59]. Furthermore, the approach in [54] can be used to cast the problem of *unstructured* decentralized $\mathcal{H}_\infty$ control of positive systems as a

semidefinite program (SDP) and [208] can be used to cast it as a Linear Program (LP). However, both the SDP and the LP formulations require a change of variables that does not preserve the structure of $F(x)$. Consequently, it is often difficult to design controllers that are feasible for a given noninvertible operator F or to impose structural constraints or penalties on x.

We show that both the $\mathcal{H}_2$ and the $\mathcal{H}_\infty$ norms are convex functions of the original optimization variable x and provide explicit expressions for the (sub)gradients. This allows us to develop efficient descent algorithms that solve regularized optimal control problems of the form (1.8).

7.2 Convexity of optimal control problems

We next establish convexity of the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ norms for systems that satisfy Assumption 3, derive a graph theoretic condition that guarantees continuous differentiability of f_∞, and develop a customized algorithm for solving optimization problem (1.8) even in the absence of differentiability.

7.2.1 Convexity of f_2 and f_∞

We first establish convexity of the $\mathcal{H}_2$ optimal control problem and provide the expression for the gradient of f_2.

Proposition 7.2.1. *Let Assumption 3 hold and let $A_{\mathrm{cl}}(x) := A + F(x)$ be a Hurwitz matrix. Then, f_2 is a convex function of x and its gradient is given by*

$$\nabla f_2(x) \;=\; 2F^\dagger(P_cP_o) \tag{7.2}$$

where P_c and P_o are the controllability and observability gramians of the closed-loop system (7.1)*,*

$$A_{\mathrm{cl}}(x)\,P_c \;+\; P_c\,A_{\mathrm{cl}}^T(x) \;+\; BB^T \quad = \quad 0 \tag{7.3a}$$

$$A_{\mathrm{cl}}^T(x)\,P_o \;+\; P_o\,A_{\mathrm{cl}}(x) \;+\; C^TC \quad = \quad 0. \tag{7.3b}$$

Proof. We first establish convexity of $f_2(x)$ and then derive its gradient. The square of the $\mathcal{H}_2$ norm is given by,

$$f_2(x) \;=\; \begin{cases} \left\langle C^TC,\, P_c \right\rangle, & A_{\mathrm{cl}}(x) \text{ Hurwitz} \\ \qquad\infty, & \text{otherwise} \end{cases}$$

where the controllability gramian P_c of the closed-loop system is determined by the solution to Lyapunov equation (7.3a). For Hurwitz $A_{\mathrm{cl}}(x)$, P_c can be expressed as,

$$P_c \;=\; \int_0^\infty \mathrm{e}^{A_{\mathrm{cl}}(x)t} B B^T \mathrm{e}^{A_{\mathrm{cl}}^T(x)t}\,\mathrm{d}t.$$

Substituting into $\langle C^T C,\, P_c \rangle$ and rearranging terms yields,

$$f_2(x) \;=\; \int_0^\infty \|C\,\mathrm{e}^{A_{\mathrm{cl}}(x)t} B\|_F^2\,\mathrm{d}t \;=\; \int_0^\infty \sum_{i,\,j} \left(c_i^T\,\mathrm{e}^{A_{\mathrm{cl}}(x)t}\,b_j\right)^2 \mathrm{d}t$$

where c_i^T is the ith row of C and b_j is the jth column of B. From [58, Lemma 3] it follows that $c^T \mathrm{e}^{A_{\mathrm{cl}}(x)t}\,b$ is a convex function of u for nonnegative vectors c and b. Since the range of this function is $\mathbb{R}_+$ and $(\cdot)^2$ is nondecreasing over $\mathbb{R}_+$, the composition rules for convex functions [123] imply that $(c_i^T \mathrm{e}^{A_{\mathrm{cl}}(x)t}\,b_j)^2$ is convex in x. Convexity of $f_2(x)$ follows from the linearity of the sum and integral operators.

To derive ∇f_2, we form the associated Lagrangian,

$$\mathcal{L}(x, P_c, P_o) \;=\; \langle C^T C,\, P_c \rangle \;+\; \langle P_o,\, A_{\mathrm{cl}}(x) P_c \;+\; P_c A_{\mathrm{cl}}^T(x) \;+\; B B^T \rangle$$

where P_o is the Lagrange multiplier associated with equality constraint (7.3a). Taking variations of $\mathcal{L}$ with respect to P_o and P_c yields Lyapunov equations (7.3a) and (7.3b) for controllability and observability gramians, respectively. Using $A_{\mathrm{cl}}(x) = A + F(x)$ and the adjoint of F, we rewrite the Lagrangian as

$$\mathcal{L}(x, P_c, P_o) \;=\; 2\,\langle F^\dagger(P_c P_o),\, x \rangle \;+\; \langle C^T C,\, P_c \rangle \;+\; \langle P_o,\, A P_c \;+\; P_c A^T \;+\; B B^T \rangle\,.$$

Taking the variation of $\mathcal{L}$ with respect to x yields (7.2). □

Remark 7.2.1. *Convexity of the quadratic cost*

$$\int_0^T \psi^T(t)\, C^T C\, \psi(t)\,\mathrm{d}t$$

also holds over a finite or infinite time horizon for a piecewise constant x. This follows from [58, Lemma 4] and suggests that an approach inspired by the Model Predictive Control (MPC) framework can be used to compute a time-varying optimal control input for a finite horizon problem.

Remark 7.2.2. *The expression for ∇f_2 in Proposition 7.2.1 remains valid for any linear system*

and any linear operator $F\colon \mathbb{R}^m \to \mathbb{R}^{n\times n}$. However, convexity of f_2 holds under Assumption 3 and is not guaranteed in general.

We now establish the convexity of the $\mathcal{H}_\infty$ control problem and provide expression for the subdifferential set of f_∞.

Proposition 7.2.2. *Let Assumption 3 hold and let $A_{\rm cl}(x) := A + F(x)$ be a Hurwitz matrix. Then, f_∞ is a convex function of x and its subdifferential set is given by*

$$\partial f_\infty(x) \;=\; \Big\{ \sum_i \alpha_i\, F^\dagger\big(A_{\rm cl}^{-1}(x)\, B\, v_i\, w_i^T\, C A_{\rm cl}^{-1}(x)\big) \;\mid w_i^T (C A_{\rm cl}^{-1}(x) B)\, v_i = f_\infty(x),\; \alpha \in \mathcal{P} \Big\} \tag{7.4}$$

where $F^\dagger$ is the adjoint of the operator F and $\mathcal{P}$ is the simplex, $\mathcal{P} := \{\alpha_i \mid \alpha_i \geq 0, \sum_i \alpha_i \,=\, 1\}$.

Proof. We first establish convexity of $f_\infty(x)$ and then derive the expression for its subdifferential set. For positive systems, the $\mathcal{H}_\infty$ norm achieves its largest value at $\omega = 0$ [54] and from (1.3b) we thus have $f_\infty(x) = \bar\sigma(-C A_{\rm cl}^{-1}(x) B)$. To show convexity of $f_\infty(x)$, we show that $-C A_{\rm cl}^{-1}(x) B$ is a convex and nonnegative function of x, that $\bar\sigma(M)$ is a convex and nondecreasing function of a nonnegative argument M, and leverage the composition rules for convex functions [123].

Since $A_{\rm cl}(x)$ is Hurwitz, its inverse can be expressed as

$$-A_{\rm cl}^{-1}(x) \;=\; \int_0^\infty \mathrm{e}^{A_{\rm cl}(x)t}\, \mathrm{d}t. \tag{7.5}$$

Convexity of $c_i \mathrm{e}^{A_{\rm cl}(x)t}\, b_j$ [58, Lemma 3] and linearity of integration implies that each element of the matrix

$$-C\, A_{\rm cl}^{-1}(x)\, B \;=\; C \int_0^\infty \mathrm{e}^{A_{\rm cl}(x)t}\, \mathrm{d}t\, B$$

is convex in x and, by part (a) of Lemma 7.1.1, nonnegative.

The largest singular value $\bar\sigma(M)$ is a convex function of the entries of M [123],

$$\bar\sigma(M) \;=\; \sup_{\|w\|=1,\, \|v\|=1} w^T M v \tag{7.6}$$

and Lemma 7.1.2 implies that the principal singular vectors v_i and w_i that achieve the supremum in (7.6) are nonnegative for $M \geq 0$. Thus,

$$w_i^T (M \,+\, \beta\, \mathrm{e}_j \mathrm{e}_k^T)\, v_i \;\geq\; w_i^T M v_i$$

for any $\beta \geq 0$, thereby implying that $\bar\sigma(M)$ is nondecreasing over $M \geq 0$. Since each element of

$-CA_{\rm cl}^{-1}(x)B \geq 0$ is convex in x, these properties of $\bar{\sigma}(\cdot)$ and the composition rules for convex functions [123] imply convexity of $f_\infty(x) = \bar{\sigma}(-CA_{\rm cl}^{-1}(x)B)$.

To derive $\partial f_\infty(u)$, we note that the subdifferential set of the supremum over a set of differentiable functions,

$$f(x) \;=\; \sup_{i \,\in\, \mathcal{I}} \; f_i(x)$$

is the convex hull of the gradients of the functions that achieve the supremum [209, Theorem 1.13],

$$\partial f(x) \;=\; \Big\{ \sum_{j \,\mid\, f_j(x) \,=\, f(x)} \alpha_j \nabla f_j(x) \,\mid\, \alpha \in \mathcal{P} \Big\}$$

Thus, the subgradient of $\bar{\sigma}(M)$ with respect to M is given by

$$\partial\, \bar{\sigma}(M) \;=\; \Big\{ \textstyle\sum_j \alpha_j w_j v_j^T \mid w_j^T M v_j \;=\; \bar{\sigma}(M),\; \alpha \;\in\; \mathcal{P} \Big\}.$$

Finally, the matrix derivative of M^{-1} with respect to M in conjunction with the chain rule yields (7.4). □

Remark 7.2.3. *The adjoint of the linear operator F, introduced in Assumption 3, with respect to the standard inner product is $F^\dagger(x) = F_x^T \operatorname{diag}(x)$. For positive systems, Lemma 7.1.1 implies that the gramians P_c and P_o are nonnegative matrices. Thus, the diagonal of the matrix P_cP_o is positive and it follows that f_2 is a monotone function of the diagonal matrix $F(x)$. Similarly, $-A_{\rm cl}^{-1}(x)B\, v_i$ and $-w_i^T CA_{\rm cl}^{-1}(x)$ are nonnegative and thus f_∞ is also a monotone function of $F(x)$.*

7.2.2 Differentiability of the $\mathcal{H}_\infty$ norm

In general, the $\mathcal{H}_\infty$ norm is a nondifferentiable function of the control parameter x. Even though, under Assumption 3, the decentralized $\mathcal{H}_\infty$ optimal control problem (1.8) for positive systems is convex, it is still difficult to solve because of the lack of differentiability of f_∞. Nondifferentiable objective functions often necessitate the use of subgradient methods, which can converge slowly to the optimal solution.

In what follows, we prove that f_∞ is a continuously differentiable function of x for weakly connected $\mathcal{G}(A)$. Then, by noting that f_∞ is nondifferentiable only when $\mathcal{G}(A)$ contains disconnected components, we develop a method for solving (1.8) that outperforms the standard subgradient algorithm.

Differentiability under weak connectivity

To show the result, we first require two technical lemmas.

Lemma 7.2.4. *Let $M \geq 0$ be a matrix whose main diagonal is strictly positive and whose associated graph $\mathcal{G}(M)$ is weakly connected. Then, the graphs associated with $\mathcal{G}(MM^T)$ and $\mathcal{G}(M^TM)$ have self loops and are strongly connected.*

Proof. Positivity of the main diagonal of M implies that if M_{ij} is nonzero, then $(M^TM)_{ij}$ and $(MM^T)_{ij}$ are nonzero; by symmetry, $(M^TM)_{ji}$ and $(MM^T)_{ji}$ are also nonzero. Thus, $\mathcal{G}(M^TM)$ and $\mathcal{G}(MM^T)$ contain all the edges (i,j) in $\mathcal{G}(M)$ as well as their reversed counterparts (j,i). Since $\mathcal{G}(M)$ is weakly connected, $\mathcal{G}(M^TM)$ and $\mathcal{G}(MM^T)$ are strongly connected. The presence of self loops follows directly from the positivity of the main diagonal of M. □

Lemma 7.2.5. *Let $M \geq 0$ be a matrix whose main diagonal is strictly positive and whose associated graph $\mathcal{G}(M)$ is weakly connected. Then, the principal singular value and the principal singular vectors of M are unique.*

Proof. Note that $\mathcal{G}(M^k)$ has an edge from i to j if M contains a directed path of length k from i to j [6, Lemma 1.32]. Since $\mathcal{G}(MM^T)$ and $\mathcal{G}(M^TM)$ are strongly connected with self loops, Lemma 7.2.4 implies the existence of $\bar{k}$ such that $(M^TM)^k > 0$ and $(MM^T)^k > 0$ for all $k \geq \bar{k}$, and Perron Theorem [180, Theorem 8.2.11] implies that $(M^TM)^k$ and $(MM^T)^k$ have unique maximum eigenvalues for all $k \geq \bar{k}$.

The eigenvalues of $(M^TM)^k$ and $(MM^T)^k$ are related to the singular values of M by,

$$\lambda_i((M^TM)^k) \;=\; \lambda_i((MM^T)^k) \;=\; (\sigma_i(M))^{2k}$$

and the eigenvectors of $(M^TM)^k$ and $(MM^T)^k$ are determined by the right and the left singular vectors of M, respectively. Since the principal eigenvalue and eigenvectors of these matrices are unique, the principal singular value and the associated singular vectors of M are also unique. □

Theorem 7.2.6. *Let Assumption 3 hold, let $A_{\mathrm{cl}}(x) := A + F(x)$ be a Hurwitz matrix, and let matrices B and C have strictly positive main diagonals. If the graph $\mathcal{G}(A)$ associated with A is weakly connected, $f_\infty(x)$ is continuously differentiable.*

Proof. Lemma 7.1.1 implies that $\mathrm{e}^{A_{\mathrm{cl}}(x)} \geq 0$. From [6, Lemma 1.32], $\mathcal{G}(M^k)$ has an edge from i to j if there is a directed path of length k from i to j in $\mathcal{G}(M)$. Weak connectivity of $\mathcal{G}(A)$ implies weak connectivity of $\mathcal{G}(\tilde{A})$, $\mathcal{G}(\tilde{A}^k)$, $\mathrm{e}^{A_{\mathrm{cl}}(x)t}$ and, by (7.5), of $\mathcal{G}(-A_{\mathrm{cl}}^{-1}(x))$.

Since $A_{\mathrm{cl}}(x)$ is Hurwitz and Metzler, its main diagonal must be strictly negative; otherwise, $\frac{\mathrm{d}}{\mathrm{d}t}\psi_i \geq 0$ for some ψ_i, contradicting stability and thus the Hurwitz assumption. Equation (7.5) and Lemma 7.1.1 imply $A_{\mathrm{cl}}^{-1}(x) \leq 0$ and, since $A_{\mathrm{cl}}(x)$ is Metzler, $A_{\mathrm{cl}}^{-1}(x)A_{\mathrm{cl}}(x) = I$ can only hold if the main diagonal of $-A_{\mathrm{cl}}^{-1}(x)$ is strictly positive.

Moreover, since the diagonals of B and C are strictly positive, $\mathcal{G}(-CA_{\mathrm{cl}}^{-1}(x)B)$ is weakly connected and the diagonal of $-CA_{\mathrm{cl}}^{-1}(x)B$ is also strictly positive. Thus, Lemma 7.2.5 implies that the principal singular value and singular vectors of $-CA_{\mathrm{cl}}^{-1}(x)B$ are unique, that (7.4) is unique for each stabilizing x, and that $f_\infty(x)$ is continuously differentiable. □

Nondifferentiability for disconnected $\mathcal{G}(A)$

Theorem 7.2.6 implies that under a mild assumption on B and C, f_∞ is only nondifferentiable when the graph associated with A has disjoint components. Since the proximal operator of f_∞ does not readily admit an explicit expression, neither the existing proximal methods discussed in Section 2.2.3 nor the novel algorithms we develop in Part I may be directly applied to solve (1.8). Instead, subgradient methods [209] must be employed. Although proximal subgradient algorithms [210] can be applied to problems of the form (1.8) where *both* f and g are nondifferentiable and g has a cheaply computable proximal operator, subgradient methods require a large number of iterations to converge.

To a large extent, subgradient methods are inefficient because they do not guarantee descent of the objective function. However, under the following mild assumption, the subgradient of f_∞, ∂f_∞, can be conveniently expressed and a descent direction can be obtained by solving a linear program.

Assumption 4. Without loss of generality, let $A_{\mathrm{cl}}(x)$ be permuted such that

$$A_{\mathrm{cl}}(x) \;=\; \mathrm{blkdiag}\left(A_{\mathrm{cl}}^1(x), \ldots, A_{\mathrm{cl}}^m(x)\right)$$

is block diagonal and let $\mathcal{G}(A_{\mathrm{cl}}^i(x))$ be weakly connected for every i. Moreover, the matrices $B = \mathrm{blkdiag}\left(B^1, \ldots, B^m\right)$ and $C = \mathrm{blkdiag}\left(C^1, \ldots, C^m\right)$ are block diagonal and partitioned conformably with the matrix $A_{\mathrm{cl}}(x)$.

Theorem 7.2.7. *Let Assumptions 3 and 4 hold and let $A_{\mathrm{cl}}(x) := A + F(x)$ be a Hurwitz matrix. Then,*

$$f_\infty(x) \;=\; \max_i \; f_\infty^i(x) \tag{7.7a}$$

where $f_\infty^i(x) := \bar{\sigma}(C^i(A_{\mathrm{cl}}^i(x))^{-1}B^i)$. Moreover, every element of the subgradient of $f_\infty(x)$ can be

expressed as the convex combination of a finite number of vectors $G_j := \nabla f_\infty^j(x)$ *corresponding to the gradients of the functions* $f_\infty^j(x)$ *that achieve the maximum in* (7.7a)*, i.e.,* $f_\infty(x) = f_\infty^j(x)$*,*

$$\partial f_\infty(x) \;=\; \{G\alpha \mid \alpha \in \mathcal{P}\} \tag{7.7b}$$

where the columns of G *are given by* G_j *and* $\mathcal{P}$ *is the simplex.*

Proof. Since $A_{\mathrm{cl}}(x)$ is a block diagonal matrix, so is $A_{\mathrm{cl}}^{-1}(x)$ and Assumption 4 implies that $-CA_{\mathrm{cl}}^{-1}(x)B = \mathrm{blkdiag}\,(-C^i(A_{\mathrm{cl}}^i(x))^{-1}B^i)$ is also block diagonal. Thus,

$$f_\infty(x) \;=\; \bar{\sigma}(-CA_{\mathrm{cl}}^{-1}(x)B) \;=\; \max_i\; \bar{\sigma}(-C^i(A_{\mathrm{cl}}^i(x))^{-1}B^i)$$

which proves (7.7a). Theorem 7.2.6 implies that each $f_\infty^i(x)$ is continuously differentiable which establishes (7.7b) by [209, Theorem 1.13]. □

When g is differentiable, we leverage the above convenient expression for ∂f_∞ to select an element of the subdifferential set which, with an abuse of terminology, we call the *optimal* subgradient. The optimal subgradient is guaranteed to be a descent direction for (1.8) and it is defined as the member of $\partial(f_\infty(x) + \gamma g(x))$ that solves

$$\begin{aligned}
\underset{v,\,\alpha}{\text{minimize}} \qquad & \max_j\; (v^T(G_j \,+\, \gamma\nabla g(x))) && \text{(7.8a)}\\
\text{subject to} \qquad & v \,=\, -(G\alpha \,+\, \gamma\nabla g(x)),\quad \alpha \,\in\, \mathcal{P} && \text{(7.8b)}\\
& v^T(G_j \,+\, \gamma\nabla g(x)) \,<\, 0,\quad \text{for all } j && \text{(7.8c)}
\end{aligned}$$

where F and f^j are defined as in Theorem 7.2.7. While (7.8a) minimizes the directional derivative of (1.8) in the search direction v, constraints (7.8b) and (7.8c) ensure that $v \in \partial(f_\infty(x) + \gamma\nabla g(x))$ and that it is a descent direction, respectively.

Problem (7.8) can be solved efficiently because it is a linear program. Moreover, the optimality condition for (1.8), $\partial f_\infty(x) + \gamma\nabla g(x) \ni 0$, can be checked by solving a linear program to verify the existence of an $\alpha \in \mathcal{P}$ such that $G\alpha + \gamma\nabla g(x) = 0$.

Customized algorithm

Ensuring a descent direction enables principled rules for step-size selection and makes problem (1.8) with nondifferentiable g tractable via the augmented-Lagrangian-based approaches.

Introducing an auxiliary variable z and deriving the proximal augmented Lagrangian as described in Section 2.3.1 yields

$$\mathcal{L}_\mu(x, z^\star_\mu(x,y); y) \;=\; f_\infty(x) \;+\; M_{\mu g}(x \,+\, \mu y) \;-\; \tfrac{\mu}{2}\,\|y\|^2.$$

Since f_∞ is the only nondifferentiable component of the proximal augmented Lagrangian, the optimal subgradient (7.8) can be used to minimize $\mathcal{L}_\mu(x; y^k)$ over x to facilitate the MM algorithm in Section (2.3) [107].

7.3 Leader selection in directed networks

We now consider the special case of system (7.1), in which the matrix A is given by a graph Laplacian, and study the leader selection problem for directed consensus networks. The question of how to optimally assign a predetermined number of nodes to act as leaders in a network of dynamical systems with a given topology has recently emerged as a useful proxy for identifying important nodes in a network [77–84]. Even though significant theoretical and algorithmic advances for undirected networks have been made, the leader selection problem in directed networks remains open. We first discussed this problem in Section 1.2.1; we describe it in more detail here.

7.3.1 Problem formulation

We describe consensus dynamics and state the problem.

