Structure, Analysis, and Synthesis of First-Order Algorithms

Jared Miller[1], Carsten Scherer[1], Fabian Jakob[2], Andrea Iannelli[2]

March 25, 2026

Abstract

Optimization algorithms can be interpreted through the lens of dynamical systems as the interconnection of linear systems and a set of subgradient nonlinearities. This dynamical systems formulation allows for the analysis and synthesis of optimization algorithms by solving robust control problems. In this work, we use the celebrated internal model principle in control theory to structurally factorize convergent composite optimization algorithms into suitable network-dependent internal models and core subcontrollers. As the key benefit, we reveal that this permits us to synthesize optimization algorithms even if information is transmitted over networks featuring dynamical phenomena such as time delays, channel memory, or crosstalk. Design of these algorithms is achieved under bisection in the exponential convergence rate either through a nonconvex local search or by alternation of convex semidefinite programs. We demonstrate factorization of existing optimization algorithms and the automated synthesis of new optimization algorithms in the networked setting.

1 Introduction

Composite optimization problems minimize a sum of functions $(f_i)_{i=1}^s$, as in $\beta^* \in \operatorname{argmin}_\beta \sum_{i=1}^s f_i(\beta)$. These composite problems arise in applications such as regression [1], image denoising [2], matrix completion [3] and robust model predictive control [4]. These problems may also be solved over networks featuring dynamics such as time delays, channels with memory, or crosstalk [5, 6]. Such network dynamics could cause algorithms that are nominally convergent to no longer converge to an optimizer β^*.

First-order optimization algorithm are procedures that find an optima β^* of the composite optimization problem by using gradient/subdifferential evaluations of ∂f_i, without needing higher order information such as Hessians $\nabla^2 f_i$ [7]. First-order optimization algorithms can be understood through a system-theoretic lens as the interconnection between memoryless nonlinearities ∂f_i and a linear system [8, 9, 10]. An example of such a composite optimization algorithm is Davis-Yin Splitting [11] with $s = 3$, where f_2 is smooth (has Lipschitz gradients). Davis-Yin Splitting with scalar parameters (γ, λ) has an algorithmic law of

$$\omega_k = (I + \gamma\partial f_1)^{-1}(\zeta_k), \quad \zeta_{k+\frac{1}{2}} = 2\omega_k - \zeta_k - \gamma\partial f_2(\omega_k), \quad \zeta_{k+1} = \zeta_k + \lambda(I + \gamma\partial f_3)^{-1}(\zeta_{k+\frac{1}{2}}) - \lambda\omega_k. \quad (1)$$

J. Miller and C. Scherer are with the Chair of Mathematical Systems Theory, Department of Mathematics, University of Stuttgart, Stuttgart, Germany ({jared.miller, carsten.scherer}@imng.uni-stuttgart.de)

F. Jakob and A. Iannelli are with the Institute for Systems Theory and Automatic Control, University of Stuttgart, Germany (e-mail: {fabian.jakob, andrea.iannelli}@ist.uni-stuttgart.de).

J. Miller and C. Scherer are funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2075 – 390740016. J. Miller and C. Scherer acknowledge the support by the Stuttgart Center for Simulation Science (SimTech). F. Jakob acknowledges the support of the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

Application of Davis-Yin Splitting in (1) towards solving a composite optimization problem with variable $\beta \in \mathbb{R}^c$ can be expressed as the following interconnection of subgradient nonlinearities $(\partial f_1, \partial f_2, \partial f_3)$ and a discrete-time linear system with time index k, state x, input w, and output z:

$$\begin{pmatrix} x_{k+1} \\ \hline z_k^1 \\ z_k^1 \\ z_k^1 \end{pmatrix} = \left[\left(\begin{array}{c|ccc} 1 & -\gamma\lambda & -\gamma\lambda & -\gamma\lambda \\ \hline 1 & -\gamma & 0 & 0 \\ 1 & -\gamma & 0 & 0 \\ 1 & -2\gamma & -\gamma & -\gamma \end{array} \right) \otimes I_c \right] \begin{pmatrix} x_k \\ \hline w_k^1 \\ w_k^2 \\ w_k^3 \end{pmatrix}, \qquad \begin{pmatrix} w_k^1 \\ w_k^2 \\ w_k^3 \end{pmatrix} \in \begin{pmatrix} \partial f_1(z_k^1) \\ \partial f_2(z_k^2) \\ \partial f_3(z_k^3) \end{pmatrix}. \tag{2}$$

If the optimizer β^* is unique and the algorithm in (2) is convergent, then the steady-state output sequence satisfies $\lim_{k\to\infty}(z_k^1, z_k^2, z_k^3) = (\beta^*, \beta^*, \beta^*)$ independent of the initial state x_0.

The problem of optimization algorithm design in this setting can be posed as choosing a linear time-invariant system to connect to the subgradient nonlinearity. The performance of an optimization algorithm can be judged by its convergence rate to the optima β^* (e.g. exponential convergence). This design task is made more difficult when convergence to β^* must be ensured in the presence of network dynamics.

Convergent algorithms must satisfy the following two properties:

1. **Consensus:** the same point β^* must be sent to every operator ∂f_i $(\forall i: \ z_i = \beta^*)$.
2. **Optimality:** The point β^* is optimal in the sense of $0 \in \sum_{i=1}^s \partial f_i(z_i)$.

The Consensus requirement also emerges in distributed optimization [12], in which each of the separate communicating and computing agents must return the same value β^*. Obedience of the Consensus and Optimality constraints imposes structural requirements on the linear system associated with any convergent optimization algorithm [13, 14]. These structural requirements can be interpreted as regulation and disturbance rejection phenomena respectively [15, 16, 17, 18] in the sense of control theory. Linear systems that satisfy the Consensus and Optimality requirements when interconnected with a network satisfy an affine relation between their state-space matrices induced by a regulator equation. The search over linear systems when designing algorithms can therefore be restricted to linear systems that satisfy the relation, thus reducing the degrees of freedom in the design procedure.

We will show that regulation theory also gives insights into structural properties of optimization algorithms. Using an internal model principle [15, 17], we can factorize the linear system part of convergent optimization algorithms over networks into the cascade of a core subcontroller Σ_{core} containing the algorithm parameters, and an internal model $\Sigma_{\min}$ that depends only on the network. Because the Davis-Yin algorithm can directly interface the subgradients $(\partial f_1, \partial f_2, \partial f_3)$ without encountering network dynamics, this particular internal model $\Sigma_{\min}$ is an integrator that depends only on the dimensions s and c. The internal-model-based cascade factorization applied to the $s = 3$ Davis-Yin algorithm in (2) is

$$\Sigma_{\text{core}}: \qquad \begin{pmatrix} \tilde{u}_k^{1,1} \\ \hdashline \tilde{u}_k^{2,1} \\ \tilde{u}_k^{2,2} \\ \tilde{u}_k^{2,3} \end{pmatrix} = \left[\left(\begin{array}{ccc} -\gamma\lambda & -\gamma\lambda & -\gamma\lambda \\ \hdashline -\gamma & 0 & 0 \\ -\gamma & 0 & 0 \\ -2\gamma & -\gamma & -\gamma \end{array} \right) \otimes I_c \right] \begin{pmatrix} w_k^1 \\ w_k^1 \\ w_k^3 \end{pmatrix}, \tag{3a}$$

$$\Sigma_{\min}: \qquad \begin{pmatrix} x_{k+1} \\ \hline z_k^1 \\ z_k^2 \\ z_k^3 \end{pmatrix} = \left[\left(\begin{array}{c|c:ccc} 1 & 1 & 0 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{array} \right) \otimes I_c \right] \begin{pmatrix} x_k \\ \hline \tilde{u}_k^{1,1} \\ \hdashline \tilde{u}_k^{2,1} \\ \tilde{u}_k^{2,2} \\ \tilde{u}_k^{2,3} \end{pmatrix}. \tag{3b}$$

The internal model factorization allows us to synthesize optimization algorithms over networks by solving structured robust control problems. These robust control problems arise by abstracting the subgradient nonlinearity into an uncertainty, and describing the uncertainty using the frameworks of Passivity and Integral Quadratic Constraints (IQCs) [19]. This principled search over Σ_{core} also certifies well-posedness and exponential convergence rates of the resultant algorithm.

1.1 Literature Review

The application of regulation theory and the internal model principle towards understanding optimization algorithms is a continuation of existing unifications between control theoretic tools and optimization (systems theory of algorithms [10, 20, 21]). Early use of dynamical systems towards the analysis and solution of optimization algorithms include Lyapunov interpretations of algorithmic convergence [22, Section 10.4], provision of stability guarantees in numerical integration [23], and solution of sorting and linear programming [24]. The work in [25] provides a systems-theoretic perspective on linear and nonlinear rootfinding algorithms. Dynamical-systems tools that have been used for the analysis and synthesis of optimization algorithms include gradient flows [26, 27, 28, 29], Integral Quadratic Constraints (IQC) [30, 19], dissipativity [31, 32], and passivity [33]. The internal model principle was used in [34, 35] to perform perfect tracking in online optimization. We note that Scaled Relative Graphs [36], originally used to analyze convergence of optimization algorithms, have made the leap to dynamical systems for the analysis and control of nonlinear systems [37]. Optimization algorithms can be deployed as controllers for cyberphysical systems [38], such as in model predictive control [39, 4].

The first application of IQCs for the analysis of optimization algorithms occurred in [40] to analyze the asymptotic exponential convergence rate of optimization algorithms. Properties of the oracle such as (strong) monotonicity, cocoercivity, and Lipschitzness can be modeled as pointwise (static) IQC constraints. When the oracle is the subdifferential of a convex function, the subgradient inequality induces an infinite set of dynamic O'Shea-Zames-Falb multipliers [41, 42], each of which expresses a dissipation relation that is valid for all compatible oracles. Algorithm convergence can be tested through the solution of semidefinite programming problems, involving a convex search over coefficients that parameterize the multipliers. The convergence rate of the algorithm can be found through bisection by finding a minimal rate (up to numerical tolerance) at which the stability test fails to certify convergence. IQC theory was used to derive the Triple Momentum scheme [43], which was proven to be optimal among all first-order and fixed-stepsize algorithms for strongly convex optimization in [44]. The optimality of Triple Momentum was numerically certified using IQCs in [45]. Conservatism in this IQC-based scheme arises through the finite order of the filters used: increasing the filter order raises computational complexity but may decrease the certified convergence rate. While determining conditions for which the true algorithm convergence rate is returned as the filter order increases remains an open problem, the IQC-based approach can (numerically) certify the optimality of algorithms such as gradient descent and triple momentum at order-1 filters. Extensions of IQC analysis beyond static optimization algorithms include smooth games [46], parameter-varying optimization [47], and online optimization [48]. IQC and dissipativity tools were used in [49] to determine conditions under which individual optimization algorithms could be tied together into a distributed optimization algorithm under a fixed Laplacian interconnection.

Synthesis of algorithms is a challenging task, and this difficulty is increased when composite optimization problem must be solved over a network. The IQC-analysis method in [40, 50] designed optimization algorithms based on a hyperparameter search: choose algorithm parameters such that the IQC-derived programs yield a minimal convergence rate. By modeling the algorithm design task as a robust control problem, classical control methods such as H_∞ optimization can be leveraged to generate controllers [51]. The work in [52, 53, 54] use frequency-domain methods to synthesize optimization algorithms in the absence of network dynamics. The Consensus and Optimality properties are interpreted as tangental interpolation constraints of transfer function matrices [55]. Nevalinna-Pick interpolation theory [56, 57] is then used to generate algorithms that satisfy the Consensus and Optimality constraints. This frequency-domain framework is extended to equality-constrained algorithms in [58] and distributed optimization algorithms in [53].

In the case of a single oracle (e.g. unconstrained optimization), the single-input-single-output structure of the controller leads to a possible commutation between the network and the filters. This allows for convex synthesis of the algorithm and the filter coefficients, even in the networked setting [59]. The commutation programs were first presented using the Youla-Kucera-Jabr [60, 61] parameterization of all stabilizing controllers (algorithms) in [62], and later reformulated with direct state-space arguments in [63]. Algorithm design has been conducted for classes of switched systems (e.g. time-varying delays in networks) in [64] through an alternating search between switch-mode-dependent controllers and filter coefficients. The work in [59, 62, 63, 64] perform synthesis only in the case of a single oracle. The work in [65] uses Lyapunov arguments and general sector conditions to generate optimization algorithms for fixed multipliers in the convex and saddle-point case. The saddle point case considered in [65] includes a class of equality-constrained strongly convex problems, which can be treated as a two-oracle problem.

The Performance Estimation Problem (PEP) approach is an alternative method for the analysis and synthesis of optimization algorithms [66]. Given a set of variables corresponding to a finite number of points, scalars, and vectors, the PEP approach uses interpolation constraints to describe conditions for which a convex function exists with values at the points equal to the scalars and gradients equal to the vectors. Queries such as the gradient error, residual, or function decrease can be exactly answered in a finite-horizon case by solving semidefinite programming problems [67]. The PEP framework extends to establishing contraction rates for monotone inclusion schemes [68]. Optimal variable-step and fixed-step rules for unconstrained gradient descent of smooth convex functions were synthesized using these interpolation constraints in [44]. The work in [69] generates optimal step sizes through the branch-and-bound-based global optimization of nonconvex quadratic programs. Interpolation constraints can be used in asymptotic analysis of optimization algorithms: the works in [70, 71, 14] use Lyapunov arguments to certify convergence rates in which the uncertainty is described by interpolation constraints.

1.2 Contributions

We first provide a condition for convergence of a well-posed optimization algorithm over a dynamical network: the algorithmic interconnection must satisfy a Regulator Equation and a Robust Stability property. The Regulator Equation may be verified by solving a linear system of equations. Certification of the Robust Stability property is a challenging robust control problem involving system dynamics, and will be approximated using a sequence of convex programs.

We then prove a state-space internal model principle [15] for optimization algorithms, ensuring that any convergent networked optimization algorithm can be split into a core subcontroller Σ_{core} and a network-dependent internal model $\Sigma_{\min}$ as in (3). Design of the overall algorithm involves searching over Σ_{core} satisfying a structural constraint induced by the regulator equation.

We analyze existing algorithms and synthesize new algorithms by solving linear matrix inequalities (LMIs). Analysis and synthesis both involve bisection over the exponential convergence rate. At each fixed convergence rate, Analysis involves searching over Zames-Falb filter coefficients that define valid relations satisfied by subgradient nonlinearities. We propose two paths for algorithm synthesis to find Σ_{core}

at a fixed convergence rate. The first path involves a nonconvex search over reduced-order controllers and filter coefficients, leading to low-order and highly structured controllers. The second path incorporates a full-order internal model into system dynamics, and performs an alternating convex search over valid filter coefficients parameterizing the IQC/passivity specification. In each of these paths, feasibility of the resulting robust control problem ensures convergence of the optimization algorithm to the optimal point.

The specific contributions of this work are

- A two-part condition for convergence of optimization algorithms (Robust Stability, Regulator Equation).
- A structural state-space factorization of convergent first-order optimization algorithms based on the internal model principle with examples.
- A synthesis framework for generating optimization algorithms over networks based on internal models and linear matrix inequalities.
- Demonstration of algorithm synthesis in the dynamical networked setting.

Table 1 lists the main theoretical contributions of the paper.

Table 1: Main results of the paper

Lemma 2.1	Sufficient condition for well-posedness of optimization algorithms
Theorem 3.1	Two-part condition for convergence of algorithms
Theorem 4.1	Controller order bound based on well-posedness and information constraints
Theorem 4.2	Factorization of optimization algorithms using the Internal Model Principle
Theorem 5.1	LMI-based exponential convergence rate analysis of well-posed algorithms
Proposition 5.3	LMI-based synthesis of certified well-posed exponentially converging algorithms

This paper has the following structure. Section 2 reviews preliminaries of notation, linear systems, convex analysis, and optimization algorithms. Section 3 provides a condition for the convergence of optimization algorithms using robust stability and a regulator equation. Section 4 identifies structural constraints on optimization algorithms and performs algorithm factorizations using an internal model principle. Section 5 performs automated synthesis of optimization algorithms based on solving nonconvex or alternating-convex programs arising from robust control. Section 6 demonstrates our approach on example networked optimization tasks. Section 7 concludes the paper.

2 Preliminaries

We first review preliminaries of notation, linear systems, and optimization algorithms.

2.1 Notation and Linear Algebra

The set of natural numbers including zero is $\mathbb{N} = \{0, 1, 2, \ldots\}$. The set of integers between a and b inclusive is $\{a, \ldots, b\}$. The set of complex numbers, real numbers, and integers are respectively $\mathbb{C}, \mathbb{R}$, and $\mathbb{Z}$. The extended real line is $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The n-dimensional real Euclidean space is $\mathbb{R}^n$. The n-dimensional positive orthant is $\mathbb{R}^n_> := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i > 0 \; \forall i \in \{1, \ldots, n\}\}$.

The set of $n \times m$ matrices is $\mathbb{R}^{n\times m}$. The set of n-dimensional symmetric matrices is $\mathbb{S}^n$ with $\mathbb{S}^n \subset \mathbb{R}^{n\times n}$. An empty matrix will be denoted as $[\cdot]$ and an empty set as $\varnothing$.

The transpose of a matrix $M \in \mathbb{R}^{m\times n}$ is $M^\top$. The rank of M is $\mathrm{rank}(M)$. The null space of M is $\mathrm{null}(M)$, and its range (column) space is $\mathrm{ran}(M)$. If M is square ($m = n$), the symmetric part of $M \in \mathbb{R}^{n\times n}$ is $\mathrm{Sym}(M) = \frac{1}{2}(M + M^\top)$ and its inverse is M^{-1}. The spectral radius of a square matrix M is the maximal absolute value of any eigenvalue of M, and is denoted as $\rho(M)$.

M is Schur if $\rho(M) < 1$. M is anti-Schur if all its eigenvalues $\lambda \in \mathbb{C}$ satisfy $|\lambda| \geq 1$. If M is symmetric, its minimal and maximal eigenvalues are $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$, respectively.

The all zeros matrix is $0_{m\times n}$, the all ones vector is $\mathbf{1}_s$, and the identity matrix is I_s. Subscripts will be omitted if the dimensions are clear from context.

Given a vector v, the square matrix constructed with v along its main diagonal and zeros everywhere else is $\mathrm{diag}(v)$. The block diagonal concatenation of a set of matrices is $\mathrm{blkdiag}(\cdot, \ldots, \cdot)$.

2.2 Linear Systems

A discrete time signal $x : \mathbb{N} \to \mathbb{R}^n$ is a time-indexed sequence $(x_k)_{k\in\mathbb{N}}$. The unit shift operator $\mathbf{z}$ acts on a sequence x as $\mathbf{z}x = \mathbf{z}(x_k)_{k\in\mathbb{N}} = (x_{k+1})_{k\in\mathbb{N}}$ and $(\mathbf{z}x)_0 = 0$. A linear system G is a mapping from an input signal $u : \mathbb{N} \to \mathbb{R}^{n_u}$ and an initial state $x_0 \in \mathbb{R}^n$ to an output signal $y : \mathbb{N} \to \mathbb{R}^{n_y}$ as $G : (x_0, u) \to y$. In this paper, linear systems G are described with matrices $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ in the state-space as

$$G : \begin{pmatrix} x_{k+1} \\ \hline y_k \end{pmatrix} = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \begin{pmatrix} x_k \\ \hline u_k \end{pmatrix} \qquad \text{with an abbreviation of} \qquad y = \left[\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right] u. \tag{4}$$

The transfer matrix of the system G is the rational function $T(\mathbf{z}) = \mathcal{C}(\mathbf{z}I - \mathcal{A})^{-1}\mathcal{B} + \mathcal{D}$ for $\mathbf{z} \in \mathbb{C}$. If a given rational function $T(\mathbf{z})$ is represented as $T(\mathbf{z}) = \mathcal{C}'(sI - \mathcal{A}')^{-1}\mathcal{B}' + \mathcal{D}'$, the tuple $(\mathcal{A}', \mathcal{B}', \mathcal{C}', \mathcal{D}')$ is a realization of $T(\mathbf{z})$. Realizations are generally nonunique.

For linear systems $G_1 : (x_0^1, u) \to y$ and $G_2 : (x_0^2, y) \to z$ with respective representations $(\mathcal{A}_1, \mathcal{B}_1, \mathcal{C}_1, \mathcal{D}_1)$, $(\mathcal{A}_2, \mathcal{B}_2, \mathcal{C}_2, \mathcal{D}_2)$, the cascade interconnection of G_1 and G_2 is $G_2G_1 : ((x_0^1, x_0^2), u) \to z,\ z = G_2(G_1(u))$. The block-diagonal system formed by $G_1 : (x_0^1, u) \to y$ and $G_2 : (x_0^2, y) \to z$ with $y_1 = G_1(x_0^1, u_1),\ y_2 = G_2(x_0^2, u_2)$ is $\mathrm{blkdiag}(G_1, G_2) : ((x_0^1, x_0^2), (u_1, u_2)) \to (y_1, y_2)$. The well-posed feedback interconnection between systems $G_1 : (x_0^1, (w, u)) \to (z, y)$ and $G_2 : (x_0^2, (y, a)) \to (u, b)$ along their shared signals (u, y) is the *star product* $G_1 \star G_2 : ((x_0^1, x_0^2), (w, a)) \to (z, b)$. Well-posedness ensures that the shared signals (u, y) are unique given (x^1, x^2, w, a), and this is true if $I - \mathcal{D}_2\mathcal{D}_1$ is invertible. We refer to Appendix A for state-space formulas of these interconnections based on representations of G_1 and G_2.

The representation $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ of G is stable if $\mathcal{A}$ is Schur. This holds iff all trajectories of (4) for $u = 0$ and any $x_0 \in \mathbb{R}^n$ satisfy $\lim_{k\to\infty} x_k = 0$. The representation $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ of G is controllable (stabilizable) or observable (detectable) if

$$\mathrm{ran}\begin{pmatrix} \mathcal{A} - \lambda I & \mathcal{B} \end{pmatrix} = n \ \text{ or } \ \mathrm{null}\begin{pmatrix} \mathcal{A} - \lambda I \\ \mathcal{C} \end{pmatrix} = \{0\} \tag{5}$$

holds for all $\lambda \in \mathbb{C}$ ($\lambda \in \mathbb{C}$ with $|\lambda| \geq 1$), respectively, This will also be expressed by saying that $(\mathcal{A}, \mathcal{B})$ is controllable (stabilizable) or $(\mathcal{A}, \mathcal{C})$ is observable (detectable).

The order of a representation $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ is the dimension of $\mathcal{A}$ (which equals the number of states). The order of the transfer matrix of a system (called McMillan degree) is the minimum number of states required for a state-space realization thereof. A realization $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ is minimal iff it is controllable and observable.

Let us finally recall that a controlled interconnection is described by

$$P : \quad \begin{pmatrix} x_{k+1} \\ \hline z_k \\ y_k \end{pmatrix} = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \end{array}\right) \begin{pmatrix} x_k \\ \hline w_k \\ u_k \end{pmatrix}, \qquad K : \quad \begin{pmatrix} \xi_{k+1} \\ \hline u_k \end{pmatrix} = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \xi_k \\ \hline y_k \end{pmatrix}. \tag{6}$$

Here P is the plant and K is the controller. They share the control input u and the measurement output y. This special star product is well-posed if $E := I - D_2 D_c$ is invertible. Well-posedness permits elimination the common signals (u, y), to arrive at the description

$$\begin{pmatrix} x_{k+1} \\ \xi_k \\ \hline z_k \end{pmatrix} = \underbrace{\left[\left(\begin{array}{cc|c} A & 0 & B_1 \\ 0 & 0 & 0 \\ \hline C_1 & 0 & D_1 \end{array} \right) + \begin{pmatrix} 0 & B_2 \\ I & 0 \\ \hline 0 & D_{12} \end{pmatrix} \begin{pmatrix} A_c + B_c E^{-1} D_2 C_c & B_c E^{-1} \\ (I + D_c E^{-1} D_2) C_c & D_c E^{-1} \end{pmatrix} \left(\begin{array}{cc|c} 0 & I & 0 \\ C_2 & 0 & D_{21} \end{array} \right) \right]}_{\left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array} \right)} \begin{pmatrix} x_k \\ \xi_k \\ \hline w_k \end{pmatrix}. \quad (7)$$

We refer to the interconnection in (6) as $P \star K$, with the tacit understanding that it is well-posed and described with the matrices in (7). The eliminated signals (y, u) can be recovered as

$$\begin{pmatrix} y_k \\ u_k \end{pmatrix} = \begin{pmatrix} I & -D_2 \\ -D_c & I \end{pmatrix}^{-1} \begin{pmatrix} C_2 x_k + D_{21} w_k \\ C_c \xi_k \end{pmatrix}. \quad (8)$$

Moreover, K internally stabilizes P if $P \star K$ is well-posed and if $\mathcal{A}$ is Schur stable. There exists a controller K which internally stabilizes P iff (A, B_2) is stabilizable and (A, C_2) is detectable.

2.3 Convex Analysis

We review and define classes of functions, subdifferentials, and operators that are used to define optimization problems and optimization algorithms.

2.3.1 Function Classes

A function $f : \mathbb{R}^c \to \bar{\mathbb{R}}$ is proper if there exists a point $x \in \mathbb{R}^c$ with $f(x) < \infty$, and if $f(x) > -\infty$ for all $x \in \mathbb{R}^c$ [72, Page 5]. The domain of a proper function f is the set $\mathrm{dom}(f) = \{x \in \mathbb{R}^c \mid f(x) < \infty\}$ and f is *globally defined* if $\mathrm{dom}(f) = \mathbb{R}^c$. A proper f is convex if $\forall (x_1, x_2) \in \mathbb{R}^c \times \mathbb{R}^c, \alpha \in [0, 1] : f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$, and it is *closed* if $\{x \in \mathbb{R}^c \mid f(x) \leq \alpha\}$ is closed for each $\alpha \in \mathbb{R}$. A function that is proper, convex, and closed will be referred to as being p.c.c.

A function $f : \mathbb{R}^c \to \bar{\mathbb{R}}$ is m-strongly convex with $m > 0$ if $f - \frac{m}{2}\|\cdot\|_2^2$ is convex. f is L-smooth with $L < \infty$ if it is differentiable at every point $x \in \mathbb{R}^c$ and if $\forall x, x' \in \mathbb{R}^c : \|\nabla f(x) - \nabla f(x')\| \leq L\|x' - x\|_2$. Define $\mathbf{q}(x) = \frac{1}{2}\|x\|_2^2$ for $x \in \mathbb{R}^c$. Given $m \in \mathbb{R}$, the set $\mathcal{S}_{m,\infty}$ comprises all functions $f : \mathbb{R}^c \to \bar{\mathbb{R}}$ such that $f - m\mathbf{q}$ is p.c.c. Moreover, $\mathcal{S}_{m,L}$ for $L \in \mathbb{R}$ consists of all $f \in \mathcal{S}_{m,\infty}$ such that $L\mathbf{q} - f$ is p.c.c.. Nonemptiness of $\mathcal{S}_{m,L}$ requires that $m \leq L$, and the only members of $\mathcal{S}_{m,m}$ are the functions $f = m\mathbf{q} + g$ with affine $g : \mathbb{R}^c \to \mathbb{R}$. If referring to the class $\mathcal{S}_{m,L}$, we will tacitly assume that $-\infty < m < L \leq \infty$ from now on. For $L = \infty$, this is an extension of existing definitions in [73, 74] to the case where f is extended real-valued, and to [75, 65] when f is nonsmooth.

For $f \in \mathcal{S}_{m,L}$, $\mathrm{dom}(f) \neq \varnothing$ is convex, since it equals $\mathrm{dom}(f - m\mathbf{q})$. Note that $\mathcal{S}_{0,\infty}$ is the set of p.c.c. functions, and any $f \in \mathcal{S}_{m,\infty}$ for $m > 0$ is m-strongly convex [76]. If $L < \infty$, all $f \in \mathcal{S}_{m,L}$ are globally defined (since $f - m\mathbf{q}$ and $L\mathbf{q} - f$ are proper) and continuously differentiable [63, Lemma 3]; if $m > 0$ then f is L-smooth and m-strongly-convex.

The indicator function of a set $\mathcal{Z} \in \mathbb{R}^c$ is defined as $I_{\mathcal{Z}}(x) = 0$ if $x \in \mathcal{Z}$ and $I_{\mathcal{Z}}(x) = \infty$ if $x \notin \mathcal{Z}$. If $\mathcal{Z} \neq \varnothing$ is closed and convex, then $I_{\mathcal{Z}} \in \mathcal{S}_{0,\infty}$. The affine hull of a set $\mathcal{Z}$ is the smallest affine space containing $\mathcal{Z}$ (intersection of all affine spaces including $\mathcal{Z}$). Its relative interior is the interior of $\mathcal{Z}$ viewed as a subset of its affine hull [72, Section 2.H].

Figure 1 visualizes the contours of four example functions inside $\mathcal{S}_{m,L}$ within the domain $[-3, 3]^2$.

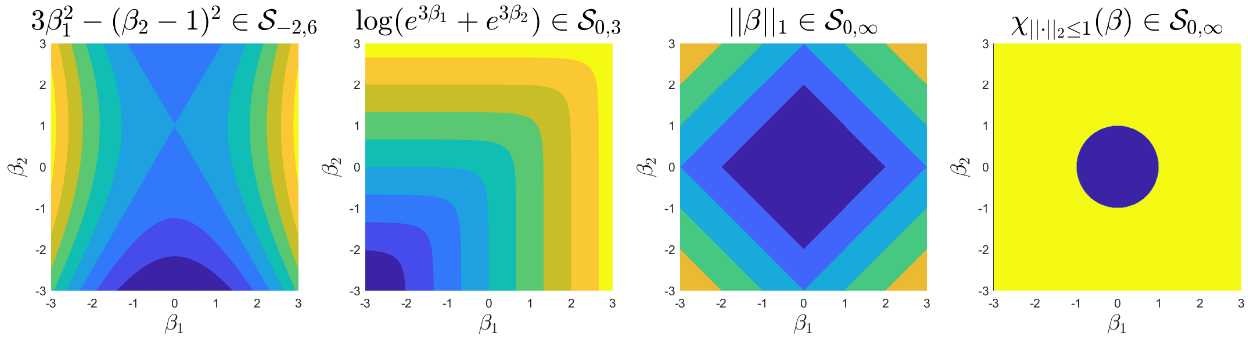


Figure 1: Contours plotted of functions in $\mathcal{S}_{m,L}$ with $c = 2$.

2.3.2 Subdifferentials

The subdifferential of a p.c.c. function f at $x \in \mathbb{R}^c$ is defined as

$$\partial f(x) = \{g \in \mathbb{R}^c \mid f(x') \geq f(x) + g^\top (x' - x) \ \forall x' \in \mathbb{R}^c\}. \tag{9}$$

The Frechét subdifferential of a function $f \in \mathcal{S}_{m,L}$ is $\partial f := \partial(f - m\mathbf{q}) + mI$, where $\partial(f - m\mathbf{q})$ is the standard subdifferential of the p.c.c.function $f - m\mathbf{q}$ ([77, Definition 6.2, (iii)] using [78, Proposition 1.107 (i)]). $\mathcal{S}^0_{m,L}$ is the subset of functions $f \in \mathcal{S}_{m,L}$ which are zero-centered as $0 = f(0)$ and $0 \in \partial f(0)$.

As examples, $\partial I_{\mathcal{Z}}$ is the normal cone operator for $\mathcal{Z}$, and $\partial f(x) = \{\nabla f(x)\}$ holds for all $x \in \mathbb{R}^c$ if $f \in \mathcal{S}_{m,L}$ and $L < \infty$. The identity function $I : \mathbb{R}^c \to \mathbb{R}^c$, $I(x) = x$, satisfies $I = \partial \mathbf{q}$.

Figure 2 plots an example of the subdifferential of functions in $\mathcal{S}_{m,L}$ for the m-parameterized function

$$g_m(\beta) := \max\left[\tfrac{1}{2}|\beta - \tfrac{1}{2}|,\ 3|\beta - \tfrac{1}{2}| - 4\right] + \tfrac{m}{2}\|\beta + 1.1\|_2^2, \tag{10}$$

in which $g_m \in \mathcal{S}_{m,\infty}$ for all $m \in \mathbb{R}$. The blue (top) and orange (bottom) curves draw the convex $g_0(\beta)$ and the nonconvex $g_{-1}(\beta)$, respectively. The black arrows visualize vectors in the respective multivalued subdifferentials $\partial g_m(\beta)$ at $\beta = 2.1$.

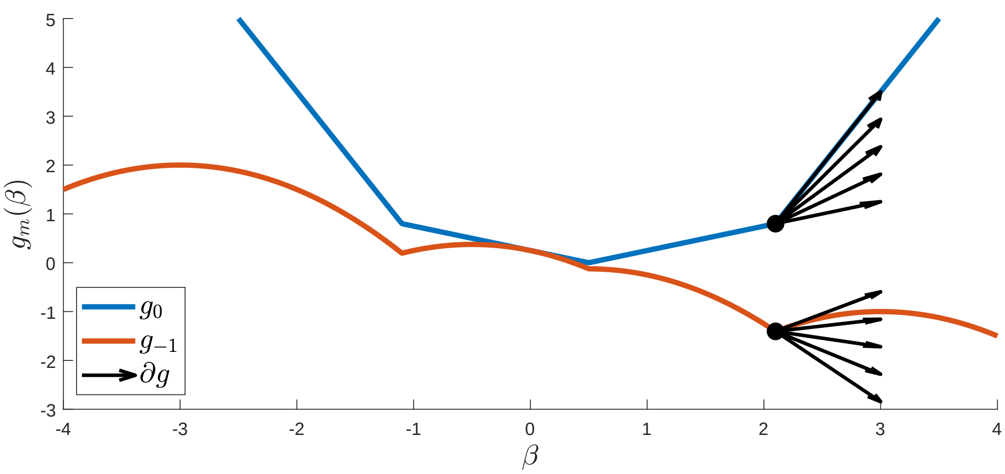


Figure 2: Functions g_m in (10) and vectors inside $\partial g_m(2.1)$ for $m \in \{-1, 0\}$.

2.3.3 Operators

An operator is a set-valued map $F : \mathbb{R}^c \rightrightarrows \mathbb{R}^c$ [72, 76]. Its inverse is defined as $F^{-1}(x) := \{y \in \mathbb{R}^c \mid x \in F(y)\}$ for $x \in \mathbb{R}^c$. A matrix $D \in \mathbb{R}^{c\times c}$ is identified with the linear map $x \mapsto Dx$ for any $x \in \mathbb{R}^c$. In this paper, $(I + DF)^{-1}$ is referred to as a generalized resolvent, and $(F^{-1} - D)^{-1}$ is referred to as the generalized Yosida operator. If $D = -\lambda I$ with $\lambda > 0$, a generalized Yosida operator is related to the standard resolvent $(I + \lambda F)^{-1}$ through the inverse resolvent identity [72, Lemma 12.14] $(F^{-1} + \lambda I)^{-1} = \lambda^{-1}[I - (I + \lambda F)^{-1}]$. F is said to be globally single-valued if, for every $x \in \mathbb{R}^c$, $F(x)$ consists of exactly one element, which is also denoted as $F(x)$. F is said to be globally continuous if F is globally single-valued and if the corresponding map $F : \mathbb{R}^c \to \mathbb{R}^c$, $x \mapsto F(x)$ is continuous.