Consensus dynamics

The weighted directed network $\mathcal{G}(L)$ with n nodes and the graph Laplacian L obeys consensus dynamics in which each node i updates its state ψ_i using relative information exchange with its neighbors,

$$\dot{\psi}_i \;=\; -\sum_{j\,\in\,\mathcal{N}_i} L_{ij}(\psi_i \,-\, \psi_j) \;+\; d_i.$$

Here, $\mathcal{N}_i := \{j \,|\, (i,j) \in \mathcal{E}\}$, $L_{ij} \geq 0$ is a weight that quantifies the importance of the edge from node j to node i, d_i is a disturbance, and the aggregate dynamics are [67],

$$\dot{\psi} \;=\; -L\psi \;+\; d$$

where L is the graph Laplacian of the directed network [211].

The graph Laplacian always has an eigenvalue at zero that corresponds to a right eigenvector of all ones, $L\mathbb{1} = 0$. If this eigenvalue is simple, all node values ψ_i converge to a constant $\bar{\psi}$ in the absence of an external input d. When $\mathcal{G}(L)$ is balanced and conncted, $\bar{\psi} = \frac{1}{n}\,\mathbb{1}^T\psi(0)$ is the average of the initial node values. In general, $\bar{\psi} = w^T\psi(0)$, where w is the left eigenvector of L corresponding to zero eigenvalue, $w^T L = 0$. If $\mathcal{G}(L)$ is not strongly connected, L may have additional eigenvalues at zero and the node values converge to distinct groups whose number is equal to or smaller than the multiplicity of the zero eigenvalue.

Leader selection

In consensus networks, the dynamics are governed by relative information exchange and the node values converge to the network average. In the leader selection paradigm [80], certain "leader" nodes are additionally equipped with *absolute* information which introduces negative feedback on the states of these nodes. If suitable leader nodes are present, the dynamical generator becomes a Hurwitz matrix and the states of all nodes asymptotically converge to zero.

An example is given by a kinematic model of vehicles where ψ represents the position vector of a formation. Relative information exchange corresponds to maintaining constant distances between neighboring vehicles and the leader nodes may have access to absolute information from GPS units.

The node dynamics in a network with leaders is

$$\dot{\psi}_i \;=\; -\sum_{j\,\in\,\mathcal{N}_i} L_{ij}(\psi_i \,-\, \psi_j) \;-\; x_i\,\psi_i \;+\; d_i$$

where $x_i \geq 0$ is the weight that node i places on its absolute information. The node i is a leader if $x_i > 0$, otherwise it is a follower. The aggregate dynamics can be written as,

$$\dot{\psi} \;=\; -(L \,+\, \mathrm{diag}\,(x))\,\psi \;+\; d$$

and placed in the form (7.1) by taking $A = -L$, $B = C = I$, and $F(x) = -\,\mathrm{diag}\,(x)$. We evaluate the performance of a leader vector $x \in \mathbb{R}^n$ using the $\mathcal{H}_2$ or $\mathcal{H}_\infty$ performance metrics f_2 or f_∞, respectively. We note that this system is marginally stable in the absence of leaders and much work on consensus networks focuses on driving the *deviations* from the average node values to zero [97] as we discussed in the edge addition example for directed consensus networks in Section 2.4. Instead, we here focus on driving the node values themselves to zero.

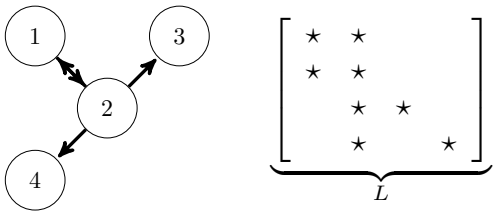


Figure 7.1: A directed network and the sparsity pattern of the corresponding graph Laplacian. This network is stabilized if and only if either node 1 or node 2 are made leaders.

We formulate the combinatorial problem of selecting N leaders to optimize either $\mathcal{H}_2$ or $\mathcal{H}_\infty$ norm as follows.

Problem 1. Given a network with a graph Laplacian L and a fixed leader weight κ, find the optimal set of N leaders that solves

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & \mathbb{1}^T x \;=\; N\,\kappa, \quad x_i \in \{0, \kappa\} \end{aligned}$$

where f is one of the performance metrics described in Section 1.2.2, with $A = -L$, $B = C = I$, and $F(x) = \operatorname{diag}(x)$.

In [78, 79], the authors derive explicit expressions for leaders in undirected networks. However, these expressions are efficient only for very few or very many leaders. Instead, we follow [80] and develop an algorithm which relaxes the integer constraint to obtain a lower bound on Problem 1 and use greedy heuristics to obtain an upper bound.

Considering leader selection in directed networks adds the challenge of ensuring stability. At the same time, we can leverage existing results on leader selection in undirected networks to derive efficient upper bounds on Problem 1.

7.3.2 Stability for directed networks

For a vector of leader weights x to be feasible for Problems 1, it must stabilize system (7.1); i.e., $-(L + \operatorname{diag}(x))$ must be a Hurwitz matrix. When $\mathcal{G}(L)$ is undirected and connected, any leader will stabilize (7.1). However, this is not the case for directed networks. For example, making node 1 or 2 a leader stabilizes the network in Fig. 7.1, but making nodes 3 or 4 a leader does not. Theorem 7.3.1 provides a necessary and sufficient condition for stability.

Theorem 7.3.1. *Let L be a weighted directed graph Laplacian and let $x \geq 0$. The matrix $-(L + \operatorname{diag}(x))$ is Hurwitz if and only if $w \circ x \neq 0$ for all nonzero w with $w^T L = 0$, where $\circ$ is the elementwise product.*

Proof. ($\Leftarrow$) If $w \circ x = 0$, $w^T \operatorname{diag}(x) = 0$. If, in addition, $w^T L = 0$, we have

$$-w^T(L + \operatorname{diag}(x)) = 0 \tag{7.9}$$

and therefore zero is an eigenvalue of $-(L + \operatorname{diag}(x))$.

($\Rightarrow$) Since the graph Laplacian L is row stochastic and $\operatorname{diag}(x)$ is diagonal and nonnegative, the Gershgorin circle theorem [180] implies that the eigenvalues of $-(L+\operatorname{diag}(x))$ are at most 0. To show that $-(L+\operatorname{diag}(x))$ is Hurwitz, we show that it has no eigenvalue at zero. Assume there exists a nonzero w such that (7.9) holds. This implies that either $w^T L = w^T \operatorname{diag}(x) = 0$ or that $w^T L = -w^T \operatorname{diag}(x)$. The first case is not possible because, by assumption, $w^T \operatorname{diag}(x) = (w \circ x)^T \neq 0$ for any w such that $w^T L = 0$. If the second case is true, then $w^T L v = -w^T \operatorname{diag}(x) v$ must also hold for all v. However, if we take $v = \mathbb{1}$, then $w^T L \mathbb{1} = 0$ but $-w^T \operatorname{diag}(x)\mathbb{1}$ is nonzero. This completes the proof. □

Remark 7.3.2. *Only the set of leader nodes is relevant to the question of stability. If x does not stabilize* (7.1)*, no positive weighting of the vector of leader nodes, $\alpha \circ x$ with $\alpha \in \mathbb{R}^n_{++}$, will stabilize* (7.1)*. Similarly if x stabilizes* (7.1)*, every $\alpha \circ x$ will.*

Corollary 7.3.2.1. *If $\mathcal{G}(L)$ is strongly connected, any choice of leader node will stabilize* (7.1)*.*

Proof. Since the graph Laplacian associated with a strongly connected graph is irreducible, the Perron-Frobenius theorem [180] implies that the left eigenvector associated with $-L$ is positive. Thus, $w \circ x \neq 0$ for any nonzero x and system (7.1) is stable by Theorem 7.3.1. □

Remark 7.3.3. *The condition in Theorem 7.3.1 requires that there is a path from the set of leader nodes to every node in the network. This can be enforced by extracting disjoint "leader subsets" $\mathcal{S}_j$ which are not influenced by the rest of the network, i.e., $(v^j)^T L = 0$ where $v^j_i = 1$ if $i \in \mathcal{S}_j$ and $v^j_i = 0$ otherwise, and which are each strongly connected components of the original network. Stability is guaranteed if there is at least one leader node in each such subset $\mathcal{S}_j$; e.g., for the network in Fig. 7.1, there is one leader subset $\mathcal{S}_1 = \{1, 2\}$. By Corollary 7.3.2.1, $\mathcal{S}_1$ contains all nodes when the network is strongly connected.*

7.3.3 Bounds for Problem 1

To approach combinatorial Problem 1, we derive bounds on its optimal objective value. These bounds can also be used to implement a branch-and-bound approach [212].

Lower bound

By relaxing the combinatorial constraint in Problem 1, we obtain a convex program,

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \kappa \mathcal{P}_N \end{aligned}$$

where $\kappa \mathcal{P}_N := \{x \mid \sum_i x_i = N\kappa,\, x_i \leq \kappa\}$ is the "capped" simplex. Using a recent result on efficient projection onto $\mathcal{P}_N$ [213], this problem can be solved efficiently via proximal gradient methods [131] to provide a lower bound on Problem 1.

When $\mathcal{G}(L)$ is not strongly connected, additional constraints can be added to enforce the condition in Theorem 7.3.1 and thus guarantee stability. Let the sets $\mathcal{S}_j$ denote "leader subsets" from which a leader must be chosen, as discussed in Remark 7.3.3. Then, the convex problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & x \in \kappa \mathcal{P}_N, \\ & \sum_{i \in \mathcal{S}_j} x_i \geq \kappa, \text{ for all } j \end{aligned} \tag{7.10}$$

relaxes the combinatorial constraint and guarantees stability. We denote the resulting lower bounds on the optimal values of the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ versions of Problem 1 with N leaders by $f_2^{\text{lb}}(N)$ and $f_\infty^{\text{lb}}(N)$, respectively.

Upper bounds for Problem 1

If k denotes the number of subsets $\mathcal{S}_j$, a stabilizing candidate solution to Problem 1 can be obtained by "rounding" the solution to (7.10) by taking the N leaders to contain the largest element from each subset $\mathcal{S}_j$ and $N - k$ largest remaining elements. The greedy swapping algorithm proposed in [80] can further tighten this upper bound.

Recent work on leader selection in undirected networks can also provide upper bounds for Problem 1 when $\mathcal{G}(L)$ is balanced. The symmetric component of the Laplacian of a balanced graph, $L_s := \frac{1}{2}(L + L^T)$, is the Laplacian of an undirected network. The exact optimal leader set

for an undirected network can be efficiently computed when N is either small or large [78, 79]. Since the performance of the symmetric component of a system provides an upper bound on the performance of the original system, these sets of leaders will have better performance with L than with L_s for both the $\mathcal{H}_2$ (Corollary 6.1.3.1) and $\mathcal{H}_\infty$ norms (Proposition 6.1.2).

Even when L does not represent a balanced network, f_2 and f_∞ are respectively upper bounded by the trace and the maximum eigenvalue of $\frac{1}{2}(L_s + F(x))^{-1}$. For small numbers of leaders, they can be efficiently computed using rank-one inversion updates. A similar approach was used in [78, 79] to derive optimal leaders for undirected networks. Moreover, this approach always yields a stabilizing set of leaders (Lemma 6.1.1).

7.3.4 Additional comments

We now provide additional discussion on interesting aspects of Problem 1. We first consider the gradients of f_2 and f_∞.

Remark 7.3.4. *When $F(x) = -\operatorname{diag}(x)$, we have $\nabla f_2 = -2\operatorname{diag}(P_c P_o)$. The matrix $P_c P_o$ often appears in model reduction and $(\nabla f_2(x))_i$ corresponds to the inner product between the ith columns of P_c and P_o.*

Remark 7.3.5. *When $F(x) = -\operatorname{diag}(x)$, $(\partial f_\infty(x))_i$ is given by the product of $-\mathrm{e}_i^T A_{\mathrm{cl}}^{-1}(x)v$ and $w^T A_{\mathrm{cl}}^{-1}(x)\mathrm{e}_i$. The former quantifies how much the forcing which causes the largest overall response of system (7.1) affects node i, and the latter captures how much the forcing at node i affects the direction of the largest output response.*

The optimal leader sets for balanced graphs are interesting because they are invariant under reversal of all edge directions.

Proposition 7.3.1. *Let $\mathcal{G}(L)$ be balanced, let $\hat{L} := L^T$ so that $\mathcal{G}(\hat{L})$ contains the reversed edges of the graph $\mathcal{G}(L)$, and let $\hat{f}_2$ and $\hat{f}_\infty$ denote the performance metrics (7.13) with $A = -\hat{L}$, $F(x) = -\operatorname{diag}(x)$, and $B = C = I$ as in Problem 1. Then, $f_2(x) = \hat{f}_2(x)$ and $f_\infty(x) = \hat{f}_\infty(x)$.*

Proof. The controllability gramian of (7.1) defined with $A_{\mathrm{cl}} = -(L + \operatorname{diag}(x))$ solves Lyapunov equation (7.3a), $-(L + \operatorname{diag}(x))P_c - P_c(L + \operatorname{diag}(x))^T + I = 0$, and is also the observability gramian $\hat{P}_o$ of (7.1) defined with $A_{\mathrm{cl}} = -(\hat{L} + \operatorname{diag}(x)) = -(L^T + \operatorname{diag}(x))$ that solves (7.3b). By definition of the $\mathcal{H}_2$ norm, $\hat{f}_2(x) = \operatorname{trace}(\hat{P}_o) = \operatorname{trace}(P_c) = f_2(x)$. Since $\bar{\sigma}(M) = \bar{\sigma}(M^T)$, $\hat{f}_\infty(x) = \bar{\sigma}(-(\hat{L} + \operatorname{diag}(x))^{-1}) = \bar{\sigma}(-(L + \operatorname{diag}(x))^{-1}) = f_\infty(x)$. □

This invariance is intriguing because the space of balanced graphs is spanned by cycles. In [214], the authors explored how undirected cycles affect undirected consensus networks.

Proposition 7.3.1 suggests that directed cycles also play a fundamental role in directed consensus networks.

7.4 Computational experiments for leader selection

Here, we illustrate our approach to Problem 1 with N leaders and the weight $\kappa = 1$. The "rounding" approach that we employ is described in Section 7.3.3.

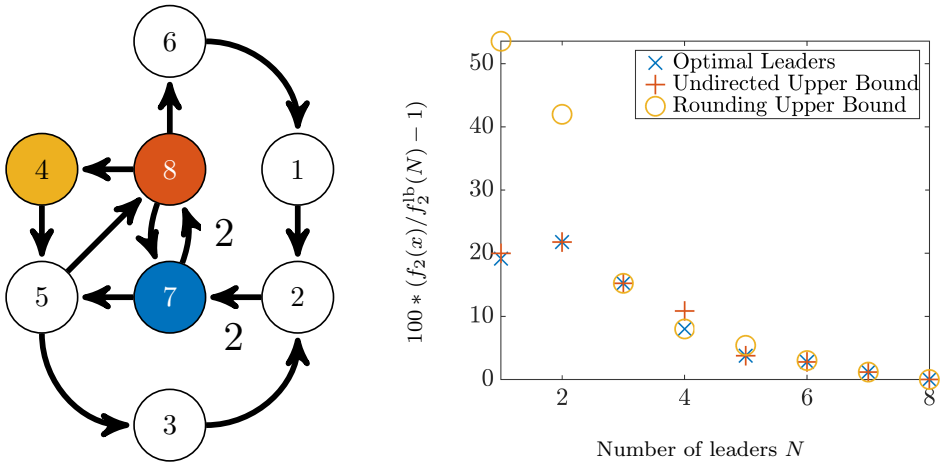
7.4.1 Example: Bounds on leader selection for a small network

For the directed network in Fig. 7.2a, let the edges from node 2 to node 7 and from node 7 to node 8 have an edge weight of 2 and let all other edges have unit edge weights. We compare the optimal set of leaders, determined by exhaustive search, to the set of leaders obtained by (i) "rounding" the solution to relaxed problem (7.10); and by (ii) the optimal selection for the undirected version of the graph via [78, 79], as discussed in Section 7.3.3. In Fig. 7.2a, **blue node** 7 represents the optimal single leader, **yellow node** 4 represents the single leader selected by "rounding", and **red node** 8 represents the optimal single leader for the undirected network. In Fig. 7.2b we show the $\mathcal{H}_2$ performance for 1 to 8 leader nodes resulting from different methods. Since in general, we do not know the optimal performance *a priori*, we plot performance degradation (in percents) relative to the lower bound on Problem 1 obtained by solving problem (7.10).

Figure 7.2b shows that neither "rounding" (**yellow** $\circ$) nor the optimal selection for undirected networks (**red** $+$) achieve unilaterally better $\mathcal{H}_2$ performance (performance of the optimal leader sets are shown in **blue** $\times$). While the procedure for the undirected networks selects better sets of 1, 2, and 5 leaders relative to "rounding", identifying them is expensive except for large or small number of leaders [78,79] and "rounding" identifies a better set of 4 leaders. This suggests that, when possible, *both* sets of leaders should be computed and the one that achieves better performance should be selected.

7.4.2 Example: Leaders in the neural network of the worm C. Elegans

We now consider the network of neurons in the brain of the worm C. Elegans with 297 nodes and 2359 weighted directed edges. The data was compiled by [215] from [216]. Inspired by the use of leader selection as a proxy for identifying important nodes in a network [77–80, 82], we employ this framework to identify important neurons in the brain of C. Elegans.



(a) Balanced Network with 8 nodes, 12 edges.

(b) Performance of optimal leaders and two leader selection techniques.

Figure 7.2: $\mathcal{H}_2$ performance of optimal leader set (**blue** $\times$) and upper bounds resulting from "rounding" (**yellow** $\circ$) and the optimal leaders for the undirected network (**red** $+$). Performance is shown as a percent increase in f_2 relative to $f_2^{\text{lb}}(N)$.

Three nodes in the network have zero in-degree, i.e., they are not influenced by the rest of the network. Thus, as discussed in Remark 7.3.3, there are three "leader subsets", each comprised of one of these nodes. Theorem 7.3.1 implies that system (7.1) can only be stable if each of these nodes are leaders.

In Figs. 7.4c and 7.4d, we show f_2 and f_∞ resulting from "rounding" the solution to problem (7.10) to select the additional 1 to 294 leaders. Performance is plotted as an increase (in percents) relative to the lower bound $f^{\text{lb}}(N)$ obtained from (7.10). This provides an upper bound on suboptimality of the identified set of leaders. While this value does not provide information about how $f(x)$ changes with the number of leaders, Remark 7.2.3 implies that it monotonically decreases with N.

For both f_2 and f_∞ performance metrics, Figs. 7.4c and 7.4d illustrates that the upper bound is loosest for 25 leaders (1.56% and 0.48%, respectively). As seen in Fig. 7.2b from the previous example, whose small size enabled exhaustive search to solve Problem 1 exactly, the upper bound on suboptimality is not tight and the exact optimal solution to Problem 1 can differ by as much 21.75% from the lower bound. This suggests that "rounding" selects very good sets of leaders for the C. Elegans example.

In Figs. 7.4a and 7.4b, we show the network with ten identified f_2 and f_∞ optimal leaders. The size of the nodes is related to their out-degree and the thickness of the edges is related to

the weight. The **red** $\diamond$ marks nodes that *must* be leaders and the **blue** $\circ$ marks the 7 additional leaders selected by "rounding".

7.4.3 Example: Ranking college football teams

Inspired by the recent use of graph theoretic tools for ranking athletic teams, we consider the problem of ranking college football teams. Here, we do not consider that each leader has the same weight and instead use a sparsity penalty in (1.8) to select sparse subsets of nodes (teams).

Due to the number of teams in the top division of college football (128) and the relative scarcity of games between them (around 13 per team), ranking these teams is an underdetermined problem. The current practice of ranking by a committee is clearly subject to bias. Recently, graph theoretic measures, such as average path length from each node, have been explored for the purpose of objectively ranking teams or athletes [217, 218].

We used the scores of college football games from the $2015-2016$ season collected from [219] to generate a network. If team A beat team B, a edge was placed from A to B with a weight equal to the score difference in the game. There were 203 teams (nodes) and 863 games (edges) in our data set. We select the top N teams by identifying N $\mathcal{H}_2$ optimal leaders. We selected sparse sets of leaders by solving (1.8) with g as the ℓ_1 norm and increasing γ until its solution had the desired number of nonzero elements. Interestingly, the metrics we use are biased against selecting leaders which are close in the network. In this context, such proximity would correspond to teams who have many common opponents.

Fig. 7.3 shows the network generated. The large connected component in the center represents the teams in the top division. Since our dataset included games played between teams from the top division and lower divisions but not from games played between teams of the lower divisions, there are a number of topgoraphically isolated nodes.

Table 7.1 shows sets of 2, 4, 6, and 8 leaders with the corresponding end-of-season rankings from the Associated Press (AP) [220]. This approach selected teams which agree well with the AP rankings with the notable exception of Southern Illinois. This team played only one game in the our set – a close loss to a poorly ranked team. We ascribe this anomaly to the the topological isolation of this team.

Figure 7.3: Network of games in the $2015-2016$ College Football season. The central connected component corresponds to the top division, and the distal nodes are lower division teams for whom there are less data.

7.5 Computational experiments for combination drug therapy

System (7.1) also arises in the modeling of combination drug therapy [89,90,221–223], which we first discuss in Section 1.2.1. It provides a model for the evolution of populations of mutants of the HIV virus ψ in the presence of a combination of drugs x. The HIV virus is known to be present in the body in the form of different mutant strands; in (7.1), the ith component of the state vector ψ represents the population of the ith HIV mutant. The diagonal entries of the matrix A represent the net replication rate of each mutant, and the off diagonal entries of A, which are all nonnegative, represent the rate of mutation from one mutant to another. The control input x_k is the dose of drug k and the kth column of the matrix F_x in $F(x) = \operatorname{diag}(F_x x)$ specifies at what rate drug k kills each HIV mutant.

7.5.1 Example: Simple problem with nondifferentiable f_∞

The mutation patterns of viruses need not be connected. In Fig. 7.5, we show a sample mutation network with 2 disconnected components. For this network, the $\mathcal{H}_\infty$ norm is nondifferentiable when $x_1 = x_2$. Nondifferentiability and the lack of an efficiently computable proximal operator

γ	0.225	0.6	1	10
Teams	8	6	4	2
AP #1	Alabama	Alabama	Alabama	Alabama
AP #4	Ohio State	Ohio State	Ohio State	Ohio State
AP #2	Clemson	Clemson	Clemson	
AP #8	Houston	Houston	Houston	
AP #3	Stanford	Stanford		
AP NR	S. Illinois	S. Illinois		
AP #12	Michigan			
AP #10	Mississippi			

Table 7.1: Leaders selected for different values of γ.

necessitates the use of subgradient methods for solving

$$\underset{x}{\text{minimize}} \quad f_\infty(x) \; + \; x^T x.$$

As shown in Fig. 7.6 with $h(x) := f_\infty(x) + x^T x$, subgradient methods are not descent methods so small constant or a divergent series of diminishing step-sizes must be employed.

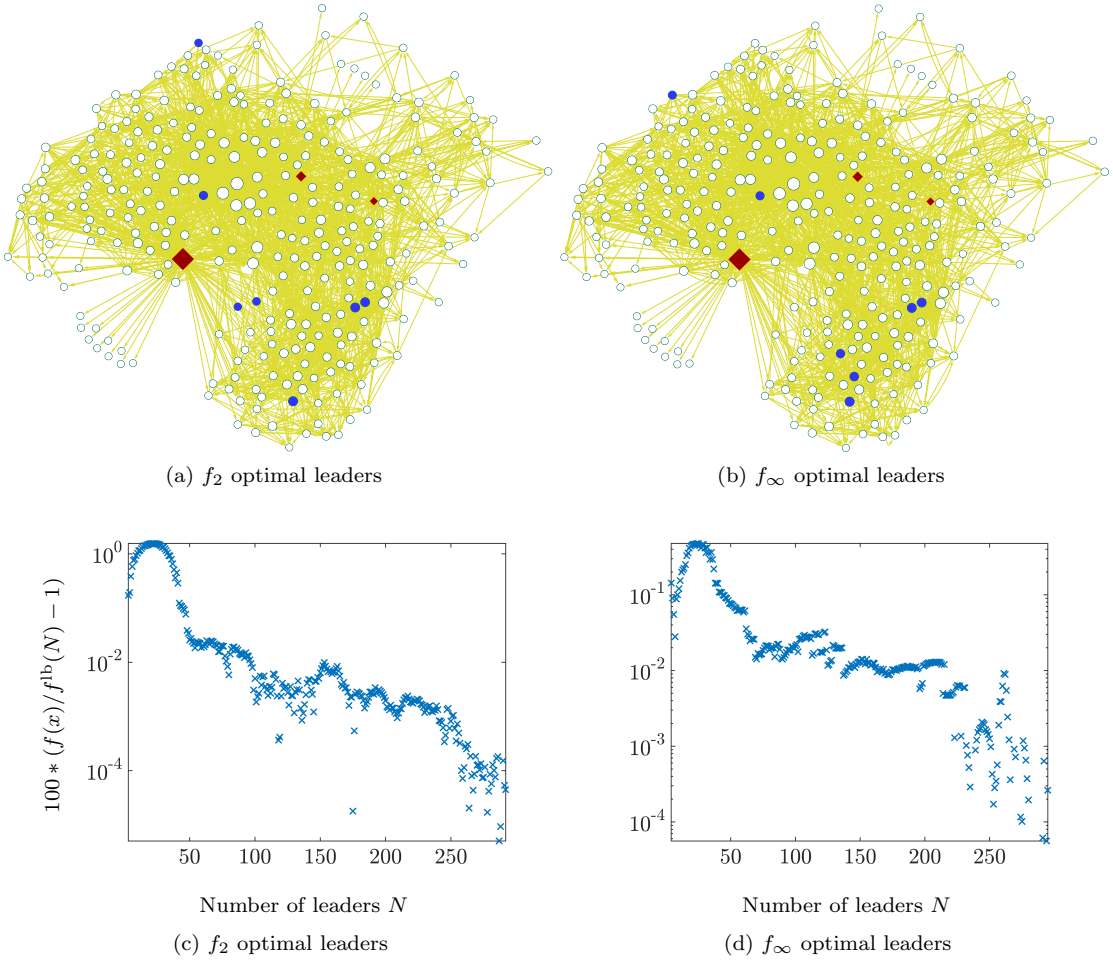
We compare the performance of the subgradient method with a constant step-size of 10^{-2} (blue) and a diminishing step-size $\frac{7\times 10^{-2}}{k}$ (red) with our optimal subgradient method in which the step-size is chosen via backtracking to ensure descent of the objective function (yellow). We show the objective function value with respect to iteration number in Fig. 7.6a and the iterates x^k in the (x_1, x_2)-plane in Fig. 7.6b.

The subgradient methods were run for 1000 iterations as there is no principled stopping criterion. Our optimal subgradient method identified the optimal point, to an accuracy of 10^{-4} (i.e., there was a $v \in \partial(f_\infty(x) + x^T x)$ such that $\|v\| \leq 10^{-4}$) in 23 iterations.

7.5.2 Example: Real world drug therapy problem

Following [89, 90] and using data from [200], we study a system with 35 mutants and 5 drugs. The sparsity pattern of the matrix A, shown in Fig. 7.7, corresponds to the mutation pattern and replication rates of 33 mutants and $F(x)$ specifies the effect of drug therapy. Two mutants are not shown in Fig. 7.7a as they have no mutation pathways to or from other mutants.

Several clinically relevant properties, such as maximum dose or budget constraints, may be directly enforced in our formulation. Other combinatorial conditions can be promoted via convex penalties, such as drug j requiring drug i via $x_i \geq x_j$ or mutual exclusivity of drugs i and j via $x_i + x_j \leq 1$. We design optimal drug doses using two convex regularizers g.



(a) f_2 optimal leaders

(b) f_∞ optimal leaders

(c) f_2 optimal leaders

(d) f_∞ optimal leaders

Figure 7.4: C. Elegans neural network with $N = 10$ (a) f_2 and (b) f_∞ leaders along with the (c) f_2 and (d) f_∞ performance of varying numbers of leaders N relative to $f^{\mathrm{lb}}(N)$. In all cases, leaders are selected via "rounding".

Budget constraint

We impose a unit budget constraint on the drug doses and solve the f_2 and f_∞ problems using proximal gradient methods [131, 134]. These can be cast in the form (1.8), where g is the indicator function associated with the probability simplex $\mathcal{P}$. Table 7.2 contains the optimal doses and illustrates the tradeoff between $\mathcal{H}_2$ and $\mathcal{H}_\infty$ performance.

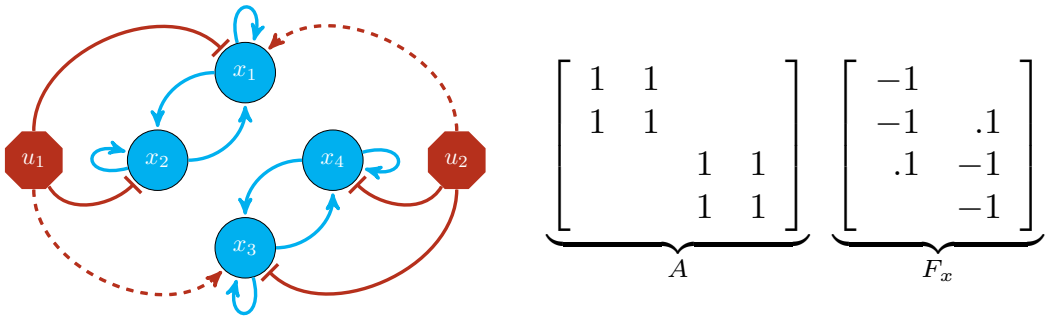


Figure 7.5: A directed network and corresponding A matrix for a virus with 4 mutants and 2 drugs. For this system, f_∞ is nondifferentiable.

Antibody	$\mathbf{x_2^\star}$	$x_\infty^\star$
3BC176	0.5952	0.9875
PG16	0	0
45-46G54W	0.2484	0.0125
PGT128	0.1564	0
10-1074	0	0

$f_2(x_2^\star)$	0.6017
$f_2(x_\infty^\star)$	1.1947
$f_\infty(x_2^\star)$	0.1857
$f_\infty(x_\infty^\star)$	0.1084

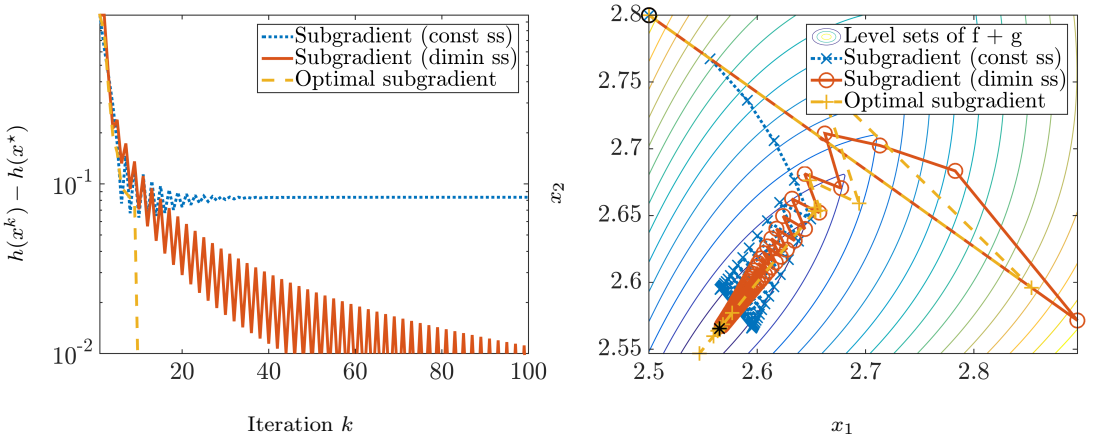
Table 7.2: Optimal budgeted doses and $\mathcal{H}_2/\mathcal{H}_\infty$ performance.

Sparsity-promoting framework

Although the above budget constraint is naturally sparsity-promoting, in Algorithm 5 we augment a quadratically regularized optimal control problem with a reweighted ℓ_1 norm [121] to select a homotopy path of successively sparser sets of drugs. We then perform a 'polishing' step to design the optimal doses of the selected set of drugs. We use 50 logarithmically spaced increments of the regularization parameter γ between 0.01 and 10 to identify the drugs and then replace the weighted ℓ_1 penalty with a constraint to prescribe the selected drugs. In Fig. 7.8, we show performance degradation (in percents) relative to the optimal dose that uses all 5 drugs with $B = C = I$ and $R = I$.

7.6 Time-varying controllers

While references [89, 90, 112, 115, 222] and the previous section either assume a constant control signal or use heuristics to introduce time dependence, we show that such a constant input is in fact optimal for the induced power norm. We cast the optimal synthesis problem of constant



(a) Descent of objective function.

(b) Iterates in the (x_1, x_2)-plane.

Figure 7.6: Comparison of different algorithms starting from initial condition $[2.5\ 2.8]^T$. The algorithms are the subgradient method with a constant step-size (**dotted blue** $\cdots$), the subgradient method with a diminishing step-size (**solid red** —) and our optimal subgradient method where the step-size is chosen via backtracking to ensure descent of the objective function (**dashed yellow - - -**)

Algorithm 5 Sparsity-promoting algorithm for N drugs

Set $\gamma > 0$, $R \succ 0$, $w = \mathbb{1}$, $\varepsilon > 0$

while card$(x_\gamma) > N$ **do**

$$x_\gamma \;=\; \underset{x}{\operatorname{argmin}} \;\; f(x) \;+\; x^T R\, x \;+\; \gamma \sum_i w_i \,|x_i|$$

increase γ, set $w_i \;=\; 1/(x_i \;+\; \varepsilon)$

end

$$x_N^\star \;=\; \underset{x}{\operatorname{argmin}} \;\; f(x) \;+\; x^T R\, x \quad \text{subject to} \;\; \operatorname{sp}(x) \;\subseteq\; \operatorname{sp}(x_\gamma)$$

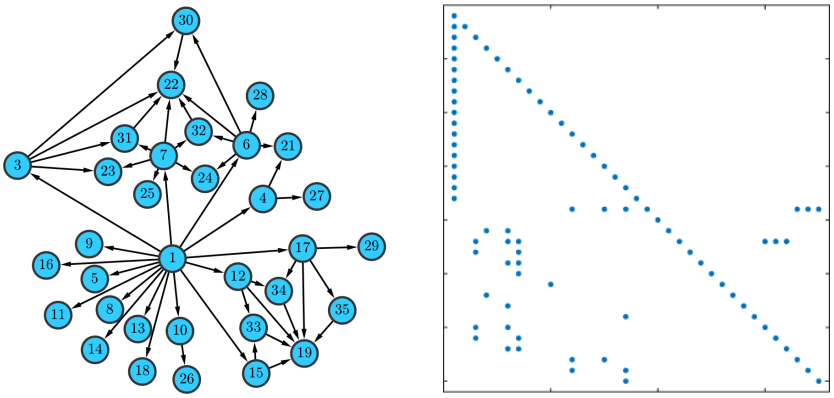
---

control inputs as a finite-dimensional non-smooth convex optimization problem and develop an algorithm for designing the optimal controller.

7.6.1 Preliminaries

The space of square integrable signals is denoted by $\mathcal{L}_2$. The inner product in this space is given by

$$\langle u,\, v\rangle_2 \;:=\; \int_0^\infty u^T(t)\, v(t)\, \mathrm{d}t$$



(a) Network of HIV mutation pathways

(b) Sparsity pattern of A

Figure 7.7: Mutation pattern of HIV.

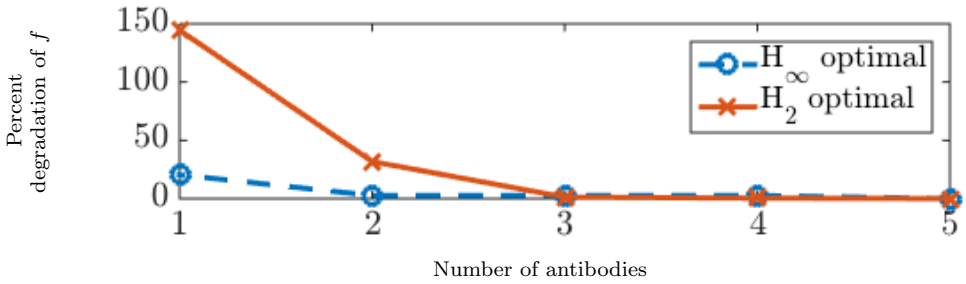


Figure 7.8: Percent performance degradation for $\mathcal{H}_2$ (**solid red** —×—) and $\mathcal{H}_\infty$ (**dashed blue** – – ○ – –) performance relative to using all 5 drugs.

with the associated norm $\|v\|_2^2 = \langle v, v\rangle_2$. The $\mathcal{L}_2$ signals are those for which $\|v\|_2^2$ is finite and the *locally* $\mathcal{L}_2$ signals are those for which $\int_\Omega v^T(t)\, v(t)\, \mathrm{d}t$ is finite over any compact set Ω, e.g., $t \in [0, T]$ for a finite T. The power semi-norm of a signal v is

$$\|v\|_{\text{pow}}^2 \;:=\; \limsup_{T\to\infty}\; \frac{1}{T}\int_0^T v^T(t)\, v(t)\, \mathrm{d}t. \tag{7.11}$$

The space of trigonometric polynomials is defined as

$$\mathcal{T} \;:=\; \left\{ g : \mathbb{R} \to \mathbb{R}^n \;\middle|\; g(t) = \sum_{k=1}^{N} \alpha_k \mathrm{e}^{\mathrm{j}\lambda_k t},\; \lambda_k \in \mathbb{R},\, \alpha_k \in \mathbb{C} \right\},$$

where j is the imaginary unit.

The closure of $\mathcal{T}$ in the space of locally $\mathcal{L}_2$ integrable functions with bounded power norm

with respect to the metric $\|f - g\|_{\text{pow}}$ is given by the space of Besicovitch almost periodic functions. Since $\|\cdot\|_{\text{pow}}$ is a seminorm, we consider $\mathcal{B}_2$, the space of Besicovitch almost periodic functions modulo functions with zero power norm [224]. Factoring out signals with zero power norm is natural in this setting because our performance metric is the power norm of a regulated output. The inner product associated with this Hilbert space is [225, 226]

$$\langle u,\, v\rangle \;:=\; \limsup_{T\to\infty}\; \frac{1}{T}\int_0^T u^T(t)\, v(t)\,\mathrm{d}t$$

and the norm $\|\cdot\|_{\text{pow}}$. The mean of a signal $v \in \mathcal{B}_2$

$$M(v) \;:=\; \lim_{T\,\to\,\infty}\; \frac{1}{T}\int_0^T v(t)\,\mathrm{d}t$$

is well defined for every $v \in \mathcal{B}_2$. Furthermore, each $v \in \mathcal{B}_2$ can be decomposed uniquely as $v = \bar{v} + \tilde{v}$, where $\bar{v}$ is a constant signal given by $\bar{v} = M(v)$ and $M(\tilde{v}) = 0$. Note that the inner product between a constant signal $\bar{v}$ and a zero-mean signal $\tilde{v}$ is zero, i.e., $\langle \bar{v},\, \tilde{v}\rangle = 0$.

The space $\mathcal{B}_2$ contains all bounded $\mathcal{L}_2$ signals, periodic signals, and almost periodic signals. At the same time, it alleviates challenges arising from the fact that the space of signals with bounded power norm is not a Hilbert space; for additional discussion see [227].

7.6.2 Problem formulation

Consider bilinear positive system (7.1) where x is now allowed to be time-varying and $x(t) \in \mathbb{R}^m$,

$$\dot{\psi} \;=\; (A \,+\, F(x(t)))\,\psi \,+\, B\,d. \tag{7.12a}$$

For given control and disturbance signals $x \in \mathcal{B}_2$ and $d \in \mathcal{B}_2$, we associate the performance output,

$$\zeta_{x,d} \;=\; \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix} \zeta \;+\; \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} x \tag{7.12b}$$

with (7.12a), where $Q \succeq 0$ and $R \succ 0$ are the state and control weights. Under Assumption 3, (7.12) is a positive system. This implies that for every control input x, every nonnegative disturbance d, and every nonnegative initial condition $\psi(0)$, the state ψ and the output $\zeta_{x,d}$ of system (7.12) remain nonnegative at all times.

The induced power norm of a stable system (7.12) is,

$$f(x) := \begin{cases} \sup\limits_{\|d\|_{\text{pow}}^2 \leq 1} \|\zeta_{x,d}\|_{\text{pow}}^2, & \text{system (7.12) stable} \\ \infty, & \text{otherwise} \end{cases} \tag{7.13}$$

and it quantifies the response to the worst case persistent disturbance d. Since system (7.12) is nonlinear, we cannot separate the initial condition and input response and we must define our notion of stability. We assume that $\psi(0) = 0$ and consider stability in a bounded input-bounded output sense.

Definition 6. A control signal $\bar{x}$ is stabilizing for nonlinear system (7.12) if for every bounded $d \in \mathcal{B}_2$, $\zeta_{\bar{x},d}$ is bounded and has finite power, $\|\zeta_{\bar{x},d}\|_{\text{pow}} < \infty$.

For unstable open-loop systems (7.12), there may be no stabilizing control input x in $\mathcal{L}_2$. Thus, the $\mathcal{L}_2$-induced gain does not provide a suitable measure of input-output amplification for (7.12) and $f(x)$ represents an appropriate generalization of the $\mathcal{H}_\infty$ norm for this class of bilinear positive systems.

We now formulate the optimal control problem.

Problem 2. Design a stabilizing bounded control signal $x \in \mathcal{B}_2$ to minimize $f(x)$ for bilinear positive system (7.12).

7.6.3 Solution to the optimal control problem

In this section, we prove that a constant control input solves Problem 2.

Since $f(x)$ is given by (7.13), any $x^\star$ which solves Problem 2 satisfies $\|\zeta_{x^\star,d}\|_{\text{pow}}^2 \leq f(x^\star) \leq f(x)$ for all $x \in \mathcal{B}_2$ and $d \in \mathcal{B}_2$. In particular, $\|\zeta_{x^\star,d}\|_{\text{pow}}^2 \leq f(\bar{x})$ where $\bar{x}$ is a constant control input. As shown in [54], for constant control inputs the worst-case disturbance $\bar{d}$ is also constant, i.e., $f(\bar{x}) = \|\zeta_{\bar{x},\bar{d}}\|_{\text{pow}}^2$. In what follows, we show that $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ is a convex function of x, that a constant $\bar{x}$ minimizes it, and, thus, that a constant control input solves Problem 2.

We first establish convexity of $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$.

Lemma 7.6.1. *Let $d(t) = \bar{d}$ be a constant non-negative disturbance. Then, under Assumption 3, the power norm of the output $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ is a convex function of $x \in \mathcal{B}_2$.*

Proof. We first show that

$$\psi^T(t)Q\psi(t) + x^T(t)Rx(t) \tag{7.14}$$

is a convex function of $x \in \mathcal{L}_2[0,t]$ for any t and then extend this statement to complete the proof. The matrix R is positive definite so the second term on the right-hand side of (7.14) is a convex function of $x(t)$. By [58, Theorem 2] every component of $\psi(t)$ is a convex function of the control input $x \in \mathcal{L}_2[0,t]$. Since the matrix $Q \succeq 0$ is positive semidefinite and has nonnegative entries, $\psi(t)^T Q \psi(t)$ is convex and nondecreasing in the elements of $\psi(t)$. Thus, the composition rules for convex functions [123, Section 3.2.4] imply that $\psi^T(t) Q \psi(t)$ is a convex function of $u \in \mathcal{L}_2[0,t]$.