If $f \in \mathcal{S}_{m,L}$ then $\partial f : \mathbb{R}^c \rightrightarrows \mathbb{R}^c$ is a set-valued map. If $m > 0$, then f is strongly convex, which implies that $(\partial f)^{-1}$ is globally continuous. For $f \in \mathcal{S}_{0,\infty}$, we also recall that $(I + \gamma\partial f)^{-1}$ is the proximal mapping $z \mapsto \operatorname{argmin}_{z\in\mathbb{R}^c} f(z) + \frac{1}{2\gamma}\|x - z\|_2^2$. The set of subdifferentials to $\mathcal{S}_{m,L}$ is $\partial\mathcal{S}_{m,L} = \{\partial f \mid f \in \mathcal{S}_{m,L}\}$. We refer to Appendix B for further details.

2.4 Composite Optimization Problems

Given $s \in \mathbb{N}$ functions $f_i \in \mathcal{S}_{m_i,L_i}$ for $i \in \{1, \dots, s\}$, a composite optimization problem amounts to finding a point $\beta^* \in \mathbb{R}^c$ with

$$\beta^* \in \operatorname*{argmin}_{\beta\in\mathbb{R}^c} \sum_{i=1}^{s} f_i(\beta). \tag{11}$$

An example of composite optimization arises from constrained minimization of a function $f \in \mathcal{S}_{m_f,L_f}$ with $L_f < \infty$ on a closed convex set $\mathcal{Z} \subset \mathbb{R}^c$. If we choose $f_1 = f$ and $f_2 = I_{\mathcal{Z}} \in \mathcal{S}_{0,\infty}$, we infer that $f_1(x) + f_2(x) = f(x) < \infty$ holds iff $x \in \mathcal{Z}$, and hence (11) is equivalent to $\beta^* \in \operatorname{argmin}_{x\in\mathcal{Z}} f(x)$.

This paper studies algorithms solving the following zero inclusion problem.

Problem 1 (Composite Optimization)**.** *Given $s \in \mathbb{N}$ functions $f_i \in \mathcal{S}_{m_i,L_i}$ for $i \in \{1, \dots, s\}$, find a point $\beta^* \in \mathbb{R}^c$ which satisfies*

$$0 \in \sum_{i=1}^{s} \partial f_i(\beta^*). \tag{12}$$

If $m_1 + \cdots + m_s > 0$, Problem 1 has a unique optimal solution since the cost $\sum_{i=1}^{s} f_i$ is proper, closed and strongly convex. By Corollary B.1, any solution β^* satisfying (12) is a solution of (11). If the domains of the functions are compatible according to

$$\operatorname{relint}(\operatorname{dom}(f_1)) \cap \cdots \cap \operatorname{relint}(\operatorname{dom}(f_s)) \neq \varnothing, \tag{13}$$

then a solution β^* of (11) also satisfies (12). Again by Corollary B.1 and without constraints on $m_1, \dots, m_s$, any local solution of (11) also solves Problem 1. For these reasons, we refer to (12) as an optimality condition for the composite optimization problem (11).

With the function $f(z) := f_1(z^1) + \cdots + f_s(z^s)$ for $z = (z^1, \dots, z^s) \in \mathbb{R}^{sc}$, the optimality condition (12) reads as $0 \in (\mathbf{1}_s \otimes I_c)^\top F(\mathbf{1}_s \otimes \beta^*)$ involving the operator $F := \partial f$. Observe that F can be considered as a concatenation of the subdifferentials $\partial f_1, \dots, \partial f_s$, since

$$F(z) = \begin{pmatrix} F_1(z^1) \\ \vdots \\ F_s(z^s) \end{pmatrix} = \begin{pmatrix} \partial f_1(z^1) \\ \vdots \\ \partial f_s(z^s) \end{pmatrix} := \left\{ \begin{pmatrix} g_1 \\ \vdots \\ g_s \end{pmatrix} \mid g_i \in \partial f_i(z^i),\ i \in \{1, \dots, s\} \right\}. \tag{14}$$

In this paper, we work with the following classes of operators.

Definition 1. *Given vectors* $m = (m_1, \ldots, m_s)$ *and* $L = (L_1, \ldots, L_s)$*, let* $\mathcal{O}_{m,L}$ *be the class of all operators* F *in* (14) *for which* (12) *has a unique solution* $\beta^* \in \mathbb{R}^c$ *and there exists a unique* $w^* \in F(\mathbf{1}_s \otimes \beta^*)$ *with* $(\mathbf{1}_s \otimes I_c)^\top w_* = 0$*. Moreover,* $\mathcal{O}^0_{m,L}$ *is the subset of all operators* $F \in \mathcal{O}_{m,L}$ *satisfying* $0 \in \partial f_i(0)$ *and* $f_i(0) = 0$ *for* $i \in \{1, \ldots, s\}$*, implying that* $0 \in F(0)$*.*

2.5 Optimization Algorithms

Time-independent iterative schemes to solve Problem 1 can be expressed as a linear time-invariant system G represented by $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ interconnected with the static map $F \in \mathcal{O}_{m,L}$ as

$$\left(\begin{array}{c} x_{k+1} \\ \hline z_k \end{array}\right) = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \left(\begin{array}{c} x_k \\ \hline w_k \end{array}\right), \qquad w_k \in F(z_k). \tag{15}$$

This feedback interconnection is depicted as a block-diagram shown in Figure 3.

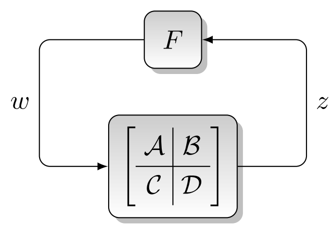


Figure 3: Standard algorithmic interconnection in (15)

Examples of optimization algorithms are Gradient Descent $x_{k+1} = x_k - \alpha \nabla f(x_k)$ represented as

$$\left(\begin{array}{c} x_{k+1} \\ \hline z_k \end{array}\right) = \left[\left(\begin{array}{c|c} 1 & -\alpha \\ \hline 1 & 0 \end{array}\right) \otimes I_c\right] \left(\begin{array}{c} x_k \\ \hline w_k \end{array}\right), \qquad w_k = \nabla f(z_k), \tag{16}$$

and the Davis-Yin algorithm [11] in (2).

The following definition of a fixed point of an algorithm is standard.

Definition 2. *A* ***fixed point*** *of* (15) *is a tuple* (x^*, z^*, w^*) *satisfying*

$$\left(\begin{array}{c} x^* \\ \hline z^* \end{array}\right) = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \left(\begin{array}{c} x^* \\ \hline w^* \end{array}\right), \qquad w^* \in F(z^*). \tag{17}$$

This leads us to the following notion of algorithm convergence used in this paper.

Definition 3. *The algorithmic interconnection in* (15) *with* $F \in \mathcal{O}_{m,L}$ *is a* ***convergent optimization algorithm*** *for Problem 1 with solution* β^* *if there exists a fixed point* (x^*, w^*, z^*) *of* (15) *which satisfies the following three properties:*

$$\textit{Consensus:} \qquad z^* = \mathbf{1}_s \otimes \beta^*, \tag{18a}$$

$$\textit{Optimality:} \qquad (\mathbf{1}_s \otimes I_c)^\top w^* = 0, \tag{18b}$$

$$\textit{Attractivity:} \qquad \forall x_0 \in \mathbb{R}^n : \lim_{k \to \infty} (x_k, w_k, z_k) = (x^*, w^*, z^*). \tag{18c}$$

Remark 2.1. *Our class of considered problems in the definition of $\mathcal{O}_{m,L}$ and the attractivity property in* (18c) *requires that there is a unique tuple (β^*, w^*) with $w^* \in F(\mathbf{1}_s \otimes \beta^*)$, $(\mathbf{1}_s^\top \otimes I_c)w^* = 0$. A sufficient condition for w^* to be unique given the existence of a unique β^* solving Problem 1 is if at most one index $i \in \{1, \ldots, s\}$ satisfies $L_i = \infty$. If $(\mathcal{A}, \mathcal{C})$ is detectable, then uniqueness of x^* follows, because there is at most one x^* such that*

$$\begin{pmatrix} \mathcal{A} - I \\ \mathcal{C} \end{pmatrix} x^* = \begin{pmatrix} -\mathcal{B}w^* \\ \mathbf{1}_s \otimes \beta^* - \mathcal{D}w^* \end{pmatrix}. \tag{19}$$

2.6 Well-Posedness and Algorithm Execution

The algorithm (15) involves the implicit constraint $z_k = \mathcal{C}x_k + \mathcal{D}w_k$, $w_k \in F(z_k)$. This translates into the equivalent inclusions $\mathcal{C}x_k + \mathcal{D}w_k \in F^{-1}(w_k)$ and $w_k \in (F^{-1} - \mathcal{D})^{-1}(\mathcal{C}x_k)$, which motivates the following definition of well-posedness.

Definition 4. *The algorithm* (15) *is* ***well-posed*** *if the generalized Yosida operator $H := (F^{-1} - \mathcal{D})^{-1}$ is globally continuous.*

If (15) is well-posed, it can be rewritten as an explicit recursion

$$x_{k+1} = \mathcal{A}x_k + \mathcal{B}H(\mathcal{C}x_k) \;\text{ with outputs }\; w_k = H(\mathcal{C}x_k), \;\; z_k = \mathcal{C}x_k + \mathcal{D}H(\mathcal{C}x_k) \;\text{ for all }\; k \in \mathbb{N}. \tag{20}$$

Well-posedness implies that trajectories $(x_k, w_k, z_k)_{k\in\mathbb{N}}$ exist and are unique for every $x_0 \in \mathbb{R}^n$. In the sequel, we refer to the algorithm as $F \star G$ (represented by (15)), with the tacit understanding that it is well-posed (and can also be expressed as (20)). We note that $F \star G$ may be well-posed but not convergent. For example, the subgradient descent scheme $z_{k+1} \in \; z_k - \gamma\partial f(z_k)$ is not well-posed if $\gamma \neq 0$ and if f is not differentiable everywhere. On the other hand, well-posedness implies uniqueness of the fixed point in convergent algorithms.

Proposition 2.1. *If the algorithm* (15) *is well-posed and there exists some vector x^* with $\lim_{k\to\infty} x_k = x^*$ for all initial conditions x_0, then* (15) *admits a unique fixed point (x^*, z^*, w^*) and all its trajectories satisfy* $\lim_{k\to\infty}(x_k, w_k, z_k) = (x^*, w^*, z^*)$.

Proof. Since H in (20) is continuous, by taking the limit $k \to \infty$ we infer $x^* = \mathcal{A}x^* + \mathcal{B}H(\mathcal{C}x^*)$ and $\lim_{k\to\infty} w_k = H(\mathcal{C}x^*) =: w^*$ as well as $\lim_{k\to\infty} z_k = \mathcal{C}x^* + \mathcal{D}H(\mathcal{C}x^*) =: z^*$. Hence $w^* = \mathcal{A}x^* + \mathcal{B}w^*$ and $\mathcal{C}x^* \in H^{-1}(w^*) = F^{-1}(w^*) - \mathcal{D}w^*$, i.e., $z^* = \mathcal{C}x^* + \mathcal{D}w^* \in F^{-1}(w_*)$ and thus $w^* \in F(z^*)$. This shows that (x^*, w^*, z^*) is a fixed point of (15) and $\lim_{k\to\infty}(x_k, w_k, z_k) = (x^*, z^*, w^*)$.

Any other fixed point $(\bar{x}^*, \bar{w}^*, \bar{z}^*)$ of (15) satisfies $\bar{x}^* = \mathcal{A}\bar{x}^* + \mathcal{B}\bar{w}^*$, $\bar{z}^* = \mathcal{C}\bar{x}^* + \mathcal{D}\bar{w}^*$ and $\bar{w}^* \in F(\bar{z}^*)$; hence $\mathcal{C}\bar{x}^* \in F^{-1}(\bar{w}^*) - \mathcal{D}\bar{w}^*$ and thus $\bar{w}^* = H(\mathcal{C}\bar{x}^*)$, which gives $\bar{x}^* = \mathcal{A}\bar{x}^* + \mathcal{B}H(\mathcal{C}\bar{x}^*)$ as well as $\bar{z}^* = \mathcal{C}\bar{x}^* + H(\mathcal{C}\bar{x}^*)$. If we initialize a trajectory of (20) as $x_0 = \bar{x}^*$, it hence satisfies $x_k = \bar{x}^*$ for all $k \in \mathbb{N}$; by assumption, we have $\lim_{k\to} x_k = x_*$ which implies $\bar{x}^* = x^*$ and hence also $(\bar{x}^*, \bar{w}^*, \bar{z}^*) = (x^*, w^*, z^*)$. ☐

Unfortunately, the evaluation of H may be as challenging as solving the original optimization problem (12). This difficulty could be due to the possible requirement of solving implicit nonlinear fixed point equations. If $\mathcal{D}$ admits a lower-triangular block structure (possibly after a permutation of $(f_1, \ldots, f_s)$), the evaluation of H can be achieved with the generalized resolvents of ∂f_i for $i \in \{1, \ldots, s\}$. To perform this algorithm evaluation, we partition $\mathcal{B} = (\mathcal{B}_i)$, $\mathcal{C} = (\mathcal{C}_i)$ and $\mathcal{D} = (\mathcal{D}_{ij})$ from (15) into $c \times c$ blocks to form the representation

$$\left(\begin{array}{c|c}\mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D}\end{array}\right) = \left(\begin{array}{c|cccc}\mathcal{A} & \mathcal{B}_1 & \mathcal{B}_2 & \dots & \mathcal{B}_s \\ \hline \mathcal{C}_1 & \mathcal{D}_{11} & 0 & \dots & 0 \\ \mathcal{C}_2 & \mathcal{D}_{21} & \mathcal{D}_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{C}_s & \mathcal{D}_{s1} & \mathcal{D}_{s2} & \dots & \mathcal{D}_{ss}\end{array}\right), \tag{21}$$

and we define the generalized Yosida operators as $H_i := (\partial f_i^{-1} - \mathcal{D}_{ii})^{-1}$ for each $i \in \{1, \dots, s\}$.

Lemma 2.1. *An algorithm* (15) *with $\mathcal{D}$ having a block-lower-triangular structure as partitioned in* (21) *is well-posed if the diagonal blocks $\mathcal{D}_{ii}$ satisfy*

$$\operatorname{Sym}((\sigma_i L_i \mathcal{D}_{ii} - \sigma_i I)^\top (I - m_i \mathcal{D}_{ii})) \prec 0 \quad \textit{with} \quad \sigma_i := (L_i - m_i)^{-1} \quad \textit{for} \quad i \in \{1, \dots, s\}. \tag{22}$$

If $L_i = \infty$, then $\sigma_i L_i = 1$ and $\sigma_i = 0$, which implies that (22) specializes to

$$\mathcal{D}_{ii} + \mathcal{D}_{ii}^\top - 2m_i \mathcal{D}_{ii}^\top \mathcal{D}_{ii} \prec 0. \tag{23}$$

The recursion in (20) *admits the explicit description starting from an initial condition $x_0 \in \mathbb{R}^n$ as follows: Determine*

$$w_k^i = H_i\left(\mathcal{C}_i x_k + \textstyle\sum_{j=1}^{i-1} \mathcal{D}_{ij} w_k^j\right) \quad \textit{for } i = 1, 2, \dots, s \textit{ in ascending order} \tag{24}$$

and set $x_{k+1} = \mathcal{A}x_k + \sum_{i=1}^s \mathcal{B}_i w_k^i$.

Proof. The relation $w \in (F^{-1} - \mathcal{D})^{-1}(z)$ is equivalent to $z = x + \mathcal{D}w$, $w \in F(z)$. In a partition of the vectors (w, z) compatible to the index $i \in \{1, \dots, s\}$ of the blocks of $\mathcal{D}$ in (21), the vector z can be expressed as

$$z^i = \mathcal{C}_i x^i + \textstyle\sum_{j=1}^{i-1} \mathcal{D}_{ij} w^j + \mathcal{D}_{ii} w^i, \; w^i \in F_i(z^i) \quad \text{for} \quad i \in \{1, \dots, s\}. \tag{25}$$

Since $F \in \mathcal{O}_{m,L}$ implies $F_i = \partial f_i \in \partial \mathcal{S}_{m_i, L_i}$, Corollary B.4 assures that $H_i = (F_i^{-1} - \mathcal{D}_{ii})^{-1}$ is globally continuous under condition (22). Therefore, (25) is equivalent to

$$w^i = H_i\left(\mathcal{C}_i x + \textstyle\sum_{j=1}^{i-1} \mathcal{D}_{ij} w^j\right) \quad \text{for} \quad i \in \{1, \dots, s\}. \tag{26}$$

This representation of $w \in (F^{-1} - \mathcal{D})^{-1}(z)$ proves that $H = (F^{-1} - \mathcal{D})^{-1}$ is globally continuous, and hence the algorithm is well-posed. Moreover, (26) leads to (24) and hence to an explicit algorithm description. □

We emphasize a few special cases based on Corollary B.3:

1. If ∂f_i is globally continuous (e.g. for $L_i < \infty$ in which $\partial f_i = \nabla f_i$), then

$$H_i(v) = \begin{cases} \nabla f_i(v) & \text{if } \mathcal{D}_{ii} = 0, \\ \nabla f_i((I - \mathcal{D}_{ii} \nabla f_i)^{-1}(v)) & \text{if } \mathcal{D}_{ii} \neq 0. \end{cases} \tag{27}$$

2. If $\partial f_i(\beta)$ is not single-valued at some $\beta \in \mathbb{R}^c$, then L_i is not finite. In this case, we can evaluate the generalized Yosida operator H_i using the generalized resolvent $(I + \mathcal{D}_{ii} \partial f_i)^{-1}$ as $H_i = \mathcal{D}_{ii}^{-1}[I - (I + \mathcal{D}_{ii} \partial f_i)^{-1}]$, noting that (23) implies invertibility of $\mathcal{D}_{ii}$.

3. If $\mathcal{D}_{ii}$ is diagonal with $\mathcal{D}_{ii} = \lambda_i I$ and $L_i = \infty$, then condition (23) requires that $\lambda_i \in (-\infty, 0) \cup (1/m_i, \infty)$ if $m_i > 0$, $\lambda_i \in (-\infty, 0)$ if $m_i = 0$, and $\lambda_i \in (1/m_i, 0)$ if $m_i < 0$.

2.7 Kronecker Structure

Most first-order optimization algorithms have a Kronecker structure [79].

Definition 5. *A system in* (15) *has* ***Kronecker Structure*** *if there exists matrices* $(\mathcal{A}^a, \mathcal{B}^a, \mathcal{C}^a, \mathcal{D}^a)$ *with*

$$\left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) = \left(\begin{array}{c|c} \mathcal{A}^a & \mathcal{B}^a \\ \hline \mathcal{C}^a & \mathcal{D}^a \end{array}\right) \otimes I_c. \tag{28}$$

Examples are the Davis-Yin algorithm (1) and Gradient descent (16). The procedure in [14, Proposition 2] is an instance of (24) when the algorithm is Kronecker structured. We will use the superscript a to denote matrices that have a Kronecker structure (e.g. $\mathcal{A}^a$ with $\mathcal{A}^a \otimes I_c = \mathcal{A}$ in (28)). Furthermore, we will use the Kronecker product to denote Kronecker-structured linear systems, using the notation

$$G = G^a \otimes I : \left(\begin{array}{c} x_{k+1} \\ \hline y_k \end{array}\right) = \left(\left(\begin{array}{c|c} \mathcal{A}^a & \mathcal{B}^a \\ \hline \mathcal{C}^a & \mathcal{D}^a \end{array}\right) \otimes I_c\right) \left(\begin{array}{c} x_k \\ \hline u_k \end{array}\right) \quad \text{abbreviated as} \quad y = \left(\left[\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right] \otimes I_c\right) u. \tag{29}$$

2.8 Algorithms over Networks

The algorithm (15) involves a linear system G directly interconnected to the operator oracle F. Design of optimization algorithms may also occur in settings where the oracle F is not directly accessible. Example instances include time delays, packet drops, and network cross-talk [6]. We model this setting by assuming that G is given by the interconnection of a network (or plant)

$$P: \qquad \left(\begin{array}{c} x^N_{k+1} \\ \hline z_k \\ y_k \end{array}\right) = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \end{array}\right) \left(\begin{array}{c} x^N_k \\ \hline w_k \\ u_k \end{array}\right), \tag{30}$$

with a controller K with n_y inputs and n_u outputs as described by

$$K: \qquad \left(\begin{array}{c} \xi_{k+1} \\ \hline u_k \end{array}\right) = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \left(\begin{array}{c} \xi_k \\ \hline y_k \end{array}\right). \tag{31}$$

If $E := I - D_2 D_c$ is invertible, we recall that this interconnection is well-posed and admits a state-space description with $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ defined by (7), which is referred to as $P \star K$.

For $G = P \star K$, the algorithmic interconnection $F \star G = F \star (P \star K)$ is visualized by block-diagrams in Figure 4. These reveal how the (given and fixed) network P bridges the oracle F and the (to be designed controller) K to generate the algorithm $F \star (P \star K)$. The specific choice

$$P^0: \qquad \begin{pmatrix} z_k \\ y_k \end{pmatrix} = \begin{pmatrix} 0 & I_{cs} \\ I_{cs} & 0 \end{pmatrix} \begin{pmatrix} w_k \\ u_k \end{pmatrix} \tag{32}$$

recovers a direct interconnection, since then $P^0 \star K = K$ and G is just equal to the controller K.

Example 2.1. *The Kronecker-structured proximal point method* $z_{k+1} = (I + \gamma \partial f)^{-1}(z_k)$ *for* $f \in \mathcal{S}_{m,L}$ *and* $\gamma > 0$ *may be described as an interconnection* $\partial f \star K_{prox}$ *with*

$$K_{prox}: \qquad \left(\begin{array}{c} x_{k+1} \\ \hline z_k \end{array}\right) = \left(\begin{array}{c|c} I & -\gamma I \\ \hline I & -\gamma I \end{array}\right) \left(\begin{array}{c} x_k \\ \hline w_k \end{array}\right). \tag{33}$$

The left interconnection of Figure 5 introduces a delay of 1 time step before and after the computation of ∂f*. The right interconnection assembles the 1-step channel delays into a network* P *sitting between*

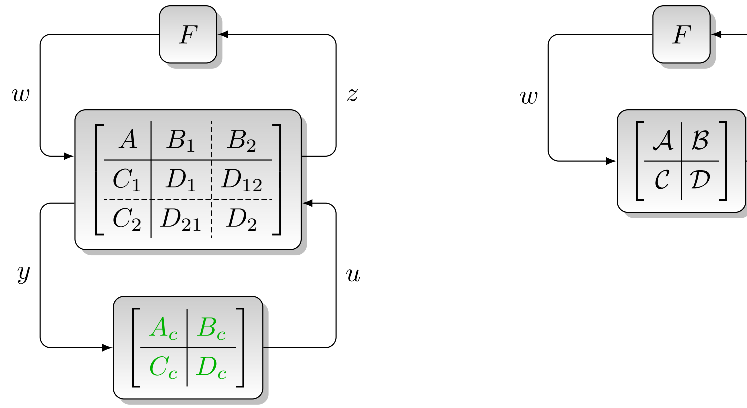


Figure 4: Block diagrams of an algorithm over a network given by $F \star (P \star K)$ (left) and its closed-loop representation $F \star G$ (right) with $G = P \star K$ represented by $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$.

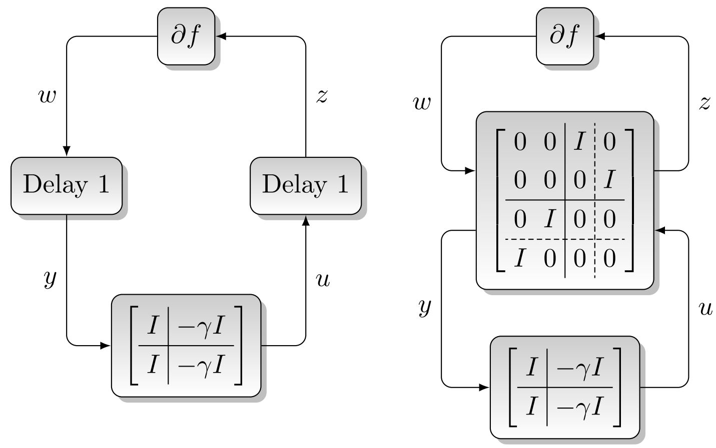


Figure 5: Proximal point method in (33) delayed by one time step (left) and the network representation (right)

∂f and K_{prox}. The interconnection $F \star (P \star K_{prox})$ can as well be described by the algorithmic procedure $z_{k+1} = z_k - \gamma \nabla f(z_{k-1})$, $z'_{k+1} = z_k - \gamma \nabla f(z_{k-1})$, in which z' is an unobservable state.

Figure 6 visualizes the execution of $\partial f \star K_{prox}$ (top) and $\partial f \star (P \star K_{prox})$ (bottom) with parameter $\gamma = 5$ when minimizing the function $f(\beta) = \|\beta - \beta_0\|_2^2$. The term $\beta_0 \in \mathbb{R}^{10}$ with $\beta^* = \beta_0$ is randomly selected. The left plot draws the error $z_k - \beta^*$ for $k \in \{0, \ldots, 14\}$ in the nominal execution $\partial f \star K_{prox}$. The convergence of the nominal execution $\partial f \star K_{prox}$ is demonstrated by the error going to zero. The error $z_k - \beta^*$ diverges to ∞ in the delayed execution $\partial f \star (P \star K_{prox})$ in the right plot. The introduction of delays therefore destabilizes the nominally convergent algorithm.

This destabilization generalizes to networks with multiple delays. Letting P^h denote a network with an h-step delay before and after ∂f, The colored curves in Figure 7 are plots of the spectral radius of the matrix $\mathcal{A}_{cl} := 1 \star (P^h \star K_{prox})$ for $h \in \{0, \ldots, 4\}$ as a function of γ. Convergence to β^0 for the function $f(\beta) = \frac{1}{2}\|\beta - \beta_0\|_2^2$ is assured if the curve stays below the black dotted line ($\rho = 1$). Without delays ($h = 0$), the proximal point algorithm is convergent if $\gamma > 0$. The $h = 1$ interconnection is convergent if $\gamma \in (0, 1)$. Increasing the delay shrinks the set of convergent γ values.

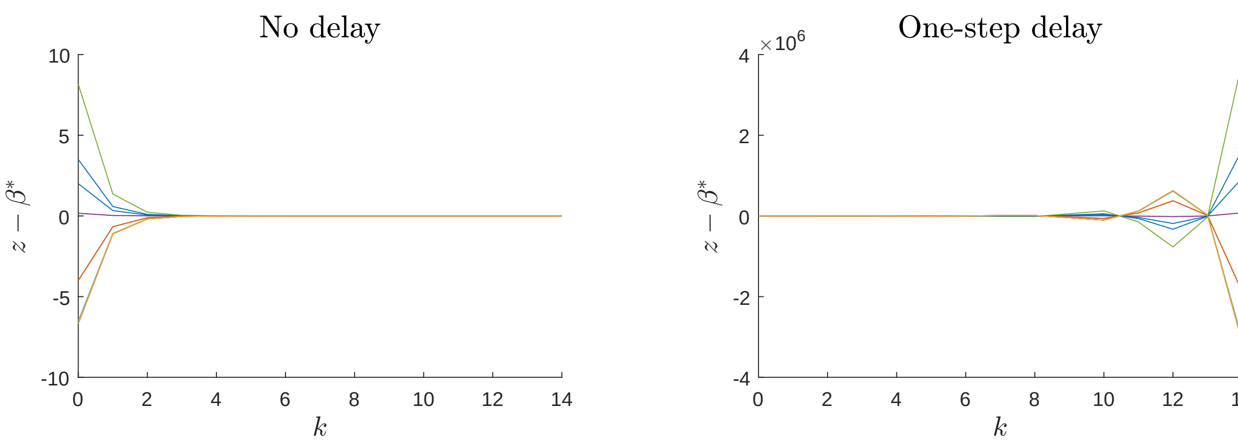


Figure 6: Nominal and one-step-delayed proximal point methods to minimize $f(\beta) = \|\beta - \beta_0\|_2^2$.

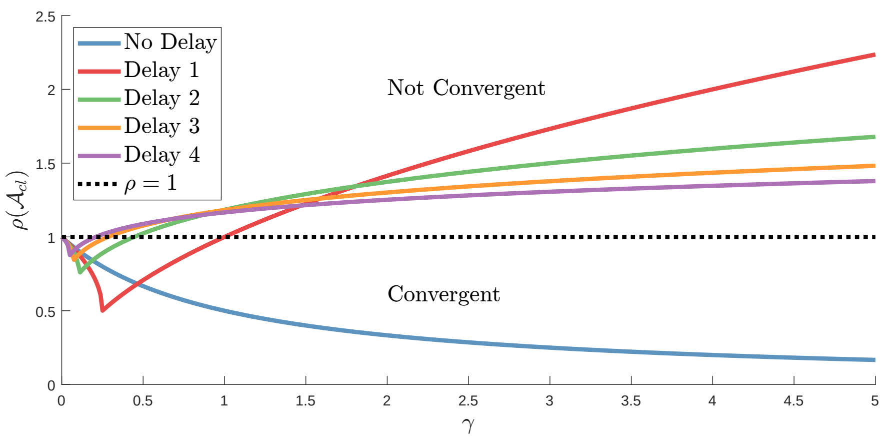


Figure 7: Spectral radius of $\mathcal{A}_{cl}$ v.s. algorithm parameter γ for the delayed version of (33)

2.9 Information Constraints

The block-sparsity pattern of $\mathcal{D}$ defines an *information structure* of the algorithm, specifying the coordinate values w_k^j of the subdifferentials that may be used to compute the other subdifferential coordinates w_k^i within the same iteration $k \in \mathbb{N}$. Descriptions and labels for the eight possible sparsity patterns of the block-lower-triangular matrices $\mathcal{D}$ for $s = 2$ are listed in Table 2, in which $\bullet$ indicates a possibly nonzero $\mathcal{D}_{ij}$ submatrix.

	Sequential	Parallel
Pure Yosida	$\begin{pmatrix} \bullet & 0 \\ \bullet & \bullet \end{pmatrix}$	$\begin{pmatrix} \bullet & 0 \\ 0 & \bullet \end{pmatrix}$
Mixed Yosida and Gradient	$\begin{pmatrix} 0 & 0 \\ \bullet & \bullet \end{pmatrix}, \begin{pmatrix} \bullet & 0 \\ \bullet & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & \bullet \end{pmatrix}, \begin{pmatrix} \bullet & 0 \\ 0 & 0 \end{pmatrix}$
Pure Gradient	$\begin{pmatrix} 0 & 0 \\ \bullet & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$

Table 2: Block-lower-triangular sparsity patterns of $\mathcal{D}$ for $s = 2$

In Table 2, an algorithm is 'parallel' if all w^i coordinates for the subgradient can be evaluated independently (block-diagonal $\mathcal{D}$), or 'sequential' if there are dependencies in w^i computation.

The algorithm $F \star (P \star K)$ may be restricted to satisfy an information constraint. An example includes the requirement that the matrix $\mathcal{D}$ from $G = P \star K$ must be block-lower-triangular, in order to admit evaluation by the procedure in (24). More generally, we choose a set $\mathcal{L}_{\text{info}} \subseteq \mathbb{R}^{n_u \times n_y}$ such that $D_c \in \mathcal{L}_{\text{info}}$ implies that the algorithm $F \star (P \star K)$ obeys a desired information constraint on $\mathcal{D}$. Imposition of convex information constraints on $\mathcal{D}$ could result in nonconvex sets sets $\mathcal{L}_{\text{info}}$, because the expression $\mathcal{D} = D_1 + D_{12}(I - D_c D_2)^{-1} D_c D_{21}$ from (7) is nonlinear in D_c if $D_2 \neq 0$.

2.10 Algorithm Synthesis

Algorithm synthesis involves the design of a controller K such that $F \star (P \star K)$ is a convergent optimization algorithm for all $F \in \mathcal{O}_{m,L}$, and that K obeys the information structure constraint expressed as $D_c \in \mathcal{L}_{\text{info}}$.

The formal description of the algorithm synthesis task is as follows:

Problem 2 (Algorithm Synthesis)**.** *Given a network P from* (30)*, an operator tuple $\mathcal{O}_{m,L}$, and a set $\mathcal{L}_{\text{info}}$, find a controller K with representation (A_c, B_c, C_c, D_c) where $D_c \in \mathcal{L}_{\text{info}}$ and such that the interconnection $F \star (P \star K)$ is a well-posed convergent optimization algorithm for any $F \in \mathcal{O}_{m,L}$.*

The goal of this paper is to derive structural conditions on any controller K which solves the synthesis problem. This permits to search over controllers K that satisfy these structural conditions to find rapidly (exponentially) converging algorithms. The structural constraints on K will be derived using principles of regulation theory for linear systems.

2.11 Regulation Theory for the Rejection of Constant Disturbances

Output regulation of linear systems involves the rejection of exogenous signals d with known dynamics. We focus on the case where the exogenous system generates constant signals $d_k = d_0$ for all $k \in \mathbb{N}$ and refer to [15, 16, 80] for further information on regulation in control theory, including non-constant disturbances.

Consider the following linear system driven by a constant disturbance $d_0 \in \mathbb{R}^{n_d}$:

$$\tilde{G} : \qquad \begin{pmatrix} x_{k+1} \\ \hline e_k \end{pmatrix} = \left(\begin{array}{c|c} \tilde{\mathcal{A}} & \tilde{\mathcal{B}} \\ \hline \tilde{\mathcal{C}} & \tilde{\mathcal{D}} \end{array}\right) \begin{pmatrix} x_k \\ \hline d_k \end{pmatrix}, \quad d_{k+1} = d_k. \tag{34}$$

Definition 6. *The system in* (34) *attains* ***output regulation*** *if $\tilde{\mathcal{A}}$ is Schur, and for all $x_0 \in \mathbb{R}^n, d_0 \in \mathbb{R}^{n_d}$, it holds that $\lim_{k\to\infty} e_k = 0$.*

A core condition for output regulation is the feasibility of the so-called regulator equation.