Since the integral preserves convexity,

$$\frac{1}{T} \int_0^T \left(\psi^T(t) Q \psi(t) \, + \, x^T(t) R x(t) \right) \mathrm{d}t$$

is a convex function of $x \in \mathcal{L}_2[0,T]$ Let P_T denote the mapping that truncates the support of a signal v to $[0,T]$. If there is T such that $P_T v$ is not $\mathcal{L}_2$ integrable, the power seminorm (7.11) of v is infinite. This implies that, for any $v \in \mathcal{B}_2$, $P_T v \in \mathcal{L}_2[0,T]$. Thus, any $v \in \mathcal{B}_2$ can be written as $\lim_{T \to \infty} P_T v$ of $P_T v \in \mathcal{L}_2[0,T]$. Since the limit and supremum preserve convexity,

$$\| \zeta_{x,\bar{d}} \|_{\mathrm{pow}}^2 \, = \, \limsup_{T \to \infty} \, \frac{1}{T} \int_0^T \left(\psi^T(t) Q \psi(t) \, + \, x^T(t) R x(t) \right) \mathrm{d}t$$

is a convex function of $x \in \mathcal{B}_2$. □

Lemma 7.6.2. *Let $\bar{x}$ be a stabilizing constant control input for* (7.12a) *and $\bar{d}$ be a constant non-negative disturbance. Then, the directional derivative of $\| \zeta_{x,\bar{d}} \|_{\mathrm{pow}}^2$ evaluated at $\bar{x}$ is zero for any bounded zero-mean variation $\tilde{x} \in \mathcal{B}_2$.*

Proof. The dynamics (7.12a) with control input $\bar{x} + \varepsilon \tilde{x}$ and constant disturbance $\bar{d}$ are

$$\dot{\psi} \, = \, (A \, + \, F(\bar{x} \, + \, \varepsilon \tilde{x}))\, \psi \, + \, B\, \bar{d}. \tag{7.15}$$

Since the unperturbed system (with $\varepsilon = 0$) is exponentially stable and boundedness of $\tilde{x}$ implies that the solution $x(t)$ is continuous in ε [183, Theorem 3.5]. Therefore, (7.15) represents a system in a regularly perturbed form [183, 228]. The Taylor series expansion can be used to write the solution to the perturbed dynamics as,

$$\psi(t) \, = \, \bar{\psi}(t) \, + \, \varepsilon\, \tilde{\psi}(t) \, + \, \mathcal{O}(\varepsilon^2) \tag{7.16}$$

where $\bar{\psi}$ is the nominal solution that solves (7.15) for $\varepsilon = 0$

$$\dot{\bar{\psi}} \;=\; (A \,+\, F(\bar{x}))\,\bar{\psi} \,+\, B\,\bar{d}. \tag{7.17a}$$

and $\tilde{\psi}$ represents the first-order correction. By [183, Theorem 10.2], $\tilde{\psi}$ is given by the solution to a differential equation corresponding to the $\mathcal{O}(\varepsilon)$ terms in the expression obtained by substituting (7.16) into equation (7.15),

$$(\dot{\bar{\psi}} \,+\, \varepsilon\,\dot{\tilde{\psi}} \,+\, \mathcal{O}(\varepsilon^2)) \;=\; \left[(A \,+\, F(\bar{x}))\,\bar{\psi} \,+\, B\,\bar{d}\right] \,+\, \varepsilon\left[(A \,+\, F(\bar{x}))\,\tilde{\psi} \,+\, F(\tilde{u})\,\bar{\psi}\right] \,+\, \mathcal{O}(\varepsilon^2).$$

Collecting the $\mathcal{O}(\varepsilon)$ terms yields the dynamics that governs the evolution of $\tilde{\psi}$,

$$\dot{\tilde{\psi}} \;=\; (A \,+\, F(\bar{x}))\,\tilde{\psi} \,+\, F(\tilde{x})\,\bar{\psi}. \tag{7.17b}$$

Both (7.17a) and (7.17b) are stable LTI systems. Thus, for the constant disturbance $d(t) = \bar{d}$, the solution $\bar{\psi}$ to (7.17a) is asymptotically constant. Since $\tilde{x}$ is zero-mean, the signal $F(\tilde{x})\,\bar{\psi}$ is also zero-mean, i.e., it has no component that contains the zero temporal frequency. Furthermore, since (7.17b) is a stable LTI system, the frequency components in $\tilde{\psi}$ must correspond to frequency components in the input, $F(\tilde{x})\,\bar{\psi}$. Thus, $\tilde{\psi}$ has no frequency component corresponding to the zero temporal frequency and therefore $\tilde{\psi}$ is also zero-mean.

Because $\bar{\psi}$ is asymptotically constant and $\tilde{\psi}$ is zero-mean, $\left\langle Q^{1/2}\bar{\psi},\, Q^{1/2}\tilde{\psi}\right\rangle = 0$. Similarly, $\left\langle R^{1/2}\bar{x},\, R^{1/2}\tilde{x}\right\rangle = 0$. Furthermore, $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2 = \left\langle Q^{1/2}\psi,\, Q^{1/2}\psi\right\rangle + \left\langle R^{1/2}x,\, R^{1/2}x\right\rangle$, and we have,

$$\begin{aligned} \|z_{(\bar{x}+\varepsilon\tilde{x}),\bar{d}}\|_{\text{pow}}^2 \;-\; \|z_{\bar{x},\bar{d}}\|_{\text{pow}}^2 \;&=\; 2\,\varepsilon\left(\left\langle Q^{1/2}\bar{\psi},\, Q^{1/2}\tilde{\psi}\right\rangle \,+\, \left\langle R^{1/2}\bar{x},\, R^{1/2}\tilde{x}\right\rangle\right) \,+\, \mathcal{O}(\varepsilon^2) \\ &=\; \mathcal{O}(\varepsilon^2). \end{aligned} \tag{7.18}$$

Thus, the first order correction to $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ evaluated at a constant control input $\bar{x}$ is zero, which completes the proof. □

Remark 7.6.3. *Lemma 7.6.2 does not need Assumption 3 and it holds for all bilinear systems of the form* (7.12a) *for which there is a stabilizing constant control input.*

Lemma 7.6.2 implies that a constant control signal $x^\star$

$$x^\star \;\in\; \underset{\bar{x}\ \text{constant}}{\operatorname{argmin}} \;\|\zeta_{\bar{x},\bar{d}}\|_{\text{pow}}^2,$$

is a stationary point of $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ under Assumption 3. Since $\|\zeta_{\bar{x},\bar{d}}\|_{\text{pow}}^2$ is a convex function of x by Lemma 7.6.1, $x^\star$ is a local and therefore global minimizer. The following theorem relates the power norm of the output of system (7.12) subject to a constant disturbance, $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$, with the worst case power norm amplification $f(x)$ and shows that the constant control signal $x^\star$ solves Problem 2.

Theorem 7.6.4. *Let Assumption 3 hold. Then, a constant control input $x(t) = x^\star$ solves Problem 2.*

Proof. Let $x^\star$ minimize $f(x)$ over the space of *constant* functions. Since (7.12a) with a constant control input is an LTI system, the maximum power-amplification coincides with the $\mathcal{H}_\infty$ norm. Moreover, because (7.12a) is a positive system, the maximal singular value of the frequency response matrix peaks at zero temporal frequency and the worst-case disturbance is a constant signal [54]. This implies that $f(x^\star) = \|\zeta_{x^\star,\bar{d}}\|_{\text{pow}}^2$ where $\bar{d} = v \geq 0$ is a constant nonnegative disturbance and v is the right principal singular vector of the matrix $-Q^{\frac{1}{2}}(A + F(x^\star))^{-1}B$.

Suppose there exists a time varying signal $\hat{x} \in \mathcal{B}_2$ such that $f(\hat{x}) < f(x^\star)$. Then, since f measures the worst-case disturbance amplification, $\hat{x}$ must also decrease the power norm of the output of system (7.12) for a constant disturbance $\bar{d} = v$, i.e.,

$$\|\zeta_{\hat{x},\bar{d}}\|_{\text{pow}}^2 \;\leq\; f(\hat{x}) \;<\; \|\zeta_{x^\star,\bar{d}}\|_{\text{pow}}^2 \;=\; f(x^\star). \tag{7.19}$$

We next show that this is not possible.

Every bounded $x \in \mathcal{B}_2$ can be written as $x = \bar{x} + \tilde{x}$ where $\bar{x}$ is constant and $\tilde{x}$ is bounded and zero-mean. Since $x^\star$ minimizes $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ over constant control inputs by definition, Lemma 7.6.2 implies that $x^\star$ is a stationary point of $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ over *all* bounded $x \in \mathcal{B}_2$. Convexity of $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$ in x by Lemma 7.6.1 thus implies that $x^\star$ is a local and therefore global minimizer of $\|\zeta_{x,\bar{d}}\|_{\text{pow}}^2$, contradicting (7.19) and completing the proof. □

Chapter 8

Actuator or sensor selection

In traditional applications, controller or observer design deals with the problem of how to use a pre-specified configuration of sensors and actuators in order to attain the desired objective. In general, the best performance is achieved by using *all* of the available sensors or actuators. However, this option may be computationally or economically infeasible. We thus consider the problem of selecting a subset of available sensors or actuators in order to gracefully degrade performance relative to the setup where all of them are used.

Typically, sensor/actuator selection and placement is performed by a designer with expert knowledge of the system. However, in large-scale applications and systems with complex interactions, it can be difficult to do this effectively. For linear time-invariant dynamical systems, we develop a framework and an efficient algorithm to systematically choose sensors and actuators via convex optimization.

Our starting point is a formulation with an abundance of potential sensors or actuators. This setup can encode information about different types or different placements of sensing and actuating capabilities. We consider the problem of selecting subsets of available options from this full model. Applications of this formulation range from placement of Phasor Measurement Units (PMUs) in power systems, to placement of sensors and actuators along an aircraft wing, to the distribution of GPS units in a formation of multi-vehicle systems.

The problem of interest is a difficult combinatorial optimization problem. Although there is a wide body of previous work in this area, most of the available literature either uses heuristic methods or does not consider dynamical models. In [128], the authors provide a convex sensor selection formulation for a problem with linear measurements. The authors of [229] select sparse subsets of sensors to minimize the Cramer-Rao bound of a class of nonlinear measurement

models. The placement of PMUs in power systems was formulated as a variation of the optimal experiment design in [230]. Actuator selection via genetic algorithms was explored in [231]. A non-convex formulation of the joint sensor and actuator placement was provided in [232,233] and it was recently applied to the linearized Ginzburg-Landau equation [234]. The leader selection problem in consensus networks can be seen as a type of a structured joint actuator and sensor selection problem which admits a convex relaxation [80] and even an analytical solution for one or two leaders [78]. However, this formulation does not extend naturally to broader classes of problems.

The sparsity-promoting framework introduced in [62–64] can be used to obtain block-sparse structured feedback and observer gains and select actuators or sensors. Indeed, algorithms developed in [64] have been used in [235] for the sensor selection in a target tracking problem. However, these algorithms have been developed for general structured control/estimation problems and they do not exploit the hidden convexity of the actuator/sensor selection problem.

In [104], the authors introduced a convex semidefinite programming (SDP) characterization of the problem formulation considered in [64] for enhancing certain forms of sparsity in the feedback gain. Although sensor and actuator selection falls into the class of problems considered by [104], generic SDP solvers cannot handle large-scale applications. Since we are interested in high-dimensional systems with many sensors/actuators, we use the alternating direction method of multipliers (ADMM) [132] and proximal gradient [131] methods to develop a customized solver that is well-suited for large problems. In contrast to standard SDP solvers, whose computational complexity scales unfavorably with the number of states/sensors/actuators, the worst case per-iteration complexity of our method scales only with the number of states. Furthermore, our algorithm performs much better than standard SDP solvers in numerical experiments.

8.1 Problem formulation

8.1.1 Actuator selection

Consider the standard state-space system first presented in (1.2),

$$\begin{aligned} \dot{\psi} &= A\,\psi \;+\; B_1\,d \;+\; B_2\,u \\ \zeta &= \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix} \psi \;+\; \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} u \end{aligned}$$

where d is a zero-mean white stochastic process with covariance V_d and the pair (A, B_2) is controllable. The optimal $\mathcal{H}_2$ controller minimizes the steady-state variance,

$$\lim_{t \to \infty} \mathbb{E}\left(\psi^T(t)\, Q\, \psi(t) \;+\; u^T(t)\, R\, u(t)\right)$$

where $Q = Q^T \succeq 0$ specifies a weight on the system states, and $R = R^T \succ 0$ specifies the penalty on the control input. The global optimal controller for this problem is a state feedback law of the form $u = -X\psi$. Although this controller is readily computed by solving the corresponding algebraic Riccati equation, it typically uses all input channels and thus all available actuators.

We are interested in designing an optimal controller which uses a subset of the available actuators. This will be accomplished by augmenting the $\mathcal{H}_2$ performance index with a term that promotes row-sparsity of the feedback gain matrix X. The resulting problem can be cast as a semidefinite program and thus solved efficiently for small problems.

SDP Formulation

Under the state feedback control law, $u = -X\psi$, the closed-loop system is given by

$$\begin{aligned} \dot{\psi} &= (A - B_2 X)\,\psi \;+\; B_1\, d \\ z &= \begin{bmatrix} Q^{1/2} \\ -R^{1/2} X \end{bmatrix} \psi. \end{aligned} \tag{8.1}$$

The $\mathcal{H}_2$ norm of system (8.1) is determined by

$$f_2(X) \;=\; \operatorname{trace}\left(Q\, P \;+\; X^T R\, X\, P\right)$$

where $P = P^T \succ 0$ is the controllability gramian of the closed loop system,

$$(A - B_2 X)\, P \;+\; P\,(A - B_2 X)^T \;+\; B_1\, V_d\, B_1^T \;=\; 0.$$

Since P is positive definite and therefore invertible, the standard change of coordinates $Z := XP$ can be used to express $f(X)$ in terms of P and Z [49],

$$f_2(P, Z) \;=\; \operatorname{trace}\left(Q\, P \;+\; P^{-1}\, Z^T R\, Z\right)$$

and to bring the optimal $\mathcal{H}_2$ problem into the following form

$$\begin{aligned} \underset{P,\,Z}{\text{minimize}} \quad & f_2(P,Z) \\ \text{subject to} \quad & AP + PA^T - B_2Z - Z^TB_2^T + V = 0 \\ & P \succ 0 \end{aligned} \tag{8.2}$$

with $V := B_1V_dB_1^T$. By taking the Schur complement of $P^{-1}Z^TRZ$ this problem can be expressed as an SDP [49]. Finally, the optimal feedback gain can be recovered by $X = ZP^{-1}$. In what follows, we use this SDP characterization to introduce the actuator selection problem.

Sparsity Structure

When the ith row of X is identically equal to zero, the ith control input is not used. Therefore, obtaining a control law which uses only a subset of available actuators can be achieved by promoting row-sparsity of X. Our developments are facilitated by the equivalence between the row-sparsity of X and Z; the ith row of Z is equal to zero if and only if the ith row of $Z = XP$ is equal to zero [104].

Drawing on the group-sparsity paradigm [122], we augment (8.2) with a sparsity-promoting penalty on the rows of Z,

$$\begin{aligned} \text{minimize} \quad & f_2(P,Z) \;+\; \gamma\, g_2(Z) \\ \text{subject to} \quad & AP + PA^T - B_2Z - Z^TB_2^T + V = 0 \\ & P \succ 0. \end{aligned} \tag{8.3}$$

where $\gamma > 0$ specifies the importance of sparsity relative to $\mathcal{H}_2$ performance and

$$g_2(Z) \;:=\; \sum_{i\,=\,1}^{m} w_i \,\|\mathrm{e}_i^T Z\|_2$$

is a row-sparsity-promoting penalty function in which w_i are nonzero weights and e_i is the ith unit vector.

This problem can still be cast as an SDP and standard solvers can be used to find its solution. However, since generic SDP solvers do not exploit the structure of (8.3), they do not scale gracefully with problem dimension.

8.1.2 Sensor selection

The sensor selection problem can be approached in a similar manner. Consider a linear time-invariant system,

$$\begin{aligned} \dot{\psi} &= A_s\,\psi + B_1\,d \\ y &= C\,\psi + \eta \end{aligned}$$

where d and η are zero-mean white stochastic processes with covariances V_d and V_η, respectively, and (A_s, C) is an observable pair. The observer,

$$\begin{aligned} \dot{\hat{\psi}} &= A_s\,\hat{\psi} + L\,(y - \hat{y}) \\ &= A_s\,\hat{\psi} + L\,C\left(\psi - \hat{\psi}\right) + L\,\eta \end{aligned}$$

estimates the state x from the noisy measurements y using a linear injection term with an observer gain L. For the Hurwitz matrix $A_s - LC$, the zero-mean estimate of ψ is given by $\hat{\psi}$ and the dynamics of the estimation error $\tilde{\psi} := \psi - \hat{\psi}$ are governed by

$$\dot{\tilde{\psi}} = (A_s - L\,C)\,\tilde{\psi} + B_1\,d - L\,\eta. \tag{8.4}$$

The variance amplification from the noise sources d and η to the estimation error $\tilde{\psi}$ is determined by

$$f_o(L) = \text{trace}\,(P_o\,B_1\,V_d\,B_1^T + P_o\,L\,V_\eta\,L^T) \tag{8.5}$$

where P_o is the observability gramian of the error system (8.4),

$$(A_s - L\,C)^T\,P_o + P_o\,(A_s - L\,C) + I = 0.$$

The Kalman filter gain L, resulting from the observer Riccati equation, provides an optimal observer with the smallest variance amplification.

Our objective is to design a Kalman filter which uses a subset of the available sensors. This can be achieved by enhancing column-sparsity of the observer gain L. Since the change of coordinates $Z_s := P_o L$ preserves column sparsity of L, we formulate the sensor selection problem

as

$$\begin{array}{ll} \underset{P_o,\, Z_s}{\text{minimize}} & f_o(P_o, Z_s) \;+\; \gamma\, g_2(Z_s^T) \\ \text{subject to} & A_s^T P_o + P_o A_s - C^T Z_s^T - Z_s C + I = 0 \\ & P_o \succ 0 \end{array} \tag{8.6}$$

where

$$f_o(P_o, Z_s) \;=\; \text{trace}\left(P_o\, B_1\, V_d\, B_1^T \;+\; P_o^{-1}\, Z_s\, V_\eta\, Z_s^T\right).$$

We note that the sensor selection problem (8.6) can be obtained from the actuator selection problem (8.3) by setting the problem data in (8.3) to

$$\begin{array}{lllllll} A &=& A_s^T, & B_2 &=& C^T, & Q \;=\; B_1\, V_d\, B_1^T \\ V &=& I, & R &=& V_\eta \end{array} \tag{8.7}$$

and recovering the variables $P_o = P$ and $Z_s = Z^T$.

8.2 Customized algorithm

We next develop an efficient algorithm for solving the actuator selection problem (8.3); the solution to the sensor selection problem (8.6) can be obtained by mapping it to (8.3) via (8.7). The challenges in solving the optimization problem (8.3) arise from

- the positive definite constraint;
- the linear Lyapunov-like constraint;
- the non-smoothness of the sparsity-promoting term.

To ensure positive definiteness of P, we use projected descent techniques to optimize over the positive definite cone. We dualize the linear constraint and split (8.3) into two simpler subproblems over P and Z via the alternating direction method of multipliers (ADMM). This splitting separates the objective function component associated with the positive definite constraint (the subproblem over P) from the non-differentiable component (the subproblem over Z). Since the subproblem over Z is efficiently solvable, our method scales well with the number of sensors or actuators.

We employ a projected version of Newton's method to solve the P-subproblem. The Newton search direction is obtained via a conjugate gradient algorithm. The Z-subproblem is solved

with a proximal method.

8.2.1 Alternating direction method of multipliers

To employ ADMM, we first form the augmented Lagrangian corresponding to the optimization problem (8.3),

$$\mathcal{L}_\mu(P, Z; Y) \;:=\; f_2(P, Z) \;+\; \gamma\, g_2(Z) \;+\; \langle Y,\, h(P, Z)\rangle \;+\; \tfrac{1}{2\mu}\|h(P, Z)\|_F^2$$

where $h(X, Y) = 0$ is the linear Lyapunov-like constraint and

$$h(P, Z) \;:=\; AP \;+\; PA^T \;-\; B_2 Z \;-\; Z^T B_2^T \;+\; V.$$

For this problem, the ADMM iteration discussed in Section 2.2.3 [132] is,

$$\begin{aligned} P^{k+1} &= \operatorname*{argmin}_P \;\; \mathcal{L}_\mu(P,\, Z^k,\, Y^k) \\ Z^{k+1} &= \operatorname*{argmin}_Z \;\; \mathcal{L}_\mu(P^{k+1},\, Z,\, Y^k) \\ Y^{k+1} &= Y^k \;+\; \tfrac{1}{\mu}\, h(P^{k+1},\, Z^{k+1}). \end{aligned}$$

The stopping criteria depend on the primal residual, which quantifies how well P^k and Z^k satisfy the linear constraint, and the dual residual, which quantifies the difference between Z^k and Z^{k-1}. We refer the reader to [132] for details.

We note that the proximal augmented Lagrangian reformulation of $\mathcal{L}_\mu$ described in Section 2.3 is not possible here because the linear constraint involves nondiagonal operators which act on both P and Z.

8.2.2 P-minimization step

Minimization of $\mathcal{L}_\mu$ over P is step (2.6a) in ADMM and can be equivalently expressed as

$$\begin{aligned} \operatorname*{minimize}_P \quad & f(P, Z^k) \;+\; \tfrac{1}{2\mu}\,\|h(P, Z^k) \;+\; \mu\, Y^k\|_F^2 \\ \text{subject to} \quad & P \;\succ\; 0. \end{aligned} \tag{8.8}$$

We use a projected version of Newton's method to solve a sequence of quadratic approximations to (8.8). In what follows, $\{P^l\}$ denotes the sequence of inner iterations that converges to the solution of (8.8).

Newton's method

At each inner iteration P^l, Newton's method does a line search from P^l in the direction $\tilde{P}^l$ which minimizes the quadratic approximation of (8.8),

$$\tilde{P}^l \;:=\; \underset{\tilde{P}}{\operatorname{argmin}}\; \tfrac{1}{2}\Big\langle H_{P^l}(\tilde{P}),\, \tilde{P}\Big\rangle \;+\; \Big\langle \nabla_P \mathcal{L}_\mu(P^l, Z^k, Y^k),\, \tilde{P}\Big\rangle \tag{8.9}$$

where $H_{P^l}(\tilde{P})$ is a linear function of $\tilde{P}$ that contains information about the Hessian of $\mathcal{L}_\mu$. The gradient of $\mathcal{L}_\mu$ with respect to P is given by

$$\nabla_P \mathcal{L}_\mu(P^l, Z^k, Y^k) \;:=\; Q \;-\; P^{-1} Z^T R\, Z P^{-1} + A^T\big(Y \;+\; \tfrac{1}{\mu}\, h(P,Z)\big) \;+\; \big(Y \;+\; \tfrac{1}{\mu}\, h(P,Z)\big) A.$$

The quadratic term used in (8.9) is,

$$\tfrac{1}{2}\,\mathrm{vec}(\tilde{P})^T (\nabla_P^2 \mathcal{L}_\mu)\,\mathrm{vec}(\tilde{P})$$

where $\nabla_P^2 \mathcal{L}_\mu$ is the Hessian of the objective function in (8.8). It can be more conveniently expressed as $\frac{1}{2}\Big\langle H(\tilde{P}),\, \tilde{P}\Big\rangle$ where H_P is the linear operator

$$H_P(\tilde{P}) \;:=\; H_{1,P}(\tilde{P}) \;+\; \tfrac{1}{\mu}\, H_2(\tilde{P}). \tag{8.10}$$

The first term in (8.10) comes from the performance index,

$$H_{1,P}(\tilde{P}) \;\;:=\;\; P^{-1} Z^T R\, Z P^{-1}\, \tilde{P}\, P^{-1} \;+\; P^{-1}\, \tilde{P}\, P^{-1} Z^T R\, Z P^{-1}$$

and the second term comes from the constraint penalties,

$$H_2(\tilde{P}) \;:=\; A^T A\, \tilde{P} \;+\; \tilde{P}\, A^T A \;+\; A\, \tilde{P}\, A \;+\; A^T \tilde{P}\, A^T.$$

Solving the linear equation

$$H_{P^l}(\tilde{P}) \;+\; \nabla_P \mathcal{L}_\mu(P^l, Z^k, Y^k) \;=\; 0$$

yields the Newton direction $\tilde{P}^l$ which is computed using the conjugate gradient method [135].

Projection

The set $\mathcal{C} := \{P \,|\, P \succeq \varepsilon I\}$ approximates the positive definite cone $P \succ 0$ for small $\varepsilon > 0$. Once the Newton direction is determined, the step

$$P \;=\; \mathcal{P}_{\mathcal{C}}\left(P^l \;+\; \alpha\,\tilde{P}^l\right)$$

is taken, where $\mathcal{P}_{\mathcal{C}}$ is the projection on the set $\mathcal{C}$ and the step size α is chosen using an Armijo backtracking search.

To project the symmetric matrix M onto $\mathcal{C}$, its eigenvalues are projected onto the set $\lambda_i \geq \epsilon$. From the eigenvalue decomposition $M = U\,\mathrm{diag}(\lambda)\,U^T$, where λ is a vector of the eigenvalues and U is a matrix of the corresponding eigenvectors, the projection is $\mathcal{P}_{\mathcal{C}}(M) = U\;\mathrm{diag}\left(\max\left(\lambda, \epsilon\right)\right)U^T$.

While Netwon's method reduces the number of required steps, computing the search directions can be prohibitively expensive for large-scale systems. In Section 8.4, we will explore a proximal gradient algorithm for solving problem (8.3).

8.2.3 Z-minimization step

The Z-minimization step is equivalent to solving,

$$\underset{Z}{\text{minimize}} \quad \gamma\, g_2(Z) \;+\; \mathrm{trace}\left((P^{k+1})^{-1}Z^T R Z\right) \;+\; \tfrac{1}{2\mu}\|h(P^{k+1}, Z) \;+\; \mu\, Y^k\|_F^2. \tag{8.11}$$

The objective function is the sum of a quadratic term and a separable sum of ℓ_2 norms: a problem form commonly referred to as group LASSO. We employ the proximal gradient method [131] to solve this problem. The gradient of the smooth part of (8.11) evaluated at $\bar{Z}$ is

$$2RZP^{-1} + \tfrac{2}{\mu}\left(B_2^T Z^T B_2^T + B_2^T B_2 Z\right) \;-\; 2B_2^T\left(Y + \tfrac{1}{\mu}(AP + PA^T + V)\right).$$

The proximal operator for g_2 is the block shrinkage operator,

$$\mathcal{S}_\alpha(\mathrm{e}_i^T Z) \;=\; \begin{cases} \left(1 \;-\; \alpha/\|\mathrm{e}_i^T Z\|_2\right)\mathrm{e}_i^T Z, & \|\mathrm{e}_i^T Z\|_2 \;>\; \alpha \\ 0, & \|\mathrm{e}_i^T Z\|_2 \;\leq\; \alpha. \end{cases}$$

8.2.4 Iterative reweighting

Inspired by [121], we employ an iterative reweighing scheme to select the weights w_i in the sparsity-promoting term $\sum_i w_i \|a_i\|_2$ to obtain sparser structures at lower values of γ. The authors in [121] noted that if $w_i = 1/\|a_i\|_2$, then there is an exact correspondence between the weighted ℓ_1 norm and the cardinality function. However, implementing such weights requires *a priori* knowledge of the values $\|a_i\|_2$ at the optimal a. Consequently, we implement a reweighting scheme in which we run the algorithm multiple times for each value of γ and update the weights as,

$$w_i^{j+1} \;=\; \frac{1}{\| \, \mathrm{e}_i^T Z \, \|_2 \;+\; \epsilon} \tag{8.12}$$

where $\epsilon > 0$ ensures that the update is always well-defined.

8.3 Example: Mass-spring system

We use a simple mass-spring system to illustrate the utility of our algorithm in the sensor selection problem. This system has a clear intuitive interpretation and its dimension can be easily scaled while retaining the problem structure. Consider a series of masses connected by linear springs. With unit masses, unit spring constants and no friction, the dynamics of each mass are described by

$$\ddot{p}_i \;=\; -(p_i \;-\; p_{i+1}) \;-\; (p_i \;-\; p_{i-1}) \;+\; d_i$$

where p_i is the position of the ith mass. When the first and last masses are affixed to rigid bodies, the aggregate dynamics are given by

$$\begin{bmatrix} \dot{p} \\ \dot{v} \end{bmatrix} \;=\; \begin{bmatrix} 0 & I \\ -T & 0 \end{bmatrix} \begin{bmatrix} p \\ v \end{bmatrix} \;+\; \begin{bmatrix} 0 \\ I \end{bmatrix} d$$

where p, v, and d are the position, velocity and disturbance vectors, and T is a Toeplitz matrix with 2 on the main diagonal and -1 on the first super- and sub-diagonals. The possible sensor outputs are the position and velocity vectors.

8.3.1 Algorithm speed and computational complexity

The complexity of solving the sensor selection SDP with interior point methods is $O((n+r)^6)$, where n is the dimension of the state-space and r is the number of sensors. In our algorithm,

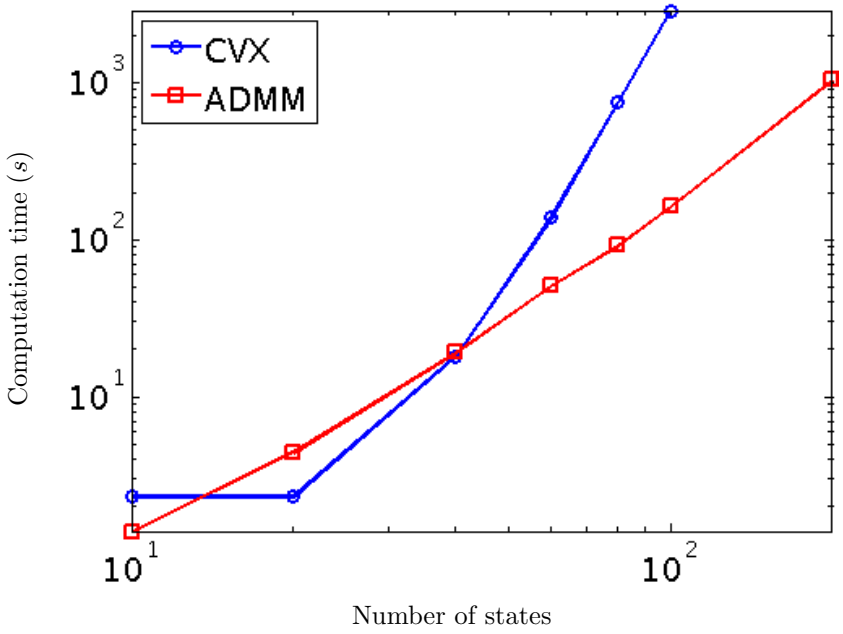


Figure 8.1: Scaling of computation time with the number of states for CVX and for ADMM for mass-spring system with $\gamma = 100$ and position and velocity outputs. Empirically, we observe that CVX scales roughly with n^6 while ADMM scales roughly with n^3.

the greatest cost is incurred by computing the Newton direction in the P-minimization step. Since $P \in \mathbb{R}^{n\times n}$, the worst case complexity of computing Newton direction is $O(n^6)$. This is because each conjugate gradient step takes $O(n^3)$ operations and, in general, n^2 conjugate gradient steps are required to obtain convergence. In well-conditioned problems, the conjugate gradient method achieves high accuracy much faster which significantly reduces computational complexity. The Z-minimization step has a computational cost of only $O(nr)$, so the overall cost per ADMM iteration is $O(n^6)$ unless $r \geq n^5$.

Figure 8.1 shows the time required by ADMM and by CVX [203, 236] to solve (8.6) with $\gamma = 100$ for mass-spring systems of increasing sizes. Our algorithm scales favorably here and it scales much better when only the number of sensors is varied. Figure 8.2 shows scaling with *just* the number of outputs. As the number of sensors increases, CVX's computation time increases while the computation time of ADMM barely changes.

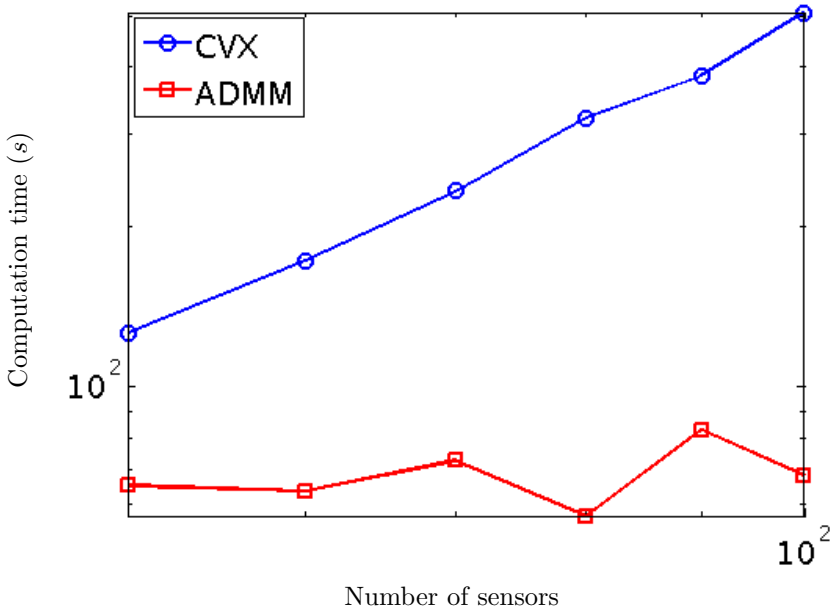


Figure 8.2: Scaling of computation time with the number of sensors for CVX and for ADMM for mass-spring system with $\gamma = 100$ and $n = 50$. Outputs are random linear combinations of the states.

8.3.2 Sensor selection

We consider a system with 20 masses (40 states) and potential position and velocity measurements for each mass. As γ increases, sparser observer structures are uncovered at the cost of compromising quality of estimation. The tradeoff between the number of sensors and the variance of the estimation error is shown in Figure 8.3.

Figures 8.4 and 8.5 show the position and velocity sensor topologies identified by the ADMM algorithm as γ is increased. To save space, only novel topologies are shown. The selected sensor configurations have symmetric topology, which is expected for this example. Notably, velocity measurements are, in general, more important but the locations of the most important position and velocity sensors differ.

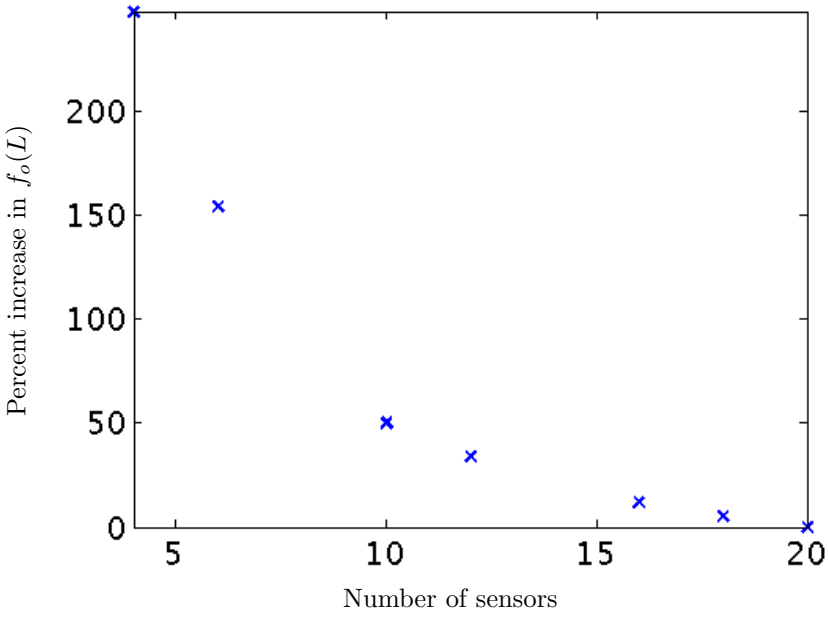


Figure 8.3: Percent increase in $f_o(L)$ in terms of the number of sensors.

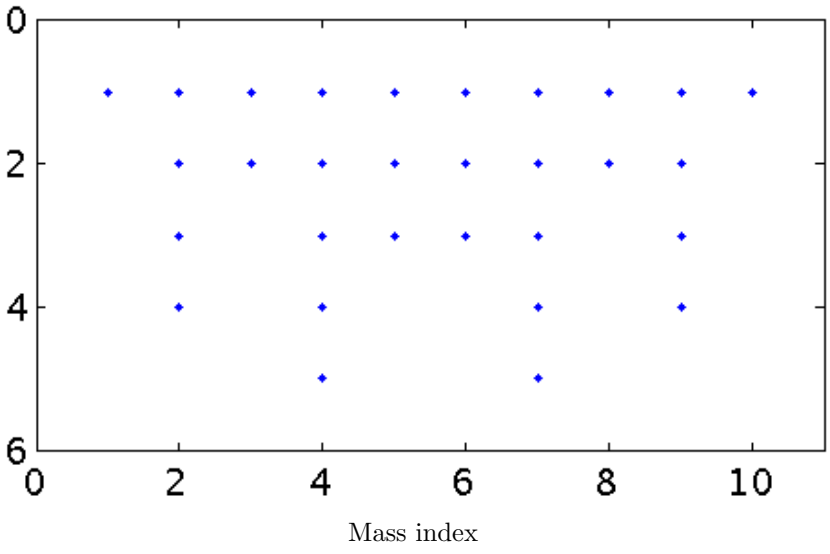


Figure 8.4: Retained position sensors as γ increases. A blue dot indicates that the position of the corresponding mass is being measured. The top row shows the densest sensor topology, and the bottom row shows the sparsest.

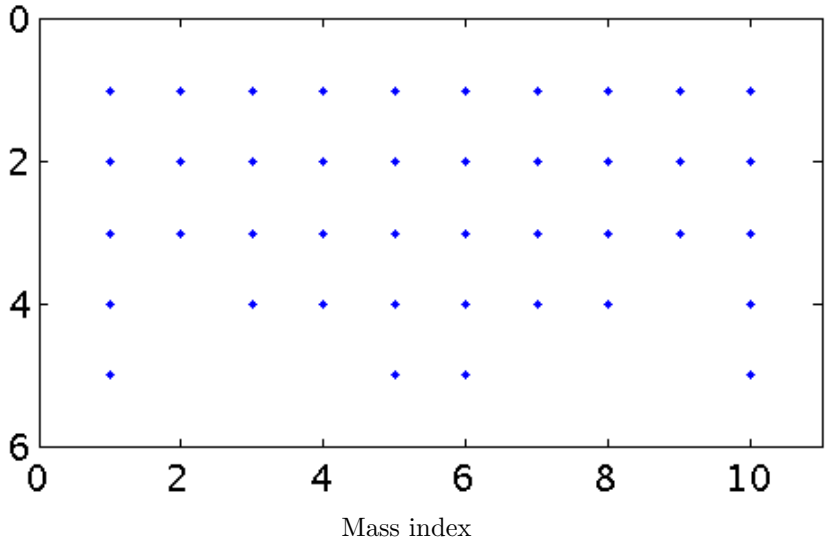


Figure 8.5: Retained velocity sensors as γ increases. A blue dot indicates that the velocity of the corresponding mass is being measured. The top row shows the densest sensor topology, and the bottom row shows the sparsest.

8.3.3 Iterative reweighting

Figure 8.6 illustrates the utility of iterative reweighting. When constant sparsity-promoting weights are used, large values of γ are required to identify sparse structures. Here, we run our ADMM algorithm 3 times for each value of γ, updating the weights using (8.12) and retaining them as we increase γ.

8.4 Proximal gradient method

When the system is open-loop stable, it is possible to express the objective function solely in terms of Z and to solve the regularized problem using proximal gradient descent. For simple problems, such as the mass-spring system, this shows good performance relative to ADMM.

8.4.1 Elimination of P

The linear constraint can be abstractly written as,

$$\mathcal{A}(P) \; - \; \mathcal{B}(Z) \; + \; V \; = \; 0.$$

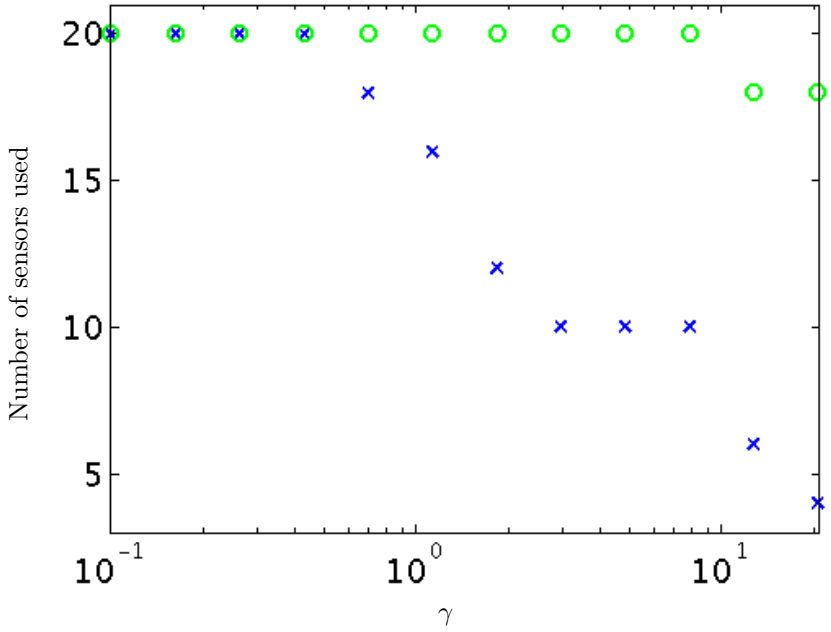


Figure 8.6: Number of sensors versus γ for a scheme which uses iterative reweighting and for the scheme which uses constant weights. Iterative reweighting promotes sparser structures earlier.

where

$$\mathcal{A}(P) \;:=\; AP \;+\; PA^T$$

$$\mathcal{B}(Z) \;:=\; B_2 Z \;+\; Z^T B_2^T .$$

We assume that the linear map $\mathcal{A}$ is invertible, which occurs if and only if the matrix A has no eigenvalues with real part 0 and no eigenvalues of A have opposite real part; i.e, no eigenvalue of A has a real part which is the negative of the real part of any other eigenvalue of A. We can then write,

$$P \;=\; \mathcal{A}^{-1}\left(\mathcal{B}(Z) \;-\; V\right).$$

8.4.2 Gradient

For notational convenience, we partition the smooth part of the objective function into two components,

$$f_a(P) \;:=\; \langle Q,\, P\rangle$$

$$f_b(P,Z) \;:=\; \left\langle P^{-1},\, Z^T R Z\right\rangle$$

Gradient of f_a

Using the explicit expression for P and the properties of linear operators, the first component of the objective function can be rewritten in terms of Z,

$$\langle Q,\, P\rangle \;=\; \left\langle Q,\, \mathcal{A}^{-1}\left(\mathcal{B}(Z) \;-\; V\right)\right\rangle \;=\; -\left\langle V,\, \mathcal{A}^{-\dagger}(Q)\right\rangle \;+\; \left\langle Z,\, \mathcal{B}^{\dagger}(\mathcal{A}^{-\dagger}(Q))\right\rangle$$

where $\dagger$ indicates the adjoint of a linear operator, and

$$\begin{aligned}\mathcal{A}^{\dagger}(P) &= A^T P \;+\; PA\\ \mathcal{B}^{\dagger}(P) &= 2B_2^T P.\end{aligned}$$

Therefore,

$$\nabla_Z f_a \;=\; -\,2B2^T Q^{ad}$$

where

$$A^T\, Q^{ad} \;+\; Q^{ad}\, A \;+\; Q \;=\; 0.$$

Gradient of f_b

The gradient of f_b can be computed via perturbation analysis. Note that,

$$(\bar{P} \;+\; \varepsilon\,\tilde{P})^{-1} \;=\; \bar{P}^{-1} \;-\; \varepsilon\bar{P}^{-1}\,\tilde{P}\,\bar{P}^{-1} \;+\; O(\varepsilon^2).$$

and that since P is specified by a linear constraint,

$$\mathcal{A}(\bar{P} \;+\; \varepsilon\,\tilde{P}) \;-\; \mathcal{B}(\bar{Z} \;+\; \varepsilon\,\tilde{Z}) \;+\; V \;=\; 0,$$

the variation of P as a result of a variation in Z is given by

$$\tilde{P} \;=\; \mathcal{A}^{-1}(\mathcal{B}(\tilde{Z})).$$

Perturbing f_b,

$$\left\langle (P \;+\; \varepsilon\,\tilde{P})^{-1},\, (Z \;+\; \varepsilon\,\tilde{Z})^T\, R\,(Z \;+\; \varepsilon\,\tilde{Z})\right\rangle$$

and collecting the order ε terms, $2\left\langle \tilde{Z},\, R\bar{Z}\bar{P}^{-1}\right\rangle - \left\langle \tilde{Z},\, \mathcal{B}^{\dagger}(\mathcal{A}^{-\dagger}(\bar{P}^{-1}\bar{Z}^T R\bar{Z}\bar{P}^{-1}))\right\rangle$. From this, it is clear that

$$\nabla_Z f_b = 2RZP^{-1} - \mathcal{B}^{\dagger}(\mathcal{A}^{-\dagger}(P^{-1}Z^T RZP^{-1}))$$

and so $\nabla f(Z) \;=\; 2B_2^T(R^{ad} \;-\; Q^{ad}) \;+\; 2RZP^{-1}$ where

$$\begin{aligned} A^T\,Q^{ad} \;+\; Q^{ad}\,A \;+\; Q \;&=\; 0 \\ A^T\,R^{ad} \;+\; R^{ad}\,A \;+\; P^{-1}\,Z^T\,R\,Z\,P^{-1} \;&=\; 0 \\ A\,P \;+\; P\,A^T \;-\; B_2\,Z \;-\; Z^T\,B_2^T \;+\; V \;&=\; 0. \end{aligned}$$

BB stepsize initialization

We implement proximal gradient descent and initialize the step-size using the Barzilai-Borwein (BB) method, which attempts to approximate the Hessian with a scaled version of the identity matrix. The secant equation,

$$\nabla f(Y^k) \;\approx\; \nabla f(Y^{k-1}) \;+\; H(Y^k, Y^k - Y^{k-1})$$

predicts the change in the gradient $\nabla f(Y)$ as a result of a change in Y. Quasi-Newton methods use the above in conjunction with observed changes in the gradient to build an estimate $\tilde{H}$ of the Hessian and determine a search direction by solving the linear equation,

$$\tilde{H}(Y^k, \widetilde{Y}) \;=\; -\;\nabla f(Y^k). \tag{8.13}$$

BB step size selection restricts the Hessian approximation to be a scaled version of the identity,

$$\widetilde{H}(Z,\,\tilde{Z}) \;:=\; \tfrac{1}{\alpha}\,\tilde{Z}$$

with $\alpha > 0$ and minimizes the residual of the secant equation,

$$\|\nabla f(Z^k) \;-\; \nabla f(Z^{k-1}) \;-\; \tfrac{1}{\alpha}\,(Z^k \;-\; Z^{k-1})\|_F^2$$

over α to obtain,

$$\alpha_{m,0} \;=\; \frac{\|Z^k \;-\; Z^{k-1}\|_F^2}{\langle Z^{k-1} \;-\; Z^k,\, \nabla f(Z^{k-1}) \;-\; \nabla f(Z^k)\rangle} \tag{8.14}$$

We backtrack from $\alpha_{m,0}$ by selecting the smallest nonnegative integer r such that $\alpha_m = c^r\alpha_{m,0}$ with $c \in (0,1)$ such that Z^{k+1} is stabilizing (i.e., yields a positive definite $P^{k+1} = \mathcal{A}^{-1}(\mathcal{B}(Z^{k+1} - V))$) and so that Y^{Z+1} satisfies either an ISTA-like or a SpaRSA-like step size selection rule.