Lemma 2.2 (Lemma 1 [15])**.** *The system $\tilde{G}$ with representation $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \tilde{\mathcal{D}})$ achieves output regulation iff the following two conditions hold:*

1. *Stability:* $\lim_{k\to 0} x_k = 0$ *for all $x_0 \in \mathbb{R}^n$ if $d_0 = 0$.*
2. *Solvability of Regulator Equation: There exists a matrix Υ satisfying*

$$\begin{pmatrix} \tilde{\mathcal{A}} - I & \tilde{\mathcal{B}} \\ \tilde{\mathcal{C}} & \tilde{\mathcal{D}} \end{pmatrix} \begin{pmatrix} \Upsilon \\ I \end{pmatrix} = 0. \tag{35}$$

Proof. We provide this proof for the purpose of illustration. The stability condition is met iff $\tilde{\mathcal{A}}$ is Schur. If true, $(\tilde{\mathcal{A}} - I)$ has full rank and there exists a unique solution Υ to $\tilde{\mathcal{A}}\Upsilon + \tilde{\mathcal{B}} - \Upsilon = 0$ (top row of (35)). The system in (34) may be described by an extended system $\tilde{G}^{\text{ext}}$ with states (x, d). Performing the coordinate change $(\hat{x}, \hat{d}) = (x - \Upsilon d, d)$ on the system $\tilde{G}^{\text{ext}}$ shows

$$\tilde{G}^{\text{ext}}: \quad \left(\begin{array}{c} x_{k+1} \\ d_{k+1} \\ \hline e_k \end{array}\right) = \left(\begin{array}{cc} \tilde{\mathcal{A}} & \tilde{\mathcal{B}} \\ 0 & I \\ \hline \tilde{\mathcal{C}} & \tilde{\mathcal{D}} \end{array}\right) \begin{pmatrix} x_k \\ d_k \end{pmatrix} \quad \Longleftrightarrow \quad \left(\begin{array}{c} \hat{x}_{k+1} \\ \hat{d}_{k+1} \\ \hline e_k \end{array}\right) = \left(\begin{array}{cc} \tilde{\mathcal{A}} & (\tilde{\mathcal{A}} - I)\Upsilon + \tilde{\mathcal{B}} \\ 0 & I \\ \hline \tilde{\mathcal{C}} & \tilde{\mathcal{C}}\Upsilon + \tilde{\mathcal{D}} \end{array}\right) \begin{pmatrix} \hat{x}_k \\ \hat{d}_k \end{pmatrix}. \tag{36}$$

Given that $\tilde{\mathcal{A}}$ is Schur and $(\tilde{\mathcal{A}} - I)\Upsilon + \tilde{\mathcal{B}} = 0$, the state $\hat{x}_{k+1}$ satisfies $\lim_{k\to\infty} \hat{x}_k = 0$ for all initial conditions $\hat{x}_0$. The steady-state value of e is $\lim_{k\to\infty} e_k = \lim_{k\to\infty} \tilde{\mathcal{C}}\hat{x}_k + (\tilde{\mathcal{C}}\Upsilon + \tilde{\mathcal{D}})\hat{d}_k = (\tilde{\mathcal{C}}\Upsilon + \tilde{\mathcal{D}})d_0$. Hence, output regulation holds iff $\tilde{\mathcal{C}}\Upsilon + \tilde{\mathcal{D}} = 0$ since, otherwise, $\lim_{k\to\infty} e_k \neq 0$ for some $d_0 \neq 0$. □

Appendix C details the consequences of this result if (34) is a controlled interconnection.

3 Convergence of Optimization Algorithms

The networked algorithm synthesis task (Problem 2) may be posed as an instance of stabilization under regulation constraints [81, 82]. To this end, we convert the algorithmic interconnection $F \star (P \star K)$ with fixed point (x^*, w^*, z^*) into an error system $\tilde{F} \star (P \star K)$ with some $\tilde{F} \in \mathcal{O}^0_{m,L}$ having the fixed point $(0, 0, 0)$. With the regulation error $e := z - \mathbf{1}_s \otimes \beta^*$, the consensus condition of a convergent algorithm $F \star (P \star K)$ is translated into the regulation requirement $\lim_{k\to 0} e_k = 0$. The latter property needs to be guaranteed for the error system $\tilde{F} \star (P \star K)$ that is affected by unknown constant disturbances which depend on the unknown solution β^* of (12). In this fashion, algorithm converges is translated into such a regulation property and global attractivity of the fixed point $(0, 0, 0)$ of the error system $\tilde{F} \star (P \star K)$.

3.1 Error Formulation

To capture the optimality condition (18b), we introduce a so-called consensus matrix.

Definition 7. *A matrix N is a **consensus matrix** if*

$$N \in \mathbb{R}^{sc\times(s-1)c}, \qquad \operatorname{rank}(N) = (s-1)c \qquad \text{and} \qquad \operatorname{ran}(N) = \operatorname{null}((\mathbf{1}_s \otimes I_c)^\top). \tag{37}$$

A particular choice is $N = \begin{pmatrix} I_{s-1} \\ -\mathbf{1}_s^\top \end{pmatrix} \otimes I_c$, such as $N = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \otimes I_c$ for $s = 3$. When $s = 1$, the consensus matrix is omitted by setting $N = [\cdot]$.

Solutions β^* of Problem 1 can then be characterized by

$$0 \in F(\mathbf{1}_s \otimes \beta^*) - N\hat{w}^* \quad \text{for some} \quad \hat{w}^* \in \mathbb{R}^{(s-1)c}. \tag{38}$$

Let us partition $g := N\hat{w}^* \in \mathbb{R}^{sc}$ as $g = \operatorname{col}(g^1, \ldots, g^s)$ with $g^i \in \partial f_i(\beta^*)$ and associate to f_i the zero-centered function $\tilde{f}_i(z) := f_i(z + \beta^*) - f_i(\beta^*) - (g^i)^\top z$ with $\tilde{f}_i \in \mathcal{S}^0_{m,L}$ for $i \in \{1, \ldots, s\}$. Then the function $\tilde{f}(z) := \sum_{i=1}^s \tilde{f}_i(z^i)$ for $z = (z^1, \ldots, z^s) \in \mathbb{R}^{sc}$ leads to the error operator $\tilde{F} := \partial\tilde{f} \in \mathcal{O}^0_{m,L}$, which satisfies

$$\tilde{F}(\tilde{z}) = F(\tilde{z} + \mathbf{1}_s \otimes \beta^*) - N\hat{w}^* \quad \text{for} \quad \tilde{z} \in \mathbb{R}^{sc}. \tag{39}$$

Given any trajectory of the algorithm $F \star (P \star K)$, we can introduce the error signals

$$\tilde{z}_k := z_k - \mathbf{1}_s \otimes \beta^* \quad \text{and} \quad \tilde{w}_k := w_k - N\hat{w}^* \tag{40}$$

and note that they are related as $\tilde{w}_k \in F(z_k) - N\hat{w}^* = F(\tilde{z}_k + \mathbf{1}_s \otimes \beta^*) - N\hat{w}^* = \tilde{F}(\tilde{z}_k)$ for $k \in \mathbb{N}$. Any trajectory of $F \star (P \star K)$ (described by (30)-(31) and $w_k \in F(z_k)$) therefore generates a trajectory of the following error system

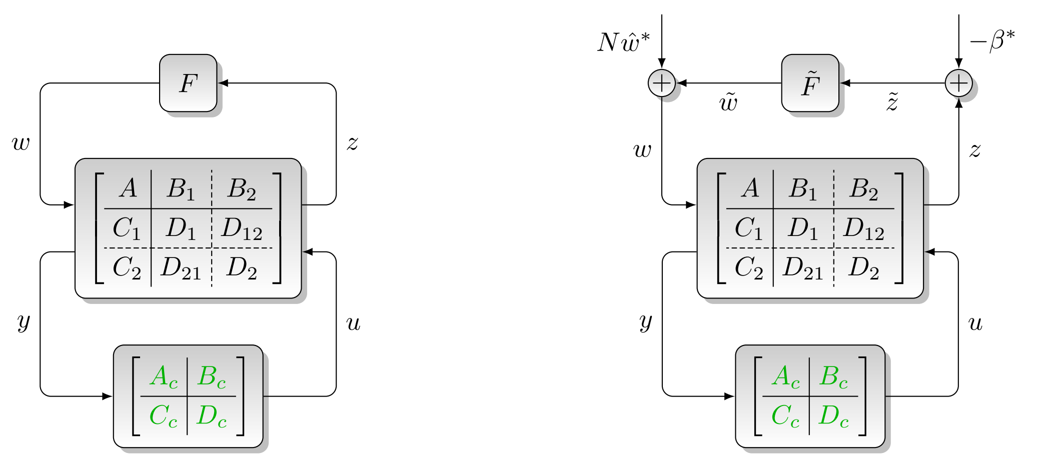
$$\tilde{F}: \qquad \tilde{w}_k \in \tilde{F}(\tilde{z}_k), \tag{41a}$$

$$P_e: \qquad \left(\begin{array}{c} x_{k+1}^N \\ \hline \tilde{z}_k \\ \hdashline e_k \\ \hdashline y_k \end{array}\right) = \left(\begin{array}{c|c:cc:c} A & B_1 & 0 & B_1 N & B_2 \\ \hline C_1 & D_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_1 & D_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_2 & D_{21} & 0 & D_{21} N & D_2 \end{array}\right) \left(\begin{array}{c} x_k^N \\ \hline \tilde{w}_k \\ \hdashline -\beta^* \\ \hat{w}^* \\ \hdashline u_k \end{array}\right), \tag{41b}$$

$$K: \qquad \left(\begin{array}{c} \xi_{k+1} \\ \hline u_k \end{array}\right) = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \left(\begin{array}{c} \xi_k \\ \hline y_k \end{array}\right). \tag{41c}$$

Here e_k is a copy of $\tilde{z}_k$ and interpreted as an error output, while $(-\beta^*, \hat{w}^*)$ is interpreted as a constant disturbance input. We refer to this interconnection as $\tilde{F} \star (P_e \star K)$. If the input disturbance vanishes and if we ignore the error output, this is identical to $P \star K$ (described by (30)-(31)) interconnected with $w_k \in \tilde{F}(z_k)$. As a consequence, $P_e \star K$ is well-posed iff $P \star K$ is well-posed, given that well-posedness of $P \star K$ requires that $I - D_2 D_c$ is invertible. Similarly, in view of Proposition B.5 and (39), $\tilde{F} \star (P_e \star K)$ is well-posed iff $F \star (P \star K)$ is well-posed.

If the well-posed algorithm $F \star (P \star K)$ converges to the fixed point $(x^*, N\hat{w}^*, 1 \otimes \beta^*)$, we infer all trajectories of the error system satisfy $\lim_{k\to\infty} e_k = 0$, despite being affected by the (unknown) constant disturbance input $(-\beta^*, \hat{w}^*)$. This hints at the relevance of the error system for embedding algorithm convergence into regulation theory. The transition to the error system is depicted in Figure 8.



(a) Interconnection with $F \in \mathcal{O}_{m,L}$ (b) Interconnection with $\tilde{F} \in \mathcal{O}_{m,L}^0$

Figure 8: Original interconnection (left) and error system (right) with exogenous disturbance $(-\beta^*, N\hat{w}^*)$

3.2 Test Quadratics and Assumptions

The interconnections $F\star(P\star K)$ and (41) involve the nonlinearities $F \in \mathcal{O}_{m,L}$ and $\tilde{F} \in \mathcal{O}_{m,L}^0$, and convergence or output regulation must occur for all nonlinearities in the respective classes. This leads to a problem of nonlinear robust output regulation. Since we are confronted with structured uncertainties, neither nonlinear nor linear output regulation theory [18, 16] can be applied to designing robustly converging algorithms.

To develop conditions for $F \star (P \star K)$ to converge for all $F \in \mathcal{O}_{m,L}$, we first investigate convergence of $F^t \star (P \star K)$ for affine maps F^t that arise from considering the composite optimization problem (12) with $f^t(z^1, \ldots, z^s) := \sum_{i=1}^s f_i^t(z^i)$ involving the quadratic functions

$$f_i^t(\beta) := \frac{m_i}{2}\|\beta\|_2^2 + b_i^\top \beta \quad \text{where} \quad b_i \in \mathbb{R}^c \quad \text{for} \quad i \in \{1, \ldots, s\}. \tag{42}$$

With $\mathbf{m} := \mathrm{diag}(m) \otimes I_c$ and $b := \mathrm{col}(b_1, \ldots, b_s)$, we infer $F^t(z) = \nabla f^t(z) = \mathbf{m}z + b$ and $\tilde{F}^t(\tilde{z}) = \mathbf{m}\tilde{z}$. We refer to these choices as test quadratics.

For all test quadratics, we note that (12) reads as

$$\beta^* \sum_{i=1}^s m_i + \sum_{i=1}^s b_i = 0, \tag{43}$$

which motivates us to assume $\sum_{i=1}^s m_i \neq 0$ to ensure that Problem 1 has a unique solution. This implies that all test quadratics satisfy $F^t \in \mathcal{O}_{m,L}$ and hence $\tilde{F}^t \in \mathcal{O}^0_{m,L}$, respectively.

If $E(\mathbf{m}) := I - D_1\mathbf{m}$ is invertible, $\tilde{F}^t \star P_e$ described by (41a) and (41b) has the state-space representation

$$P^{\mathbf{m}}: \quad \begin{pmatrix} x_{k+1}^N \\ \hline e_k \\ \hdashline y_k \end{pmatrix} = \left(\begin{array}{c|c:c} A^{\mathbf{m}} & B_1^{\mathbf{m}} & B_2^{\mathbf{m}} \\ \hline C_1^{\mathbf{m}} & D_1^{\mathbf{m}} & D_{12}^{\mathbf{m}} \\ \hdashline C_2^{\mathbf{m}} & D_{21}^{\mathbf{m}} & D_2^{\mathbf{m}} \end{array}\right) \begin{pmatrix} x_k^N \\ \hline d \\ \hdashline u_k \end{pmatrix} \quad \text{where} \quad d = \begin{pmatrix} -\beta^* \\ \hat{w}^* \end{pmatrix} \tag{44}$$

with the matrices

$$\left(\begin{array}{c|cc} A^{\mathbf{m}} & B_1^{\mathbf{m}} & B_2^{\mathbf{m}} \\ \hline C_1^{\mathbf{m}} & D_1^{\mathbf{m}} & D_{12}^{\mathbf{m}} \\ C_2^{\mathbf{m}} & D_{21}^{\mathbf{m}} & D_2^{\mathbf{m}} \end{array}\right) := \left(\begin{array}{c|cc:c} A & 0 & B_1 N & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_2 & 0 & D_{21} N & D_2 \end{array}\right) + \begin{pmatrix} B_1 \\ \hline D_1 \\ D_{21} \end{pmatrix} \mathbf{m}E(\mathbf{m})^{-1} \left(\begin{array}{c|cc:c} C_1 & (\mathbf{1}_s \otimes I_c) & D_1 N & D_{12} \end{array} \right). \tag{45}$$

This illustrates that solving the algorithm design problem for $F^t \star (P \star K)$ with test quadratics boils down to solving a nominal linear regulation problem for $P^{\mathbf{m}} \star K$. To guarantee that $P^{\mathbf{m}}$ can be internally stabilized with some K, we require that $(A^{\mathbf{m}}, B_2^{\mathbf{m}})$ is stabilizable and $(A^{\mathbf{m}}, C_2^{\mathbf{m}})$ is detectable.

Let us collect all assumptions as follows.

Assumption 1. *The vector m in $\mathcal{O}_{m,L}$ satisfies $\sum_{i=1}^s m_i \neq 0$.*

Assumption 2. *The matrix $E(\mathbf{m}) = I - D_1\mathbf{m}$ is invertible.*

Assumption 3. *The pair $(A^{\mathbf{m}}, B_2^{\mathbf{m}})$ is stabilizable and the pair $(A^{\mathbf{m}}, C_2^{\mathbf{m}})$ is detectable.*

3.3 Convergence of Algorithms

We are now ready to formulate the first main contribution of this paper.

Theorem 3.1. *Suppose we are given a network P, a consensus matrix N, and a class of subgradient operators $\mathcal{O}_{m,L}$ such that Assumptions 1-3 hold. If the controller K generates a well posed convergent algorithm $F \star (P \star K)$ for all $F \in \mathcal{O}_{m,L}$, then the following is satisfied:*

1. ***Robust Stability:*** *$I - D_2 D_c$ is invertible and $\tilde{F} \star (P \star K)$ described by $w_k \in \tilde{F}(z_k)$ and*

$$\begin{pmatrix} x_{k+1}^N \\ \hline z_k \\ \hdashline y_k \end{pmatrix} = \left(\begin{array}{c|c:c} A & B_1 & B_2 \\ \hline C_1 & D_1 & D_{12} \\ \hdashline C_2 & D_{21} & D_2 \end{array}\right) \begin{pmatrix} x_k^N \\ \hline w_k \\ \hdashline u_k \end{pmatrix}, \tag{46a}$$

$$\begin{pmatrix} \xi_{k+1} \\ \hline u_k \end{pmatrix} = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \xi_k \\ \hline y_k \end{pmatrix} \tag{46b}$$

is well-posed and satisfies $\lim_{k\to\infty}(x_k, \xi_k) = 0$ *for all initial states* (x_0, ξ_0) *and all* $\tilde{F} \in \mathcal{O}^0_{m,L}$.

2. *Solvability of* ***Regulator Equations:*** *There exists a solution* $(\Pi, \Theta, \Gamma, \Phi)$ *of*

$$\left(\begin{array}{c|cc:c} A & 0 & B_1 N & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_2 & 0 & D_{21} N & D_2 \end{array}\right) \left(\begin{array}{c} \Pi \\ \hline I \\ \hdashline \Gamma \end{array}\right) = \left(\begin{array}{c} \Pi \\ \hline 0 \\ \hdashline \Phi \end{array}\right), \tag{47a}$$

$$\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} \Theta \\ \Phi \end{pmatrix} = \begin{pmatrix} \Theta \\ \Gamma \end{pmatrix}. \tag{47b}$$

If Conditions 1 and 2 are satisfied, then the algorithm $F \star (P \star K)$ *described by* $w_k \in F(z_k)$ *and* (46) *is well-posed and convergent for all* $F \in \mathcal{O}_{m,L}$. *If* β^* *and* $\hat{w}^*$ *are taken with* $0 \in F(\mathbf{1}_s \otimes \beta^*) - N\hat{w}^*$, *all its trajectories satisfy*

$$\lim_{k\to\infty} \begin{pmatrix} x_k^N \\ \xi_k \\ \hline y_k \\ u_k \end{pmatrix} = \begin{pmatrix} \Pi \\ \Theta \\ \hline \Phi \\ \Gamma \end{pmatrix} \begin{pmatrix} -\beta^* \\ \hat{w}^* \end{pmatrix}, \qquad \lim_{k\to\infty} \begin{pmatrix} z_k \\ w_k \end{pmatrix} = \begin{pmatrix} z^* \\ w^* \end{pmatrix} = \begin{pmatrix} \mathbf{1}_s \otimes \beta^* \\ N\hat{w}^* \end{pmatrix}. \tag{48}$$

Proof. See Appendix D. It also comprises a proof that Condition 2 is invariant under the choice of N. □

Condition 1 is related to the operator class $\mathcal{O}^0_{m,L}$, while Condition 2 is independent from $\mathcal{O}_{m,L}$. The regulator equation (47a) only involves the description of the network P and the consensus matrix N (but not the controller K). It admits a solution iff

$$\operatorname{ran} \begin{pmatrix} 0 & -B_1 N \\ \mathbf{1}_s \otimes I_c & -D_{21} N \end{pmatrix} \subset \operatorname{ran} \begin{pmatrix} A - I & B_2 \\ C_1 & D_{12} \end{pmatrix}.$$

This permits us to compute (Π, Γ) and to define Φ in order to satisfy (47a).

The regulator equation (47b) imposes constraints on the description of the controller K. Given K and a solution (Π, Γ, Φ) of (47a), the controller regulator equation (47b) has a solution Θ iff

$$\operatorname{ran} \begin{pmatrix} -B_c \Phi \\ \Gamma - D_c \Phi \end{pmatrix} \subset \operatorname{ran} \begin{pmatrix} A_c - I \\ C_c \end{pmatrix}. \tag{49a}$$

The Robust Stability condition is significantly more involved to verify than the Regulator Equation condition, because (46a) is a statement about global attractivity of the fixed point 0 for all operators in $\mathcal{O}^0_{m,L}$. In this work, we numerically certify Robust Stability through sufficient antipassivity conditions based on Zames-Falb filters. These antipassivity-based analysis and synthesis methods have previously been employed in [62, 63] for the case of $s = 1$. We extend this methodology to the case of $s > 1$, with a novel proof showing that a successful antipassivity-based verification of Robust Stability implies algorithm well-posedness and convergence.

Remark 3.1. *If the Robust Stability condition fails, there does exist an operator* $\tilde{F} \in \mathcal{O}^0_{m,L}$ *such that* $\tilde{F} \star (P \star K)$ *is not convergent. If the Regulator Equation has no solution, there does exist a test quadratic* F^t *such that that* $F^t \star (P \star K)$ *is not convergent.*

3.4 Convergence Example with Channel Memory

To demonstrate the Convergence conditions in Theorem 3.1, we consider composite optimization in Problem 1 with $s = 2$, mediated by a dynamical network with channel memory. The network P_h^α is defined by parameters $\alpha \in \mathbb{R}$ and $h \in \mathbb{N}$. The output $w^1 \in \partial f_1(z^1)$ is filtered by $(\mathbf{z} - \alpha)^{-h}$ before arriving at the controller ($y^1 = (\mathbf{z} - \alpha)^{-h} w^1$). Figure 9 visualizes the filtering process (left) and an instance of a network model with P_3^α (right).

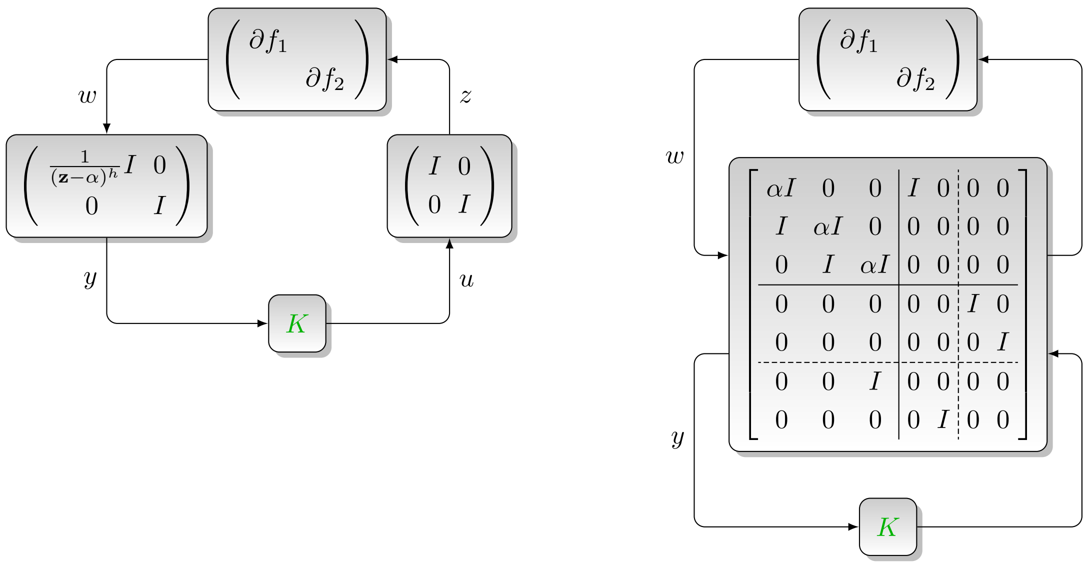


Figure 9: Network P_h^α (left) and the specific state-space description of P_3^α (right)

The regulator equation in (47a) has no solution when $\alpha = 1$ and $h > 0$. As a result, it is then **never** possible to find a controller K such that $F \star (P_h^1 \star K)$ is convergent for any $F \in \mathcal{O}_{m,L}$ (under Assumptions 1-3). When $\alpha \neq 1$, the network regulator equation in (47a) for P_h^α has the unique solution given by

$$\Pi = \begin{pmatrix} 0 & (1-\alpha)^{-1} \\ 0 & (1-\alpha)^{-2} \\ \vdots & \vdots \\ 0 & (1-\alpha)^{-h} \end{pmatrix} \otimes I_c, \qquad \Gamma = \begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix} \otimes I_c, \qquad \Phi = \begin{pmatrix} 0 & (1-\alpha)^{-h} \\ 0 & -1 \end{pmatrix}, \tag{50}$$

with $\Pi = [\cdot]$ if $h = 0$. We consider a class of controllers K described by a network with $\alpha \in \mathbb{R}$, $h \in \mathbb{N}$, and algorithm parameters $b \in \mathbb{R}^3$ as in

$$K_h^\alpha[b]: \qquad \begin{pmatrix} \xi_{k+1} \\ \hline u_k^1 \\ u_k^2 \end{pmatrix} = \left[\left(\begin{array}{c|cc} 1 & (1-\alpha)^h b_0 & b_0 \\ \hline 1 & (1-\alpha)^h b_2 & 0 \\ 1 & (1-\alpha)^h (b_1 + b_2) & b_1 \end{array} \right) \otimes I_c \right] \begin{pmatrix} \xi_k \\ \hline y_k^1 \\ y_k^2 \end{pmatrix} \otimes I_c. \tag{51}$$

Then the controller regulator equation (47b) reads with $\Theta = \begin{pmatrix} -1 & -b_2 \end{pmatrix} \otimes I_c$ as

$$\left(\begin{array}{c|cc} 1 & (1-\alpha)^h b_0 & b_0 \\ \hline 1 & (1-\alpha)^h b_2 & 0 \\ 1 & (1-\alpha)^h (b_1 + b_2) & b_1 \end{array} \right) \begin{pmatrix} -1 & -b_2 \\ \hline 0 & (1-\alpha)^{-h} \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & -b_2 \\ \hline -1 & 0 \\ -1 & 0 \end{pmatrix}. \tag{52}$$

The Douglas-Rachford algorithm [83] with parameters $(\gamma, \lambda) > 0$ may be represented as a controller $K_{\mathrm{DR}}[\gamma, \lambda] := K_h^0[-\gamma\lambda, -\lambda, -\lambda]$ interconnected with P_0^0. The Douglas-Rachford algorithm K_{DR} satisfies the regulator equation (47b) when $\alpha = 0$, and this fails when $\alpha \neq 0$ and $h > 0$. Consequently, the deployment of the Douglas-Rachford algorithm as $F \star (P_h^\alpha \star K_{\text{DR}})$ will fail to converge for any $h > 0$ unless $\alpha = 0$. Convergence of the Douglas-Rachford scheme at $\alpha = 0$ depends on satisfaction of the Robust Stability requirement.

The interconnection $G = P_h^\alpha \star K_h^\alpha[b]$ has $\mathcal{D} = \begin{pmatrix} 0 & 0 \\ 0 & b_1 \end{pmatrix}$. For the class $\mathcal{O}_{(1,0),(10,\infty)}$ with $m = (1, 0)$ and $L = (10, \infty)$ and network parameters $\alpha = 0.5$ and $h = 3$, well-posedness of $F \star G$ for all $F \in \mathcal{O}_{m,L}$ is assured if $b_1 < 0$ (Lemma 2.1.) The controller $K_3^{0.5}[-4, -1, -2]$ satisfies the controller regulator equation condition but fails to guarantee robust stability. As an example, the system $\mathbf{m} \star P_3^{0.5} \star K_3^{0.5}[-4, -1, -2]$ described by

$$\mathbf{m} \star P_3^{0.5} \star K_3^{0.5}[-4, -1, -2] : \qquad \begin{pmatrix} x_{k+1}^N \\ \hdashline \xi_{k+1} \end{pmatrix} = \left[\left(\begin{array}{ccc:c} 0.5 & 0 & -0.125 & 1 \\ 1 & 0.5 & 0 & 0 \\ 0 & 1 & 0.5 & 0 \\ \hdashline 0 & 0 & -0.5 & 1 \end{array} \right) \otimes I_c \right] \begin{pmatrix} x_k^N \\ \hdashline \xi_k \end{pmatrix} \tag{53}$$

is unstable, because the $\mathcal{A}$-matrix in the description (53) has a spectral radius $1.3641 > 1$.

The controller $K_3^{0.5}[-0.04, -0.2, -0.1]$ passes both the Regulator Equations and Robust Stability tests. The regulator equations are satisfied by construction, while the Robust Stability property is computationally verified using the analysis method in the forthcoming Section 5.2. In particular, the $\mathcal{A}$-matrix in the representation of $\mathbf{m} \star P_3^{0.5} \star K_3^{0.5}[-0.04, -0.2, -0.1]$ has the spectral radius $0.9524 < 1$, and is thus stable.

Let us demonstrate composite optimization with channel memory by solving a soft-LASSO-type quadratic program [1] parameterized by a matrix $Q \in \mathbb{R}^{200 \times 200}$ and an offset $\beta_Q \in \mathbb{R}^{200}$:

$$\beta^* \in \operatorname*{argmin}_{\beta \in \mathbb{R}^{200}} \frac{1}{2} (\beta - \beta_Q)^\top Q (\beta - \beta_Q) + \chi_{\|\cdot\|_1 \leq 300}(\beta). \tag{54}$$

The matrix Q in (54) is a randomly generated symmetric matrix with eigenvalues between 1 and 10. The vector β_Q has randomly generated entries in $\{-100, \ldots, 100\}$. The optimal solution β^* satisfying $\|\beta^*\|_1 \leq 300$ is unique and has 14 nonzero entries, and the optimal value is $f(\beta^*) = 3.1685 \times 10^6$. The problem in (54) is an instance of (1) with functions $f_1(\beta) := \frac{1}{2}(\beta - \beta_Q)^\top Q(\beta - \beta_Q)$ and $f_2 = \chi_{\|\cdot\|_1 \leq 300}(\beta)$.

We solve the composite optimization algorithm in (54) using the optimization algorithm $F \star G$ with $G = P_3^{0.5} \star K_3^{0.5}[-0.04, -0.2, -0.1]$. Starting from a random initial condition in which each element of x_0 is uniformly drawn among integers $\{-150, \ldots, 150\}$, we execute the algorithm $F \star G$ using the procedure in (20). Figure 10 plots traces of the first 50 steps of this iteration. The top-left subplot draws z_k for all coordinates, and the top-right plots the error $\tilde{z}_k = z_k - \mathbf{1}_s \otimes \beta^*$. The bottom-left subplot shows the subdifferential w_k, and the bottom-right subplot visualizes the subdifferential error $\tilde{w}_k = w_k - N\hat{w}^*$. The errors $(\tilde{z}, \tilde{w})$ both are convergent to 0.

Figure 11 visualizes the convergence in (48) for the signals (x^N, ξ, y, u) over the course of the algorithmic execution. This convergence involves the unique regulator equation solutions (Π, Γ, Φ) from (50) along with $\Theta = \begin{pmatrix} -1 & -b_2 \end{pmatrix} \otimes I_c$.

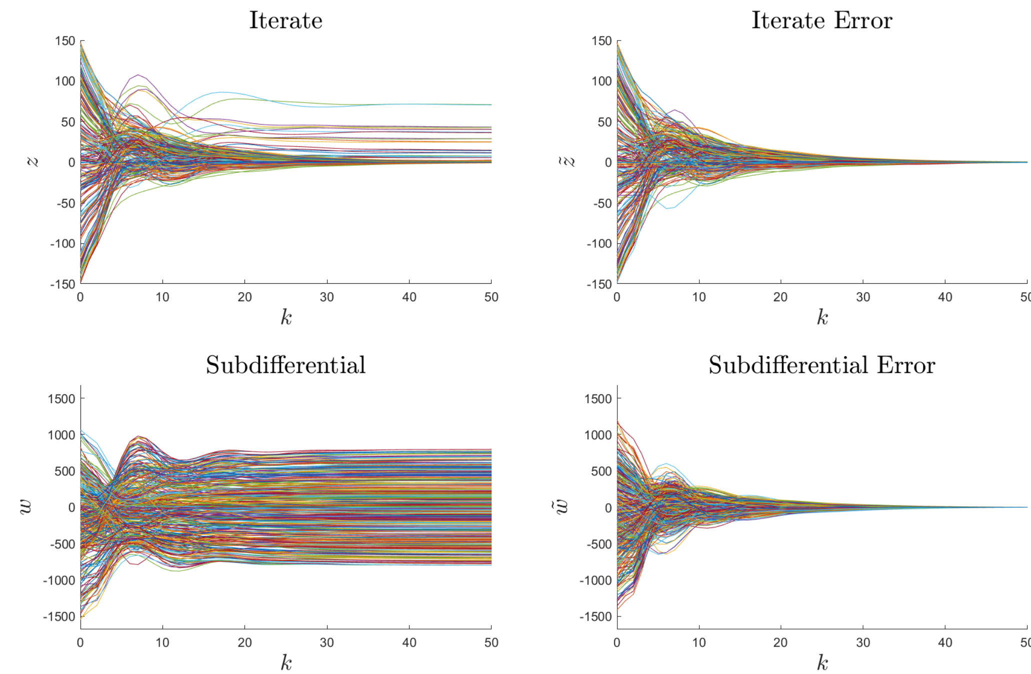


Figure 10: Signals (z, w) and errors $(\tilde{z}, \tilde{w})$ for networked execution of the quadratic program in (54)

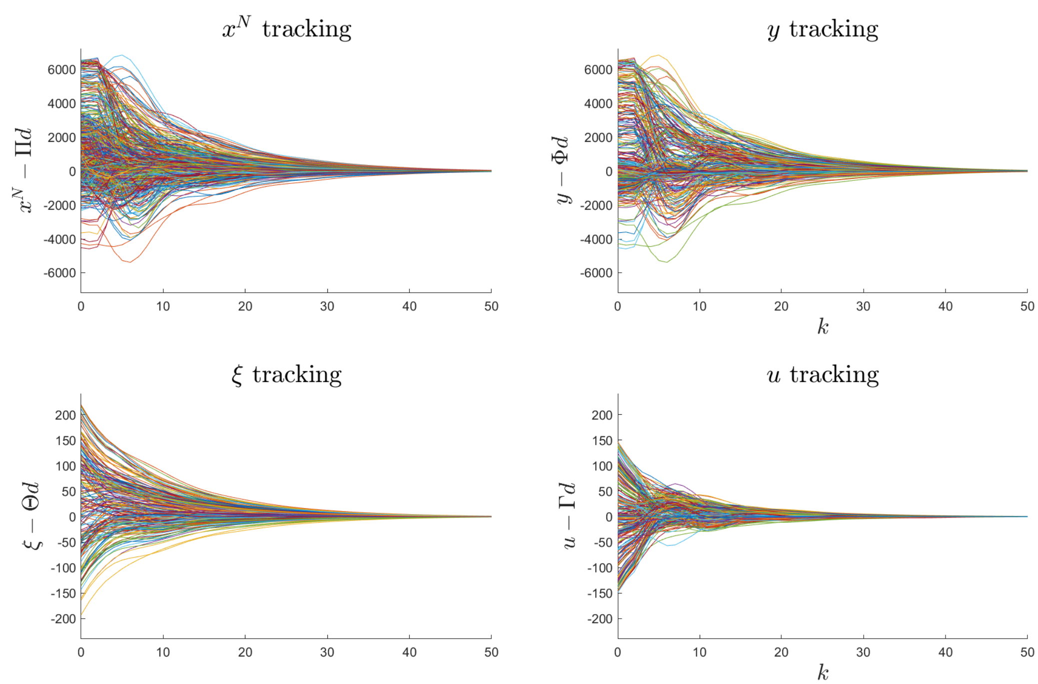


Figure 11: Tracking of (x^N, y, ξ, u) signals based on $d = (-\beta^*, w^*)$

3.5 Convergence under Direct Interconnection

In the case of direct interconnection dynamics in (32), the error system P_e^0 in (41b) is static with

$$P_e^0 : \begin{pmatrix} \tilde{z}_k \\ \hdashline e_k \\ \hdashline y_k \end{pmatrix} = \left(\begin{array}{c:cc:c} 0 & \mathbf{1}_s \otimes I_c & 0 & I_{sc} \\ \hdashline 0 & \mathbf{1}_s \otimes I_c & 0 & I_{sc} \\ \hdashline I_{sc} & 0 & N & 0 \end{array}\right) \begin{pmatrix} \tilde{w}_k \\ \hdashline -\beta^* \\ \hat{w}^* \\ \hdashline u_k \end{pmatrix}. \tag{55}$$

The corresponding test system $P^{\mathbf{m}} = \mathbf{m} \star P_e^0$ is also static with

$$P^{\mathbf{m}} : \begin{pmatrix} e_k \\ \hdashline y_k \end{pmatrix} = \left(\begin{array}{cc:c} \mathbf{1}_s \otimes I_c & 0 & I_{sc} \\ \hdashline \mathbf{m}(\mathbf{1}_s \otimes I_c) & N & \mathbf{m} \end{array}\right) \begin{pmatrix} -\beta^* \\ \hat{w}^* \\ \hdashline u_k \end{pmatrix}. \tag{56}$$

The regulator equations (47a) for P_e^0 in (55) are satisfied by

$$\Pi = [\cdot], \qquad \Gamma = -\begin{pmatrix} \mathbf{1}_s \otimes I_c & 0 \end{pmatrix}, \qquad \Phi = \begin{pmatrix} 0 & N \end{pmatrix}. \tag{57}$$

The controller regulator equation (47b) may be expressed using the parameters in (57) as

$$\begin{pmatrix} A_c - I \\ C_c \end{pmatrix} \Theta = \begin{pmatrix} 0 & -B_c N \\ -(\mathbf{1}_s \otimes I_c) & -D_c N \end{pmatrix}. \tag{58}$$

Let us relate the regulator equations in (58) to existing results in the optimization literature. The work in [14] considers the Kronecker-structured setting with $N^a \otimes I_c = N$. A system K with Kronecker-structured representation $({A_c}^a, {B_c}^a, {C_c}^a, {D_c}^a)$ forming an algorithm $F \star K$ satisfies the Fixed Point Encoding property of [14, Definition 1] (based on prior work in [84]) if the following containments hold [14, Proposition 1]:

$$\operatorname{ran} \begin{pmatrix} {B_c}^a N^a & \mathbf{0} \\ {D_c}^a N^a & -\mathbf{1}_s \end{pmatrix} \subseteq \operatorname{ran} \begin{pmatrix} I - {A_c}^a \\ -{C_c}^a \end{pmatrix}, \tag{59a}$$

$$\operatorname{null} \begin{pmatrix} {A_c}^a - I & {B_c}^a \end{pmatrix} \subseteq \operatorname{null} \begin{pmatrix} (N^a)^\top {C_c}^a & (N^a)^\top {D_c}^a \\ 0 & \mathbf{1}_s^\top \end{pmatrix}. \tag{59b}$$

Algorithm convergence in [14] is based on the verification that $({A_c}^a, {B_c}^a, {C_c}^a, {D_c}^a)$ satisfies (59a) and (59b), and $F \star K$ obeys a Lyapunov stability property based on interpolation conditions [66].

Proposition 3.1. *If a Kronecker-structured controller K with representation (A_c, B_c, C_c, D_c) satisfies the regulator equation in* (58) *and $(A_c - I \;\; B_c)$ has full row rank (e.g. if (A_c, B_c, C_c, D_c) is a minimal realization), then both conditions in* (59) *are satisfied.*

Proof. **Range-space condition:** The range-space condition in (59a) can be recognized as a Kronecker-structured and sign-adjusted version of the equation (58). If Θ^a solves the Kronecker structured instance of (58), then (59a) is certified with

$$\begin{pmatrix} I - {A_c}^a \\ -{C_c}^a \end{pmatrix} \left[\Theta^a \begin{pmatrix} 0 & 1 \\ -I_{s-1} & 0 \end{pmatrix} \right] = \begin{pmatrix} 0 & {B_c}^a N^a \\ \mathbf{1}_s & {D_c}^a N^a \end{pmatrix}. \tag{60}$$

Nullspace condition: If $(A_c - I, B_c)$ has full row rank, we note that $\dim(\operatorname{null}(A_c - I \;\; B_c)) = sc$; due to the Kronecker structure, we infer implies $\dim(\operatorname{null}({A_c}^a - I \;\; {B_c}^a)) = s$. Letting n_c be dimension of ${A_c}^a$, we can partition and rearrange (58) in the Kronecker structured setting into

$$\begin{pmatrix} {A_c}^a - I & {B_c}^a \\ {C_c}^a & {D_c}^a \end{pmatrix} \begin{pmatrix} \Theta_1^a & \Theta_2^a \\ 0 & N^a \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \mathbf{1}_s & 0 \end{pmatrix}, \tag{61}$$

in which $\Theta_1^a \in \mathbb{R}^{n_c \times 1}$ and $\Theta_2^a \in \mathbb{R}^{n_c \times (s-1)}$. Equation (61) shows ${C_c}^a \Theta_1^a = \mathbf{1}_s$, which implies that $\mathrm{rank}(\Theta_1^a) = 1$. The matrix $N^a \in \mathbb{R}^{s \times (s-1)}$ has full column rank because N is a Kronecker-structured consensus matrix (Definition 7). We therefore have

$$\dim\left(\mathrm{null}\left(\begin{array}{cc} {A_c}^a - I & {B_c}^a \end{array}\right)\right) = s, \quad \dim\left(\mathrm{ran}\begin{pmatrix} \Theta_1^a & \Theta_2^a \\ 0 & N^a \end{pmatrix}\right) = s, \quad \mathrm{ran}\begin{pmatrix} \Theta_1^a & \Theta_2^a \\ 0 & N^a \end{pmatrix} \subseteq \mathrm{null}\left(\begin{array}{cc} {A_c}^a - I & {B_c}^a \end{array}\right), \tag{62}$$

from which it follows that $\mathrm{ran}\begin{pmatrix} \Theta_1^a & \Theta_2^a \\ 0 & N^a \end{pmatrix} = \mathrm{null}\left(\begin{array}{cc} {A_c}^a - I & {B_c}^a \end{array}\right)$. We continue by noting that

$$\begin{pmatrix} (N^a)^\top {C_c}^a & (N^a)^\top {D_c}^a \\ 0 & \mathbf{1}_s^\top \end{pmatrix} \begin{pmatrix} \Theta_1^a & \Theta_2^a \\ 0 & N^a \end{pmatrix} = 0 \tag{63}$$

if using the regulator equation (61) and the relation $\mathbf{1}_s^\top N^a = 0$. Together, this implies (59b) since

$$\mathrm{null}\left(\begin{array}{cc} {A_c}^a - I & {B_c}^a \end{array}\right) = \mathrm{ran}\begin{pmatrix} \Theta_1^a & \Theta_2^a \\ 0 & N^a \end{pmatrix} \subseteq \mathrm{null}\begin{pmatrix} (N^a)^\top {C_c}^a & (N^a)^\top {D_c}^a \\ 0 & \mathbf{1}_s^\top \end{pmatrix}. \tag{64}$$

□

4 Structure of Optimization Algorithms

Controllers K forming well-posed convergent optimization algorithms $F \star (P \star K)$ satisfy structural constraints arising from the regulator equation (47b). Requiring D_c to obey an information structure as $D_c \in \mathcal{L}_{\mathrm{info}}$ imposes a lower-bound on the order of K. These controllers K also admit a factorization into a network-dependent internal model and a core subcontroller carrying the algorithm parameters.

4.1 Information Structure

Our first structural property involves an interplay between well-posedness and information constraints.

Theorem 4.1. *Under Assumptions 1-3, consider a network P, a function class $\mathcal{O}_{m,L}$, a set $\mathcal{L}_{\mathrm{info}} \subseteq \mathbb{R}^{n_u \times n_y}$ arising from the information constraint, and a given solution (Π, Γ, Φ) to (47a). Define $r_\Gamma := dim(null(\Gamma))$ and $r_{\Gamma\Phi} := dim(null(\Phi) \cap null(\Gamma))$, and let r_{info} be a lower-bound on the solution of the optimization problem*

$$r_{\mathrm{info}} = \mathit{minimize}\ \mathrm{rank}(D_c) \tag{65a}$$

$$\mathit{subject\ to}\ D_c \in \mathcal{L}_{\mathrm{info}}\ \mathit{such\ that}\ F \star (P \star K)\ \mathit{is\ well\text{-}posed\ for\ all}\ F \in \mathcal{O}_{m,L}. \tag{65b}$$

Then any controller K with representation (A_c, B_c, C_c, D_c) such that $D_c \in \mathcal{L}_{\mathrm{info}}$ and $F \star (P \star K)$ is a convergent optimization algorithm for all $F \in \mathcal{O}_{m,L}$ must have at least $r_{\mathrm{info}} + r_\Gamma - r_{\Gamma\Phi} - n_y$ states.

Proof. If C_c has τ columns, the number of states of K with the representation (A_c, B_c, C_c, D_c) is τ. Since $\tau \geq \mathrm{rank}(C_c) \geq \mathrm{rank}(C_c\Theta)$, it suffices to show $\mathrm{rank}(C_c\Theta) \geq r_{\mathrm{info}} + r_\Gamma - r_{\Gamma\Phi} - n_y$. We exploit the bottom equation in (47b) which reads $C_c\Theta = \Gamma - D_c\Phi$. Let $\Xi = (\Xi_1\ \Xi_2)$ be an invertible matrix such that $\mathrm{ran}(\Xi_2) = \mathrm{null}(\Gamma)$ and where Ξ_2 has full column rank. We infer

$$\mathrm{rank}(C_c\Theta) = \mathrm{rank}(\Gamma - D_c\Phi) = \mathrm{rank}((\Gamma - D_c\Phi)\Xi) = \mathrm{rank}\left(\begin{array}{cc} \Gamma\Xi_1 - D_c\Phi\Xi_1, & -D_c\Phi\Xi_2 \end{array}\right), \tag{66a}$$

leading to

$$\operatorname{rank}(C_c\Theta) \geq \operatorname{rank}(D_c\Phi\Xi_2) = \operatorname{rank}(\Xi_2) - \dim(\operatorname{null}(\Phi) \cap \operatorname{ran}(\Xi_2)) - \dim(\operatorname{null}(D_c) \cap \operatorname{ran}(\Phi\Xi_2)). \tag{66b}$$

Dropping $\operatorname{ran}(\Phi\Xi_2)$ in the right-most intersection and using $\operatorname{ran}(\Xi_2) = \operatorname{null}(\Gamma)$ yields

$$\operatorname{rank}(C_c\Theta) \geq \dim(\operatorname{null}(\Gamma)) - \dim(\operatorname{null}(\Phi) \cap \operatorname{null}(\Gamma)) - \dim(\operatorname{null}(D_c)) = r_\Gamma - r_{\Gamma\Phi} - \dim(\operatorname{null}(D_c)). \tag{66c}$$

Noting $\dim(\operatorname{null}(D_c)) = n_y - \operatorname{rank}(D_c)$, we obtain

$$\operatorname{rank}(C_c\Theta) \geq r_\Gamma - r_{\Gamma\Phi} - n_y + \operatorname{rank}(D_c) \geq r_{\text{info}} + r_\Gamma - r_{\Gamma\Phi} - n_y, \tag{66d}$$

which finishes the proof. □

Remark 4.1. *Finding an exact solution to the optimization problem in* (65) *may be intractable. A special case where this is possible involves the direct interconnection* (32) *where $\mathcal{L}_{\text{info}}$ enforces block-lower-triangularity of D_c. By Proposition B.6, well-posedness of $F \star K$ for all $F \in \mathcal{O}_{m,L}$ with a block-lower-triangular of D_c requires that $[D_c]_{ii}$ must be invertible for any $i \in \{1, \ldots, s\}$ with $L_i = \infty$. Letting $t := \#(L_i = \infty)$ be the number of possibly nonsmooth oracles in $\mathcal{O}_{m,L}$, any well-posed algorithm $F \star K$ must satisfy rank$(D_c) \geq tc$. Since $(r_\Gamma, r_{\Gamma\Phi}, n_y) = ((s-1)c, 0, sc)$, Theorem 4.1 leads to the lower-bound $(t-1)c$ on the number of states of K.*

This $(t-1)c$ bound is found in [85, Theorem 8.1]. The work in [85] builds upon prior results in [86], in which [86, Theorem 3.3] reports an order bound of $(s-1)c$ for the special case that $t = s$.

4.2 Internal Model Structure

Our second contribution on structure is the application of the internal model principle to obtain a decomposition of optimization algorithms. This factorization is developed under the following additional assumptions.

Assumption 4. *The matrix* $\begin{pmatrix} A - I & B_2 \\ C_1 & D_{12} \end{pmatrix}$ *has full column rank.*

Assumption 5. *The pair* $\left(\begin{pmatrix} A^{\mathbf{m}} & B_1^{\mathbf{m}} \\ 0 & I_{cs} \end{pmatrix}, \begin{pmatrix} C_2^{\mathbf{m}} & D_{21}^{\mathbf{m}} \end{pmatrix} \right)$ *is detectable.*

Assumption 4 ensures that the regulator equation in (47a) has at most one solution (Π, Γ, Φ).

Assumption 5 is standard in linear regulation theory, and serves to ensure the existence of an internal-model-based controller K such that $F^t \star (P \star K)$ is a convergent optimization algorithm for the test quadratics F^t (further motivated in the forthcoming Remark 5.1). In the case of a direct interconnection (32), the condition $\sum_{i=1}^s m_i \neq 0$ (Assumption 1) is necessary and sufficient for Assumption 5 to be satisfied.

Theorem 4.2 (Algorithm Structure)**.** *Under Assumptions 1-5, if K has a minimal realization (A_c, B_c, C_c, D_c) and $F \star (P \star K)$ is convergent for all $F \in \mathcal{O}_{m,L}$ as certified by Theorem 3.1, then there exist*

1. *a unique $(\Pi, \Theta, \Gamma, \Phi)$ solving the regulator equation* (47)*,*
2. *an internal model system Σ_{min} that depends only on (Π, Γ, Φ),*
3. *and a core subcontroller Σ_{core} with $(\Sigma_{core})_{ss}$ lying in a Θ-parameterized affine space $\mathcal{L}_{core}(\Theta)$*

such that the controller K may be factored into a cascade $K = \Sigma_{min}\Sigma_{core}$, as visualized in Figure 12.

Proof. See Appendix E for the full proof. □

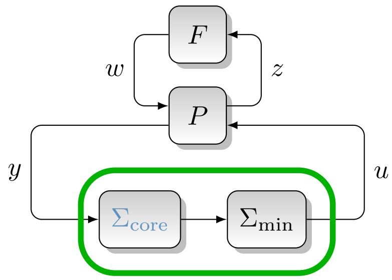


Figure 12: Factorization of the algorithmic interconnection $F \star (P \star K)$ into $F \star (P \star (\Sigma_{\text{min}} \star \Sigma_{\text{core}}))$.

Let us summarize the construction as follows. By Assumption 4, the solution (Π, Γ, Φ) of (47a) is unique. We introduce an invertible coordinate-change matrix $R \in \mathbb{R}^{sc \times sc}$ with the partitioning $R = (R_1 \; R_2)$ such that $\text{ran}(R_1) = \text{null}(\Phi)$. Using $r := \dim(\text{null}(\Phi))$, we partition the following products based on their first r columns as

$$\Pi R = \begin{pmatrix} \Pi_1 & \Pi_2 \end{pmatrix}, \qquad \Gamma R = \begin{pmatrix} \Gamma_1 & \Gamma_2 \end{pmatrix}, \qquad \Phi R = \begin{pmatrix} 0 & \Phi_2 \end{pmatrix}, \qquad \Theta R = \begin{pmatrix} \Theta_1 & \Theta_2 \end{pmatrix}. \tag{67}$$

If right-multiplying both sides of the regulator equation in (47b) by the matrix R, we get

$$\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} \Theta_1 & \Theta_2 \\ 0 & \Phi_2 \end{pmatrix} = \begin{pmatrix} \Theta_1 & \Theta_2 \\ \Gamma_1 & \Gamma_2 \end{pmatrix}. \tag{68}$$

Any Θ that solves (47b) under Assumptions 4 obeys $\text{rank}(\Theta_1) = r$ by Lemmas E.1 and E.2. It is then possible to choose an invertible matrix Q to transform the controller state as $\xi \to Q\xi$ such that

$$Q^{-1}\Theta R = \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \end{pmatrix}. \tag{69}$$

Then the internal model Σ_{min} and the core subcontroller Σ_{core} are given by

$$\Sigma_{\text{min}}: \quad \left[\begin{array}{c|c:c} I_r & I_r & 0 \\ \hline -\Gamma_1 & 0 & I_{n_u} \end{array}\right], \qquad \Sigma_{\text{core}}: \quad \left[\begin{array}{c|c} A_{\text{core}} & B_{\text{core}} \\ \hline C_{\text{core}1} & D_{\text{core}1} \\ C_{\text{core}2} & D_{\text{core}2} \end{array}\right]. \tag{70}$$

The affine space $\mathcal{L}_{\text{core}}(\Theta)$ constraining the state-space matrices of Σ_{core} is defined by the equation

$$\begin{pmatrix} A_{\text{core}} & B_{\text{core}} \\ C_{\text{core}1} & D_{\text{core}1} \\ C_{\text{core}2} & D_{\text{core}2} \end{pmatrix} \begin{pmatrix} \Theta_{22} \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \Theta_{22} \\ \Gamma_1\Theta_{12} + \Gamma_2 \end{pmatrix}, \tag{71}$$

in which the notation $\mathcal{L}_{\text{core}}(\Theta)$ refers to the dependence on $(\Theta_{12}, \Theta_{22})$ from (69). The cascade interconnection $K = \Sigma_{\text{min}}\Sigma_{\text{core}}$ of the controller hence admits the representation

$$\left[\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right] = \left[\begin{array}{c|c} Q^{-1}A_cQ & Q^{-1}B_c \\ \hline C_cQ & D_c \end{array}\right] = \left[\begin{array}{c|c:c} I_r & I_r & 0 \\ \hline -\Gamma_1 & 0 & I_{n_u} \end{array}\right] \left[\begin{array}{c|c} A_{\text{core}} & B_{\text{core}} \\ \hline C_{\text{core}1} & D_{\text{core}1} \\ C_{\text{core}2} & D_{\text{core}2} \end{array}\right]. \tag{72}$$

The controller factorization in (72) is visualized in Figure 13.

By virtue of Theorems 3.1 and 4.2, the controller synthesis task (Problem 2) can be formulated as follows.

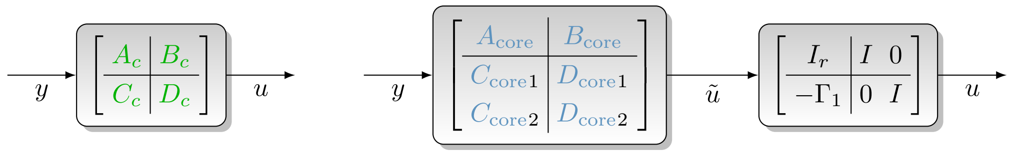


Figure 13: Factorization of K (left) into an internal model and a core controller (right)

Problem 3 (Structured Synthesis)**.** *Given a network P from* (30)*, a function class $\mathcal{O}_{m,L}$, an internal model Σ_{min}, and spaces $\mathcal{L}_{\text{info}}, \mathcal{L}_{core}(\Theta)$, find parameters $(\Theta_{12}, \Theta_{22})$ and a subcontroller Σ_{core}* (70) *with a representation in $\mathcal{L}_{core}(\Theta)$ such that the system $F \star (P \star (\Sigma_{\min}\Sigma_{core}))$ is an optimization algorithm for any $F \in \mathcal{O}_{m,L}$, and such that the information constraints are satisfied as $D_{core2} \in \mathcal{L}_{\text{info}}$.*

Example 4.1. *In the special case where Σ_{core} is static ($A_{core} = [\cdot]$, $\Theta_{22} = [\cdot]$), the constraint* (172) *reads as*

$$\mathcal{L}_{core}(\Theta): \qquad \begin{pmatrix} D_{core1} \\ D_{core2} \end{pmatrix} \Phi_2 = \begin{pmatrix} 0 \\ \Gamma_1\Theta_{12} + \Gamma_2 \end{pmatrix}. \tag{73}$$

The class $K_h^{\alpha}(b)$ in (51) *from the Channel Memory example is a specific instance of a Kronecker-structured algorithm with a static Σ_{core}. When $\alpha \neq 1$, the regulator equation in* (50) *leads to*

$$\textit{null}(\Phi^a) = \textit{ran}\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \Gamma_1^a = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \Gamma_2^a = 0, \qquad \Phi_2^a = \begin{pmatrix} (1-\alpha)^h \\ -1 \end{pmatrix}, \qquad \Theta_{12}^a = -b_2. \tag{74}$$

The algorithms in (51) *may be represented as an interconnection $K_h^{\alpha}(b) = \Sigma_{min}\Sigma_{core}$ with*

$$\Sigma_{min} = \left[\begin{array}{c|c:cc} 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right] \otimes I_c, \qquad \Sigma_{core} = \left[\begin{array}{cc} (1-\alpha)^h b_0 & b_0 \\ \hdashline (1-\alpha)^h b_2 & 0 \\ (1-\alpha)^h (b_1 + b_2) & b_1 \end{array}\right] \otimes I_c. \tag{75}$$

Appendix F discusses how Theorems 3.1 and 4.2 are situated in the literature of output regulation and the internal model principle of control theory.

The structural results in Theorems 4.1 and 4.2 not only apply to strongly convex optimization problems, but also to more general problems. As an example, a convex composite optimization problem (12) with $m_i \geq m'_i$ at each $i \in 1, \ldots, s$ and $\sum_{i=1}^{s} m'_i = 0$ may not have a unique solution β^*. Given that $\mathcal{O}_{m',L}$ satisfies $\mathcal{O}_{m',L} \supset \mathcal{O}_{m,L}$, any controller K forming an convergent optimization algorithm $F \star (P \star K)$ for all $F \in \mathcal{O}_{m',L}$ satisfies Theorems 4.1 and 4.2. These structural properties on K also apply to algorithms that solve variational inequalities and monotone inclusions [87], in which the solution to strongly convex optimization arises as a specialization to fixed point equations based on a general solution concept.

Remark 4.2. *If Assumption 4 was removed, then there could exist multiple* (Π, Φ, Γ) *solving* (47a)*. Letting r denote the minimal value of* $\dim(\text{null}(\Phi))$ *over all solutions of* (47a)*, any solution* $(\Pi_*, \Phi_*, \Gamma_*)$ *of* (47a) *with* $\dim(\text{null}(\Phi_*)) = r$ *defines a family of minimal internal models.*

4.3 Structures under Direct Interconnections

We now explore structures of optimization algorithms under a direct interconnection in (32) (such that the algorithm reads as $F \star G = F \star K$ without networked dynamics), and demonstrate factorizations of existing optimization algorithms for composite optimization. The matrices $\Gamma = -\begin{pmatrix} \mathbf{1}_s \otimes I_c & 0 \end{pmatrix}$ and $\Phi = \begin{pmatrix} 0 & N \end{pmatrix}$

from (57) with $r = \dim(\text{null}(\Phi)) = c$ and generate the internal model and affine constraint

$$\Sigma_{\min}: \quad \left[\begin{array}{c|cc} 1 & 1 & 0 \\ \hline \mathbf{1}_s & 0 & I_s \end{array}\right] \otimes I_c, \qquad \mathcal{L}_{\text{core}}(\Theta): \begin{pmatrix} A_{\text{core}} & B_{\text{core}} \\ C_{\text{core}1} & D_{\text{core}1} \\ C_{\text{core}2} & D_{\text{core}2} \end{pmatrix} \begin{pmatrix} \Theta_{22} \\ N \end{pmatrix} = \begin{pmatrix} \Theta_{22} \\ 0 \\ -(\mathbf{1}_s \otimes I_c)\Theta_{12} \end{pmatrix}. \tag{76}$$

The integrator cascade in (76) is visualized in Figure 14.

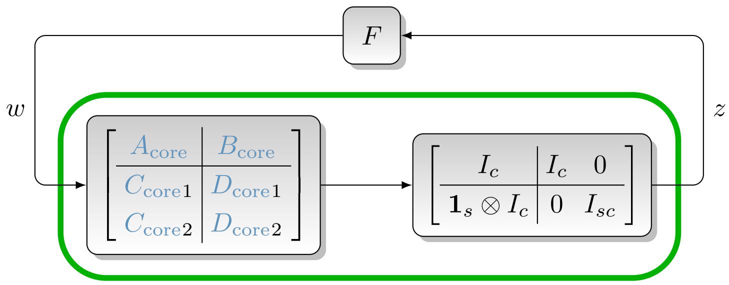


Figure 14: Decomposition of an algorithm with a direct interconnection

Existing algorithms can be decomposed using our structural results.

Example 4.2. *We perform factorizations of the Chambolle-Pock [2], Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [88], and Forward-Reflected-Backward splitting (FRBS) [89] algorithms. The following list documents the original system G, the core subcontroller Σ_{core}, and the associated matrix Θ in the regulator equation* (58) *for the original system representation G.*

The Chambolle-Pock procedure [2] with parameters (τ_1, τ_2, θ) decomposes as

$$G^a = \left[\begin{array}{cc|cc} 1 & -\tau_1 & -\tau_1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & -\tau_1 & -\tau_1 & 0 \\ 1 & \frac{1}{\tau_2} - \tau_1(1+\theta) & -\tau_1(1+\theta) & -1/\tau_2 \end{array}\right], \quad \Sigma^a_{core} = \left[\begin{array}{c|cc} 0 & 0 & 1 \\ \hline -\tau_1 & -\tau_1 & 0 \\ \hdashline -\tau_1 & -\tau_1 & 0 \\ \frac{1}{\tau_2} - \tau_1(1+\theta) & -\tau_1(1+\theta) & -\frac{1}{\tau_2} \end{array}\right], \quad \Theta^a = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [88] with parameters (q, λ) decomposes as

$$G^a = \left[\begin{array}{cc|cc} \frac{2}{1+\sqrt{q}} & -\frac{1-\sqrt{q}}{1+\sqrt{q}} & -\lambda & -\lambda \\ 1 & 0 & 0 & 0 \\ \hline \frac{2}{1+\sqrt{q}} & -\frac{1-\sqrt{q}}{1+\sqrt{q}} & 0 & 0 \\ \frac{2}{1+\sqrt{q}} & -\frac{1-\sqrt{q}}{1+\sqrt{q}} & \lambda & \lambda \end{array}\right], \qquad \Sigma^a_{core} = \left[\begin{array}{c|cc} -\frac{\sqrt{q}-1}{\sqrt{q}+1} & \lambda & \lambda \\ \hline \frac{\sqrt{q}-1}{\sqrt{q}+1} & -\lambda & -\lambda \\ \hdashline \frac{\sqrt{q}-1}{\sqrt{q}+1} & 0 & 0 \\ \frac{\sqrt{q}-1}{\sqrt{q}+1} & -\lambda & -\lambda \end{array}\right], \qquad \Theta^a = \begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}.$$

Forward-Reflected-Backward splitting (FRBS) [89] with parameter λ decomposes as

$$G^a = \left[\begin{array}{ccc|cc} 1 & 0 & \lambda & -2\lambda & -\lambda \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & \lambda & -2\lambda & -\lambda \end{array}\right], \qquad \Sigma^a_{core} = \left[\begin{array}{cc|cc} 0 & -\lambda & -2\lambda & \lambda \\ 0 & 0 & 1 & 0 \\ \hline 0 & \lambda & 2\lambda & -\lambda \\ \hdashline 0 & 0 & 0 & 0 \\ 0 & \lambda & -2\lambda & -\lambda \end{array}\right], \qquad \Theta^a = \begin{pmatrix} -1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In each case, a state-coordinate change Q exists such that the first column of $\mathcal{C}^a$ is $\mathbf{1}_s$. The similarity transformations for each example are

$$Q_{\textit{Chambolle-Pock}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad Q_{\textit{FISTA}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad Q_{\textit{FRBS}} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{77}$$

The controller Σ_{core} can then be read from the system matrices after performing the similarity transformation.

We next form an association with existing structural decompositions in [62] for gradient algorithms.

Example 4.3. *Theorem 2.1 of [62] shows that any algorithm with $\mathcal{D} = 0$ and $s = 1$ can be written as a cascade with an integrator at the input of the controller. Our work leads to an integrator internal model at the output of the controller instead, and generalizes to algorithms with $s > 1$ and/or $\mathcal{D} \neq 0$.*

Figure 15 visualizes decompositions of the Triple Momentum method [43] with parameters (α, η, δ). The Triple Momentum procedure is depicted on the left. Its integrator-at-input factorization from [62] is on the top-right, and our integrator-at-output factorization on the bottom-right.

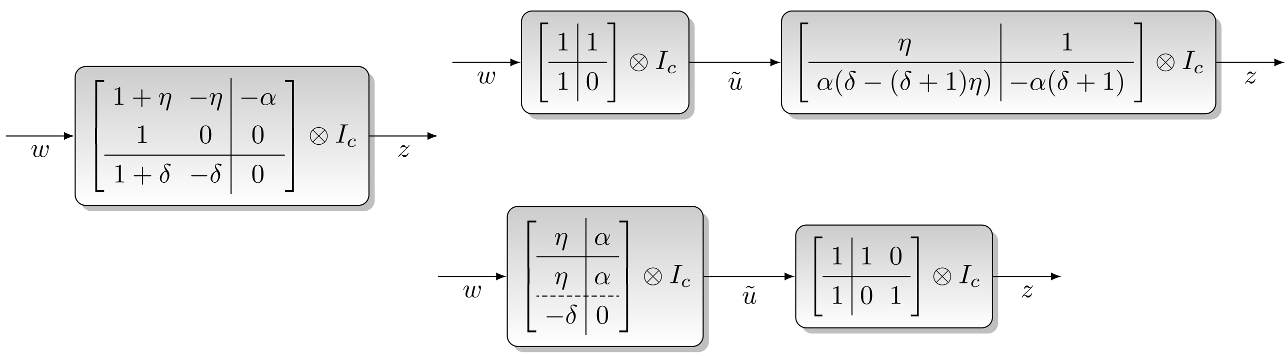


Figure 15: Factorizations of the Triple Momentum method from [43]

4.4 Static Core Subcontrollers

The Davis-Yin factorization in (3) involves a Kronecker-structured static core controller Σ_{core} (with $A_{\text{core}} = [\cdot], \Theta_{22} = [\cdot]$). The factorization in (3) is a specific instance of a general pattern involving static and Kronecker-structured Σ_{core} systems under direct interconnection dynamics (32).

Any core subcontroller Σ_{core} in this setting admits the representation

$$\Sigma_{core} = \left[\begin{array}{c} D_{\text{core}1} \\ \hdashline D_{\text{core}2} \end{array}\right] = \left[\begin{array}{c} D_{\text{core}}{}_1^a \\ \hdashline D_{\text{core}}{}_2^a \end{array}\right] \otimes I_c, \qquad \left(\begin{array}{c} D_{\text{core}}{}_1^a \\ \hdashline D_{\text{core}}{}_2^a \end{array}\right) \in \mathbb{R}^{(s+1)\times s}, \tag{78}$$

in which $D_{\text{core}}{}_1^a$ is a row vector and $D_{\text{core}}{}_2^a$ is a square matrix. The structural constraints on the entries of $D_{\text{core}}{}_1^a$ and $D_{\text{core}}{}_2^a$ from (71) are

$$\mathcal{L}_{\text{core}}(\Theta) : \begin{pmatrix} D_{\text{core}}{}_1^a N^a \\ D_{\text{core}}{}_2^a N^a \end{pmatrix} = \begin{pmatrix} 0 \\ -\mathbf{1}_s \Theta_{12}^a \end{pmatrix}. \tag{79}$$

The constraints in (79) further imply $(N^a)^\top D_{\text{core}}{}_2^a N^a = 0$.

Theorem 4.3. *If the algorithm $F \star G$ described as* (15) *with $s > 1$ has Kronecker structure and a static core controller Σ_{core}, then Σ_{core} may be described by parameters $b_0 \in \mathbb{R}$, $b_1 \in \mathbb{R}^s, b_2 \in \mathbb{R}^s$ as*

$$\Sigma_{core}{}^a = \begin{bmatrix} D_{core}{}^a_1 \\ \hdashline D_{core}{}^a_2 \end{bmatrix} = \begin{bmatrix} b_0 \mathbf{1}_s^\top \\ \hdashline b_1 \mathbf{1}_s^\top + \mathbf{1}_s b_2^\top \end{bmatrix} \otimes I_c, \qquad G = \left[\begin{array}{c|c} 1 & b_0 \mathbf{1}_s^\top \\ \hline \mathbf{1}_s & b_1 \mathbf{1}_s^\top + \mathbf{1}_s b_2^\top \end{array}\right] \otimes I_c. \tag{80}$$

If $D_{core}{}^a_2 = \mathcal{D}^a$ is a lower-triangular matrix, then the parameterization of the set of controllers describing (80) *is further restricted to a 3-dimensional subspace $(b_0, b_{11}, b_{21}) \in \mathbb{R}^3$ with $(b_{1,1:s-1} = 0$ and $b_{2,2:s} = 0)$.*

The associated Θ_{12} matrices used in the constraint (79) *are*

$$\textit{General}: \qquad \Theta_{12} = \left(-1 \;\vdots\; -b_2^\top N^a \right) \otimes I_c. \tag{81a}$$

$$\textit{Lower-Triangular } \mathcal{D}^a: \qquad \Theta_{12} = \left(-1 \;\vdots\; -b_{21} \quad 0_{1 \times s-1} \right) \otimes I_c. \tag{81b}$$

This value of Θ is dependent on the algorithm parameter b_2.

Proof. **General case:** Because $\text{null}((N^a)^\top) = \mathbf{1}_s$, the matrix $D_{\text{core}}{}^a_1$ must satisfy $(D_{\text{core}}{}^a_1)^\top \in \text{ran}(\mathbf{1}_s)$. Thus, there exists a b_0 with $D_{\text{core}}{}^a_1 = b_0 \mathbf{1}_s^\top$. The kernel of the affine mapping $D \mapsto (N^a)^\top D N^a$ is spanned by $(b_1, b_2) \in \mathbb{R}^{2s}$ as $b_1 \mathbf{1}_s^\top + \mathbf{1}_s b_2^\top$ due to $\text{null}((N^a)^\top) = \mathbf{1}_s$. Furthermore, $(b_1 \mathbf{1}_s^\top + \mathbf{1}_s b_2^\top) N = \mathbf{1}_s b_2^\top N = -\mathbf{1}_s \Theta^a_{12}$, which proves $\Theta^a_{12} = -b_2^\top N$ in (81a).

Lower triangular: The algorithm matrix $\mathcal{D}^a$ is equal to $D_{\text{core}}{}^a_2$. Lower triangularity of $D_{\text{core}}{}^a_2$ implies that $(b_{1,1:s-1} = 0$ and $b_{2,2:s} = 0)$ in the matrix $b_1 \mathbf{1}_s^\top + \mathbf{1}_s b_2^\top$, completing the proof. □

Figure 16 visualizes the factorization in the lower triangular case with the parameters $(b_0, b_{11}, b_{21}) \in \mathbb{R}^3$.