ISTA-like step size selection rule

The step size α is chosen so that the approximation of the objective function is an overestimate of the actual objective function

$$f(Z^{k+1}) \;+\; g(Z^{k+1}) \;\leq\; f(Z^k) \;+\; \left\langle \widetilde{Y},\, \nabla f(Z^k) \right\rangle \;+\; \tfrac{1}{2\alpha}\,\|\tilde{Z}\|_F^2 \;+\; g(Z^{k+1})$$

where $\tilde{Z} := Z^{k+1} - Z^k$ [131].

SpaRSA-like step size selection rule

In this approach [187], the objective function is only checked every p iterations. At that iteration, backtracking is used to ensure that,

$$f(Z^{k+1}) \;<\; \max_{m\,=\,k\,-\,p,\ldots,k} \left(f(Z^m)\right).$$

Otherwise, backtracking is just used to ensure stability.

8.4.3 Example: Damped mass-spring systems

Since this method requires open-loop stability, we must introduce some damping into the mass-spring example explored earlier. For the damped mass-spring model with N masses,

$$\begin{bmatrix} \dot{p} \\ \dot{v} \end{bmatrix} \;=\; \begin{bmatrix} 0 & I \\ T \;-\; bI & -bI \end{bmatrix} \begin{bmatrix} p \\ v \end{bmatrix} \;+\; u \;+\; d$$

where T is a Toeplitz matrix with -2 on the main diagonal and 1 on the first super- and sub-diagonals. The parameter b represents the strength of damping. Experiments were performed with $\gamma = 30$, $Q = I$ and $R = 10I$. Figure 8.7 shows the computation time for proximal gradient with both ISTA (PG) and SpaRSA (PGs) step-size selection, CVX, ADMM, and an MM algorithm with the proximal augmented Lagrangian from Chapter 2.

8.5 Example: Flexible wing aircraft

Finally, we present an example of the sensor selection algorithm applied to a practical example from aerospace engineering. One barrier in reducing aircraft weight in order to improve fuel

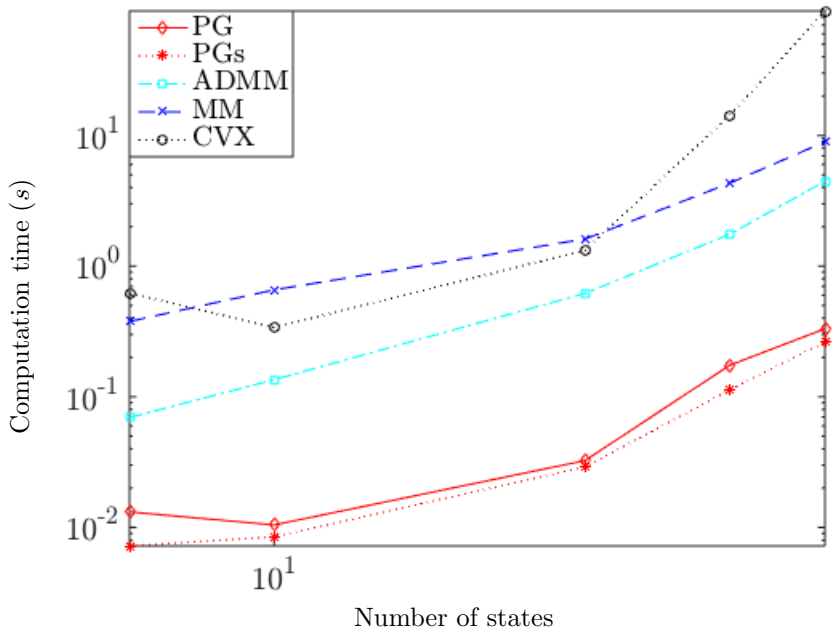


Figure 8.7: Computation time for mass-spring system damped with $b = 0.1$ and $\gamma = 30$.

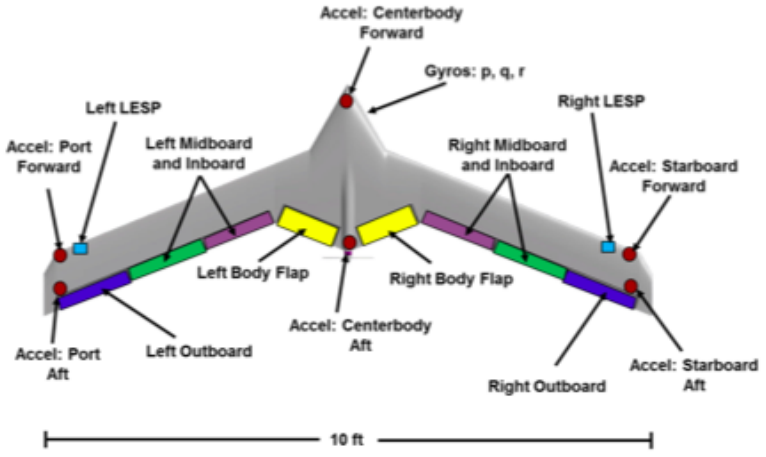


Figure 8.8: Body Freedom Flutter flexible wing testbed aircraft.

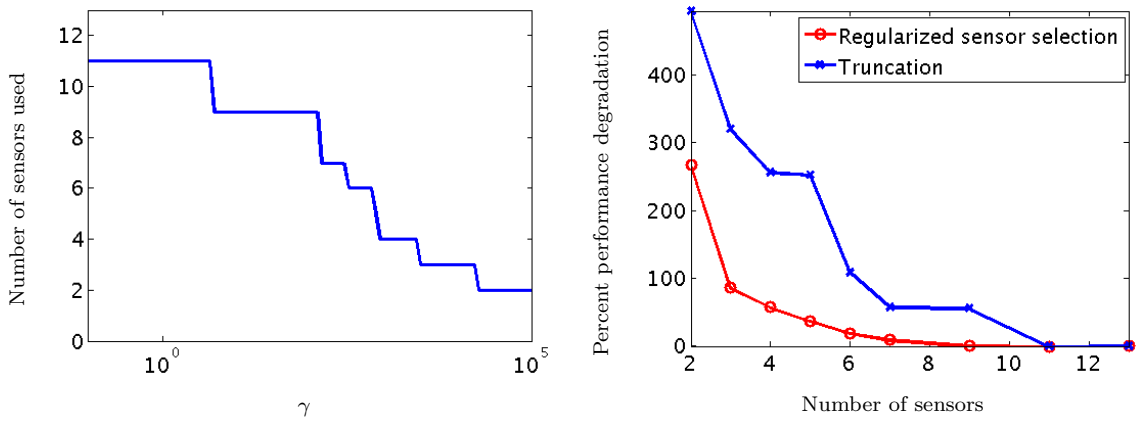


Figure 8.9: (a) Number of sensors as a function of the sparsity-promoting parameter γ; and (b) Performance comparison of the Kalman filter associated with the sets of sensors resulting from the regularized sensor selection problem and from truncation.

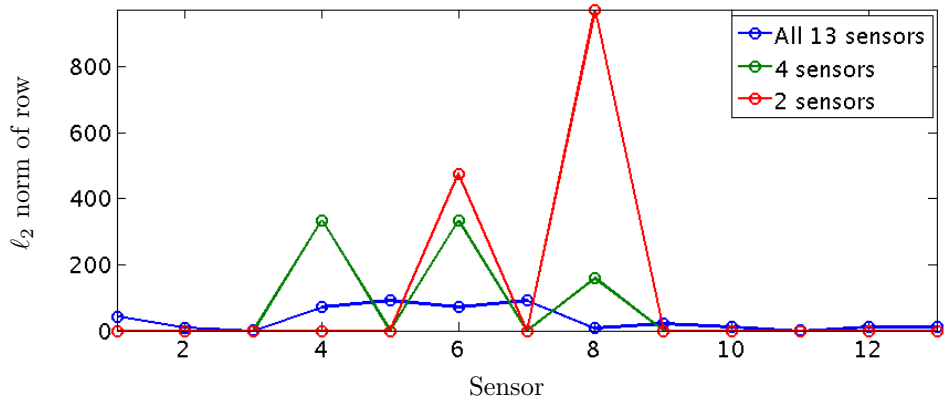


Figure 8.10: The ℓ_2 norm of each column of the Kalman gain for Kalman filters designed for different sets of sensors.

efficiency is that lighter airframes are more flexible and thus susceptible to vibrational instabilities [237]. These instabilities, known as flutter, were behind the famous Tacoma Narrows Bridge collapse and have been identified as the likely cause of the loss of NASA's Helios Prototype aircraft [238].

Recent work has sought to approach this problem by actively damping flutter instabilities [239]. Since active control requires reliable detection of instabilities, selection of sensors is an important challenge. For the Body Freedom Flutter test aircraft shown in Fig. 8.8 [240], we use the dual formulation to the actuator selection approach described in Section 8.1.2 in conjunction with iterative re-weighting algorithm to select sparse sets of sensors. Figure 8.9 shows the number of sensors as a function of the sparsity-promoting parameter γ and the performance of Kalman filter with the limited sets of sensors [105].

We compare the performance of the Kalman filter corresponding to the sensors selected by our approach to the Kalman filter associated with sensors selected by truncation. For the truncation approach, the Kalman gain matrix corresponding to a set of sensors was computed. The sensor corresponding to the row with the lowest ℓ_2 norm was discarded and the Kalman gain was recomputed for the new set of sensors. This process was repeated iteratively from the full set of sensors to a set of two sensors. Clearly, the regularized sensor selection algorithm selects better subsets of sensors than the truncation approach.

To further justify the use of this algorithm over truncation, we show the ℓ_2 norm of different columns of the Kalman gain in Fig. 8.10. The sensor which has the most effect on the state estimate changes as the number of sensors is decreased, even when each smaller set of sensors is a subset of a previous set.

Chapter 9

Conclusions and future directions

Proximal augmented Lagrangian methods

In this thesis, we have developed custom optimization algorithms for composite problems by introducing an auxiliary variable and reformulating the associated augmented Lagrangian. The resulting function, which we call the *proximal* augmented Lagrangian, is continuously differentiable and thus opens the door to many methods for solving the original problem.

The proximal augmented Lagrangian facilitates the method of multipliers by transforming the primal-minimization step into a tractable, differentiable problem. Differentiability also enables primal-dual gradient updates which, for certain problems, are convenient for a distributed implementation. Generalizations of the Jacobian for once-differentiable functions allow us to derive second order updates which lead to a globally exponentially convergent differential inclusion as well as a globally convergent algorithm with asympotitic quadratic convergence.

Structured optimal control via regularized optimization

The motivation behind developing these algorithms was to use regularization to design structured controllers. We successfully applied those algorithms to the *nonconvex* problems of edge addition in directed consensus networks and sparse optimal control. We also identified classes of systems for which structure-promoting optimal control problems can be cast as *convex* problems.

We first showed how the symmetric component of a system can be used to pose a regularized convex controller design problem. The resulting controller is structured and has stability and performance guarantees for the original system. We used this theory to inform the problems of edge addition and leader selection in directed consensus networks as well as for combination

drug therapy design for HIV treatment.

We also considered the decentralized control of positive systems. We showed that the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ optimal problems are convex in the original coordinates, allowing us to pose convex regularized problems to design structured controllers. We applied this theory to leader selection in consensus networks and combination drug therapy design for HIV. We then established that a constant controller is optimal for an induced-power performance index.

Finally, we considered the problem of choosing a limited number of actuators or sensors to effectively control or observe a system. We posed this problem as a semidefinite program and developed customized methods to solve it efficiently. We applied this approach to select sensors for a flexible wing aircraft.

Future directions

Future algorithmic research will investigate how to deal with nonconvexity. Although we can find local minima of nonconvex problems using the method of multipliers with the proximal augmented Lagrangian, we have no guarantees on the *quality* of these local minima. It would be interesting to explore techniques to discover a variety of local minima or to characterize the quality of any particular local minimum. Moreover, it would be interesting to explore the use of nonconvex regularizers, especially in distributed methods.

We will also work to develop techniques for structured time-varying or dynamic controllers. So far, we have considered exclusively *static* structured controllers. Although we show that a static controller achieves the optimal for decentralized control of positive systems with the induced power norm, we cannot expect this to hold for more general systems or performance indices. Investigating model predictive control, dynamic programming, or periodic structured control approaches would be an interesting avenue of further research.

References

[1] Enrique Fernández Cara and Enrique Zuazua Iriondo. Control Theory: History, Mathematical Achievements and Perspectives. *Boletín de la Sociedad Española de Matemática Aplicada, 26, 79-140.*, 2003.

[2] Christopher Bissell. A History of Automatic Control. In *Springer Handbook of Automation*, pages 53–69. Springer, 2009.

[3] Neculai Andrei. Modern Control Theory. *Studies in Informatics and Control*, 15(1):51, 2006.

[4] Vincent Blondel and John N Tsitsiklis. NP-hardness of some linear control design problems. *SIAM J. Control Optim.*, 35(6):2118–2127, 1997.

[5] Minyue Fu and Zhi-Quan Luo. Computational complexity of a problem arising in fixed order output feedback design. *Syst. Control Lett.*, 30(5):209–215, 1997.

[6] Francesco Bullo, Jorge Cortés, and Sonia Martínez. *Distributed Control of Robotic Networks*. Princeton University Press, 2009.

[7] Mehran Mesbahi and Magnus Egerstedt. *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[8] Dragoslav D Šiljak. *Decentralized control of complex systems*. Academic Press, New York, 1991.

[9] Bassam Bamieh, Fernando Paganini, and Munther A Dahleh. Distributed control of spatially invariant systems. *IEEE Trans. Automat. Control*, 47(7):1091–1107, 2002.

[10] Gustavo A de Castro and Fernando Paganini. Convex synthesis of localized controllers for spatially invariant system. *Automat.*, 38:445–456, 2002.

[11] Petros G Voulgaris, Gianni Bianchini, and Bassam Bamieh. Optimal $\mathcal{H}^2$ controllers for spatially invariant systems with delayed communication requirements. *Syst. Control Lett.*, 50:347–361, 2003.

[12] Raffaello D'Andrea and Geir E Dullerud. Distributed control design for spatially interconnected systems. *IEEE Trans. Automat. Control*, 48(9):1478–1495, 2003.

[13] Cédric Langbort, Ramu S Chandra, and Raffaello D'Andrea. Distributed control design for systems interconnected over an arbitrary graph. *IEEE Trans. Automat. Control*, 49(9):1502–1519, 2004.

[14] Dragoslav D Šiljak and Aleksandar I. Zečević. Control of large-scale systems: Beyond decentralized feedback. *Annual Reviews in Control*, 29(2):169–179, 2005.

[15] Bassam Bamieh and Petros G Voulgaris. A convex characterization of distributed control problems in spatially invariant systems with communication constraints. *Syst. Control Lett.*, 54(6):575–583, 2005.

[16] Michael Rotkowitz and Sanjay Lall. A characterization of convex problems in decentralized control. *IEEE Trans. Automat. Control*, 51(2):274–286, Feb 2006.

[17] Francesco Borrelli and Tamás Keviczky. Distributed LQR design for identical dynamically decoupled systems. *IEEE Trans. on Automat. Control*, 53(8):1901–1912, 2008.

[18] Javad Lavaei and Amir G Aghdam. Control of continuous-time LTI systems by means of structurally constrained controllers. *Automat.*, 44(1):141–148, 2008.

[19] Parikshit Shah and Pablo A Parrilo. $\mathcal{H}_2$ optimal decentralized control over posets: A state space solution for state-feedback. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 6722–6727, 2010.

[20] John Swigart and Sanjay Lall. An explicit state-space solution for a decentralized two-player optimal linear-quadratic regulator. In *Proceedings of the 2010 American Control Conference*, pages 6385–6390, 2010.

[21] Makan Fardad and Mihailo R Jovanović. Design of optimal controllers for spatially invariant systems with finite communication speed. *Automat.*, 47(5):880–889, May 2011.

[22] Karl Martensson and Anders Rantzer. A scalable method for continuous-time distributed control synthesis. In *Proceedings of the 2012 American Control Conference*, pages 6308–6313, 2012.

[23] Laurent Lessard and Sanjay Lall. Optimal controller synthesis for the decentralized two-player problem with output feedback. In *Proceedings of the 2012 American Control Conference*, pages 6314–6321, 2012.

[24] Aditya Mahajan, Nuno C Martins, Michael C Rotkowitz, and Serdar Yuksel. Information structures in optimal decentralized control. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 1291–1306, 2012.

[25] Parikshit Shah and Pablo A Parrilo. $\mathcal{H}_2$ optimal decentralized control over posets: A state-space solution for state-feedback. *IEEE Trans. Automat. Control*, 58(12):3084–3096, 2013.

[26] Javad Lavaei. Optimal decentralized control problem as a rank-constrained optimization. In *Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing*, pages 39–45, 2013.

[27] Ghazal Fazelnia, Ramtin Madani, and Javad Lavaei. Convex relaxation for optimal distributed control problem. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 896–903, 2014.

[28] Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. Localized LQR optimal control. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 1661–1668. IEEE, 2014.

[29] Laurent Lessard and Sanjay Lall. Optimal control of two-player systems with output feedback. *IEEE Trans. Automat. Control*, 60(8):2129–2144, 2015.

[30] Andrew Lamperski and John C Doyle. The $\mathcal{H}_2$ control problem for quadratically invariant systems with delays. *IEEE Trans. Automat. Control*, 60(7):1945–1950, 2015.

[31] Takuto Mori, Takayuki Wada, and Yasumasa Fujisaki. Structured feedback gain design via stabilizable dilation and LMIs. In *Proceedings of the 54th IEEE Conference on Decision and Control*, pages 3343–3348, 2015.

[32] Mihailo R Jovanović and Neil K Dhingra. Controller architectures: Tradeoffs between performance and structure. *Eur. J. Control*, 30:76–91, July 2016.

[33] Yuh-Shyang Wang. Localized LQR with adaptive constraint and performance guarantee. In *Proceedings of the 55th IEEE Conference on Decision and Control*, pages 2769–2776. IEEE, 2016.

[34] Yuh-Shyang Wang and Nikolai Matni. Localized LQG optimal control for large-scale systems. In *Proceedings of the 2016 American Control Conference*, pages 1954–1961. IEEE, 2016.

[35] Yuh-Shyang Wang. *A system level approach to optimal controller design for large-scale distributed systems*. PhD thesis, California Institute of Technology, 2017.

[36] Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. System level parameterizations, constraints and synthesis. In *Proceedings of the 2017 American Control Conference*, pages 1308–1315, May 2017.

[37] Yuh-Shyang Wang, Nikolai Matni, and John C Doyle. Separable and localized system level synthesis for large-scale systems. *arXiv preprint arXiv:1701.05880*, 2017.

[38] Mihailo R Jovanović and Bassam Bamieh. Componentwise energy amplification in channel flows. *J. Fluid Mech.*, 534:145–183, July 2005.

[39] Binh K Lieu and Mihailo R Jovanović. Computation of frequency responses for linear time-invariant PDEs on a compact interval. *J. Comput. Phys.*, 250:246–269, October 2013.

[40] Binh K Lieu, Mihailo R. Jovanović, and Satish Kumar. Worst-case amplification of disturbances in inertialess Couette flow of viscoelastic fluids. *J. Fluid Mech.*, 723:232–263, May 2013.

[41] Binh K Lieu. *Dynamics and control of Newtonian and viscoelastic fluids*. PhD thesis, University of Minnesota, 2014.

[42] Rashad Moarref. *Model-based control of transitional and turbulent wall-bounded shear flows*. PhD thesis, University of Minnesota, 2012.

[43] Rashad Moarref and Mihailo R Jovanović. Model-based design of transverse wall oscillations for turbulent drag reduction. *J. Fluid Mech.*, 707:205–240, September 2012.

[44] Rashad Moarref and Mihailo R Jovanović. Controlling the onset of turbulence by streamwise traveling waves. Part 1: Receptivity analysis. *J. Fluid Mech.*, 663:70–99, November 2010.

[45] Binh K Lieu, Rashad Moarref, and Mihailo R Jovanović. Controlling the onset of turbulence by streamwise traveling waves. Part 2: Direct numerical simulations. *J. Fluid Mech.*, 663:100–119, November 2010.

[46] Armin Zare, Mihailo R Jovanović, and Tryphon T Georgiou. Colour of turbulence. *J. Fluid Mech.*, 812:636–680, February 2017.

[47] Armin Zare. *Low-complexity stochastic modeling of wall-bounded shear flows*. PhD thesis, University of Minnesota, 2016.

[48] Bruce A Francis. *A Course in $\mathcal{H}_\infty$ Control Theory*. Springer-Verlag, New York, 1987.

[49] Geir E Dullerud and Fernando Paganini. *A Course in Robust Control Theory*. Springer, 2000.

[50] Yu Chi Ho and Kai-Ching Chu. Team decision theory and information structures in optimal control problems–part I. *IEEE Trans. Automat. Control*, 17(1):15–22, 1972.

[51] Petros G Voulgaris. Control of nested systems. In *Proceedings of the 2001 American Control Conference*, pages 4442–4445, 2000.

[52] Petros G Voulgaris. A convex characterization of classes of problems in control with specific interaction and communication structures. In *Proceedings of the 2001 American Control Conference*, pages 3128–3133, 2001.

[53] Lorenzo Farina and Sergio Rinaldi. *Positive Linear Systems: Theory and Applications*. John Wiley & Sons, 2011.

[54] Takashi Tanaka and Cédric Langbort. The bounded real lemma for internally positive systems and $\mathcal{H}_\infty$ structured static state feedback. *IEEE Trans. Automat. Control*, 56(9):2218–2223, 2011.

[55] Corentin Briat. Robust stability and stabilization of uncertain linear positive systems via integral linear constraints: $\mathcal{L}_1$ gain and $\mathcal{L}_\infty$ gain characterization. *Int. J. Robust Nonlin.*, 23(17):1932–1954, 2013.

[56] Yoshio Ebihara, Dimitri Peaucelle, and Denis Arzelier. $\mathcal{L}_1$ gain analysis of linear positive systems and its application. In *Proceedings of 50th IEEE Conference on Decision and Control*, pages 4029–4034. IEEE, 2011.

[57] Anders Rantzer. Distributed control of positive systems. In *Proceedings of 50th IEEE Conference on Decision and Control*, pages 6608–6611, Dec 2011.

[58] Patrizio Colaneri, Richard H Middleton, Zhiyong Chen, Danilo Caporale, and Franco Blanchini. Convexity of the cost functional in an optimal control problem for a class of positive switched systems. *Automat.*, 4(50):1227–1234, 2014.

[59] Anders Rantzer and Bo Bernhardsson. Control of convex monotone systems. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 2378–2383, 2014.

[60] Nader Motee and Ali Jadbabaie. Optimal control of spatially distributed systems. *IEEE Trans. Automat. Control*, 53(7):1616–1629, 2008.

[61] Nader Motee and Qiyu Sun. Sparsity measures for spatially decaying systems. In *Proceedings of the 2014 American Control Conference*, pages 5459–5464, 2014.

[62] Makan Fardad, Fu Lin, and Mihailo R Jovanović. Sparsity-promoting optimal control for a class of distributed systems. In *Proceedings of the 2011 American Control Conference*, pages 2050–2055, 2011.

[63] Fu Lin, Makan Fardad, and Mihailo R Jovanović. Sparse feedback synthesis via the alternating direction method of multipliers. In *Proceedings of the 2012 American Control Conference*, pages 4765–4770, Montréal, Canada, 2012.

[64] Fu Lin, Makan Fardad, and Mihailo R Jovanović. Design of optimal sparse feedback gains via the alternating direction method of multipliers. *IEEE Trans. Automat. Control*, 58(9):2426–2431, September 2013.

[65] Makan Fardad, Fu Lin, and Mihailo R Jovanović. On optimal link creation for facilitation of consensus in social networks. In *Proceedings of the 2014 American Control Conference*, pages 3802–3807, Portland, OR, 2014.

[66] George F Young, L. Scardovi, and Naomi E Leonard. Robustness of noisy consensus dynamics with directed communication. In *Proceedings of the 2010 American Control Conference*, pages 6312–6317, 2010.

[67] Lin Xiao, Stephen P Boyd, and Seung-Jean Kim. Distributed average consensus with least-mean-square deviation. *J. Parallel Distrib. Comput.*, 67(1):33–46, 2007.

[68] Daniel Zelazo and Mehran Mesbahi. Edge agreement: Graph-theoretic performance bounds and passivity analysis. *IEEE Trans. Automat. Control*, 56(3):544–555, 2011.

[69] Makan Fardad, Fu Lin, and Mihailo R Jovanović. On the optimal synchronization of oscillator networks via sparse interconnection graphs. In *Proceedings of the 2012 American Control Conference*, pages 4777–4782, Montréal, Canada, 2012.

[70] David JT Sumpter, Jens Krause, Richard James, Ian D Couzin, and Ashley JW Ward. Consensus decision making by fish. *Current Biology*, 18(22):1773–1777, 2008.

[71] Mehran Mesbahi and Fred Y Hadaegh. Formation flying control of multiple spacecraft via graphs, matrix inequalities, and switching. *J. Guid. Control Dyn.*, 24(2):369–377, 2001.

[72] Lin Xiao, Stephen P Boyd, and Sanjay Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks*, pages 63–70, 2005.

[73] Arpita Ghosh and Stephen P Boyd. Growing well-connected graphs. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 6605–6611, 2006.

[74] Fu Lin, Makan Fardad, and Mihailo R Jovanović. Identification of sparse communication graphs in consensus networks. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, pages 85–89, Monticello, IL, 2012.

[75] Tyler Summers, Iman Shames, John Lygeros, and Florian Dörfler. Topology design for optimal network coherence. In *Proceedings of the 2015 European Control Conference*, pages 575–580. IEEE, 2015.

[76] Sepideh Hassan-Moghaddam and Mihailo R Jovanović. Topology design for stochastically-forced consensus networks. *IEEE Trans. Control Netw. Syst.*, 2017. doi:10.1109/TCNS.2017.2674962.

[77] Stacy Patterson and Bassam Bamieh. Leader selection for optimal network coherence. In *Proceedings of the 49th IEEE Conference on Decision and Control*, pages 2692–2697, 2010.

[78] Katherine Fitch and Naomi E Leonard. Information centrality and optimal leader selection in noisy networks. In *Proceedings of the 52nd IEEE Conference on Decision and Control*, pages 7510–7515, 2013.

[79] Katherine Fitch and Naomi E Leonard. Joint centrality distinguishes optimal leaders in noisy networks. *IEEE Trans. Control Netw. Syst.*, 2014.

[80] Fu Lin, Makan Fardad, and Mihailo R Jovanović. Algorithms for leader selection in stochastically forced consensus networks. *IEEE Trans. Automat. Control*, 59(7):1789–1802, July 2014.

[81] Makan Fardad, Fu Lin, and Mihailo R Jovanović. Algorithms for leader selection in large dynamical networks: noise-free leaders. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, pages 7188–7193, Orlando, FL, 2011.

[82] Andrew Clark, Linda Bushnell, and Radha Poovendran. A supermodular optimization framework for leader selection under link noise in linear multi-agent systems. *IEEE Trans. Automat. Control*, 59(2):283–296, 2014.

[83] Andrew Clark, Basel Alomair, Linda Bushnell, and Radha Poovendran. *Submodularity in Dynamics and Control of Networked Systems*. Springer, 2016.

[84] Andrew Clark, Basel Alomair, Linda Bushnell, and Radha Poovendran. Submodularity in input node selection for networked linear systems: Efficient algorithms for performance and controllability. *IEEE Control Syst. Mag.*, 37(6):52–74, Dec 2017.

[85] Guodong Shi, Kin Cheong Sou, Henrik Sandberg, and Karl Henrik Johansson. A graph-theoretic approach on optimizing informed-node selection in multi-agent tracking control. *Physica D: Nonlinear Phenomena*, 267:104–111, 2014.

[86] Stacy Patterson, Neil McGlohon, and Kirill Dyagilev. Optimal k-leader selection for coherence and convergence rate in one-dimensional networks. *IEEE Trans. Control Netw. Syst.*, 2016.

[87] Neil K Dhingra, Marcello Colombino, and Mihailo R Jovanović. Leader selection in directed networks. In *Proceedings of the 55th IEEE Conference on Decision and Control*, pages 2715–2720, Las Vegas, NV, 2016.

[88] Neil K Dhingra, Marcello Colombino, and Mihailo R Jovanović. Structured decentralized control of positive systems with applications to combination drug therapy and leader selection in directed networks. *IEEE Trans. Control Netw. Syst.*, 2017. submitted.

[89] Vanessa D Jonsson, Anders Rantzer, and Richard M Murray. A scalable formulation for engineering combination therapies for evolutionary dynamics of disease. *Proceedings of the 2014 American Control Conference*, pages 2771–2778, 2014.

[90] Vanessa D Jonsson, Nikolai Matni, and Richard M Murray. Reverse engineering combination therapies for evolutionary dynamics of disease: An $\mathcal{H}_\infty$ approach. In *Proceedings of the 52nd IEEE Conference on Decision and Control*, pages 2060–2065, 2013.

[91] Makan Fardad, Fu Lin, and Mihailo R Jovanović. On the optimal design of structured feedback gains for interconnected systems. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 978–983, Shanghai, China, 2009.

[92] Fu Lin, Makan Fardad, and Mihailo R Jovanović. Augmented Lagrangian approach to design of structured optimal state feedback gains. *IEEE Trans. Automat. Control*, 56(12):2923–2929, 2011.

[93] David L Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[94] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009.

[95] Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Stat.*, pages 199–227, 2008.

[96] Tom Goldstein and Stanley Osher. The split Bregman method for ℓ_1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, 2009.

[97] Bassam Bamieh, Mihailo R Jovanović, Partha Mitra, and Stacy Patterson. Coherence in large-scale networks: dimension dependent limitations of local feedback. *IEEE Trans. Automat. Control*, 57(9):2235–2249, September 2012.

[98] Simone Schuler, Ping Li, James Lam, and Frank Allgöwer. Design of structured dynamic output-feedback controllers for interconnected systems. *International Journal of Control*, 84(12):2081–2091, 2011.

[99] Makan Fardad and Mihailo R Jovanović. On the design of optimal structured and sparse feedback gains using semidefinite programming. In *Proceedings of the 2014 American Control Conference*, pages 2438–2443, Portland, OR, 2014.

[100] Nikolai Matni and Venkat Chandrasekaran. Regularization for design. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 1111–1118, Los Angeles, CA, 2014.

[101] Nikolai Matni. Communication delay co-design in $\mathcal{H}_2$-distributed control using atomic norm minimization. *IEEE Transactions on Control of Network Systems*, 4(2):267–278, June 2017.

[102] Nikolai Matni and Venkat Chandrasekaran. Regularization for design. *IEEE Trans. Automat. Control*, 61(12):3991–4006, 2016.

[103] Nikolai Matni. *Distributed optimal control of cyber-physical systems: controller synthesis, architecture design and system identification*. PhD thesis, California Institute of Technology, 2016.

[104] Boris Polyak, Mikhail Khlebnikov, and Pavel Shcherbakov. An LMI approach to structured sparse feedback design in linear control systems. In *Proceedings of the 2013 European Control Conference*, pages 833–838, 2013.

[105] Neil K Dhingra, Mihailo R Jovanović, and Z. Q. Luo. An ADMM algorithm for optimal sensor and actuator selection. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 4039–4044, Los Angeles, CA, 2014.

[106] Xiaofan Wu and Mihailo R Jovanović. Sparsity-promoting optimal control of systems with symmetries, consensus and synchronization networks. *Syst. Control Lett.*, 103:1–8, May 2017.

[107] Neil K Dhingra, Sei Zhen Khong, and Mihailo R Jovanović. The proximal augmented Lagrangian method for nonsmooth composite optimization. *IEEE Trans. Automat. Control*, 2016. submitted; also arXiv:1610.04514.

[108] Florian Dörfler, Mihailo R Jovanović, Michael Chertkov, and Francesco Bullo. Sparse and optimal wide-area damping control in power networks. In *Proceedings of the 2013 American Control Conference*, pages 4295–4300, Washington, DC, 2013.

[109] Florian Dörfler, Mihailo R Jovanović, M. Chertkov, and F. Bullo. Sparsity-promoting optimal wide-area control of power networks. *IEEE Trans. Power Syst.*, 29(5):2281–2291, September 2014.

[110] Xiaofan Wu, Florian Dörfler, and Mihailo R Jovanović. Input-output analysis and decentralized optimal control of inter-area oscillations in power systems. *IEEE Trans. Power Syst.*, 31(3):2434–2444, May 2016.

[111] Xiaofan Wu, Florian Dörfler, and Mihailo R Jovanović. Topology identification and design of distributed integral action in power networks. In *Proceedings of the 2016 American Control Conference*, pages 5921–5926, Boston, MA, 2016.

[112] Neil K Dhingra, Marcello Colombino, and Mihailo R Jovanović. On the convexity of a class of structured optimal control problems for positive systems. In *Proceedings of the 2016 European Control Conference*, pages 825–830, Aalborg, Denmark, 2016.

[113] Sepideh Hassan-Moghaddam and Mihailo R Jovanović. An interior point method for growing connected resistive networks. In *Proceedings of the 2015 American Control Conference*, pages 1223–1228, Chicago, IL, 2015.

[114] Sepideh Hassan-Moghaddam and Mihailo R Jovanović. Customized algorithms for growing connected resistive networks. In *Proceedings of the 10th IFAC Symposium on Nonlinear Control Systems*, pages 986–991, Monterey, CA, 2016.

[115] Neil K Dhingra and Mihailo R Jovanović. Convex synthesis of symmetric modifications to linear systems. In *Proceedings of the 2015 American Control Conference*, pages 3583–3588, Chicago, IL, 2015.

[116] Neil K Dhingra, Xiaofan Wu, and Mihailo R Jovanović. Sparsity-promoting optimal control of systems with invariances and symmetries. In *Proceedings of the 10th IFAC Symposium on Nonlinear Control Systems*, pages 648–653, Monterey, CA, 2016.

[117] Makan Fardad, Fu Lin, and Mihailo R Jovanović. Design of optimal sparse interconnection graphs for synchronization of oscillator networks. *IEEE Trans. Automat. Control*, 59(9):2457–2462, September 2014.

[118] Neil K Dhingra and Mihailo R Jovanović. A method of multipliers algorithm for sparsity-promoting optimal control. In *Proceedings of the 2016 American Control Conference*, pages 1942–1947, Boston, MA, 2016.

[119] Neil K Dhingra, Sei Zhen Khong, and Mihailo R Jovanović. A second order primal-dual algorithm for non-smooth convex composite optimization. In *Proceedings of the 56th IEEE Conference on Decision and Control*, Melbourne, Australia, 2017.

[120] Neil K Dhingra, Sei Zhen Khong, and Mihailo R Jovanović. A second order primal-dual method for nonsmooth convex composite optimization. *IEEE Trans. Automat. Control*, 2017. submitted; also arXiv:1709.01610.

[121] Emmanuel J Candès, Mike B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl*, 14:877–905, 2008.

[122] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Series B Stat. Methodol.*, 68(1):49–67, 2006.

[123] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[124] Athanasios C Antoulas. *Approximation of Large-Scale Dynamical Systems*. SIAM, 2005.

[125] Armin Zare, Yongxin Chen, Mihailo R Jovanović, and Tryphon T Georgiou. Low-complexity modeling of partially available second-order statistics: theory and an efficient matrix completion algorithm. *IEEE Trans. Automat. Control*, 62(3):1368–1383, March 2017.

[126] Armin Zare, Mihailo R Jovanović, and Tryphon T Georgiou. Alternating direction optimization algorithms for covariance completion problems. In *Proceedings of the 2015 American Control Conference*, pages 515–520, Chicago, IL, 2015.

[127] Armin Zare, Mihailo R Jovanović, and Tryphon T Georgiou. Completion of partially known turbulent flow statistics. In *Proceedings of the 2014 American Control Conference*, pages 1680–1685, Portland, OR, 2014.

[128] Siddharth Joshi and Stephen P Boyd. Sensor selection via convex optimization. *IEEE Trans. Signal Process.*, 57(2):451–462, 2009.

[129] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational Mathematics*, 12(6):805–849, 2012.

[130] Neal Parikh and Stephen P Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013.

[131] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[132] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learning*, 3(1):1–124, 2011.

[133] Dimitri P Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.

[134] Dimitri P Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[135] Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer, 2006.

[136] Huan Li and Zhouchen Lin. Accelerated proximal gradient methods for nonconvex programming. In *Adv. Neural Inf. Process Syst.*, pages 379–387, 2015.

[137] Mingyi Hong, Zhi-Quan Luo, and Meisam Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM J. Optimiz.*, 26(1):337–364, 2016.

[138] Ralph T Rockafellar and Roger J-B Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009.

[139] Frank H Clarke. *Optimization and Nonsmooth Analysis*, volume 5. SIAM, 1990.

[140] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Math. Program.*, 146(1-2):459–494, 2014.

[141] Yin Wang, Jose A Lopez, and Mario Sznaier. Sparse static output feedback controller design via convex optimization. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 376–381, 2014.

[142] Tankred Rautert and Ekkehard W Sachs. Computational design of optimal output feedback controllers. *SIAM J. Optimiz.*, 7(3):837–852, 1997.

[143] Jonathan Barzilai and Jonathan M Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8(1):141–148, 1988.

[144] Andrei Patrascu and Ion Necoara. Efficient random coordinate descent algorithms for large-scale structured nonconvex optimization. *J. Global Optim.*, 61(1):19–46, 2015.

[145] Kenneth J Arrow, Leonid Hurwicz, and Hirofumi Uzawa. *Studies in Linear and Non-Linear Programming*. Stanford University Press, 1958.

[146] Jing Wang and Nicola Elia. A control perspective for centralized and distributed convex optimization. In *Proceedings of the 50th IEEE Conference on Decision and Control and the 10th European Control Conference*, pages 3800–3805, 2011.

[147] Angelia Nedic and Asuman Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Automat. Control*, 54(1):48–61, 2009.

[148] Diego Feijer and Fernando Paganini. Stability of primal–dual gradient dynamics and applications to network optimization. *Automat.*, 46(12):1974–1981, 2010.

[149] Ashish Cherukuri, Enrique Mallada, and Jorge Cortés. Asymptotic convergence of constrained primal–dual dynamics. *Syst. Control Lett.*, 87:10–15, 2016.

[150] Ashish Cherukuri, Enrique Mallada, Steven Low, and Jorge Cortes. The role of convexity on saddle-point dynamics: Lyapunov function and robustness. *arXiv preprint arXiv:1608.08586*, 2016.

[151] Ashish Cherukuri, Bahman Gharesifard, and Jorge Cortes. Saddle-point dynamics: conditions for asymptotic stability of saddle points. *SIAM Journal on Control and Optimization*, 55(1):486–511, 2017.

[152] Laurent Lessard, Benjamin Recht, and Andrew Packard. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM J. Optimiz.*, 26(1):57–95, 2016.

[153] Bin Hu and Peter Seiler. Exponential decay rate conditions for uncertain linear systems using integral quadratic constraints. *IEEE Trans. Automat. Control*, 61(11):3631–3637, 2016.

[154] Ross Boczar, Laurent Lessard, and Benjamin Recht. Exponential convergence bounds using integral quadratic constraints. In *Proceedings of the 54th IEEE Conference on Decision and Control*, pages 7516–7521. IEEE, 2015.

[155] Ross Boczar, Laurent Lessard, Andrew Packard, and Benjamin Recht. Exponential stability analysis via integral quadratic constraints. *arXiv preprint arXiv:1706.01337*, 2017.

[156] Anders Rantzer. On the Kalman-Yakubovich-Popov lemma. *Systems & Control Letters*, 28(1):7–10, 1996.

[157] Paul Tseng and Sangwoon Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.*, 117(1-2):387–423, 2009.

[158] Richard H Byrd, Jorge Nocedal, and Figen Oztoprak. An inexact successive quadratic approximation method for ℓ_1 regularized optimization. *Math. Program.*, pages 1–22, 2015.

[159] Jason D Lee, Yuekai Sun, and Michael Saunders. Proximal Newton-type methods for convex optimization. In *Adv. Neural Inf. Process. Syst.*, pages 836–844, 2012.

[160] Cho-Jui Hsieh, Inderjit S Dhillon, Pradeep K Ravikumar, and Matyas A Sustik. Sparse inverse covariance matrix estimation using quadratic approximation. In *Adv. Neural Inf. Process. Syst.*, pages 2330–2338, 2011.

[161] Cho-Jui Hsieh, Matyas A Sustik, Inderjit S Dhillon, and Pradeep Ravikumar. QUIC: Quadratic approximation for sparse inverse covariance estimation. *J. Mach. Learn. Res.*, 15:2911–2947, 2014.

[162] Philip E Gill and Daniel P Robinson. A primal-dual augmented Lagrangian. *Comput. Optim. Appl.*, 51(1):1–25, 2012.

[163] Paul Armand, Joël Benoist, Riadh Omheni, and Vincent Pateloup. Study of a primal-dual algorithm for equality constrained minimization. *Comput. Optim. Appl.*, 59(3):405–433, 2014.

[164] Paul Armand and Riadh Omheni. A globally and quadratically convergent primal-dual augmented Lagrangian algorithm for equality constrained optimization. *Optim. Method. Softw.*, pages 1–21, 2015.