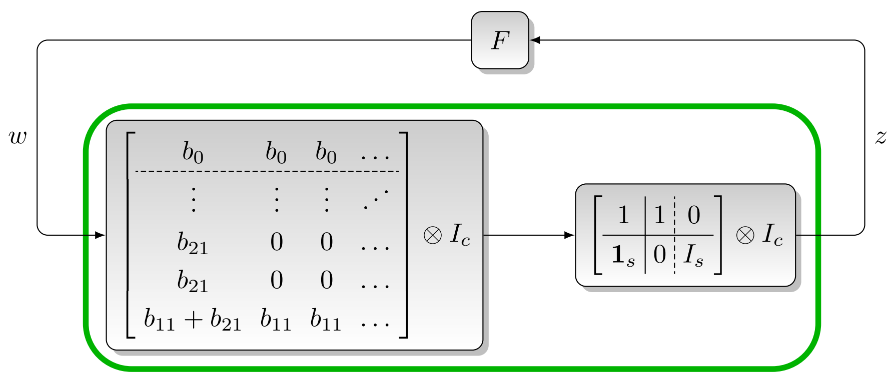


Figure 16: Lower-triangular structured algorithm factorization from Theorem 4.3

Corollary 4.1. *In the lower-triangular Kronecker structured case, all main-diagonal elements of $\mathcal{D}_{ii}$ with $i = \{2, \ldots, s-1\}$ are zero. Ensuring well-posedness of this s-dependent family of algorithms requires that $L_i < \infty$ for all i between 2 and $s-1$ by Proposition B.6.*

Table 3 summarizes the parameters in (80) for existing optimization algorithms. We highlight that the Davis-Yin factorization in (3) is a specific instance of the structure in (80).

5 Synthesis of Optimization Algorithms

Problem 3 offers a method to synthesize optimization algorithms in the networked setting by searching over Σ_{core}. Any core subcontroller Σ_{core} with a representation in $\mathcal{L}_{\text{core}}(\Theta)$ (which is expressed as $(\Sigma_{\text{core}})_{ss} \in$

Algorithm	Parameters	s	b_0	b_{11}	b_{21}
Gradient Descent	α	1	$-\alpha$	0	0
Proximal Point Method	λ	1	$-\lambda$	$-\lambda$	0
Projected Gradient Descent	α	2	$-\alpha$	$-\alpha$	$-\alpha$
Douglas Rachford [83]	γ, λ	2	$-\gamma\lambda$	$-\gamma$	$-\gamma$
Forward-Backward [90]	γ	2	$-\gamma$	$-\gamma$	0
Davis-Yin [11]	γ, λ	3	$-\gamma\lambda$	$-\gamma$	$-\gamma$

Table 3: Parameters describing block-lower-triangular, Kronecker-structured algorithms $F \star G$ with $\mathcal{A} = [.]$

$\mathcal{L}_{\text{core}}(\Theta)$) satisfies the Regulator Equation condition for convergence in Theorem 3.1. In this section, we use LMIs to verify the Robust Stability condition of Theorem 3.1. Our objective in control synthesis is to ensure that $F \star (P \star K)$ is exponentially convergent for all $F \in \mathcal{O}_{m,L}$ with minimal worst-case convergence rate. All synthesis programs are formulated in the strongly convex setting ($\sum_{i=1}^{s} m_i > 0$), which is stronger than Assumption 1.

The algorithm synthesis can be accomplished in a nonconvex or an alternating convex manner. This division is based on how to incorporate the regulation constraints. The nonconvex approach involves the cascade interconnection $K = \Sigma_{\text{core}}\Sigma_{\text{min}}$. Design of Σ_{core} under the constraint $(\Sigma_{\text{core}})_{ss} \in \mathcal{L}_{\text{core}}(\Theta)$ is generally NP-hard [91], because affine constraints are imposed on the elements of $(A_{\text{core}}, B_{\text{core}}, C_{\text{core}1}, C_{\text{core}2})$. However, imposition of convex constraints on $D_{\text{core}1}, D_{\text{core}2}$ may lead to convex controller design problems (e.g. if $D_2 = 0$).

The alternating convex method is based on a full-order internal model Σ_{full} such that the interconnection $K = \Sigma_{\text{full}} \star \Sigma_f$ for any system Σ_f will lead to satisfaction of the regulator equations. Alternation is performed between the convex problems of searching for a controller Σ_f to certify convergence, and searching over filter coefficients describing valid dissipation relations satisfied by uncertainties in the class $\mathcal{O}_{m,L}$. The full-order internal model methodology generally yields higher order controllers as compared to structured control, but offers an (alternatingly) convex computational formulation.

5.1 Exponential Convergence Rates

The speed of an algorithm may be judged based on exponential convergence rates. This can be used in quantify and analyze performance of algorithms, and as an objective when synthesizing fast algorithms.

Definition 8. *The convergence in Definition 3 is exponential with rate $\rho > 0$ if*

$$\exists \gamma > 0, \quad \forall x_0 \in \mathbb{R}^n,\ k \in \mathbb{N}: \qquad \max\{\|x_k - x^*\|_2, \|w_k - w^*\|_2, \|z_k - z^*\|_2\} \leq \gamma \rho^k \|x_0 - x^*\|_2. \tag{82}$$

Following [63], we now sketch how the exponential weighting of signals allows us to develop guarantees for exponential stability of the algorithmic interconnection in Theorem 3.1. Given a fixed rate $\rho > 0$, the ρ-weighting $\bar{x}$ of a sequence $x = (x_k)_{k\in\mathbb{N}}$ is defined as $\bar{x}_k := x_k \rho^{-k}$ for $k \in \mathbb{N}$. For $\tilde{F} \in \mathcal{O}^0_{m,L}$, the ρ-weighting of the signals (x^N, ξ, z, w) in the interconnection $\tilde{F} \star (P \star K)$ in Theorem 3.1 leads to

$$\begin{aligned} \bar{F}: \quad & \bar{w}_k \in \rho^{-k}\tilde{F}(\rho^k \bar{z}_k), \\ \bar{P}: \quad & \begin{pmatrix} \bar{x}_{k+1}^N \\ \hline \bar{z}_k \\ \hdashline \bar{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c} \rho^{-1}A & \rho^{-1}B_1 & \rho^{-1}B_2 \\ \hline C_1 & D_1 & D_{12} \\ \hdashline C_2 & D_{21} & D_2 \end{array}\right) \begin{pmatrix} \bar{x}_k^N \\ \hline \bar{w}_k \\ \hdashline \bar{u}_k \end{pmatrix}, \\ \bar{K}: \quad & \begin{pmatrix} \bar{\xi}_{k+1} \\ \hline \bar{u}_k \end{pmatrix} = \left(\begin{array}{c|c} \rho^{-1}A_c & \rho^{-1}B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \xi_k \\ \hline y_k \end{pmatrix}. \end{aligned} \tag{83}$$

If $G = P \star K$ is described by $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$, then the system that $\bar{G} := \bar{P} \star \bar{K}$ admits a description with $(\rho^{-1}\mathcal{A}, \rho^{-1}\mathcal{B}, \mathcal{C}, \mathcal{D})$. Hence, $\bar{F} \star (\bar{P} \star \bar{K}) = \bar{F} \star \bar{G}$ is compactly described by

$$\begin{pmatrix} \bar{x}_{k+1} \\ \hline \bar{z}_k \end{pmatrix} = \left(\begin{array}{c|c} \rho^{-1}\mathcal{A} & \rho^{-1}\mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \begin{pmatrix} \bar{x}_k \\ \hline \bar{w}_k \end{pmatrix}, \qquad \bar{w}_k \in \rho^{-k}\tilde{F}(\rho^k \bar{z}_k), \tag{84}$$

where $\bar{x}$ is defined as $x_k := \mathrm{col}(\bar{x}_k^N, \bar{\xi}_k)$ for $k \in \mathbb{N}$. Note that (84) emerges from $\tilde{F} \star G$ (described as (15) with $\tilde{F}$ instead of F, see Figure 17a) by ρ-weighting the signals (x, w, z), and both have the fixed point $(0, 0, 0)$.

We say that (84) is Lyapunov stable if is well-posed and satisfies

$$\exists \gamma > 0, \quad \forall \bar{x}_0 \in \mathbb{R}^n, \ k \in \mathbb{N}: \qquad \max\{\|\bar{x}_k\|_2, \|\bar{w}_k\|_2, \|\bar{z}_k\|_2\} \leq \gamma \|\bar{x}_0\|_2. \tag{85}$$

This implies that $\tilde{F} \star G$ is well-posed and satisfies the exponential convergence property with rate ρ:

$$\exists \gamma > 0, \quad \forall x_0 \in \mathbb{R}^n, \ k \in \mathbb{N}: \qquad \max\{\|x_k\|_2, \|w_k\|_2, \|z_k\|_2\} \leq \gamma \rho^k \|x_0\|_2. \tag{86}$$

We arrive at the following variant of Theorem 3.1 to guarantee exponential convergence of algorithms.

Corollary 5.1. *Suppose that $I - D_2 D_c$ is invertible, that (84) resulting from the ρ-weighted interconnection (83) is Lyapunov stable for all $\tilde{F} \in \mathcal{O}_{m,L}^0$, and that the regulator equation (47) has a solution. Then $F \star (P \star K)$ described by $w_k \in F(z_k)$ and (46) is well-posed and ρ-exponentially convergent for all $F \in \mathcal{O}_{m,L}$.*

5.2 Analysis

In extension of [92] for $s = 1$, the goal is to construct an LMI-based method to guarantee that (84) is Lyapunov stable. We use the matrices

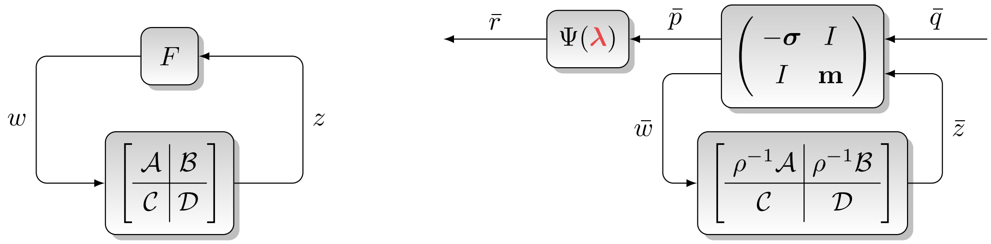
$$\mathbf{m} := \mathrm{diag}(m) \otimes I_c, \qquad \boldsymbol{\sigma} := \mathrm{diag}(L - m)^{-1} \otimes I_c, \qquad \Sigma_{m,L} := \left(\begin{array}{c:c} -\boldsymbol{\sigma} & I_{cs} \\ \hdashline I_{cs} & \mathbf{m} \end{array}\right) \tag{87}$$

and recall that any $\tilde{F} \in \mathcal{O}_{m,L}^0$ satisfies the dissipation relations in Lemma B.3. With the corresponding coefficients collected as $\lambda_i = (\lambda_0^i, \ldots, \lambda_{\nu_{\max}}^i)$ for $i \in \{1 \ldots, s\}$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_s)$, we can define the filters

$$\Psi^i(\lambda^i) = \left[\begin{array}{cc|c} 0_{1 \times \nu_{\max}-1} & 0 & 1 \\ I_{\nu_{\max}-1} & 0_{\nu_{\max}-1 \times 1} & 0 \\ \hline \lambda_{1:\nu_{\max}-1}^i & \lambda_{\nu_{\max}}^i & \lambda_0^i \end{array}\right] \otimes I_c \qquad \text{and} \qquad \Psi(\boldsymbol{\lambda}) = \mathrm{blkdiag}(\Psi^1(\lambda^1), \ldots, \Psi^s(\lambda^s)), \tag{88}$$

which are linear systems with initial condition zero. If $\boldsymbol{\lambda}$ satisfies the constraints (134), the conclusion of Lemma B.3 reads as follows: $\sum_{k=0}^{T-1} \bar{q}_k^\top \bar{r}_k \geq 0$ holds for any $T > 0$, any $\tilde{F} \in \mathcal{O}_{m,L}^0$, and any signal satisfying

$$\bar{r} = \Psi(\boldsymbol{\lambda})\bar{p}, \quad \begin{pmatrix} \bar{p} \\ \bar{w} \end{pmatrix} = \Sigma_{m,L} \begin{pmatrix} \bar{q} \\ \bar{z} \end{pmatrix} \quad \text{and} \quad \bar{w}_k \in \rho^{-k}\tilde{F}(\rho^k \bar{z}_k) \quad \text{for} \quad k \in \mathbb{N}.$$



(a) Algorithm with $\tilde{F}$ (b) Filtered interconnection $G^{\Psi}(\boldsymbol{\lambda}) := \Psi(\boldsymbol{\lambda})(\Sigma_{m,L} \star \bar{G})$

Figure 17: Framework for computational algorithm analysis

Following [92, Section 3.2] and if $I - \mathbf{m}\mathcal{D}$ is invertible, we construct the filtered interconnection $G^{\Psi}(\boldsymbol{\lambda}) := \Psi(\boldsymbol{\lambda})(\Sigma_{m,L}\star\bar{G})$ as depicted in Figure 17b by standard operations; this leads to a description $(\hat{\mathcal{A}}, \hat{\mathcal{B}}, \hat{\mathcal{C}}(\boldsymbol{\lambda}), \hat{\mathcal{D}}(\boldsymbol{\lambda}))$ in which $\hat{\mathcal{C}}(\boldsymbol{\lambda})$ and $\hat{\mathcal{D}}(\boldsymbol{\lambda})$ are affine in $\boldsymbol{\lambda}$. We then arrive at the following extension of [92, Corollary 18].

Theorem 5.1. *Under Assumptions 1-5, suppose we are given a network P, a controller K such that $P \star K$ is well posed, a rate $\rho > 0$, and a maximal filter length $\nu_{\max} \in \mathbb{N}$. Then the algorithm $F \star (P \star K)$ is well-posed and ρ-exponentially convergent for all $F \in \mathcal{O}_{m,L}$ if the following conditions hold:*

1. ***Information Constraint:*** *$\mathcal{D}$ is block-lower-triangular as in (21).*
2. ***Robust Stability:*** *With $(\hat{\mathcal{A}}, \hat{\mathcal{B}}, \hat{\mathcal{C}}(\boldsymbol{\lambda}), \hat{\mathcal{D}}(\boldsymbol{\lambda}))$ representing $G^{\Psi}(\boldsymbol{\lambda}) := \Psi(\boldsymbol{\lambda})(\Sigma_{m,L} \star \bar{P} \star \bar{K})$, there exist a matrix $\mathcal{M} \succ 0$ and a coefficient vector $\boldsymbol{\lambda}$ satisfying the constraints*

$$\begin{pmatrix} \hat{\mathcal{A}} & \hat{\mathcal{B}} \\ I & 0 \end{pmatrix}^{\top} \begin{pmatrix} \mathcal{M} & 0 \\ 0 & -\mathcal{M} \end{pmatrix} \begin{pmatrix} \hat{\mathcal{A}} & \hat{\mathcal{B}} \\ I & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \hat{\mathcal{C}}(\boldsymbol{\lambda}) & \hat{\mathcal{D}}(\boldsymbol{\lambda}) \\ 0 & I \end{pmatrix}^{\top} \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \begin{pmatrix} \hat{\mathcal{C}}(\boldsymbol{\lambda}) & \hat{\mathcal{D}}(\boldsymbol{\lambda}) \\ 0 & I \end{pmatrix} \prec 0, \tag{89a}$$

$$\textstyle\sum_{\nu=0}^{\nu_{\max}} \rho^{-\nu} \lambda_{\nu}^{i} > 0, \quad \lambda_{\nu}^{i} \leq 0, \quad \forall \nu \geq 1, \quad \forall i \in \{1 \ldots, s\}. \tag{89b}$$

3. *Solvability of **Regulator Equations:** The regulator equations in (47) admit a solution.*

Proof. Due to Corollary 5.1, it suffices to prove that (84) is Lyapunov stable.

Since the "D"-matrices of P and $\bar{P}$ coincide, we observe that $\bar{P}$ satisfies Assumption 2 and $\bar{G} = \bar{P} \star \bar{K}$ is well-posed. Hence $I - \mathcal{D}\mathbf{m}$ is invertible and, therefore, $\Sigma_{m,L} \star \bar{G}$ is well-posed, having the "D"-matrix $-\boldsymbol{\sigma}I + \mathcal{D}(I - \mathbf{m}\mathcal{D})^{-1}$. Since the "D"-matrix of the state-space representation of $\Psi(\boldsymbol{\lambda})$ in (88) is $\Lambda := \mathrm{diag}(\lambda_0^1 I_c, \ldots \lambda_0^s I_c)$, we conclude that $\hat{\mathcal{D}}(\boldsymbol{\lambda}) = \boldsymbol{\Lambda}(-\boldsymbol{\sigma}I + \mathcal{D}(I - \mathbf{m}\mathcal{D})^{-1})$. Now note that (89) implies

$$\hat{\mathcal{D}}(\boldsymbol{\lambda}) + \hat{\mathcal{D}}(\boldsymbol{\lambda})^{\top} \prec -\hat{\mathcal{B}}^{\top}\mathcal{M}\hat{\mathcal{B}} \preceq 0. \tag{90}$$

Hence $\mathrm{Sym}[\boldsymbol{\Lambda}(-\boldsymbol{\sigma}I + \mathcal{D}(I - \mathbf{m}\mathcal{D})^{-1})] \prec 0$, and thus $\mathrm{Sym}[\boldsymbol{\Lambda}(\boldsymbol{\sigma}\mathbf{L}\mathcal{D} - \boldsymbol{\sigma})^{\top}(I - \mathbf{m}\mathcal{D})] \prec 0$ by Lemma B.7. In view of Condition 1, we conclude that (22) is satisfied, which implies that $(F^{-1} - \mathcal{D})^{-1}$ is globally continuous for all $F \in \mathcal{O}_{m,L}$ by Lemma 2.1. For $\tilde{F} \in \mathcal{O}_{m,L}^{0}$, global continuity of $(\tilde{F}^{-1} - \mathcal{D})^{-1}$ implies global continuity of $(\rho^{-k}(\tilde{F}(\rho^{k}.)^{-1} - \mathcal{D})^{-1}$ for every $k \in \mathbb{N}$ by Proposition B.5, which in turn guarantees that (84) is well-posed. By following routine dissipation arguments [63, 92], the LMI (89) guarantees that (84) is Lyapunov stable. □

The *identity filter* $\Psi(\boldsymbol{\lambda})(\mathbf{z}) = I$ obtained with $\lambda_0^i = 1$ and $\lambda_\nu^i = 0$ for $\nu \in \{1, \ldots, \nu_{\max}\}$ and $i \in \{1, \ldots, s\}$ always satisfies the condition in (89b). Due to the KYP Lemma, feasibility of the LMI in (89b) can be interpreted as certifying strict anti-passivity of the system in Figure 17b, or strict negative-realness of the corresponding transfer matrix [62, 92].

5.3 Structured Synthesis with Minimal Internal Models

The interconnection for structured synthesis to solve Problem 2 is depicted in Figure 18. For a fixed rate $\rho > 0$, it involves the joint search over filter coefficients $\boldsymbol{\lambda}$, controllers Σ_{core}, and regulator equation parameters Θ satisfying the subspace constraint $\mathcal{L}_{\text{core}}(\Theta)$ in (71).

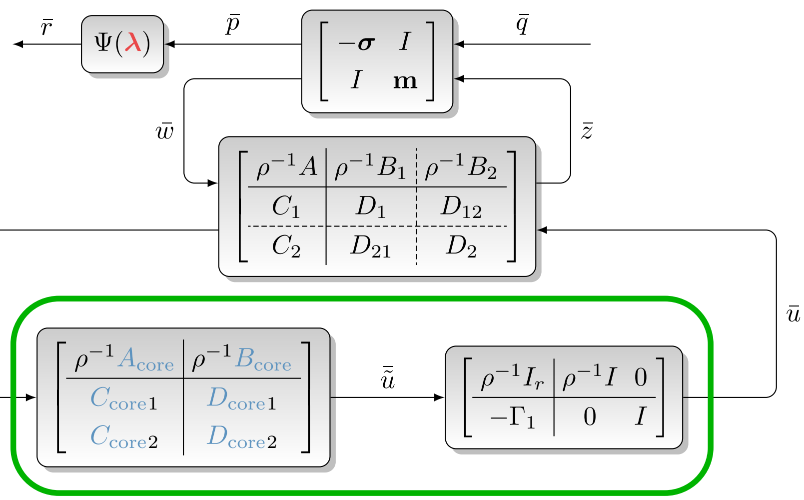


Figure 18: Filtering framework for structured synthesis with a minimal internal model

To ensure well-posedness of the synthesized optimization algorithm, we require the following assumption on the given information structure $\mathcal{L}_{\text{info}}$ from Section 2.9.

Assumption 6. *The information structure constraint $\mathcal{L}_{\text{info}} \subseteq \mathbb{R}^{n_u \times n_y}$ is taken to ensure that $\mathcal{D} = D_1 + D_{12}D_c(I - D_2D_c)^{-1}D_{21}$ is block-lower-triangular for all $D_c \in \mathcal{L}_{\text{info}}$.*

When K is formed by the interconnection $K = \Sigma_{\text{min}}\Sigma_{\text{core}}$, then $D_c = D_{\text{core}2}$, and the information constraint can be posed as $D_{\text{core}2} \in \mathcal{L}_{\text{info}}$. Furthermore, the matrix $D_{\text{core}2}$ is unchanged by the exponential weighting in $\bar{\Sigma}_{\text{core}}$ as compared to Σ_{core}.

Problem 4 (Structured Synthesis)**.** *Let $\mathcal{O}_{m,L}$ be a class of operators, $\mathcal{L}_{\text{info}}$ be an information structure set, and P be a network model such that Assumptions 1-6 hold. For a rate $\rho > 0$ and a core controller controller dimension n_ξ, find matrices $(\Theta_{12}) \in \mathbb{R}^{r \times sc - r}$ and $(\Theta_{22}) \in \mathbb{R}^{n_\xi - r \times sc - r}$, a core subcontroller Σ_{core} with a state-space representation $(\Sigma_{core})_{ss}$, and filter coefficients $\boldsymbol{\lambda}$ such that the following hold:*

1. *The constraints of $D_{core2} \in \mathcal{L}_{\text{info}}$ and $(\Sigma_{core})_{ss} \in \mathcal{L}_{core}(\Theta)$ are obeyed and $\bar{P} \star (\bar{\Sigma}_{min}\bar{\Sigma}_{core})$ is well-posed.*
2. *The LMI in (89) is feasible for the system $\hat{G}(\boldsymbol{\lambda}) = \Psi(\boldsymbol{\lambda})(\Sigma_{m,L} \star \bar{P} \star (\bar{\Sigma}_{min}\bar{\Sigma}_{core}))$.*

By repeatedly solving Problem 5, the rate ρ may be minimized through bisection.

Proposition 5.1. *For any solution of Problem 4, the controller $K := \Sigma_{min}\Sigma_{core}$ leads to a well-posed algorithm $F \star (P \star K)$ which converges exponentially with rate ρ for all $F \in \mathcal{O}_{m,L}$.*

Proof. Due to Theorem 4.2, the choice of K guarantees that the regulator equations (47) admit a solution. Then the claim follows by applying Theorem 5.1. ☐

Appendix G explains how to use tools from robust control and nonconvex optimization to search for a solution of Problem 4.

5.4 Alternating Convex Synthesis with Full-Order Internal Models

The structural constraints in Condition 2 of Problem 4 are induced by working with an internal model of small order r, which renders the design of the subcontroller non-convex. In this section we propose a full-order internal model, which leads to controllers and algorithms with larger state-dimensions, but which opens the way for (alternating) convex algorithm design algorithms.

Definition 9. *For a network P and a solution (Π, Γ, Φ) of the plant regulator equation (47a), a **full-order** internal model Σ_{full} and a **unstructured** subcontroller Σ_f are defined as*

$$\Sigma_{full}: \quad \left[\begin{array}{c|c:cc} I_{cs} & 0 & I & 0 \\ \hline -\Gamma & 0 & 0 & I \\ \hdashline \Phi & I & 0 & 0 \end{array}\right] \qquad \text{and} \qquad \Sigma_f := \left[\begin{array}{c|c} A_f & B_f \\ \hline C_f^1 & D_f^1 \\ \hdashline C_f^2 & D_f^2 \end{array}\right]. \tag{91}$$

Proposition 5.2. *For a network P and a solution (Π, Γ, Φ) of the plant regulator equation (47a), the full-order internal model Σ_{full} and any subcontroller Σ_f define a controller $K = \Sigma_{full} \star \Sigma_f$ with a representation such that the controller regulator equation (47b) is satisfied with $\Theta = [-I_{cs}\ 0]^\top$.*

Proof. The interconnection $K = \Sigma_{\text{full}} \star \Sigma_f$ is well-posed and has the state-space representation

$$K = \left[\begin{array}{cc|c} I_{cs} - D_{f1}\Phi & C_{f1} & D_{f1} \\ -B_f\Phi & A_f & B_f \\ \hline -\Gamma - D_{f2}\Phi & C_{f2} & D_{f2} \end{array}\right]. \tag{92}$$

By inspection, we infer

$$\left(\begin{array}{cc|c} I_{cs} - D_{f1}\Phi & C_{f1} & D_{f1} \\ -B_f\Phi & A_f & B_f \\ \hline -\Gamma - D_{f2}\Phi & C_{f2} & D_{f2} \end{array}\right) \left(\begin{array}{c} -I_{cs} \\ 0 \\ \hline \Phi \end{array}\right) = \left(\begin{array}{c} -I_{cs} \\ 0 \\ \hline \Gamma \end{array}\right). \tag{93}$$

□

The systems Σ_{core} and Σ_f can be related to each other as follows.

Theorem 5.2. *Let (Π, Θ, Φ) be a solution to the regulator equation (47), R be a basis-change matrix as in (67), $\Sigma_{\min}$ be an internal model constructed from (70), and Σ_{full} be a full-order internal model from (91).*

We have the following correspondences:

1. *Given any full-order subcontroller Σ_f, there exists a system Σ_{core} with $(\Sigma_{core})_{ss} \in \mathcal{L}_{core}(\Theta)$ such that $(\Sigma_{\min}\Sigma_{core})_{ss} = (\Sigma_{full} \star \Sigma_f)_{ss}$.*
2. *Given any Σ_{core} with $(\Sigma_{core})_{ss} \in \mathcal{L}_{core}(\Theta)$, there exists a Σ_f such that $(\Sigma_{\min}\Sigma_{core})_{ss} = (\Sigma_{full} \star \Sigma_f)_{ss}$.*

Proof. See Appendix H. □

Any controller Σ_f can be factorized into a $\Sigma_{\text{core}} \in \mathcal{L}_{\text{core}}(\Theta)$ and the model $\Sigma_{\min}$. Any controller Σ_{core} can be lifted into a plant Σ_f in which the acquired $\Sigma_{\text{full}} \star \Sigma_f$ may have a non-minimal representation. Theorem 5.2 ensures that no conservatism is introduced when searching over controllers Σ_f as compared to performing a search over Σ_{core} via Problem 4.

Remark 5.1. *Theorem 5.2 justifies the requirement of Assumption 5 in full-order and structured synthesis of optimization algorithms. The interconnection* $\tilde{F}^t \star (P \star \Sigma_{full}) = \mathbf{m} \star (P \star \Sigma_{full}) = (\mathbf{m} \star P) \star \Sigma_{full} = P^{\mathbf{m}} \star \Sigma_{full}$ *(extended test system) formed by using the test quadratics* (42) *admits the following description with* (45)*:*

$$P^{\mathbf{m}} \star \Sigma_{full} : \qquad \left(\begin{array}{c:c|cc} A^{\mathbf{m}} & -B_2^{\mathbf{m}}\Gamma & 0 & B_2^{\mathbf{m}} \\ \hdashline 0 & I_{sc} & I_{sc} & 0 \\ \hline C_1^{\mathbf{m}} & -D_{12}^{\mathbf{m}}\,\Gamma & 0 & D_{12}^{\mathbf{m}} \\ C_2^{\mathbf{m}} & \Phi - D_2^{\mathbf{m}}\,\Gamma & 0 & D_2^{\mathbf{m}} \end{array}\right). \tag{94}$$

A state-coordinate change with $\begin{pmatrix} I & -\Pi \\ 0 & I \end{pmatrix}$ *and using the regulator equation* (47a) *(following Appendix D.1) leads to the following equivalent representation* $P^{\mathbf{m}}_{full}$ *of* $P^{\mathbf{m}} \star \Sigma_{full}$*:*

$$P^{\mathbf{m}}_{full} : \qquad \left(\begin{array}{c:c|cc} A^{\mathbf{m}} & B_1^{\mathbf{m}} & \Pi & B_2^{\mathbf{m}} \\ \hdashline 0 & I_{sc} & I_{sc} & 0 \\ \hline C_1^{\mathbf{m}} & D_1^{\mathbf{m}} & 0 & D_{12}^{\mathbf{m}} \\ C_2^{\mathbf{m}} & D_{21}^{\mathbf{m}} & 0 & D_2^{\mathbf{m}} \end{array}\right). \tag{95}$$

The existence of a controller Σ_f *that internally stabilizes* $P^{\mathbf{m}}_{full}$ *is guaranteed iff* $\left(\begin{pmatrix} A^{\mathbf{m}} & B_1^{\mathbf{m}} \\ 0 & I_{sc} \end{pmatrix}, \begin{pmatrix} \Pi & B_2^{\mathbf{m}} \\ I_{sc} & 0 \end{pmatrix}\right)$ *is stabilizable and* $\left(\begin{pmatrix} A^{\mathbf{m}} & B_1^{\mathbf{m}} \\ 0 & I_{sc} \end{pmatrix}, \begin{pmatrix} C_1^{\mathbf{m}} & D_1^{\mathbf{m}} \end{pmatrix}\right)$ *is detectable. The former is assured by Assumption 3 since* $(A^{\mathbf{m}}, B_2^{\mathbf{m}})$ *is stabilizable, while the latter is exactly enforced by Assumption 5.*

Given any structured controller Σ_{core} *satisfying* $(\Sigma_{core})_{ss} \in \mathcal{L}_{core}(\Theta)$ *with internal model* Σ_{min}*, Theorem 5.2 ensures that there exists a* Σ_f *with* $K = \Sigma_{min}\Sigma_{core} = \Sigma_{full} \star \Sigma_f$*. Therefore, Assumption 5 must be satisfied to ensure the existence of a structured controller* $(\Sigma_{core})_{ss} \in \mathcal{L}_{core}(\Theta)$ *in the synthesis of convergent optimization algorithms.*

Given a network P and a full-order model Σ_{full} from (91), the original plant to be controlled by Σ_f is $P \star \Sigma_{\text{full}}$. Figure 19 visualizes the overall system setup for ρ-exponentially weighted synthesis. The goal of full-order synthesis is to choose a subcontroller Σ_f such that the system formed by interconnection with input $\bar{q}$ and output $\bar{r}$ satisfies (89).

To enable a convex design of Σ_f, we introduce the following assumption.

Assumption 7. *Given a network* P *and an information constraint set* $\mathcal{L}_{\text{info}}$ *for* D_c*, there exists a convex set* $\mathcal{L}^D \subseteq \mathbb{R}^{n_u \times n_y}$ *with an LMI representation such that*

$$X(I - D_2 X)^{-1} \in \mathcal{L}_{\text{info}} \quad \textit{holds for all} \quad X \in \mathcal{L}^D. \tag{96}$$

As an example, if $\mathcal{L}_{\text{info}}$ only enforces block-lower-triangularity of D_c and $D_2 \in \mathcal{L}_{\text{info}}$, then Assumption 7 is respected with $\mathcal{L}^D := \mathcal{L}_{\text{info}}$. If $D_2 = 0$ and $\mathcal{L}_{\text{info}}$ is a set with an LMI representation, we can as well choose $\mathcal{L}^D := \mathcal{L}_{\text{info}}$ to satisfy Assumption 7.

The full-order controller synthesis problem reads as follows.

Problem 5 (Full-Order Synthesis)**.** *Let* $\mathcal{O}_{m,L}$ *be a class of operators,* $\mathcal{L}^D$ *an information structure set derived from* $\mathcal{L}_{\text{info}}$*, and* P *a network model such that Assumptions 1-7 hold. For a rate* $\rho > 0$ *and with the full-order internal model* Σ_{full} *from* (91)*, find a controller* Σ_f *and filter coefficients* $\boldsymbol{\lambda}$ *such that the following conditions hold:*

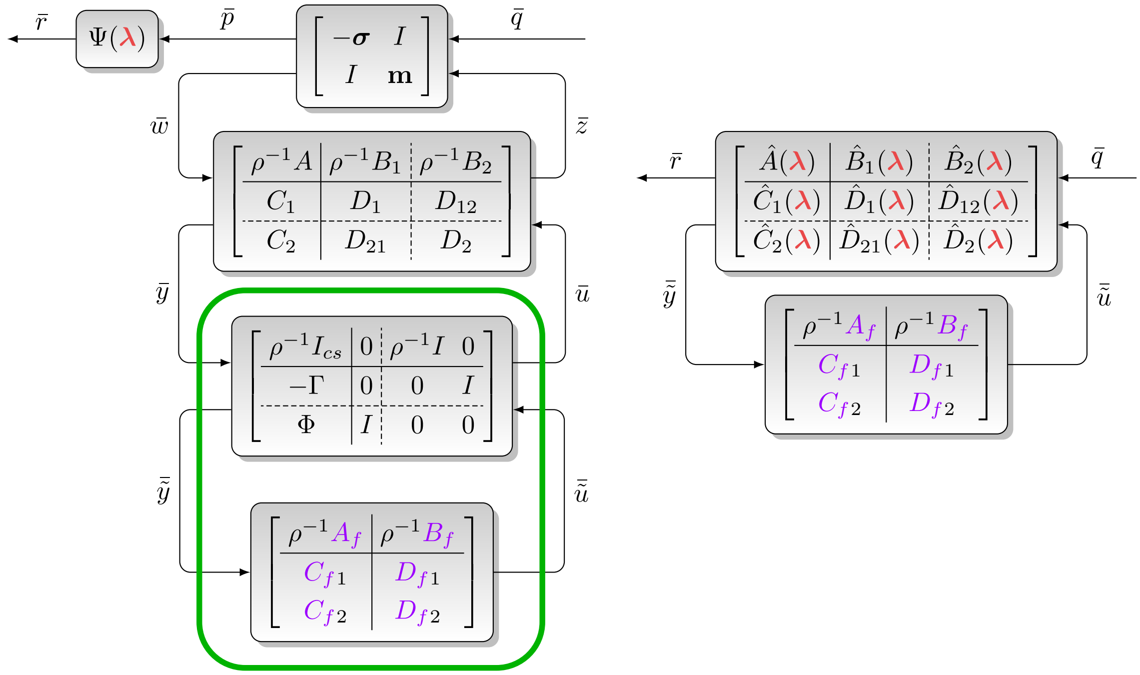


Figure 19: Filtering framework for synthesis with full-order models

1. *The constraint* $D_{f2} \in \mathcal{L}^D$ *is obeyed.*
2. *The LMI in* (89) *is feasible for the system* $\hat{G}(\lambda) = \Psi(\lambda)(\Sigma_{m,L} \star \bar{P} \star (\bar{\Sigma}_{full} \star \bar{\Sigma}_f))$.

For fixed filter coefficients λ, the design of Σ_f in Problem 5 can be convexified as described in Appendix I. For a fixed controller Σ_f, the search over λ is convex. This permits the minimization of ρ by the alternating solution of convex programs for (λ, Σ_f) and bisection in ρ.

Proposition 5.3. *For any solution of Problem 5, the controller* $K := \Sigma_{full} \star \Sigma_f$ *leads to a well-posed algorithm* $F \star (P \star K)$ *which converges exponentially with rate* ρ *for all* $F \in \mathcal{O}_{m,L}$.

Proof. Due to Proposition 5.2, the choice of K guarantees that the regulator equations (47) admit a solution. Then the claim follows by applying Theorem 5.1. □

6 Numerical Examples

Code to generate the following examples is publicly available[1]. All code was written in MATLAB (R2023b). For structured synthesis, the nonconvex optimization problems were solved using `hinfstruct` with fixed matrices Θ_{22} [93]. For full-order internal models, the linear matrix inequalities were solved using LMILab [94] from the MATLAB Robust Control Toolbox. Synthesis and controller recovery in the provided experiments always impose Kronecker structure.

6.1 Two Operator Splitting over Networks

Our first example involves constrained optimization (Problem 1 for $m = (m_1, 0)$, $L = (L_1, \infty)$, $0 < m_1 < L_1 < \infty$), with $f_2 = I_{\mathcal{Z}}$ for a closed convex set $\mathcal{Z}$. Hence the resolvent $(I + \gamma\partial f_2)^{-1}$ is the projection

[1] https://github.com/jarmill/composite_opt

operator onto $\mathcal{Z}$. This scenario will be explored with a specific network and with a parameter sweep. The information constraint $\mathcal{L}_{\text{info}}$ enforces that $D_{\text{core}2}$ has the block-sparsity pattern $\begin{pmatrix} \bullet & 0 \\ \bullet & \bullet \end{pmatrix}$, in which $\bullet$ denotes possibly nonzero entries. This block-sparsity pattern therefore allows for proximal evaluations of both ∂f_1 and ∂f_2.

6.1.1 Single Delay

This scenario includes network dynamics of one delay before and after the ∂f_1 oracle and zero delays before and after ∂f_2. The transfer matrix of the corresponding network dynamics P is

$$P : \left(\begin{array}{cc|cc} 0 & 0 & \mathbf{z}^{-1} & 0 \\ 0 & 0 & 0 & 1 \\ \hline \mathbf{z}^{-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right) \otimes I_c. \tag{97}$$

With $f_1 \in \mathcal{S}_{1,5}$, we solve Problem 5 to create a Kronecker-structured optimization algorithm after four Synthesis-Analysis iterations. Table 4 reports the rates ρ obtained by bisection, with the top row showing the values after designing subcontrollers Σ_f, while the bottom row depicting those after analysis by searching filter parameters $\boldsymbol{\lambda}$. Along these alternating steps, we note that the rates are monotonically decreasing as the interation number increases.

	Iter. 1	Iter. 2	Iter. 3	Iter. 4
Synthesis	0.8428	0.8254	0.8147	0.8079
Analysis	0.8395	0.8228	0.8131	0.8065

Table 4: Certified bounds on ρ for $\nu_{\max} = 3$ and during four Synthesis/Analysis iterations

The subcontroller Σ_f has $10c$ states. Attaching the $2c$-state full-order internal model leads to the controller $K = \Sigma_{\text{full}} \star \Sigma_{\text{f}}$ with $12c$ states, and hence to an algorithm where $G = P \star K$ has $14c$ states.

Figure 20 plots an execution of the algorithm with $\rho < 0.8065$ if minimizing a strongly convex quadratic function for $c = 7$ with the unconstrained optimum β_Q and an L_1 norm constraint as in

$$\beta^* = \operatorname*{argmin}_{\beta \in \mathbb{R}^7} \frac{1}{2}(\beta - \beta_Q)^\top Q(\beta - \beta_Q) + \chi_{\|\cdot\|_1 \leq 110}(\beta). \tag{98}$$

The matrix $Q = Q^\top$ is positive definite with eigenvalues between 1 and 5. The nonzero initial state $x_0 \in \mathbb{R}^{98}$ is randomly drawn from a uniform distribution of integers between $\{-100, \dots, 100\}$. The optimal solution of this example is $\beta^* = (-37.7643, 0, -67.3062, 0, 4.9294, 0, 0)^\top$. The top plots of Figure 20 show z_k, w_k, and x_k, respectively, over the course of $T = 40$ iterations starting from x_0. In the top-left plot, each curve is a trace of z_k^i for $k \in \{1, \dots, 40\}$ in each coordinate $i \in \{1, \dots, 7\}$. Similar traces are shown for the other plots of w_k and x_k. The bottom-left plot shows the elementwise consensus error $|z_k^i - z_{avg,k}|$ for $i \in \{1, 2\}$ with respect to the average $z_{avg,k} := \frac{1}{2}(z_k^1 + z_k^2)$, the bottom-center plot displays the elementwise optimality condition $|w^1 + w^2|$, while the bottom-right plot shows the gap between the function value and the optimal value $f^* = 1.4523 \times 10^4$.

6.1.2 Parameter Sweep

We now perform a parameter sweep over L_1, the delays in the networks, and the sparsity patterns of the controllers. We increase L_1 from 1 to 500 with fixed $m_1 = 1$, and we introduce a delay of $h_z \in \mathbb{N}$ time steps

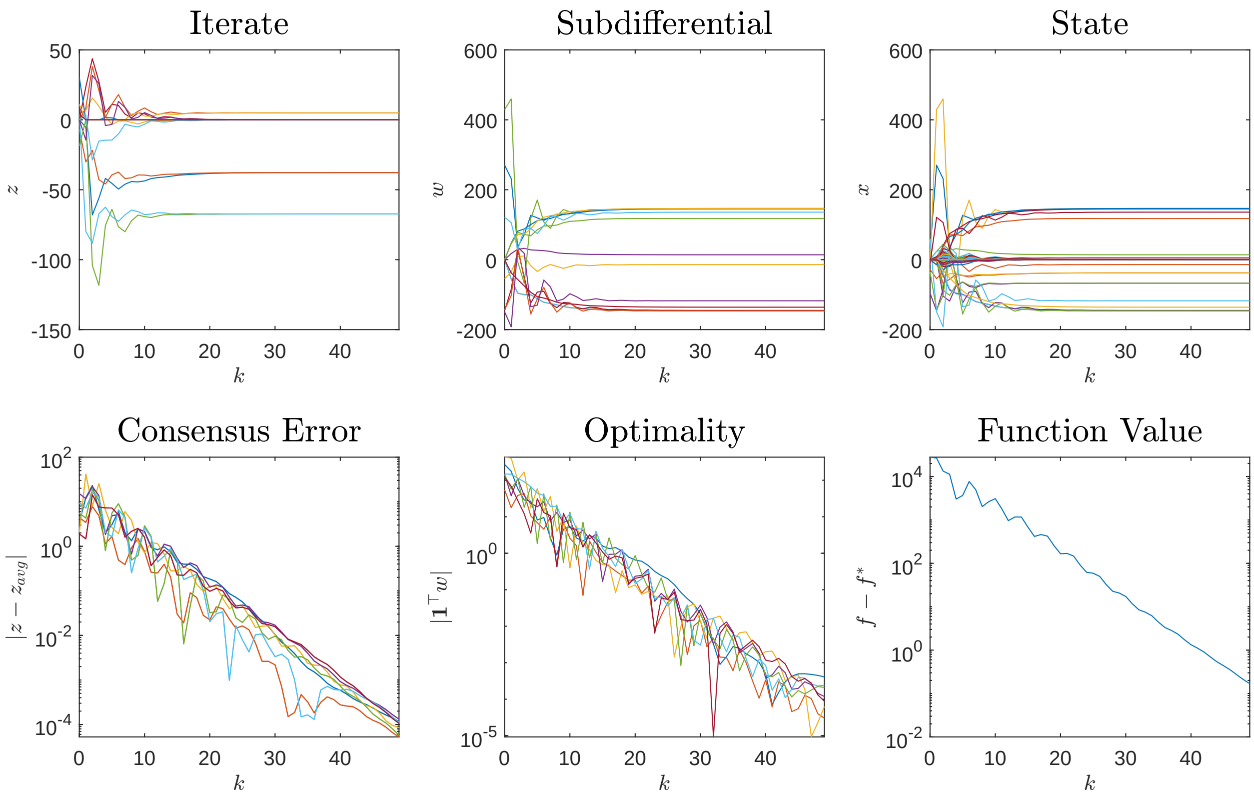


Figure 20: Quadratic program (98) solved with designed algorithm for $\rho < 0.8065$ under 1-step delays

before and $h_w \in \mathbb{N}$ time steps after the evaluation of ∂f_1, as modeled by a network with the transfer matrix

$$P^h : \left(\begin{array}{cc|cc} 0 & 0 & \mathbf{z}^{-h_w} & 0 \\ 0 & 0 & 0 & 1 \\ \hline \mathbf{z}^{-h_z} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right) \otimes I_c. \tag{99}$$

Synthesis is performed in each case with identity filters, and no subsequent Analysis/Synthesis iterations are performed after the initial synthesis. Figure 21 plots the rate bounds ρ as a function of L_1/m_1. The before-∂f_1 delay h_z increases from 0 to 3 vertically in the plot, and the after-∂f_1 delay increases from 0 to 3 horizontally in the plot. The title of each subplot lists the tuple $h = (h_z, h_w)$. The colors of the curves match with the those of the employed sparsity patterns of $D_{\text{core}2}$ in the following list:

$$\mathcal{L}_{\text{info}} : \quad \begin{pmatrix} \bullet & 0 \\ \bullet & \bullet \end{pmatrix}, \quad \begin{pmatrix} \bullet & 0 \\ 0 & \bullet \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ \bullet & \bullet \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & \bullet \end{pmatrix}. \tag{100}$$

The dotted black line displays the threshold $\rho = 1$, since algorithm convergence is only confirmed if the curve stays below this line. The dotted green curves in Figure 21 are the rates computed by searching over static subcontrollers (Theorem 4.3). If compared to the full dynamic algorithms, the static ones tend to have worse LMI-certified convergence rates for $L_1/m_1 \leq 10$, whereas the oppositie is true for $L_1/m_1 \geq 100$.

Algorithms which allow for proximal evaluation of both ∂f_1 and ∂f_2 have the smallest worst-case convergence rate among all algorithms generated by convex synthesis, because they offer the least restrictive sparsity patterns among block-lower-triangular-constrained D_c matrices.

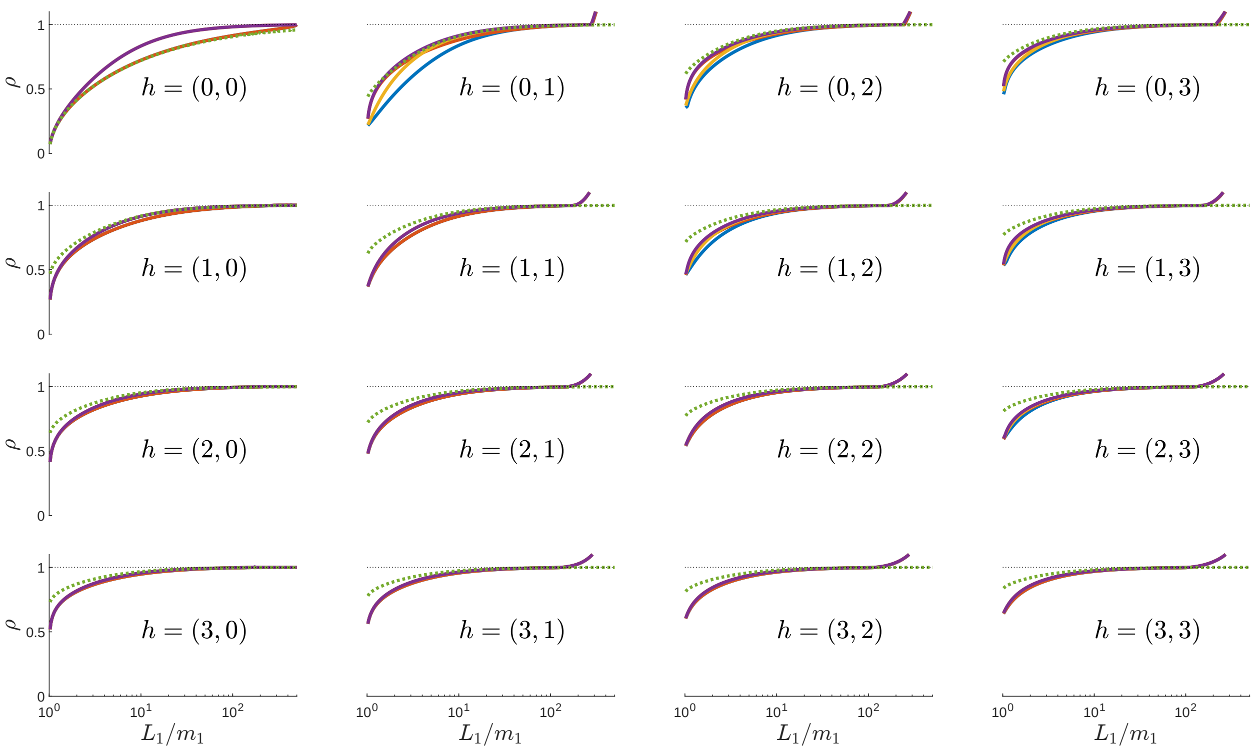


Figure 21: Sweeps over L_1/m_1, delays, and sparsity patterns with identity filters

6.1.3 Unstable Dynamics

We finish the two-operator example by considering the same operator class $\mathcal{O}_{m,L}$ for the following network with an unstable pole $\mathbf{z} = -1.1$ and sparsity pattern imposed on D_c:

$$P : \left(\begin{array}{cc|cc} 0 & 0 & \frac{1}{\mathbf{z}+1.1} & 0 \\ 0 & 0 & \frac{1}{\mathbf{z}-0.3} & 2 \\ \hline 1 & \frac{1}{\mathbf{z}} & \frac{2}{\mathbf{z}^2} & \frac{1}{\mathbf{z}-0.9} \\ 1 & 1-\frac{1}{\mathbf{z}^3} & 0 & 1 \end{array}\right) \otimes I_c, \qquad \mathcal{L}_{\text{info}} : \begin{pmatrix} \bullet & 0 \\ \bullet & \bullet \end{pmatrix}. \tag{101}$$

The plant regulator equation in (47) have the unique solution

$$\Pi = \begin{pmatrix} 0 & -\mathbf{1}_4 \\ -1 & 0 \\ -3 & 0 \\ -4.2 & 0 \\ -4.2 & 0 \\ 10 & 0 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} -2.1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \Phi = \begin{pmatrix} 5.8 & 0 \\ 0 & 1 \end{pmatrix}. \tag{102}$$

The unstable channel dynamics are stabilized by the synthesized controller. Figure 22 displays a trajectory solving the optimization problem in (98), in the same manner as Figure 20.

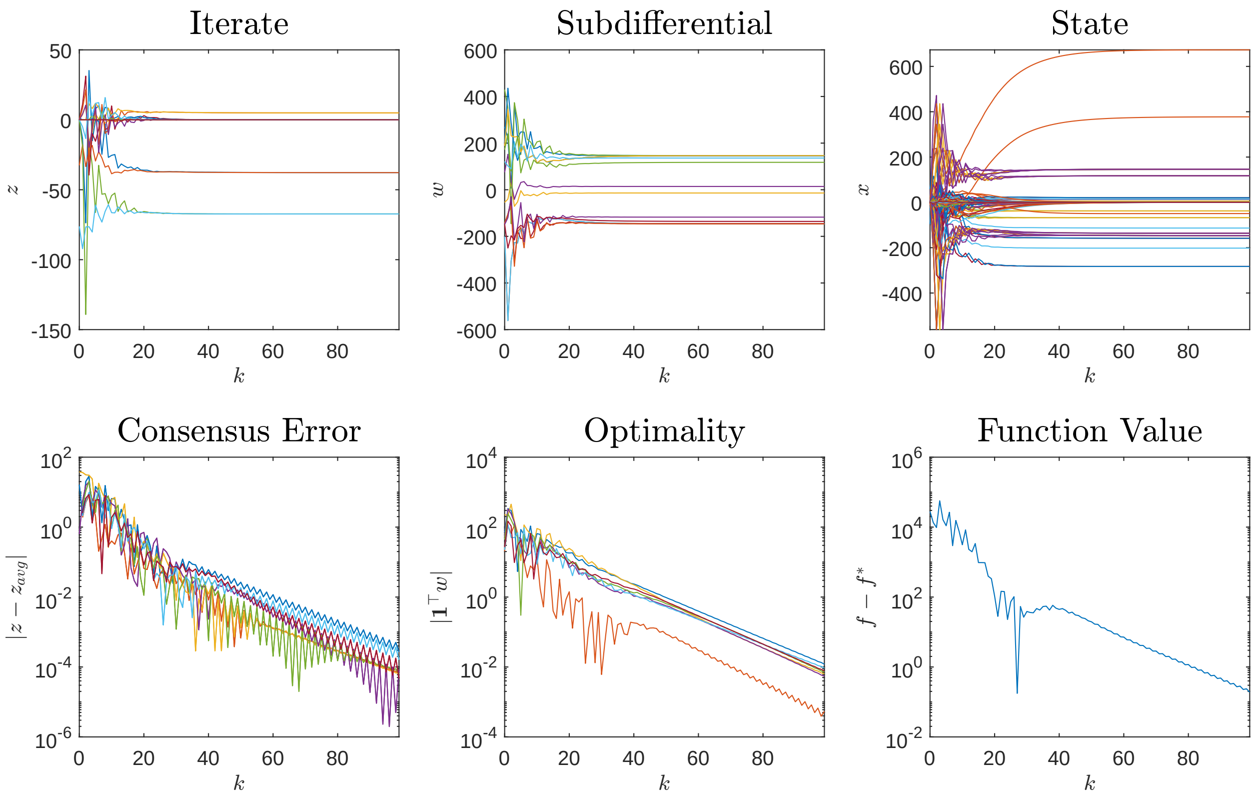


Figure 22: Problem (98) solved with designed algorithm for $\rho < 0.9476$ and the unstable network in (101)

6.2 Case Study: Image Denoising with Transmission Delay

We provide an example of image denoising using total variation regularization. The spatial dimensions of the image are n_x and n_y, and the image has n_c color channels (three for RGB). The values of each pixel are normalized to lie within $[0, 1]$. The goal is to recover a clean image $X^\star \in [0,1]^{n_x \times n_y \times n_c}$ from a noisy observation $Y = X^\star + \omega$ corrupted by additive noise ω. We formulate the recovery of the original image $X^\star$ as an optimization problem in the variable $\beta := \mathrm{vec}(X)$:

$$\min_{\beta \in \mathbb{R}^{n_x \times n_y \times n_c}} \quad \underbrace{\frac{\lambda_{\text{ridge}}}{2}\|\beta\|_2^2}_{f_1(\beta)} + \underbrace{\frac{1}{2}\|\mathrm{vec}(Y) - \beta\|_2^2}_{f_2(\beta)} + \underbrace{\lambda_{\text{TV}}\mathrm{TV}(\beta) + \chi_{[0,1]}(\beta)}_{f_3(\beta)}. \tag{103}$$

Here, $\mathrm{TV}(\beta)$ denotes the total variation of the image defined as

$$\mathrm{TV}(\beta) = \sum_{c=1}^{n_c}\sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\sqrt{(\Delta_x X_{i,j,c})^2 + (\Delta_y X_{i,j,c})^2} \tag{104}$$

and with the difference operators

$$(\Delta_x X)_{i,j,c} = \begin{cases} X_{i+1,j,c} - X_{i,j,c}, & i < n_x, \\ 0, & i = n_x, \end{cases} \qquad (\Delta_y X)_{i,j,c} = \begin{cases} X_{i,j+1,c} - X_{i,j,c}, & j < n_y, \\ 0, & j = n_y. \end{cases} \tag{105}$$

We demonstrate this approach on the denoising of a pepper-garlic test image, as shown in Figure 23. The original image X^* is in Figure 23a. The noise-corrupted image $Y = X^* + \omega$ is in Figure 23b, in which each $\omega_{i,j,c}$ is i.i.d. randomly sampled from an normal distribution with mean 0.01 and standard deviation 0.2.

The denoised image β^* found by solving (103) with Total Variation regularization parameter $\alpha_{\mathrm{TV}} = 0.15$ is in Figure 23c.

(a) Original Image X^* (b) Noisy Image Y (c) Denoised Image β^*

Figure 23: Sample pepper-garlic images used for problem (103).

We solve the total-variation denoising problem in (103) using a two-operator splitting algorithm with $f_1 \in \mathcal{S}_{1,1}$, and $f_2 \in \mathcal{S}_{0,\infty}$. The function f_3 admits the proximal operator

$$\mathrm{prox}_{\gamma f_2}(z) = \mathrm{Proj}_{[0,1]}\left[\mathrm{prox}_{\gamma \mathrm{TV}}(z)\right],$$

which follows from adjusting the first-order optimality condition of $\mathrm{prox}_{\gamma\mathrm{TV}}$ as given in [95]. The proximal operator of TV itself is computed with an inner iteration given in [95].

We consider the case where the noisy image Y is only available remotely, involving a delay of $h \in \mathbb{N}$ steps before and after the evaluation of ∇f_1. The network is an instance of (99) with a symmetric delay $h = h_z = h_w$. Perfect transmission means to choose $h = 0$, and delayed transmission involves $h > 0$. Our reference algorithm is the Douglas-Rachford [83] with parameters $(\gamma, \lambda) = (1, 2)$ with a description of

$$K_{DR} = \left[\begin{array}{c|cc} 1 & -2 & -2 \\ \hline 1 & -1 & 0 \\ 1 & -2 & -1 \end{array}\right]. \tag{106}$$

We solve (103) by applying the Douglas-Rachford algorithms for different values of delays $h \in \{0, \dots, 5\}$. We compare these results to an algorithm that is specifically synthesized for the given delay-h transmission network model using minimal and full-order internal models. The results are summarized in Table 5.

Table 5: Convergence rates ρ for the denoising problem (103) with different transmission delays h using the Davis-Yin splitting method and our synthesized algorithms (structured and full-order).

Delay h	Convergence rate ρ		
	Douglas-Rachford	Syn. (structured synthesis)	Syn. (full-order synthesis)
0	0.0134	.01	0.064
1	2.4149	0.5932	0.2092
2	1.4159	0.7223	0.3483
3	1.6184	0.7915	0.4497
4	1.3481	0.8365	0.5249
5	1.4199	0.8568	0.5827

While all methods perform comparably well in the absence of delays, convergence of the Douglas-Rachford splitting scheme cannot be certified once delays are introduced. In contrast, our synthesized algorithms (both of full and reduced order) retain convergence rates well below one across all tested delay values. We observe an advantage of using full over reduced order algorithms in terms of smaller convergence rates.

Figure 24 plots the image iterates z_k^2 (input to ∂f_2) generated by the Douglas-Rachford scheme as k increases (vertical) and as h increases (horizontal). The denoised image is recovered only at $h = 0$ (first column), while at all other delays h the Douglas-Rachford algorithm is nonconvergent. Figure 25 plots the same iterates z_k^2 generated by the synthesized algorithms with full-order internal models. The denoised image is found in the presence of delays.

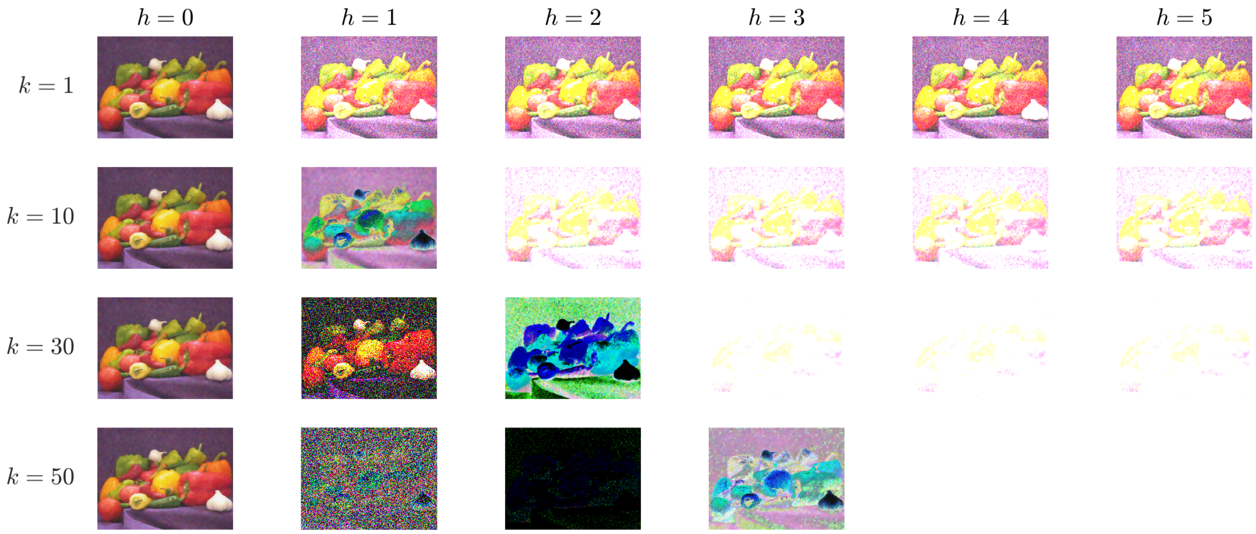


Figure 24: Douglas-Rachford algorithm vs. delay h and iteration k. Nonconvergent when $h > 0$.

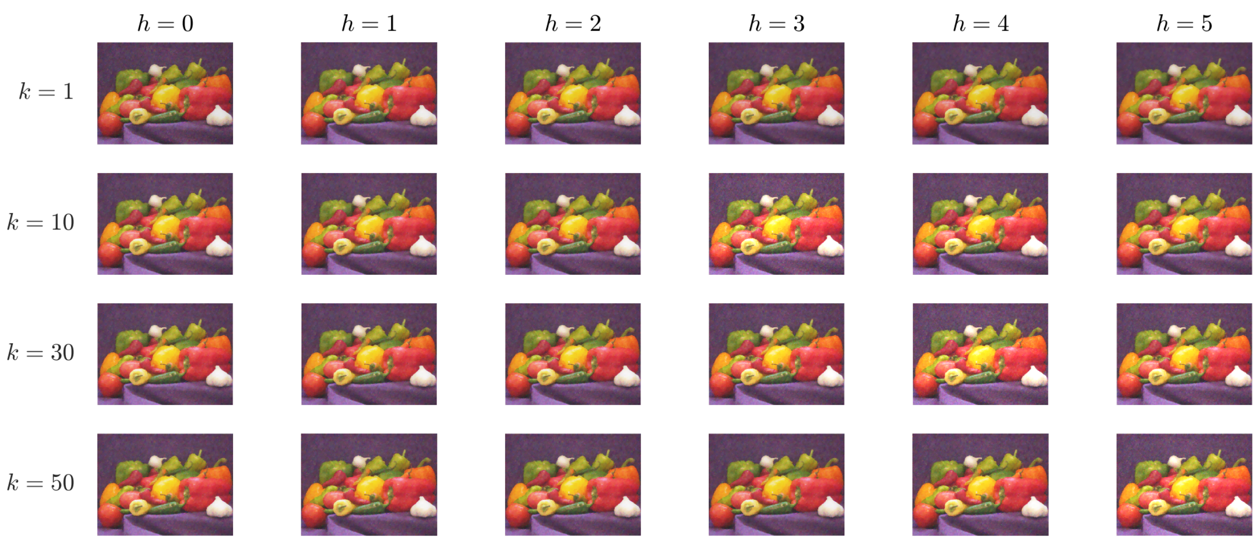


Figure 25: Full-order synthesized algorithm vs. delay h and iteration k. Convergent when $h > 0$.

Figure 26 plots another example of image denoising. The original image X^* of an Uluguru forest tree frog is on the top-left, and the additively-noise-corrupted image Y is shown on the the top-right. A delay of $h = 4$ is imposed in the network (99). The Douglas-Rachford algorithm (bottom-left image) and our synthesized algorithm (bottom-right image) are run for 15 iterations. The Douglas-Rachford algorithm over this time-delayed network is unstable, and generates a severely corrupted image after 50 iterations. Our synthesized algorithm is convergent, and the beginnings of a denoising process is visible in the bottom-right image around the green background, the green skin, the purple leaf, and the black eye.

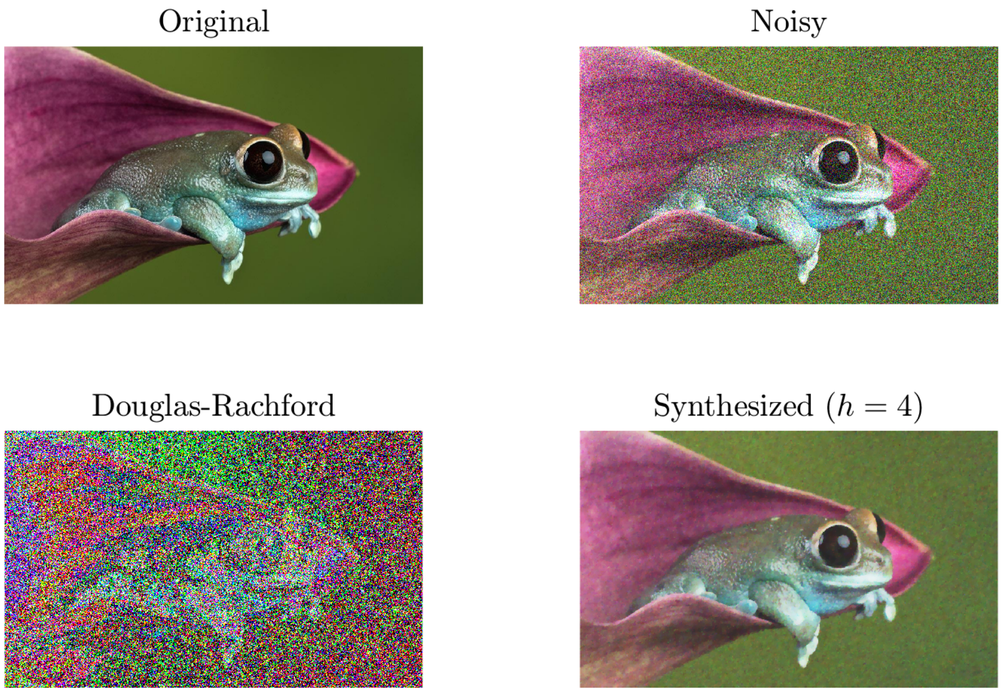


Figure 26: Denoising of an Uluguru forest tree frog image after 50 iterations with $h = 4$.

In particular, representations of the full-order and reduced-order algorithms for a delay of $h = 1$ are

$$K_{\text{full}}^{h=1} : \left(\begin{array}{cccc|cc} 0.716 & 1.387 & -0.352 & -9.192 & 0.965 & 1.249 \\ 0.487 & -0.255 & 0.601 & 57.624 & -0.231 & -0.718 \\ -0.117 & -0.238 & 0.065 & 1.811 & -0.167 & -0.0503 \\ -0.003 & -0.005 & 0.002 & 0.048 & -0.004 & -0.001 \\ \hline -1.304 & -0.798 & -0.572 & -75.342 & -0.478 & 0 \\ -0.795 & -1.401 & 0.269 & 1.496 & -0.969 & -1.000 \end{array}\right) \otimes I_c, \quad K_{\text{red}}^{h=1} : \left(\begin{array}{c|cc} 1 & -0.196 & -0.196 \\ \hline 1 & 0.131 & 0 \\ 1 & -0.483 & -0.352 \end{array}\right) \otimes I_c.$$

6.3 Composite Optimization Algorithms for Six Operators

We demonstrate composite optimization for $s = 6$ functions and a network of order $9c$ described by

$$P(\mathbf{z}) = \left(\begin{array}{cccccc|cccccc} 0 & 0 & 0 & 0 & 0 & 0 & \frac{0.5}{\mathbf{z}-0.5} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{\mathbf{z}+0.4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{z}^{-1} & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline \frac{0.5}{\mathbf{z}-0.5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \mathbf{z}^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{-1}{\mathbf{z}+0.4} & \mathbf{z}^{-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{z}^{-2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \otimes I_c. \tag{107}$$

The parameters are $m = (0, 1, -1, 1, 1, 0)$ and $L = (5, 2, 1, 6, 1, \infty)$. Since $\sum_{i=1}^{6} m_i = 1$, the overall optimization problem is 1-strongly convex.

Both for the network P in (107) and the direct interconnection P^0 in (32), Table 6 reports the guaranteed rates ρ computed by synthesis with identity filters and a full-order internal model (91), together with the respective sparsity patterns $\mathcal{L}_{\text{info}}$ of D_c.

Table 6: Sparsity restrictions $\mathcal{L}_{\text{info}}$ on D_c and rates obtained by synthesis

Pattern $\mathcal{L}_{\text{info}}$	$\begin{pmatrix} \bullet & 0 & \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \end{pmatrix}$	$\begin{pmatrix} \bullet & 0 & 0 & 0 & 0 & 0 \\ \bullet & \bullet & 0 & 0 & 0 & 0 \\ \bullet & \bullet & \bullet & 0 & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \end{pmatrix}$	$\begin{pmatrix} \bullet & 0 & 0 & 0 & 0 & 0 \\ 0 & \bullet & 0 & 0 & 0 & 0 \\ 0 & 0 & \bullet & 0 & 0 & 0 \\ 0 & 0 & 0 & \bullet & 0 & 0 \\ 0 & 0 & 0 & 0 & \bullet & 0 \\ 0 & 0 & 0 & 0 & 0 & \bullet \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \bullet \end{pmatrix}$
ρ with P in (107)	0.9334	0.9355	0.9394	0.9559
ρ with P^0 in (32)	0.7432	0.7681	0.7681	0.8821

All the listed sparsity patterns ensure that $\mathcal{D}$ is block-lower-triangular when the controller is interconnected with P in (107), while this is not true for the first sparsity pattern and P^0.

7 Conclusion

A first-order networked composite optimization algorithms can be interpreted as an interconnection of a static map $F \in \mathcal{O}_{m,L}$, a given network P, and a to-be-designed controller K. In this paper, we unveil the modeling power of this interconnection-based paradigm. We first show that the convergence of the the corresponding well-posed algorithmic interconnection $F \star (P \star K)$ for all $F \in \mathcal{O}_{m,L}$ requires satisfaction of Robust Stability and Regulator Equation conditions. These necessary conditions for algorithmic convergence induce structural constraints on any controller K, in that it must satisfy a minimal order bound based on information constraints $\mathcal{L}_{\text{info}}$, and that it must factorize into the cascade of a core subcontroller Σ_{core} and an internal model $\Sigma_{\min}$. Moreover, we address how these structural properties permit the design of novel composite optimization algorithms with certified exponential convergence rates, even under the presence of network dynamics. The results are illustrated with several examples, include the tailored design of algorithms for image denoising under delayed gradient information transmission.

Possible extensions include generalizations of algorithm synthesis from strongly convex to merely convex problems with guarantees of sublinear convergence, the use of integral quadratic constraint synthesis to generate algorithms that can solve solve variational inclusion problems, and the reduction of the conservatism of the employed uncertainty characterizations for the class $\mathcal{O}_{m,L}$. Finally, we aim at developing convex methods for the joint search of stability filter coefficients and subcontrollers.

Acknowledgements

The authors would like to thank Lennart Dalhues, Dennis Gramlich, Emna Ayadi, Manuel Zobel, Adrien Taylor, Manu Upadhyaya, Matthias Müller, Michael Mühlenbach, Timm Faulwasser, Nicola Bastianello, Jaap Eising, Zhiyu He, Niklas Schmid, and Mathias Hudoba de Badyn for discussions about structures and synthesis of optimization algorithms.