[165] Stephen Becker and Jalal Fadili. A quasi-Newton proximal splitting method. In *Adv. Neural Inf. Process. Syst.*, pages 2618–2626, 2012.

[166] Liqun Qi and Jie Sun. A nonsmooth version of Newton's method. *Math. Program.*, 58(1):353–367, 1993.

[167] Stephen M Robinson. Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity. In *Nonlinear Analysis and Optimization*, pages 45–66. Springer, 1987.

[168] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer Science & Business Media, 2013.

[169] Godfrey Harold Hardy and Edward Maitland Wright. *An Introduction to the Theory of Numbers*, volume 5. Oxford University Press, 1979.

[170] Kaifeng Jiang, Defeng Sun, and Kim-Chuan Toh. A partial proximal point algorithm for nuclear norm regularized matrix least squares problems. *Math. Program. Comp.*, 6(3):281–325, 2014.

[171] Fanwen Meng, Defeng Sun, and Gongyun Zhao. Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. *Math. Program.*, 104(2):561–581, 2005.

[172] Joseph B Kruskal. Two convex counterexamples: A discontinuous envelope function and a nondifferentiable nearest-point mapping. *Proceedings of the American Mathematical Society*, 23(3):697–703, 1969.

[173] Jason D Lee, Yuekai Sun, and Michael A Saunders. Proximal Newton-type methods for minimizing composite functions. *SIAM J. Optimiz.*, 24(3):1420–1443, 2014.

[174] Dimitri P Bertsekas. Projected Newton methods for optimization problems with simple constraints. *SIAM J. Control Opt.*, 20(2):221–246, 1982.

[175] Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal. Statist. Soc B.*, 58(1):267–288, 1996.

[176] Panagiotis Patrinos, Lorenzo Stella, and Alberto Bemporad. Forward-backward truncated Newton methods for large-scale convex composite optimization. 2014. arXiv:1402.6655.

[177] Lorenzo Stella, Andreas Themelis, and Panagiotis Patrinos. Forward–backward quasi-Newton methods for nonsmooth optimization problems. *Computational Optimization and Applications*, 67(3):443–487, 2017.

[178] Andreas Themelis, Lorenzo Stella, and Panagiotis Patrinos. Forward-backward envelope for the sum of two nonconvex functions: Further properties and nonmonotone line-search algorithms. 2016. arXiv:1606.06256.

[179] James M Ortega and Werner C Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. SIAM, 2000.

[180] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[181] Walter Rudin. *Principles of Mathematical Analysis*, volume 3. McGraw-Hill, Inc., 1964.

[182] Francisco Facchinei. Minimization of SC^1 functions and the Maratos effect. *Oper. Res. Lett.*, 17(3):131–137, 1995.

[183] Hassan K Khalil. *Nonlinear Systems*. Prentice Hall, Third edition, 1996.

[184] Tecla De Luca, Francisco Facchinei, and Christian Kanzow. A semismooth equation approach to the solution of nonlinear complementarity problems. *Math. Program.*, 75(3):407–439, 1996.

[185] Andrew R Conn, Nicholas IM Gould, and Philippe Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28(2):545–572, 1991.

[186] Jerome Friedman, Trevor Hastie, and Rob Tibshirani. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, 33(1):1, 2010.

[187] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Processing*, 57(7):2479–2493, 2009.

[188] Seung-Jean Kim, Kwangmoo Koh, Michael Lustig, Stephen Boyd, and Dimitry Gorinevsky. An interior-point method for large-scale ℓ_1-regularized least squares. *IEEE J. Sel. Topics Signal Process*, 1(4):606–617, 2007.

[189] Junfeng Yang and Yin Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM J. Sci. Comput.*, 33(1):250–278, 2011.

[190] David M Zoltowski, Neil K Dhingra, Fu Lin, and Mihailo R Jovanović. Sparsity-promoting optimal control of spatially-invariant systems. In *Proceedings of the 2014 American Control Conference*, pages 1261–1266, Portland, OR, 2014.

[191] Ju Swift and Pierre C Hohenberg. Hydrodynamic fluctuations at the convective instability. *Physical Review A*, 15(1):319, 1977.

[192] Ralph T Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of Operations Research*, 1:97–116, 1976.

[193] Bahman Gharesifard and Jorge Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Trans. Automat. Control*, 59(3):781–786, 2014.

[194] Wei Shi, Qing Ling, Gang Wu, and Wotao Yin. EXTRA: An exact first-order algorithm for decentralized consensus optimization. *SIAM J. Optimiz.*, 25(2):944–966, 2015.

[195] Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997.

[196] Bassam Bamieh and Mohammed Dahleh. Exact computation of traces and $\mathcal{H}^2$ norms for a class of infinite dimensional problems. *IEEE Trans. Automat. Control*, 48(4):646–649, 2003.

[197] Franziska Michor, Yoh Iwasa, and Martin A Nowak. Dynamics of cancer progression. *Nature Reviews Cancer*, 4(3):197–205, 2004.

[198] Bissan Al-Lazikani, Udai Banerji, and Paul Workman. Combinatorial drug therapy for cancer in the post-genomic era. *Nature Biotechnology*, 30(7):679–692, 2012.

[199] Daniel S Rosenbloom, Alison L Hill, S Alireza Rabi, Robert F Siliciano, and Martin A Nowak. Antiretroviral dynamics determines hiv evolution and predicts therapy outcome. *Nature Medicine*, 18(9):1378–1385, 2012.

[200] Florian Klein, Ariel Halper-Stromberg, Joshua A Horwitz, Henning Gruell, Johannes F Scheid, Stylianos Bournazos, Hugo Mouquet, Linda A Spatz, Ron Diskin, Alexander Abadir, et al. HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature*, 492(7427):118–122, 2012.

[201] Lloyd N Trefethen and Mark Embree. *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton University Press, 2005.

[202] JAC Weideman and Satish C Reddy. A MATLAB differentiation matrix suite. *ACM Transactions on Mathematical Software*, 26(4):465–519, December 2000.

[203] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.0 beta. `http://cvxr.com/cvx`, September 2013.

[204] Marcello Colombino and Roy S Smith. Convex characterization of robust stability analysis and control synthesis for positive linear systems. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, Los Angeles, California, 2014.

[205] Marcello Colombino and Roy S Smith. A convex characterization of robust stability for positive and positively dominated linear systems. *IEEE Trans. Automat. Control*, 61(7):1965–1971, July 2016.

[206] Marcello Colombino. *Robust and Decentralized Control of Positive Systems: a Convex Approach*. PhD thesis, ETH Zürich, 2016.

[207] Joel E Cohen. Convexity of the dominant eigenvalue of an essentially nonnegative matrix. *Proc. Am. Math. Soc.*, 81(4):657–658, 1981.

[208] Anders Rantzer. On the Kalman-Yakubovich-Popov lemma for positive systems. *IEEE Trans. Automat. Control*, 61(5):1346–1349, 2016.

[209] Naum Z Shor. *Minimization Methods for Non-differentiable Functions*, volume 3. Springer Science & Business Media, 2012.

[210] José YB Cruz. On proximal subgradient splitting method for minimizing the sum of two nonsmooth convex functions. *Set-Valued and Variational Analysis*, 25(2):245–263, 2017.

[211] Dragoš M Cvetković, Michael Doob, and Horst Sachs. *Spectra of Graphs: Theory and Application*, volume 87. Academic Pr, 1980.

[212] Eugene L Lawler and David E Wood. Branch-and-bound methods: a survey. *Operations Research*, 14(4):699–719, 1966.

[213] Weiran Wang and Canyi Lu. Projection onto the capped simplex. *arXiv preprint arXiv:1503.01002*, 2015.

[214] Daniel Zelazo, Simone Schuler, and Frank Allgöwer. Performance and design of cycles in consensus networks. *Syst. Control Lett.*, 62(1):85–96, 2013.

[215] Duncan J Watts and Stephen H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.

[216] John G White, Eileen Southgate, J Nichol Thomson, and Sydney Brenner. The structure of the nervous system of the nematode Caenorhabditis Elegans: the mind of a worm. *Phil. Trans. R. Soc. Lond*, 314:1–340, 1986.

[217] Shun Motegi and Naoki Masuda. A network-based dynamical system for competitive sports. *Scientific Reports*, 2, 2012.

[218] Juyong Park and Mark EJ Newman. A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):P10014, 2005.

[219] Sports-Reference. Sports Reference – College Football 2015 schedule and results. `http://www.sports-reference.com/cfb/years/2015-schedule.html`, 2016. Accessed: 2016-03-15.

[220] Associated-Press. The AP Top 25 Poll. `http://collegefootball.ap.org/poll`, 2016. Accessed: 2016-03-15.

[221] Esteban Hernandez-Vargas, Patrizio Colaneri, Richard Middleton, and Franco Blanchini. Discrete-time control for switched positive systems with application to mitigating viral escape. *Int. J. Robust Nonlin.*, 21(10):1093–1111, 2011.

[222] Vanessa D Jonsson, Nikolai Matni, and Richard M Murray. Synthesizing combination therapies for evolutionary dynamics of disease for nonlinear pharmacodynamics. In *Proceedings of the 53rd Conference on Decision and Control*, pages 2352–2358. IEEE, 2014.

[223] Vanessa D Jonsson. *Robust Control of Evolutionary Dynamics*. PhD thesis, California Institute of Technology, 2016.

[224] Abram S Besicovitch. *Almost Periodic Functions*. Dover Publications, 1954.

[225] Constantin Corduneanu. *Almost Periodic Oscillations and Waves*. Springer Science & Business Media, 2009.

[226] Christian Constanda and Maria E Perez. *Integral Methods in Science and Engineering*. Springer, 2010.

[227] Jorge Mari. A counterexample in power signals space. *IEEE Trans. Automat. Control*, 41(1):115–116, 1996.

[228] Petar Kokotović, Hassan K Khalil, and John O'Reilly. *Singular perturbation methods in control: analysis and design*. SIAM, 1999.

[229] Venkat Roy, Sundeep P Chepuri, and Geert Leus. Sparsity-enforcing sensor selection for DOA estimation. In *Proceedings of the 2013 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pages 340–343, 2013.

[230] Vassilis Kekatos, Georgios B Giannakis, and Bruce Wollenberg. Optimal placement of phasor measurement units via convex relaxation. *IEEE Trans. Power Syst.*, 27(3):1521–1530, 2012.

[231] James L Rogers. A parallel approach to optimum actuator selection with a genetic algorithm. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, pages 14–17, 2000.

[232] Shinji Kondoh, Chikayoshi Yatomi, and Koichi Inoue. The positioning of sensors and actuators in the vibration control of flexible systems. *JSME Int. J., Series III*, 33(2):145–152, 1990.

[233] Kazuhiko Hiramoto, Hitoshi Doki, and Goro Obinata. Optimal sensor/actuator placement for active vibration control using explicit solution of algebraic Riccati equation. *J. Sound Vib.*, 229(5):1057–1075, 2000.

[234] Kevin K Chen and Clarence W Rowley. $\mathcal{H}_2$ optimal actuator and sensor placement in the linearised complex Ginzburg-Landau system. *J. Fluid Mech.*, 681:241–260, 2011.

[235] Engin Masazade, Makan Fardad, and Pramod K Varshney. Sparsity-promoting extended Kalman filtering for target tracking in wireless sensor networks. *IEEE Signal Processing Letters*, 19:845–848, 2012.

[236] Michael Grant and Stephen P Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[237] Martin J Brenner, Richard C Lind, and David F Voracek. *Overview of Recent Flight Flutter Testing Research at NASA Dryden*, volume 4792. National Aeronautics and Space Administration, Office of Management, Scientific and Technical Information Program, 1997.

[238] Thomas E Noll, John M Brown, Marla E Perez-Davis, Stephen D Ishmael, Geary C Tiffany, and Matthew Gaier. Investigation of the Helios prototype aircraft mishap volume I mishap report. *NASA*, 1, 2004.

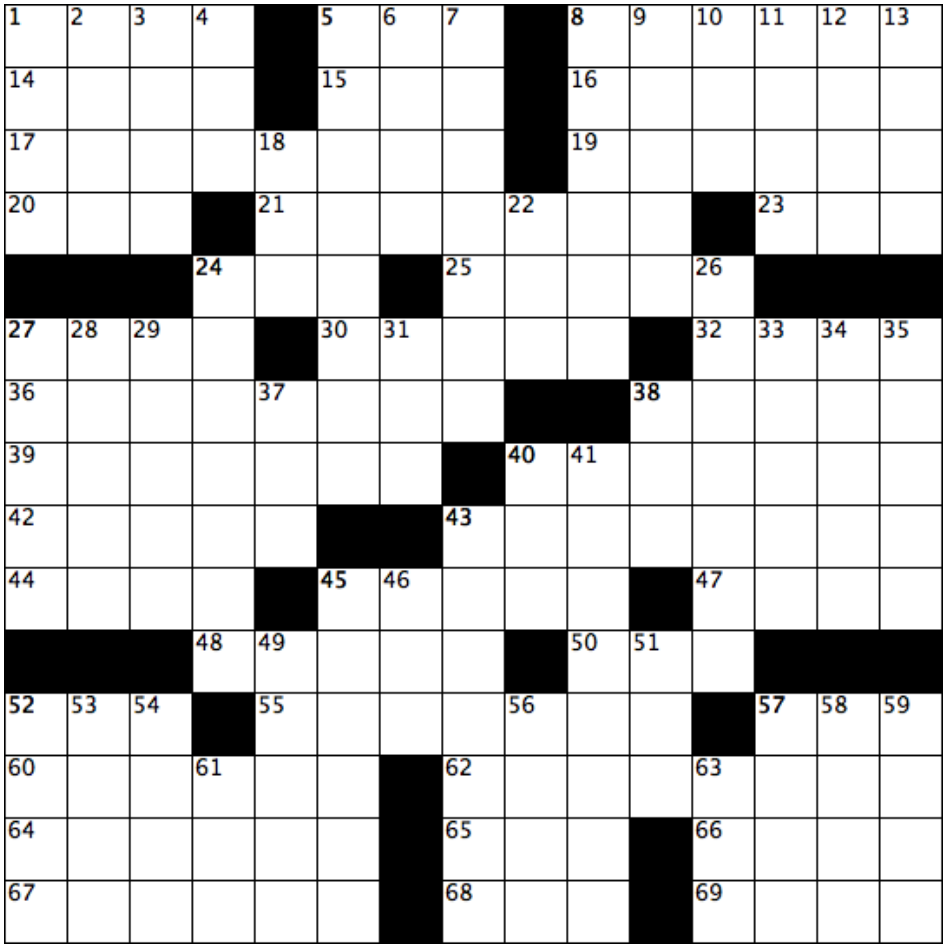
[239] Jeffrey M Barker and Gary J Balas. Comparing linear parameter-varying gain-scheduled control techniques for active flutter suppression. *J. Guid. Control Dyn.*, 23(5):948–955, 2000.

[240] Claudia Moreno, Peter Seiler, and Gary J Balas. Linear parameter varying model reduction for aeroservoelastic systems. In *Proceedings of the AIAA Atmospheric Flight Mechanics Conference*, page 4859, 2012.

Appendix A

Crossword

Since the subject of my thesis is mathematical in nature, I thought it appropriate that I leave something as an exercise to the reader.



Across

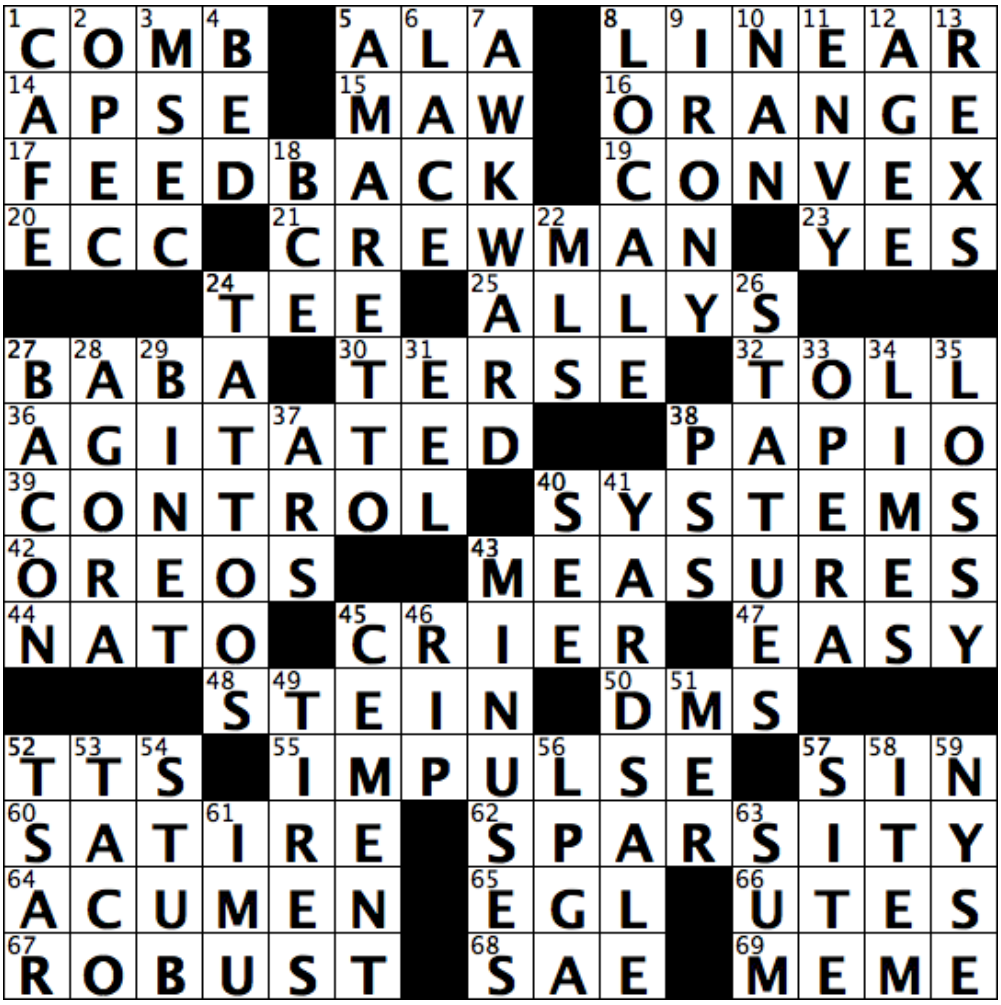
1. Dirac ___, i.e. a 55-across train
5. In the style of
8. Preserves additivity and homogeneity
14. Basilica part
15. Gaping hole
16. Non-rhyming fruit
17. It closes the loop
19. As a set which admits a supporting hyperplane at each point on its boundary
20. Systems theory-focused Elsevier gathering across the pond, abbr.
21. Sailor
23. ___ We Can!, 2008 Obama-Biden slogan
24. Stand for a golfer
25. McBeal, et. al
27. Ali who stole from forty thieves
30. Gruffly succinct
32. Road usage fee
36. All hot and bothered, in a bad way
38. Baboon genus
39. With 40-across, the subject area of this thesis
40. See 39-across
42. Sandwich cookies
43. Generalizations of volume, such as those of Lebesgue or Borel
44. Western Cold War alliance
45. Town official in charge of making announcements
47. Trivial
48. 2016 Green Party candidate Jill
50. Private Twitter communiques
52. Some Audi roadsters
55. Dirac delta function
57. Act whose wages is death
60. A Modest Proposal, e.g.
62. A property promoted by ℓ_1 regularization (e.g., in the LASSO problem)
64. Ability
65. For profit diamond grading agcy
66. Salt Lake City collegiate athletes
67. Functions as desired despite uncertainty
68. Professional organization whose members design cars
69. Viral phenomenon

Down

1. Small eatery with caffeinated beverages
2. Oil cartel
3. 1/60,000 min.
4. Crash site?
5. Almond-flavored Italian liqueur
6. Delicate fabric
7. Maladroit
8. Place
9. Concept famously misunderstood by Alanis Morissette in a 1996 hit single
10. What you get when you multiply 0 by Inf in Matlab
11. Green emotion
12. Author James
13. Property of the green dinosaur in Toy Story
18. Abbr. for ancient dates
22. American pro. football assn.
24. Snapbacks and _____
26. Some sculptures
27. Philosopher Francis
28. Ancient Greek public square
29. French psychologist Alfred who co-created IQ testing
31. Unagi
33. Carmen or Don Giovanni
34. Fruit famously used by the British Navy to prevent scurvy
35. As a transmission line with non-negligible resistnace
37. ___ longa, vita brevis
38. Local frequency controller for a generator (abbr.)
40. Call, as a poker bet
41. Springtime household liquidation
43. Opposite of plusses
45. Type of shoes favored for maritime Mafia executions
46. Hamilton formerly of the Detroit Pistons
49. Components of the Michelin Man's body
51. French sea
52. Former Winter Palace resident
53. With "cat", it's a palindrome
54. Retained ticket portion
56. Org. of female drivers?
57. Location
58. Couple
59. Wall St. trading venue
61. Electronic sensing unit composed of accelerometers and gyroscopes
63. Result of addition

The solution to the crossword puzzle is on next page.

Solution


1 C 2 O 3 M 4 B ■ 5 A 6 L 7 A ■ 8 L 9 I 10 N 11 E 12 A 13 R
14 A P S E ■ 15 M A W ■ 16 O R A N G E
17 F E E D 18 B A C K ■ 19 C O N V E X
20 E C C ■ 21 C R E W 22 M A N ■ 23 Y E S
■ ■ ■ 24 T E E ■ 25 A L L Y 26 S ■ ■ ■
27 B 28 A 29 B A ■ 30 T 31 E R S E ■ 32 T 33 O 34 L 35 L
36 A G I T 37 A T E D ■ ■ 38 P A P I O
39 C O N T R O L ■ 40 S 41 Y S T E M S
42 O R E O S ■ ■ 43 M E A S U R E S
44 N A T O ■ 45 C 46 R I E R ■ 47 E A S Y
■ ■ ■ 48 S 49 T E I N ■ 50 D 51 M S ■ ■ ■
52 T 53 T 54 S ■ 55 I M P U 56 L S E ■ 57 S 58 I 59 N
60 S A T 61 I R E ■ 62 S P A R 63 S I T Y
64 A C U M E N ■ 65 E G L ■ 66 U T E S
67 R O B U S T ■ 68 S A E ■ 69 M E M E