A Background on Linear Systems

Given a linear system $G : (x^0, u) \to y$ with internal state x and matrix representation (A, B, C, D), performing a coordinate change on the state $x \mapsto T^{-1}x$ of a linear system G yields a system $\tilde{G}(T^{-1}x_0, u) \to y$. The expression of this coordinate change induces a transformation of the representing matrices as in

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{c|c} T^{-1}AT & T^{-1}B \\ \hline CT & D \end{array}\right]. \tag{108}$$

The cascade (series) interconnection between two systems G_1 and G_2 has the representations

$$\left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right] \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right] = \left[\begin{array}{cc|c} A_1 & B_1C_2 & B_1D_2 \\ 0 & A_2 & B_2 \\ \hline C_1 & D_1C_2 & D_1D_2 \end{array}\right] = \left[\begin{array}{cc|c} A_2 & 0 & B_2 \\ B_1C_2 & A_1 & B_1D_2 \\ \hline D_1C_2 & C_1 & D_1D_2 \end{array}\right]. \tag{109}$$

The block-diagonal concatenation of systems G_1 and G_2 has a representation

$$\operatorname{blkdiag}\left(\left[\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right], \left[\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right] \right) = \left[\begin{array}{cc|cc} A_1 & 0 & B_1 & 0 \\ 0 & A_2 & 0 & B_2 \\ \hline C_1 & 0 & D_1 & 0 \\ 0 & C_2 & 0 & D_2 \end{array}\right]. \tag{110}$$

Given two linear systems $G : (x^0, (w, u)) \to (z, y)$ and $G^c : (\xi^0, (y, d)) \to (u, e)$ with

$$\begin{pmatrix} x_{k+1} \\ \hline z_k \\ y_k \end{pmatrix} = \left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_1 & D_{12} \\ C_2 & D_{21} & D_2 \end{array}\right) \begin{pmatrix} x_k \\ \hline w_k \\ u_k \end{pmatrix}, \quad \begin{pmatrix} \xi_{k+1} \\ \hline u_k \\ b_k \end{pmatrix} = \left(\begin{array}{c|cc} \mathbf{A} & \mathbf{B}_1 & \mathbf{B}_2 \\ \hline \mathbf{C}_1 & \mathbf{D}_1 & \mathbf{D}_{12} \\ \mathbf{C}_2 & \mathbf{D}_{21} & \mathbf{D}_2 \end{array}\right) \begin{pmatrix} \xi_k \\ \hline y_k \\ a_k \end{pmatrix}, \tag{111}$$

the signals $(x, w, z, y, \xi, u, b, a)$ are related by a under a linear constraint formed by assignment of their common (u, y) channels,

$$\left(\begin{array}{c} x_{k+1} \\ \xi_{k+1} \\ \hline z_k \\ b_k \\ \hdashline 0 \\ 0 \end{array}\right) = \left(\begin{array}{cc|cc:cc} A & 0 & B_1 & 0 & 0 & B_2 \\ 0 & \mathbf{A} & 0 & \mathbf{B}_2 & \mathbf{B}_1 & 0 \\ \hline C_1 & 0 & D_1 & 0 & 0 & D_{12} \\ 0 & \mathbf{C}_1 & 0 & \mathbf{D}_2 & \mathbf{D}_{21} & 0 \\ \hdashline C_2 & 0 & D_{21} & 0 & -I & D_2 \\ 0 & \mathbf{C}_1 & 0 & \mathbf{D}_{12} & \mathbf{D}_1 & -I \end{array}\right) \left(\begin{array}{c} x_k \\ \xi_k \\ \hline w_k \\ a_k \\ \hdashline y_k \\ u_k \end{array}\right). \tag{112}$$

If the matrix $E = I - D_2\mathbf{D}_1$ is invertible, then the star product $G \star G^c$ is the well-posed interconnection of G and G^c along their common u and y channels. The star product admits the unique description

$$\left(\begin{array}{c} x_{k+1} \\ \xi_{k+1} \\ \hline z_k \\ b_k \end{array}\right) = \left[\left(\begin{array}{cc|cc} A & 0 & B_1 & 0 \\ 0 & \mathbf{A} & 0 & \mathbf{B}_2 \\ \hline C_1 & 0 & D_1 & 0 \\ 0 & \mathbf{C}_1 & 0 & \mathbf{D}_2 \end{array}\right) + \left(\begin{array}{cc} 0 & B_2 \\ \mathbf{B}_1 & 0 \\ \hline 0 & D_{12} \\ \mathbf{D}_{21} & 0 \end{array}\right) \begin{pmatrix} I & -D_2 \\ -D_1 & I \end{pmatrix}^{-1} \left(\begin{array}{cc|cc} C_2 & 0 & D_{21} & 0 \\ 0 & \mathbf{C}_1 & 0 & \mathbf{D}_{12} \end{array}\right)\right] \left(\begin{array}{c} x_k \\ \xi_k \\ \hline w_k \\ a_k \end{array}\right), \tag{113}$$

which may be acquired by eliminating (u_k, y_k) from (111) using a Schur complement. The common signals (u, y) can then be uniquely recovered from $((x, \xi), (w, a))$ as

$$\begin{pmatrix} y_k \\ u_k \end{pmatrix} = \begin{pmatrix} I & -D_2 \\ -\mathbf{D}_1 & I \end{pmatrix}^{-1} \begin{pmatrix} C_2 x_k + D_{21} w_k \\ \mathbf{C}_1 \xi_k + \mathbf{D}_{12} a_k \end{pmatrix}. \tag{114}$$

The star product $G_1 \star G_2$ is a property of the given representations in (111) for systems G_1, G_2. We only define the star product $G_1 \star G_2$ if E is invertible (the interconnection is well-posed). An alternative formula for computing the unique star product $G_1 \star G_2$ is

$$\left(\begin{array}{c} x_{k+1} \\ \xi_k \\ \hline z_k \\ b_k \end{array}\right) = \left[\left(\begin{array}{cc|cc} A & 0 & B_1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline C_1 & 0 & D_1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) + \left(\begin{array}{c:cc} 0 & B_2 & 0 \\ I & 0 & 0 \\ \hline 0 & D_{12} & 0 \\ 0 & 0 & I \end{array}\right) \mathbf{M}_E \left(\begin{array}{cc|cc} 0 & I & 0 & 0 \\ \hdashline C_2 & 0 & D_{21} & 0 \\ 0 & 0 & 0 & I \end{array}\right)\right] \left(\begin{array}{c} x_k \\ \xi_k \\ \hline w_k \\ a_k \end{array}\right) \tag{115}$$

with

$$\mathbf{M}_E := \begin{pmatrix} \mathbf{A} + \mathbf{B}_1 E^{-1} D_2 \mathbf{C}_1 & \mathbf{B}_1 E^{-1} & \mathbf{B}_2 + \mathbf{B}_1 E^{-1} D_2 \mathbf{D}_{12} \\ (I + \mathbf{D}_1 E^{-1} D_2)\mathbf{C}_1 & \mathbf{D}_1 E^{-1} & (I + \mathbf{D}_1 E^{-1} D_2)\mathbf{D}_{12} \\ \mathbf{C}_2 + \mathbf{D}_{21} E^{-1} D_2 \mathbf{C}_1 & \mathbf{D}_{21} E^{-1} & \mathbf{D}_2 + \mathbf{D}_{21} E^{-1} D_2 \mathbf{D}_{12} \end{pmatrix}. \tag{116}$$

If E is invertible, then the bracketed matrices in (113) and (115) are identical. The matrix $\mathbf{M}_E$ may also be expressed by using the identity $(I - D_2\mathbf{D}_{11})^{-1} D_2 = D_2(I - \mathbf{D}_{11} D_2)^{-1}$.

For completeness, we provide a proof of a folklore result in control theory involving associativity of star products. This associativity will be used to justify Assumption 2 for convergence of optimization algorithms.

Proposition A.1. *If the star products $(P_1 \star P_2) \star P_3$ and $P_2 \star P_3$ exist (form well-posed interconnections), then $P_1 \star (P_2 \star P_3)$ exists and $P_1 \star (P_2 \star P_3) = (P_1 \star P_2) \star P_3$.*

Proof. Let the systems $\{P_j\}_{j=1}^3$, $P_1 \star P_2$, and $(P_1 \star P_2) \star P_3$ have the descriptions,

$$P_j: \quad \begin{pmatrix} x_{k+1}^j \\ \hline z_k^j \\ y_k^j \end{pmatrix} = \left(\begin{array}{c|cc} A^j & B_1^j & B_2^j \\ \hline C_1^j & D_1^j & D_{12}^j \\ C_2^j & D_{21}^j & D_2^j \end{array}\right) \begin{pmatrix} x_k^j \\ \hline w_k^j \\ u_k^j \end{pmatrix}, \qquad \text{for all } \ j \in \{1,2,3\}, \tag{117a}$$

$$P_1 \star P_2: \quad \begin{pmatrix} \bar{x}_{k+1} \\ \hline \bar{z}_k^j \\ \bar{y}_k^j \end{pmatrix} = \left(\begin{array}{c|cc} \bar{A} & \bar{B}_1 & \bar{B}_2 \\ \hline \bar{C}_1 & \bar{D}_1 & \bar{D}_{12} \\ \bar{C}_2 & \bar{D}_{21} & \bar{D}_2 \end{array}\right) \begin{pmatrix} \bar{x}_k \\ \hline \bar{w}_k \\ \bar{u}_k \end{pmatrix}, \tag{117b}$$

$$(P_1 \star P_2) \star P_3: \quad \begin{pmatrix} \hat{x}_{k+1} \\ \hline \hat{z}_k^j \\ \hat{y}_k^j \end{pmatrix} = \left(\begin{array}{c|cc} \hat{A} & \hat{B}_1 & \hat{B}_2 \\ \hline \hat{C}_1 & \hat{D}_1 & \hat{D}_{12} \\ \hat{C}_2 & \hat{D}_{21} & \hat{D}_2 \end{array}\right) \begin{pmatrix} \hat{x}_k \\ \hline \hat{w}_k \\ \hat{u}_k \end{pmatrix}. \tag{117c}$$

The interconnection of (P_1, P_2, P_3) is performed using the assignments,

$$w^2 = y^1, \qquad w^3 = y^2, \qquad u^1 = z^2, \qquad u^2 = z^3. \tag{118}$$

The signals $(x^j, w^j, u^j, z^j, y^j)_{j=1}^3$ obey the linear relation

$$\begin{pmatrix} x_{k+1}^1 \\ x_{k+1}^2 \\ x_{k+1}^3 \\ \hline z_k^1 \\ y_k^3 \\ \hdashline 0 \\ 0 \\ \hdashline 0 \\ 0 \end{pmatrix} = \left(\begin{array}{ccc|cc:cc:cc} A^1 & 0 & 0 & B_1^1 & 0 & 0 & B_2^1 & 0 & 0 \\ 0 & A^2 & 0 & 0 & 0 & B_1^2 & 0 & 0 & B_2^2 \\ 0 & 0 & A^3 & 0 & B_2^3 & 0 & 0 & B_1^3 & 0 \\ \hline C_1^1 & 0 & 0 & D_1^1 & 0 & 0 & D_{12}^1 & 0 & 0 \\ 0 & 0 & C_2^3 & 0 & D_2^3 & 0 & 0 & D_{21}^3 & 0 \\ \hdashline C_2^1 & 0 & 0 & D_{21}^1 & 0 & -I & D_2^1 & 0 & 0 \\ 0 & C_1^2 & 0 & 0 & 0 & D_1^2 & -I & 0 & D_{12}^2 \\ \hdashline 0 & C_2^2 & 0 & 0 & 0 & D_{21}^2 & 0 & -I & D_2^2 \\ 0 & 0 & C_1^3 & 0 & D_2^3 & 0 & 0 & D_1^3 & -I \end{array}\right) \begin{pmatrix} x_k^1 \\ x_k^2 \\ x_k^3 \\ \hline w_k^1 \\ u_k^3 \\ \hdashline w_k^2 \\ u_k^1 \\ \hdashline w_k^3 \\ u_k^2 \end{pmatrix}. \tag{119}$$

We focus on proving invertibility properties of the lower-right D-block of (119) based on the existence of the star products $(P_1 \star P_2) \star P_3$ and $P_2 \star P_3$ from the proposition statement. Because $P_1 \star P_2$ and $P_2 \star P_3$ exist, the submatrices $\begin{pmatrix} -I & D_2^1 \\ D_1^2 & -I \end{pmatrix}$ and $\begin{pmatrix} -I & D_2^2 \\ D_1^3 & -I \end{pmatrix}$ of the bottom-right corner of (119) are both invertible. Decoupling (w^2, u^1) in (119) using a Schur complement yields

$$\left(\begin{array}{c:c:c} I & -\begin{pmatrix} 0 & D_{12}^1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -I & D_2^1 \\ D_1^2 & -I \end{pmatrix}^{-1} & 0 \\ \hdashline 0 & I & 0 \\ \hdashline 0 & -\begin{pmatrix} D_{21}^2 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -I & D_2^1 \\ D_1^2 & -I \end{pmatrix}^{-1} & I \end{array}\right) \left(\begin{array}{cc:cc:cc} D_1^1 & 0 & 0 & D_{12}^1 & 0 & 0 \\ 0 & D_2^3 & 0 & 0 & D_{21}^3 & 0 \\ \hdashline D_{21}^1 & 0 & -I & D_2^1 & 0 & 0 \\ 0 & 0 & D_1^2 & -I & 0 & D_{12}^2 \\ \hdashline 0 & 0 & D_{21}^2 & 0 & -I & D_2^2 \\ 0 & D_2^3 & 0 & 0 & D_3^1 & -I \end{array}\right) = \left(\begin{array}{cc:cc:cc} \bar{D}_1 & 0 & 0 & 0 & 0 & \bar{D}_{12} \\ 0 & D_2^3 & 0 & 0 & D_{21}^3 & 0 \\ \hdashline D_{21}^1 & 0 & -I & D_2^1 & 0 & 0 \\ 0 & 0 & D_1^2 & -I & 0 & D_{12}^2 \\ \hdashline \bar{D}_{21} & 0 & 0 & 0 & -I & \bar{D}_2 \\ 0 & D_2^3 & 0 & 0 & D_3^1 & -I \end{array}\right). \tag{120}$$

Because invertible matrices form a group under matrix multiplication, we have that

$$\begin{pmatrix} -I & \bar{D}_2 \\ D_3^1 & -I \end{pmatrix} \text{ is invertible iff } \left(\begin{array}{cc:cc} -I & D_2^1 & 0 & 0 \\ D_1^2 & -I & 0 & D_{12}^2 \\ \hdashline D_{21}^2 & 0 & -I & D_2^2 \\ 0 & 0 & D_3^1 & -I \end{array}\right) \text{ is invertible,} \tag{121}$$

and invertibility in (121) is met because $(P_1 \star P_2) \star P_3$ exists in the proposition statement. We apply another Schur complement to decouple (w^3, u^2) as

$$\left(\begin{array}{c:c:c} I & 0 & -\begin{pmatrix} 0 & \bar{D}_{12} \\ D^3_{21} & 0 \end{pmatrix}\begin{pmatrix} -I & \bar{D}_2 \\ D^3_1 & -I \end{pmatrix}^{-1} \\ \hdashline 0 & I & 0 \\ \hdashline 0 & 0 & I \end{array}\right)\left(\begin{array}{cc:cc:cc} \bar{D}_1 & 0 & 0 & 0 & 0 & \bar{D}_{12} \\ 0 & D^3_2 & 0 & 0 & D^3_{21} & 0 \\ \hdashline D^1_{21} & 0 & -I & D^1_2 & 0 & 0 \\ 0 & 0 & D^2_1 & -I & 0 & D^2_{12} \\ \hdashline \bar{D}_{21} & 0 & 0 & 0 & -I & \bar{D}_2 \\ 0 & D^3_2 & 0 & 0 & D^1_3 & -I \end{array}\right) = \left(\begin{array}{cc:cc:cc} \hat{D}_1 & \hat{D}_{12} & 0 & 0 & 0 & 0 \\ \hat{D}_{21} & \hat{D}_2 & 0 & 0 & 0 & 0 \\ \hdashline D^1_{21} & 0 & -I & D^1_2 & 0 & 0 \\ 0 & 0 & D^2_1 & -I & 0 & 0 \\ \hdashline \bar{D}_{21} & 0 & 0 & 0 & -I & \bar{D}_2 \\ 0 & D^3_2 & 0 & 0 & D^1_3 & -I \end{array}\right). \tag{122}$$

The $\hat{D}$ terms of the star product $(P_1 \star P_2) \star P_3$ from (117c) are visible in the top-right corner of (122). We can use the invertibility relation in (121) to generate the star product $P_1 \star (P_2 \star P_3)$ by eliminating the upper sub-block $\begin{pmatrix} 0 & 0 \\ 0 & D^2_{12} \end{pmatrix}$, instead of eliminating the lower sub-block $\begin{pmatrix} D^1_{21} & 0 \\ 0 & 0 \end{pmatrix}$ to obtain $(P_1 \star P_2) \star P_3$. Due to invertibility of all matrices involved in the well-posedness conditions, we therefore have $(P_1 \star P_2) \star P_3 = P_1 \star (P_2 \star P_3)$. ☐

B Properties of Functions in $\mathcal{S}_{m,L}$

This appendix collects several results of the class $\mathcal{S}_{m,L}$ as defined in Section 2.3.

B.1 Frechét Subdifferentials

The Frechét subdifferential of a function $f : \mathbb{R}^c \to \bar{\mathbb{R}}$ at a point x is [78, Eq. (16.2)]

$$\partial f(x) = \left\{ g \in \mathbb{R}^c \mid \liminf_{y \to x} \frac{f(y) - f(x) - g^\top (y - x)}{\|y - x\|} \geq 0 \right\}. \tag{123}$$

If a function f is p.c.c., then its Frechét subdifferential and its standard convex subdifferential are equal.

Proposition B.1. *The Frechét subdifferential of $f \in \mathcal{S}_{m,L}$ is $\partial f = \partial(f - m\mathbf{q}) + mI$.*

Proof. By [78, Proposition 1.107], the Frechét subdifferential of $f = (f - m\mathbf{q}) + m\mathbf{q}$ is $\partial f = \partial(f - mI) + mI$, given that $\mathbf{q}$ is strictly differentiable [72, Definition 9.17] everywhere in $\mathbb{R}^c$. ☐

Proposition B.2. *Let $f \in \mathcal{S}_{m,\infty}$. If $x \in \operatorname{dom}(f)$ is a local minimum of f, then $0 \in \partial f(x)$. If $m \geq 0$ and $x \in \mathbb{R}^c$ satisfies $0 \in \partial f(x)$ then x is a global minimum of f.*

Proof. The local optimality property follows from [78, Theorem 16.2] for optima of Frechét-differentiable functions. Because f is convex if $m \geq 0$, global optimality follows. ☐

B.2 The Sum-Rule and Consequences

Proposition B.3. *If $f_1 \in \mathcal{S}_{m_1,\infty}$ and $f_2 \in \mathcal{S}_{m_2,\infty}$ then $f_1 + f_2 \in \mathcal{S}_{m_1+m_2,\infty}$ and*

$$\partial f_1(x) + \partial f_2(x) \subset \partial(f_1 + f_2)(x) \quad \text{for all} \quad x \in \mathbb{R}^c. \tag{124}$$

If $\operatorname{relint}(\operatorname{dom}(f_1)) \cap \operatorname{relint}(\operatorname{dom}(f_2)) \neq \varnothing$, then equality holds:

$$\partial f_1(x) + \partial f_2(x) = \partial(f_1 + f_2)(x) \quad \text{for all} \quad x \in \mathbb{R}^c. \tag{125}$$

Proof. This follows from the trivial identity

$$(f_1 - m_1\mathbf{q}) + (f_2 - m_2\mathbf{q}) = (f_1 + f_2) - (m_1 + m_2)\mathbf{q}.$$

Since $\partial f_1 = \partial(f_1 - m_1\mathbf{q}) + m_1 I$, $\partial f_2 = \partial(f_2 - m_2\mathbf{q}) + m_2 I$ and $\partial(f_1 + f_2) = \partial((f_1 + f_2) - (m_1 + m_2)\mathbf{q}) + (m_1 + m_2)I$ (by definition), it suffices to apply the standard rules to the sum $(f_1 - m_1\mathbf{q}) + (f_2 - m_2\mathbf{q})$ of two p.c.c. functions. □

A combination with Proposition B.2 leads to the following result.

Corollary B.1. *For $f_1 \in \mathcal{S}_{m_1,\infty}$ and $f_2 \in \mathcal{S}_{m_2,\infty}$ consider the optimization problem*

$$\inf_{x\in\mathbb{R}^c} [f_1(x) + f_2(x)]. \tag{126}$$

If $\mathrm{relint}(\mathrm{dom}(f_1)) \cap \mathrm{relint}(\mathrm{dom}(f_2)) \neq \varnothing$ *and* $x^* \in \mathrm{dom}(f)$ *is a local minimum of* (126)*, then*

$$0 \in \partial f_1(x^*) + \partial f_2(x^*). \tag{127}$$

If $m_1 + m_2 \geq 0$ and $x^* \in \mathbb{R}^c$ satisfies (127), then x^* is a global minimum for (126).

B.3 Generalized Subgradient Inequalities

We slightly extend [73, Theorem 5] as in [92, Lemma 12] to the case of $L = \infty$ where $f \in \mathcal{S}_{m,\infty}$ are not necessarily differentiable everywhere.

Lemma B.1. *Given $f \in \mathcal{S}_{m,L}$, define $\sigma := (L - m)^{-1} \geq 0$ (satisfying $\sigma = 0$ and $\sigma L = 1$ for $L = \infty$) and*

$$V(x, g) := f(x) - m\mathbf{q}(x) - \sigma\mathbf{q}(g - mx) \quad \text{for} \quad x, g \in \mathbb{R}^c. \tag{128}$$

Then for all $x_1, x_2 \in \mathbb{R}^c$ and $g_1 \in \partial f(x_1)$, $g_2 \in \partial f(x_2)$, the following inequalities hold:

$$V(x_1, g_1) - V(x_2, g_2) \leq (g_1 - mx_1)^\top[(\sigma L x_1 - \sigma g_1) - (\sigma L x_2 - \sigma g_2)], \tag{129a}$$

$$0 \leq [(g_1 - mx_1) - (g_2 - mx_2)]^\top[(\sigma L x_1 - \sigma g_1) - (\sigma L x_2 - \sigma g_2)]. \tag{129b}$$

Proof. When $L < \infty$, (129a) results from multiplying equations (15)-(17) in [73, Theorem 5] by σ (as performed in [92, Lemma 12]). When $L = \infty$, (129a) reduces to

$$[f(x_1) - m\mathbf{q}(x_1)] - [f(x_2) - m\mathbf{q}(x_2)] \leq (g_1 - mx_1)^\top(x_1 - x_2) \tag{130}$$

for all $x_1, x_2 \in \mathbb{R}^c$ and $g_1 \in \partial f(x_1)$. Since $g_1 - mx_1 \in \partial(f - m\mathbf{q})(x_1)$ (by the definition of the generalized subgradient), (130) just is the standard subgradient inequality for the p.c.c. function $f - m\mathbf{q}$. Finally, note that (129b) follows by swapping (x_1, g_1) and (x_2, g_2) in (129a) and subtracting the respective right-hand-sides. □

The terms in (129) can be weighted by scalars as follows.

Corollary B.2. *Given constants $\alpha_j > 0$, points $\bar{x}_j \in \mathbb{R}^c$, and vectors $g_j \in \alpha_i^{-1}\partial f(\alpha_i \bar{x}_i)$ for $j \in \{1, 2\}$, the following inequalities hold:*

$$V(\alpha_1\bar{x}_1, \alpha_1\bar{g}_1) - V(\alpha_2\bar{x}_2, \alpha_2\bar{g}_2) \leq \alpha_1(\bar{g}_1 - m\bar{x}_1)^\top[\alpha_1(\sigma L\bar{x}_1 - \sigma g_1) - \alpha_2(\sigma L\bar{x}_2 - \sigma\bar{g}_2)], \tag{131a}$$

$$0 \leq [\alpha_1(\bar{g}_1 - m\bar{x}_1) - \alpha_2(\bar{g}_2 - m\bar{x}_2)]^\top[\alpha_1(\sigma L\bar{x}_1 - \sigma\bar{g}_1) - \alpha_2(\sigma L\bar{x}_2 - \sigma\bar{g}_2)]. \tag{131b}$$

Proof. This holds by Lemma B.1 under the association $\bar{x}_j = \alpha_j x_j$ and $\bar{g}_j = \alpha_j g_j$ for $j \in \{1, 2\}$. □

Proposition B.4. *Given functions $f_i \in \mathcal{S}_{m_i, L_i}$ with $\sigma_i = 1/(L_i - m_i)$, constants $\lambda_i, \alpha_{ij} > 0$, and points (x^i_j, g^i_j) with $g^i_j \in \alpha_{ij}^{-1} \partial f_i(\alpha_{ij} x^i_j)$ for each $i \in \{1, \ldots, s\}$ and $j \in \{1, 2\}$, the following inequality is satisfied:*

$$0 \leq \textstyle\sum_{i=1}^{s} \lambda_i \left([\alpha_{i1}(g^i_1 - m_i x^i_1) - \alpha_{i2}(g^i_2 - m_i x^i_2)]^\top [\alpha_{i1}(\sigma_i L_i x^i_1 - \sigma_i g^i_1) - \alpha_{i2}(\sigma_i L_i x^i_2 - \sigma_i g^i_2)]\right). \tag{132}$$

Proof. The inequality in (132) holds by taking conic combinations of terms in (131b) as weighted by $\{\lambda_i\}_{i=1}^s$. □

B.4 Dissipation Relations

The following consequence is an extension of [92, Lemma 14] to the case where $f \in \mathcal{S}^0_{m,\infty}$ is not globally defined. It uses Corollary B.2 with the choices $\alpha_1 = 1,\ \alpha_2 = \rho^k$ to generate a dissipation relation.

Lemma B.2. *Let $f \in \mathcal{S}^0_{m,L}$, and suppose that $w_k \in \partial f(z_k)$ holds for all $k \in \mathbb{N}$. Given a rate $\rho > 0$ and defining the terms $\bar{p}_k := \rho^{-k}(\sigma L z_k - w_k)$, $\bar{q}_k := \rho^{-k}(w_k - m z_k)$, the following dissipation relations hold for all $T, \ell \in \mathbb{N}$, subject to $T - \ell > 0$:*

$$\textstyle\sum_{k=0}^{T-1} \bar{q}_k^\top \bar{p}_k \geq 0, \qquad \textstyle\sum_{k=0}^{T-\ell} \bar{q}_k^\top (\bar{p}_k - \rho^\ell \bar{p}_{k-\ell}) \geq 0. \tag{133}$$

Lemma B.2 can be used to define valid Zames-Falb filters.

Lemma B.3. *Let $\tilde{F} \in \mathcal{O}^0_{m,L}$ and let $\rho > 0$ be an exponential convergence rate. Let $\nu_{\max} \in \mathbb{N}$ be a finite filter length, and let $(\lambda^i_\nu)_{\nu=0}^{\nu_{\max}}$ denote a set of real-valued filter coefficients for $i \in \{1, \ldots, s\}$ satisfying the following constraints:*

$$\textstyle\sum_{\nu=0}^{\nu_{\max}} \rho^{-\nu} \lambda^i_\nu > 0, \quad \lambda^i_\nu \leq 0, \quad \forall \nu \geq 1, \quad \forall i \in \{1 \ldots, s\}. \tag{134}$$

Given any sequence (z, w) satisfying $w_k \in \tilde{F}(z_k)$ for all $k \in \mathbb{N}$, we define the sequences $(\bar{z}, \bar{w})$ as $\bar{z}_k := \rho^{-k} z_k, \bar{w}_k := \rho^{-k} w_k$. These sequences satisfy $\bar{w}_k \in \rho^{-k} \tilde{F}(\rho^k \bar{z}_k)$ for all $k \in \mathbb{N}$. After defining the ρ-weighted sequences $\bar{q}, \bar{p}, \bar{r}$ indexed by $i \in \{1, \ldots, s\}$ as

$$\bar{q}^i_k := \bar{w}_k - m_i \bar{z}_k, \qquad \bar{p}^i_k := \sigma_i L_i \bar{z}_k - \sigma_i \bar{w}_k = \bar{z}_k - \sigma_i \bar{q}^i_k, \qquad \bar{r}^i_k := \textstyle\sum_{\nu=0}^{k} \lambda^i_\nu \bar{p}_{k-\nu}, \tag{135}$$

the sequences $\bar{q}, \bar{r}$ satisfy the dissipation relation

$$\forall T \in \mathbb{N},\ T > 0: \qquad \textstyle\sum_{k=0}^{T-1} \bar{q}_k^\top \bar{r}_k \geq 0. \tag{136}$$

Proof. This follows as an extension of [63, Lemma 5] to the case where the operator $\tilde{F} \in \mathcal{O}^0_{m,L}$ is not globally defined. The arguments used here are identical to those employed by [63, Lemma 5] and [92, Section 3.2], and are provided for completeness.

Lemma B.2 shows that for all $\nu \in \mathbb{R}, T \in \mathbb{N}, T > 0, i \in \{1, \ldots, s\}$, the following inequalities hold:

$$\textstyle\sum_{k=0}^{T-1} (\bar{q}^i_k)^\top \bar{p}^i_k \geq 0, \qquad \textstyle\sum_{k=0}^{T-1} (\bar{q}^i_k)^\top (\bar{p}^i_k - \rho^\nu \bar{p}^i_{k-\nu}) \geq 0. \tag{137}$$

These relations may be convexly combined with $\mu^i_0 > 0$ and $\{\mu^i_\nu\}_{\nu=1}^{\nu_{\max}} \leq 0$ for each $i \in \{1, \ldots, s\}$, to obtain for every $i \in \{1, \ldots s\}$:

$$\sum_{k=0}^{T-1} \left(\mu^i_0 (\bar{q}^i_k)^\top \bar{p}^i_k + \sum_{\nu=1}^{\nu_{\max}} \mu^i_r (\bar{q}^i_k)^\top (\bar{p}^i_k - \rho^\nu \bar{p}^i_{k-\nu}) \right) = \sum_{k=0}^{T-1} (\bar{q}^i_k)^\top \left[(\sum_{\nu=0}^{\nu_{\max}} \mu_\nu) \bar{p}^i_k \right] - \sum_{\nu=1}^{\nu_{\max}} \mu_\nu \rho^\nu (\bar{q}^i_k)^\top \bar{p}^i_{k-\nu} \geq 0. \tag{138}$$

The parameters λ defining the filtering in $\bar{r}$ are chosen as

$$\lambda^i_0 = \textstyle\sum_{\nu=0}^{\nu_{\max}} \mu^i_\nu, \qquad \lambda^i_\nu = -\rho^\nu \mu^i_\nu, \qquad \forall \nu \in \{1, \ldots r\}, i \in \{1, \ldots, s\}. \tag{139}$$

to ensure that (136) holds. Moreover, they are easily seen to satisfy (134). □

B.5 Generalized Resolvents and Yosida Operators

We first review properties of operators $H : \mathbb{R}^c \rightrightarrows \mathbb{R}^c$ [72, 76]. The graph of the operator H is $\text{gra}(H) = \{(x, y) \in \mathbb{R}^c \times \mathbb{R}^c \mid y \in H(x)\}$. H is strongly monotone with constant $m > 0$ (or just monotone for $m = 0$) if $(y_1 - y_2)^\top (x_1 - x_2) \geq m\|x_1 - x_2\|^2$ holds for all $(x_1, y_1), (x_2, y_2) \in \text{gra}(H)$. An operator H is maximal (strongly) monotone if it is (strongly) monotone and there does not exist a (strongly) monotone H' such that $\text{gra}(H) \subset \text{gra}(H')$. If H is maximal strongly monotone, then H^{-1} is globally continuous [72, Proposition 12.54]. We need the following auxiliary results to provide sufficient well-posedness conditions.

Lemma B.4. *Let $E, F, G : \mathbb{R}^c \rightrightarrows \mathbb{R}^c$ be operators, and suppose that F is at-most single-valued. Then*

$$(EF^{-1} - G)^{-1} = F(E - GF)^{-1}.$$

Proof. For any $x \in \mathbb{R}^c$, $y \in (EF^{-1} - G)^{-1}(x)$ implies $x \in (EF^{-1} - G)(y)$, and hence $x \in E(z) - G(y)$ for some $z \in F^{-1}(y)$. Because $y \in F(z)$, we have $x \in E(z) - GF(z) = (E - GF)(z)$ with $z \in (E - GF)^{-1}(x)$. This finally shows $y \in F(E - GF)^{-1}(x)$.

Now let $y \in F(E - GF)^{-1}(x)$ for any $x \in \mathbb{R}^c$. Then $y \in F(z)$ for some $z \in (E - GF)^{-1}(x)$, and thus $x \in (E - GF)(z)$. It therefore holds that $x \in E(z) - G(\tilde{y})$ for some $\tilde{y} \in F(z)$. Since F is at most single-valued, we have $\tilde{y} = y$. Using the relations $x \in E(z) - G(y)$ and $z \in F^{-1}(y)$, we infer $x \in EF^{-1}(y) - G(y) = (EF^{-1} - G)(y)$, which finally shows $y \in (EF^{-1} - G)^{-1}(x)$. □

Corollary B.3. *For an operator $F : \mathbb{R}^c \rightrightarrows \mathbb{R}^c$ and a matrix $D \in \mathbb{R}^{c \times c}$, the following holds:*

1. *If F is at most single-valued, then $(F^{-1} - D)^{-1} = F(I - DF)^{-1}$.*
2. *If F^{-1} is at most single-valued, then $(DF - I)^{-1} = F^{-1}(D - F^{-1})^{-1}$.*
3. *If D is invertible, then $(F^{-1} - D)^{-1} = D^{-1}[I - (I - DF)^{-1}]$.*

Proof. To prove 1., apply Lemma B.4 with $E = I$ and $G = D$. For 2. use Lemma B.4 for $E = D$, F replaced by F^{-1} and $G = I$. To show 3., choose $E = F^{-1}$, $F = D$ and $G = I$ in Lemma B.4 to get $(F^{-1}D^{-1} - I)^{-1} = D(F^{-1} - D)$. By the inverse-resolvent identity, the left-hand side indeed equals $I - (I - DF)^{-1}$. □

Let us finally collect a few extra facts concerning the notion of well-posedness in this paper.

Proposition B.5. *Given $F : \mathbb{R}^c \rightrightarrows \mathbb{R}^c$ and $D \in \mathbb{R}^{c \times c}$, let $(F^{-1} - D)^{-1}$ be globally continuous. Then:*

1. *If $\alpha > 0$ and $\bar{F}(x) := \alpha^{-1}F(\alpha x)$ for $x \in \mathbb{R}^c$ then $(\bar{F}^{-1} - D)^{-1}(x) = \alpha^{-1}(F^{-1} - \mathcal{D})^{-1}(\alpha x)$ for $x \in \mathbb{R}^c$. Hence $(\bar{F}^{-1} - D)^{-1}$ is globally continuous.*
2. *If $a, b \in \mathbb{R}^c$ and $\bar{F}(x) := F(x + a) - b$ for $x \in \mathbb{R}^c$, then $(\bar{F}^{-1} - D)^{-1}(x) = (F^{-1} - D)^{-1}(x + a - Db) - b$ for $x \in \mathbb{R}^c$. Hence $(\bar{F}^{-1} - D)^{-1}$ is globally continuous.*

Proof. To show 1., take any $x \in \mathbb{R}^c$ and let $y = \alpha^{-1}(F^{-1} - D)^{-1}(\alpha x)$. This is equivalent to $\alpha y = (F^{-1} - D)^{-1}(\alpha x)$ and hence to $\alpha x \in F^{-1}(\alpha y) - D(\alpha y)$, which is nothing but $x \in \alpha^{-1}F^{-1}(\alpha y) - Dy$; since $\alpha^{-1}F^{-1}(\alpha.) = \bar{F}^{-1}$, this holds iff $x \in \bar{F}^{-1}(y) - Dy$, which is finally equivalent to $y \in (\bar{F}^{-1} - D)^{-1}(x)$.

To show 2., pick any $x \in \mathbb{R}^c$ and let $y \in (F^{-1} - D)^{-1}(x + a - Db) - b$; then $y + b \in (F^{-1} - D)^{-1}(x + a - Db)$ and hence $x + a - Db \in F^{-1}(y + b) - D(y + b)$; we infer that there exists some $z \in F^{-1}(y + b)$ with $x = z - a - Dy$; hence $y + b \in F(z)$ and $x = z - a - Dy$; then $\bar{z} := z - a$ satisfies $y \in F(\bar{z} + a) - b = \bar{F}(\bar{z})$ and $x = \bar{z} - Dy$, implying $\bar{z} \in \bar{F}^{-1}(y)$ and $x = \bar{z} - Dy$, which finally gives $x \in (\bar{F}^{-1} - D)(y)$ and thus $y \in (\bar{F}^{-1} - D)^{-1}(x)$. Reversing the arguments proves that $(F^{-1} - D)^{-1}(x + a - Db)$ is not only contained in, but actually equal to $(\bar{F}^{-1} - D)^{-1}(x)$. □

Proposition B.6. *If a matrix $D \in \mathbb{R}^{c\times c}$ ensures that $H = (\partial f^{-1} - D)^{-1}$ is globally continuous for all $f \in \mathcal{S}_{m,\infty}$, then D is invertible.*

Proof. We prove this proposition by contradiction. For any $m \in \mathbb{R}$, consider the function $f(\beta) = \|\beta\|_1 + \frac{m}{2}\|\beta\|_2^2$ with $f \in \mathcal{S}_{m,\infty}$. The subdifferential of f at 0 is $\partial f(0) = \partial\|0\|_1 + m0 = \partial\|0\|_1 = [-1,1]^c$. If D is rank-deficient, there exists a vector $w \in \mathbb{R}^c$ with $\|w\|_1 = 1$ such that $Dw = 0$. We have $0, w \in \partial f(0)$, which implies $0 \in \partial f^{-1}(0)$ and $0 \in \partial f^{-1}(w)$ and hence $0 \in \partial f^{-1}(0) - D0 = \partial f^{-1}(0)$ and $0 \in \partial f^{-1}(w) - Dw = \partial f^{-1}(w)$. This shows $0 \in (\partial f^{-1} - D)^{-1}(0)$ and $w \in (\partial f^{-1} - D)^{-1}(0)$, which contradicts global single-valuedness of H. Hence H is not globally continuous. □

B.6 Guaranteeing Well-Posedness

The goal is to characterize those matrices $D \in \mathbb{R}^{sc\times sc}$ for which $(F^{-1} - D)^{-1}$ is globally continuous for all $F \in \mathcal{O}_{m,L}$. Preliminary results about invertibility and monotonicity are required before providing such a well-posedness condition.

Lemma B.5. *For any $F \in \mathcal{O}_{m,L}$ and vector m^0 such that $m - m^0 \in \mathbb{R}^s_>$, the map $F_0 := F - (\mathrm{diag}(m^0) \otimes I_c)$ has a globally continuous inverse F_0^{-1}.*

Proof. Since $f_i - m_i^0\mathbf{q} = f_i - m_i\mathbf{q} + (m_i - m_i^0)\mathbf{q}$ is p.c.c., we have $f_i \in \mathcal{S}_{m_i^0,L_i}$ and $f_i - m_i^0\mathbf{q} \in \mathcal{S}_{m_i - m_i^0,\infty}$ for each $i \in \{1,\dots s\}$. Since $m_i - m_i^0 > 0$, we infer that $(F_0)_i = \partial(f_i - m_0^i\mathbf{q})$ has a globally continuous inverse for each $i \in \{1,\dots,s\}$, which implies that F_0^{-1} is globally continuous. □

Let us now associate with $\Lambda \in \mathbb{R}^s_>$, $m \in \mathbb{R}^s$ and $L \in \mathbb{R}^s$ with $L - m \in \mathbb{R}^s_>$ the diagonal $sc \times sc$-matrices

$$\mathbf{\Lambda} := \mathrm{diag}(\Lambda) \otimes I_c \succ 0, \quad \mathbf{m} := \mathrm{diag}(m) \otimes I_c, \quad \mathbf{L} := \mathrm{diag}(L) \otimes I_c \quad \text{and} \quad \boldsymbol{\sigma} := (\mathbf{L} - \mathbf{m})^{-1}$$

as well as the $2sc \times 2sc$ matrices

$$\Psi_{m,\sigma} := \begin{pmatrix} I & -\boldsymbol{\sigma} \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ -\mathbf{m} & I \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}\mathbf{L} & -\boldsymbol{\sigma} \\ -\mathbf{m} & I \end{pmatrix}, \quad \hat{\mathbf{\Lambda}} := \begin{pmatrix} \mathbf{\Lambda} & 0 \\ 0 & I_{sc} \end{pmatrix} \quad \text{and} \quad J := \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}. \tag{140}$$

Lemma B.6. *For any $F \in \mathcal{O}_{m,L}$ and $\Lambda \in \mathbb{R}^s_>$, the map $\mathbf{\Lambda}((F - \mathbf{m})^{-1} - \boldsymbol{\sigma})$ is monotone.*

Proof. If $y_j \in \mathbf{\Lambda}((F - \mathbf{m})^{-1} - \boldsymbol{\sigma})(x_j)$ for $j \in \{1,2\}$, there exists some z_j such that $z_j \in (F - \mathbf{m})^{-1}(x_j)$ and $y_j = \Lambda(z_j - \boldsymbol{\sigma}x_j)$. Therefore $x_j \in F(z_j) - \boldsymbol{\sigma}z_j$, and thus $x_j = g_j - \boldsymbol{\sigma}z_j$ for some $g_j \in F(z_j)$. This shows

$$\begin{pmatrix} y_j \\ x_j \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}(z_j - \boldsymbol{\sigma}(g_j - \mathbf{m}z_j)) \\ g_j - \mathbf{m}z_j \end{pmatrix}. \tag{141}$$

Noting that $\boldsymbol{\sigma}\mathbf{L}x_j - \boldsymbol{\sigma}g_j = (\mathbf{L} - \mathbf{m})^{-1}\mathbf{L}x_j - \boldsymbol{\sigma}g_j = x_j + \boldsymbol{\sigma}\mathbf{m}x_j - \boldsymbol{\sigma}g_j = x_j - \boldsymbol{\sigma}(g_j - \mathbf{m}x_j)$, we infer

$$\begin{aligned}[y_1 - y_2]^\top[x_1 - x_2] &= [\mathbf{\Lambda}(z_1 - \boldsymbol{\sigma}(g_1 - \mathbf{m}z_1)) - \mathbf{\Lambda}(z_2 - \boldsymbol{\sigma}(g_2 - \mathbf{m}z_2))]^\top[(g_1 - \mathbf{m}z_1) - (g_2 - \mathbf{m}z_2)] = \\ &= \textstyle\sum_{i=1}^s \Lambda_i \left([(z_1^i - \sigma_i(g_1^i - m_i z_1^i) - (z_1^i - \sigma_i(g_2^i - m_i z_2^i)]^\top[(g_1^i - m_i z_1^i) - (g_2^i - m_i x_2^i)]\right) \\ &= \textstyle\sum_{i=1}^s \Lambda_i \left([(\sigma_i L_i z_1^i - \sigma_i g_1^i) - (\sigma_i L_i z_1^i - \sigma_i g_2^i]^\top[(g_1^i - m_i z_1^i) - (g_2^i - m_i x_2^i)]\right).\end{aligned} \tag{142}$$

In view of (132), we conclude $[y_1 - y_2]^\top[x_1 - x_2] \geq 0$ for all $y_j \in \mathbf{\Lambda}((F - \mathbf{m})^{-1} - \boldsymbol{\sigma})(x_j)$ and $j \in \{1,2\}$, thus proving monotonicity of $\mathbf{\Lambda}((F - \mathbf{m})^{-1} - \boldsymbol{\sigma})$. □

To formulate our well posedness condition, we note that (132) for $\alpha_{ij} = 1$ can be expressed as

$$\begin{pmatrix} x_1 - x_2 \\ g_1 - g_2 \end{pmatrix}^\top \Psi_{m,\sigma}^\top \hat{\mathbf{\Lambda}}^\top J \hat{\mathbf{\Lambda}} \Psi_{m,\sigma} \begin{pmatrix} x_1 - x_2 \\ g_1 - g_2 \end{pmatrix} \geq 0 \ \text{ for all } \ x_j \in \mathbb{R}^{sc}, \ g_j \in F(x_j), \ j \in \{1,2\}. \tag{143}$$

Theorem B.1. *Suppose that there exists a vector $\Lambda \in \mathbb{R}^c_>$ such that $\mathcal{D} \in \mathbb{R}^{sc\times sc}$ satisfies*

$$2\,\mathrm{Sym}([\mathbf{\Lambda}(\boldsymbol{\sigma}\mathbf{L}\mathcal{D}-\boldsymbol{\sigma})]^\top[I-\mathbf{m}\mathcal{D}]) = \begin{pmatrix}\mathcal{D}\\ I\end{pmatrix}^\top \Psi_{m,\sigma}^\top \hat{\mathbf{\Lambda}}^\top J \hat{\mathbf{\Lambda}} \Psi_{m,\sigma}\begin{pmatrix}\mathcal{D}\\ I\end{pmatrix} \prec 0. \tag{144}$$

Then $(I-\mathcal{D}F)^{-1}$ is globally continuous for all $F \in \mathcal{O}_{m,L}$.

Proof. Choose $m_0 < m$ elementwise such that (144) still holds when $(\mathbf{m}_0, \boldsymbol{\sigma}_0)$ replaces $(\mathbf{m}, \boldsymbol{\sigma})$ with $\mathbf{m}_0 = \mathrm{diag}(m_0)\otimes I_c$ and $\boldsymbol{\sigma}_0 := (\mathbf{L}-\mathbf{m}_0)^{-1}$. Since both $\Psi_0 := \Psi_{m_0,\sigma_0}$ and $\hat{\mathbf{\Lambda}}$ are invertible and $J^{-1} = J$ has exactly sc positive and sc negative eigenvalues, we can invoke the Dualization Lemma [96, Lemma 4.9] to infer that the following two conditions are equivalent:

$$\begin{pmatrix}\mathcal{D}\\ I\end{pmatrix}^\top \Psi_0^\top \hat{\mathbf{\Lambda}}^\top J \hat{\mathbf{\Lambda}} \Psi_0\begin{pmatrix}\mathcal{D}\\ I\end{pmatrix} \prec 0 \quad\text{and}\quad \begin{pmatrix} I & -\mathcal{D}\end{pmatrix}\Psi_0^{-1}\hat{\mathbf{\Lambda}}^{-1} J \hat{\mathbf{\Lambda}}^{-\top}\Psi_0^{-\top}\begin{pmatrix} I\\ -\mathcal{D}^\top\end{pmatrix} \succ 0. \tag{145}$$

If we introduce $G_1, G_2, H_1, H_2 \in \mathbb{R}^{sc\times sc}$ by

$$\begin{pmatrix}G_1\\ G_2\end{pmatrix} := \hat{\mathbf{\Lambda}}\Psi_0\begin{pmatrix}\mathcal{D}\\ I\end{pmatrix} \quad\text{and}\quad \begin{pmatrix}H_1 & H_2\end{pmatrix} := \begin{pmatrix} I & -\mathcal{D}\end{pmatrix}\Psi_0^{-1}\hat{\mathbf{\Lambda}}^{-1},$$

we observe that $H_1G_1 + H_2G_2 = 0$. Moreover, (145) reads as $G_1^\top G_2 + G_2^\top G_1 \prec 0$ and $H_1^\top H_2 + H_2^\top H_1 \succ 0$. In turn, G_1, G_2 and H_1, H_2 are invertible. We conclude $G_1G_2^{-1} + H_1^{-1}H_2 = 0$ and $G_1G_2^{-1} + [G_1G_2^{-1}]^\top \prec 0$, which shows that

$$E := H_1^{-1}H_2 \;\text{ satisfies }\; E^\top + E \succ 0. \tag{146}$$

For the map $F_0 = F - \mathbf{m}_0$, we now prove that $I - \mathcal{D}F = H_1[(E-\mathbf{\Lambda}\boldsymbol{\sigma}_0)F_0 + \mathbf{\Lambda}]$. If $x, y \in \mathbb{R}^{sc}$ satisfy $y \in (I-\mathcal{D}F)(x)$, then there exists some $g \in F(x)$ with $y = x - \mathcal{D}g$ and we indeed conclude

$$\begin{aligned} y &= \begin{pmatrix} I & -\mathcal{D}\end{pmatrix}\begin{pmatrix}x\\ g\end{pmatrix} = \begin{pmatrix} I & -\mathcal{D}\end{pmatrix}\Psi_0^{-1}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{\Lambda}}\Psi_0\begin{pmatrix}x\\ g\end{pmatrix} = \begin{pmatrix}H_1 & H_2\end{pmatrix}\hat{\mathbf{\Lambda}}\Psi_0\begin{pmatrix}x\\ g\end{pmatrix} = \\ &= H_1\begin{pmatrix} I & E\end{pmatrix}\begin{pmatrix}\mathbf{\Lambda} & 0\\ 0 & I\end{pmatrix}\begin{pmatrix} I & -\boldsymbol{\sigma}_0\\ 0 & I\end{pmatrix}\begin{pmatrix} I & 0\\ -\mathbf{m}_0 & I\end{pmatrix}\begin{pmatrix}x\\ g\end{pmatrix} = H_1\begin{pmatrix}\mathbf{\Lambda} & E-\mathbf{\Lambda}\boldsymbol{\sigma}_0\end{pmatrix}\begin{pmatrix}x\\ g-\mathbf{m}_0x\end{pmatrix} = \\ &\qquad H_1[(E-\mathbf{\Lambda}\boldsymbol{\sigma}_0)(g-\mathbf{m}_0x)+\mathbf{\Lambda}x] \in H_1[(E-\mathbf{\Lambda}\boldsymbol{\sigma}_0)F_0+\mathbf{\Lambda}](x). \end{aligned}$$

We infer $(I-\mathcal{D}F)^{-1} = [(E-\mathbf{\Lambda}\boldsymbol{\sigma}_0 I)F_0 + \mathbf{\Lambda}]^{-1}H_1^{-1}$. Because F_0^{-1} is globally continuous by Lemma B.5, we invoke Property 2. in Corollary B.3 to get

$$(I-\mathcal{D}F)^{-1} = F_0^{-1}[\mathbf{\Lambda}(F_0^{-1}-\boldsymbol{\sigma}_0 I)+E]^{-1}H_1^{-1}. \tag{147}$$

Moreover, $\mathbf{\Lambda}(F_0^{-1}-\boldsymbol{\sigma}_0 I)$ is globally continuous and also monotone by Lemma B.6, given that $F_0 \in \mathcal{O}_{m_0,L}$. Therefore $\mathbf{\Lambda}(F_0^{-1}-\boldsymbol{\sigma}_0 I)$ is maximal monotone. By (146), the map $\mathbf{\Lambda}(F_0^{-1}-\boldsymbol{\sigma}_0 I)+E$ is maximal strongly monotone, and hence admits a globally continuous inverse. Therefore, the inverses F_0^{-1}, $[\mathbf{\Lambda}(F_0^{-1}-\boldsymbol{\sigma}_0 I)+E]^{-1}$, and H_1^{-1} are all globally continuous, proving that $(I-\mathcal{D}F)^{-1}$ is globally continuous via (147). ☐

Corollary B.4. *The operator $((\partial f)^{-1}-D)^{-1}$ is globally continuous for all $f \in \mathcal{S}_{m,L}$ if $D \in \mathbb{R}^{c\times c}$ satisfies* $\mathrm{Sym}([\sigma LD-\sigma)]^\top[I-mD]) \prec 0$.

Proof. By Theorem B.1 for $s = 1$ and $\Lambda = 1$, we infer that $(I-D\partial f)^{-1}$ is globally continuous. If $\sigma > 0$, we have $L < \infty$ and, therefore, ∂f is globally continuous. Due to 1. in Corollary B.3, we conclude that $((\partial f)^{-1}-D)^{-1} = \partial f(I-D\partial f)^{-1}$ is also globally continuous. If $\sigma = 0$, the D satisfies $\mathrm{Sym}[D^\top(I-mD)] \prec 0$ and is, hence, invertible. Then Property 3. in Corollary B.3 implies that $((\partial f)^{-1}-D)^{-1} = D^{-1}[I-(I-D\partial f)^{-1}]$ is globally continuous. ☐

Lemma B.7. *If $I - \mathcal{D}\mathbf{m}$ is invertible, (144) is equivalent to* $\mathrm{Sym}[\mathbf{\Lambda}(-\boldsymbol{\sigma} + \mathcal{D}(I - \mathbf{m}\mathcal{D})^{-1})] \prec 0$.

Proof. This follows from manipulating the symmetrized matrix in (144) as

$$(\mathbf{\Lambda}\boldsymbol{\sigma}\mathbf{L}\mathcal{D} - \mathbf{\Lambda}\boldsymbol{\sigma})^\top (I - \mathbf{m}\mathcal{D}) = (\mathbf{\Lambda}\mathcal{D} - \mathbf{\Lambda}\boldsymbol{\sigma}(I - \mathbf{m}\mathcal{D}))^\top (I - \mathbf{m}\mathcal{D}) \tag{148a}$$

$$= (I - \mathbf{m}\mathcal{D})^\top (\mathbf{\Lambda}\mathcal{D}(I - \mathbf{m}\mathcal{D})^{-1} - \mathbf{\Lambda}\boldsymbol{\sigma})^\top (I - \mathbf{m}\mathcal{D}). \tag{148b}$$

□

C Controlled Regulation Theory

Based on [15, 16, 80], let us review nominal regulation theory for the interconnection of a plant $\tilde{P}$ and a controller K described as

$$\begin{pmatrix} x_{k+1} \\ \hline e_k \\ y_k \end{pmatrix} = \left(\begin{array}{c|cc} \tilde{A} & \tilde{B}_1 & \tilde{B}_2 \\ \hline \tilde{C}_1 & \tilde{D}_1 & \tilde{D}_{12} \\ \tilde{C}_2 & \tilde{D}_{21} & \tilde{D}_2 \end{array}\right) \begin{pmatrix} x_k \\ \hline d \\ u_k \end{pmatrix}, \qquad \begin{pmatrix} \xi_{k+1} \\ \hline u_k \end{pmatrix} = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \xi_k \\ \hline y_k \end{pmatrix} \tag{149}$$

and affected by the constant disturbance d. It is assumed that K internally stabilizes $\tilde{P}$ and that $\tilde{P} \star K$ is represented by $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \tilde{\mathcal{D}})$ (Section 2.2).

Now suppose that K achieves output regulation for $\tilde{P} \star K$ as discussed in Section 2.11. Then (35) admits a unique solution Υ and we can partition $\Upsilon = (\Pi^\top,\ \Theta^\top)^\top$ according to the partition of $\tilde{\mathcal{A}}$ (see (7)). A simple calculation shows that the matrices

$$\Gamma := (I - D_c\tilde{D}_2)^{-1}(C_c\Theta + D_c(\tilde{C}_2\Pi + \tilde{D}_{21})) \ \text{ and } \ \Phi := \tilde{C}_2\Pi + \tilde{D}_{21} + \tilde{D}_2\Gamma$$

generate a solution $(\Pi, \Gamma, \Phi, \Theta)$ of the *open-loop regulator equations*

$$\left(\begin{array}{c|cc} \tilde{A} & \tilde{B}_1 & \tilde{B}_2 \\ \hline \tilde{C}_1 & \tilde{D}_{11} & \tilde{D}_{12} \\ \tilde{C}_2 & \tilde{D}_{21} & \tilde{D}_2 \end{array}\right) \begin{pmatrix} \Pi \\ \hline I \\ \Gamma \end{pmatrix} = \begin{pmatrix} \Pi \\ \hline 0 \\ \Phi \end{pmatrix}, \qquad \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} \Theta \\ \Phi \end{pmatrix} = \begin{pmatrix} \Theta \\ \Gamma \end{pmatrix}. \tag{150}$$

Conversely, for any quadruple $(\Pi, \Gamma, \Phi, \Theta)$, the signal transformations

$$\hat{x}_k = x_k - \Pi d, \qquad \hat{\xi}_k = \xi_k - \Theta d, \qquad \hat{y}_k = y_k - \Phi d, \qquad \hat{u}_k = u_k - \Gamma d \tag{151}$$

in the interconnection (149) lead to

$$\begin{pmatrix} \hat{x}_{k+1} \\ \hline e_k \\ \hat{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c} \tilde{A} & \tilde{A}\Pi - \Pi + \tilde{B}_2\Gamma + \tilde{B}_1 & \tilde{B}_2 \\ \hline \tilde{C}_1 & \tilde{C}_1\Pi + \tilde{D}_{12}\Gamma + \tilde{D}_1 & \tilde{D}_{12} \\ \tilde{C}_2 & \tilde{C}_2\Pi + \tilde{D}_2 - \Phi + \tilde{D}_{21} & \tilde{D}_2 \end{array}\right) \begin{pmatrix} \hat{x}_k \\ \hline d \\ u_k \end{pmatrix}, \quad \begin{pmatrix} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A_c & A_c\Theta - \Theta + B_c\Phi & B_c \\ \hline C_c & C_c\Theta + D_c\Phi - \Gamma & D_c \end{array}\right) \begin{pmatrix} \hat{\xi}_k \\ \hline d \\ \hat{y}_k \end{pmatrix}.$$

If $(\Pi, \Gamma, \Phi, \Theta)$ is taken to satisfy the regulator equations (150), we clearly obtain

$$\begin{pmatrix} \hat{x}_{k+1} \\ \hline e_k \\ \hat{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c} \tilde{A} & 0 & \tilde{B}_2 \\ \hline \tilde{C}_1 & 0 & \tilde{D}_{12} \\ \tilde{C}_2 & 0 & \tilde{D}_2 \end{array}\right) \begin{pmatrix} x_k \\ \hline d \\ u_k \end{pmatrix}, \quad \begin{pmatrix} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A_c & 0 & B_c \\ \hline C_c & 0 & D_c \end{array}\right) \begin{pmatrix} \hat{\xi}_k \\ \hline d \\ \hat{y}_k \end{pmatrix}. \tag{152}$$

As a consequence of the regulator equations, the signals $(\hat{x}, \hat{\xi}, \hat{y}, \hat{u}, e)$ in (152) are decoupled from the exogenous input d, which implies that K achieves output regulation for $\tilde{P} \star K$. Indeed, since $\mathcal{A}$ is Schur, all trajectories of (152) satisfy $\lim_{k\to\infty}(\hat{x}_k, \hat{\xi}_k) = 0$. Invertibility of $I - \tilde{D}_2 D_c$ ensures that the internal signal $(\hat{u}, \hat{y})$ also satisfies $\lim_{k\to\infty}(\hat{u}_k, \hat{y}_k) = 0$. This implies $\lim_{k\to\infty} e_k = 0$ for any constant disturbance d. In view of (151), we can conclude for all trajectories of (149) that

$$\lim_{k\to\infty}(x_k, \xi_k, y_k, u_k) = (\Pi d, \Theta d, \Phi d, \Gamma d). \tag{153}$$

In summary, solvability of (150) is necessary and sufficient for an internally stabilizing controller K achieving output regulation for $\tilde{P} \star K$. The left-most equation in (150) involving (Π, Γ, Φ) is independent of the controller K and depends only on the plant $\tilde{P}$. The right-most equation in (150) involves only the controller K and the matrices (Θ, Γ, Φ).

D Proof of Theorem 3.1

To show the necessity of Conditions 1 and 2, we assume that $F \star (P \star K)$ is a well-posed convergent algorithm for all $F \in \mathcal{O}_{m,L}$. Hence $P \star K$ is well-posed and, therefore, $I - D_2 D_c$ is invertible. Moreover, for any $\tilde{F} \in \mathcal{O}^0_{m,L}$, the algorithm $\tilde{F} \star (P \star K)$ is well-posed and convergent since $\tilde{F} \in \mathcal{O}_{m,L}$. Because $\tilde{F}$ satisfies $0 \in \tilde{F}(0)$, this algorithm has the fixed point $(0, 0, 0)$. By Proposition 2.1, all its trajectory converge to $(0, 0, 0)$. This shows that Condition 1 (robust stability) holds true.

D.1 Proof of Necessity of Condition 2

Let us use denote the matrices describing $G = P \star K$ by $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ and represent the algorithm $F \star (P \star K)$ by (15). Moreover, we now choose test quadratics $F^t(z) = \mathbf{m}z + b$, which satisfy $F^t \in \mathcal{O}_{m,L}$ (by Assumption 1); hence $F^t \star G$ is well-posed and convergent.

First suppose $b = 0$ such that $F^t(z) = \mathbf{m}z$ and $F^t \in \mathcal{O}^0_{m,L}$. Since $F^t \star G$ is well-posed, we conclude that $I - \mathcal{D}\mathbf{m}$ is invertible and that $((F^t)^{-1} - \mathcal{D})^{-1}$ is defined by the matrix $H := \mathbf{m}(I - \mathcal{D}\mathbf{m})^{-1}$.

Indeed, suppose there exists some $z \neq 0$ with $(I - \mathcal{D}\mathbf{m})z = 0$; this implies $0 = z - \mathcal{D}y$ for $y = \mathbf{m}z \neq 0$; hence $z \in (F^t)^{-1}(y)$ and thus $0 \in ((F^t)^{-1} - \mathcal{D})(y)$, implying $y \in ((F^t)^{-1} - \mathcal{D})^{-1}(0)$; the same argument for $z = 0$ shows $0 \in ((F^t)^{-1} - \mathcal{D})^{-1}(0)$, which contradicts the fact that $((F^t)^{-1} - \mathcal{D})^{-1}$ is single-valued. By 1. in Corollary B.3, we infer $((F^t)^{-1} - \mathcal{D})^{-1} = \mathbf{m}(I - \mathcal{D}\mathbf{m})^{-1}$.

We conclude that $\mathbf{m} \star G$ is well-posed as a star-product of linear systems and that (20) reads as

$$x_{k+1} = (\mathcal{A} + \mathcal{B}(I - \mathbf{m}\mathcal{D})^{-1}\mathbf{m}\mathcal{C})x_k, \quad w_k = H\mathcal{C}x_k, \quad z_k = (I + \mathcal{D}H)\mathcal{C}x_k \quad \text{for all} \quad k \in \mathbb{N}. \tag{154}$$

By convergence of $F^t \star G$ described by (15) with F^t replacing F, we infer that any of its trajectories (x_k, w_k, z_k) converges to a fixed point (x^*, w^*, z^*) for $k \to \infty$; clearly, $(0, 0, 0)$ is a fixed point of the algorithm since $F^t(0) = 0$; since fixed points are unique (Proposition 2.1), we infer $(x^*, w^*, z^*) = (0, 0, 0)$ and hence conclucde $\lim_{k\to\infty} x_k = 0$. Since (154) is another representation of the algorithm $F^t \star G$, we infer that all trajectories of (154) satisfy $\lim_{k\to\infty} x_k = 0$, which in turn shows that $\mathcal{A} + \mathcal{B}(I - \mathbf{m}\mathcal{D})^{-1}\mathbf{m}\mathcal{C}$ is Schur.

Moreover, $\mathbf{m} \star P$ is well-posed by Assumption 2. Since $\mathbf{m} \star G$ and $P \star K$ are well-posed, we infer that the matrices describing $\mathbf{m} \star G = \mathbf{m} \star (P \star K)$ and $(\mathbf{m} \star P) \star K = P^{\mathbf{m}} \star K$ are identical; since $\mathcal{A} + \mathcal{B}(I - \mathbf{m}\mathcal{D})^{-1}\mathbf{m}\mathcal{C}$ is Schur, we conclude that K internally stabilizes $P^{\mathbf{m}}$.

Let us now choose arbitrary $\beta^* \in \mathbb{R}^c$ and $\hat{w}^* \in \mathbb{R}^{(s-1)c}$ and define $F^t(z) = \mathbf{m}z + b$ with

$$b := N\hat{w}^* - \mathbf{m}(\mathbf{1}_s \otimes \beta^*).$$

By construction, $\hat{w}^*$ is the unique vector satisfying $F^t(\mathbf{1}_s \otimes \beta^*) - N\hat{w}^* = 0$, which is in turn used in the definition (39) of $\tilde{F}^t$. Since $(\mathbf{1}_s \otimes I_c)^\top N = 0$, we infer $(\mathbf{1}_s \otimes I_c)^\top \mathbf{m}(\mathbf{1}_s \otimes \beta^*) + (\mathbf{1}_s \otimes I_c)^\top b = 0$, which is nothing but (43); hence β^* is the unique solution to Problem 1. By convergence of $F^t \star (P \star K)$ and hence of $(\tilde{F}^t \star P) \star K = P^{\mathbf{m}} \star K$, we conclude that K achieves regulation for the plant $P^{\mathbf{m}}$ described by (44).

As seen in Section C, we conclude that there exits a solution $(\Pi, \Gamma, \Phi, \Theta)$ of the regulator equations

$$(\Omega_0 + \Omega_{\mathbf{m}}) \begin{pmatrix} \Pi \\ \hline I \\ \Gamma \end{pmatrix} = \begin{pmatrix} \Pi \\ \hline 0 \\ \Phi \end{pmatrix}, \qquad \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \Theta \\ \Phi \end{pmatrix} = \begin{pmatrix} \Theta \\ \Gamma \end{pmatrix}, \tag{155}$$

where we introduce the following abbreviations for the matrices describing $P^{\mathbf{m}}$:

$$\Omega_0 := \left(\begin{array}{c|cc:c} A & 0 & B_1 N & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_2 & 0 & D_{21} N & D_2 \end{array}\right), \qquad \Omega_{\mathbf{m}} := \begin{pmatrix} B_1 \\ \hline D_1 \\ \hdashline D_{21} \end{pmatrix} \mathbf{m} E(\mathbf{m})^{-1} \left(\begin{array}{c|cc:c} C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \end{array}\right). \tag{156}$$

Since $I + D_1 \mathbf{m} E(\mathbf{m})^{-1} = E(\mathbf{m})^{-1}$, the middle block row of $\Omega_0 + \Omega_{\mathbf{m}}$ is $E(\mathbf{m})^{-1} \left(\begin{array}{c|cc:c} C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \end{array}\right)$. Hence the middle block row of the first equation in (155) leads to

$$E(\mathbf{m})^{-1} \left(\begin{array}{c|cc:c} C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \end{array}\right) \begin{pmatrix} \Pi \\ \hline I \\ \hdashline \Gamma \end{pmatrix} = 0 \qquad \text{and hence} \qquad \Omega_{\mathbf{m}} \begin{pmatrix} \Pi \\ \hline I \\ \hdashline \Gamma \end{pmatrix} = 0. \tag{157}$$

In view of the definition of Ω_0, (155) is identical to (47), which concludes the proof.

D.2 Proof of Sufficiency of Conditions 1 and 2 for Convergence

Fix any $F \in \mathcal{O}_{m,L}$. By assumption, Problem 1 has a solution β^* and we can pick $\hat{w}^*$ with $0 \in F(\mathbf{1}_s \otimes \beta^*) - N\hat{w}^*$. This permits to define the error map $\tilde{F} \in \mathcal{O}^0_{m,L}$ by (39) and the corresponding error system (41).

Since $I - D_2 D_c$ is invertible (Condition 1), $P_e \star K$ and $P \star K$ are both well-posed. Since $\tilde{F} \star (P \star K)$ is well-posed (Condition 1), we infer that $\tilde{F} \star (P_e \star K)$ and hence also $F \star (P \star K)$ is well-posed (Section 3.1).

Let us now pick any trajectory of the algorithm $F \star (P \star K)$ described by $w_k \in F(z_k)$ and (46). The signal transformation (40) generates a trajectory of the error system (41). Now pick a solution $(\Pi, \Gamma, \Phi, \Theta)$ of the regulator equations in Condition 2. The subsequent signal transformation

$$\hat{x}^N_k = x^N_k - \Pi d, \qquad \hat{\xi}_k = \xi_k - \Theta d, \qquad \hat{y}_k = y_k - \Phi d, \qquad \hat{u}_k = u_k - \Gamma d. \tag{158}$$

applied to (41) leads, after a simple computation and with $d := \mathrm{col}(-\beta^*, \hat{w}^*)$, to a trajectory of

$$\begin{pmatrix} \hat{x}^N_{k+1} \\ \hline \tilde{z}_k \\ \hdashline e_k \\ \hdashline \hat{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c:c} A & B_1 & (A-I)\Pi + B_2\Gamma + \begin{pmatrix} 0 & B_1 N \end{pmatrix} & B_2 \\ \hline C_1 & D_1 & C_1\Pi + D_{12}\Gamma + \begin{pmatrix} \mathbf{1}_s \otimes I_c & D_1 N \end{pmatrix} & D_{12} \\ \hdashline C_1 & D_1 & C_1\Pi + D_{12}\Gamma + \begin{pmatrix} \mathbf{1}_s \otimes I_c & D_1 N \end{pmatrix} & D_{12} \\ C_2 & D_{21} & C_2\Pi + D_2\Gamma + \begin{pmatrix} 0 & D_{21} N \end{pmatrix} - \Phi & D_2 \end{array}\right) \begin{pmatrix} \hat{x}^N_k \\ \hline \tilde{w}_k \\ \hdashline d \\ \hdashline \hat{u}_k \end{pmatrix},$$

$$\begin{pmatrix} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A_c & (A_c - I)\Theta + B_c\Phi & B_c \\ \hline C_c & C_c\Theta + D_c\Phi - \Gamma & D_c \end{array}\right) \begin{pmatrix} \hat{\xi}_k \\ \hline d \\ \hdashline \hat{y}_k \end{pmatrix}, \qquad \tilde{w}_k \in \tilde{F}(\tilde{z}_k).$$

Due to the regulator equations, this simplifies to

$$\begin{pmatrix} \hat{x}^N_{k+1} \\ \hline \tilde{z}_k \\ \hdashline e_k \\ \hdashline \hat{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c:c} A & B_1 & 0 & B_2 \\ \hline C_1 & D_1 & 0 & D_{12} \\ \hdashline C_1 & D_1 & 0 & D_{12} \\ C_2 & D_{21} & 0 & D_2 \end{array}\right) \begin{pmatrix} \hat{x}^N_k \\ \hline \tilde{w}_k \\ \hdashline d \\ \hdashline \hat{u}_k \end{pmatrix}, \qquad \begin{pmatrix} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A_c & 0 & B_c \\ \hline C_c & 0 & D_c \end{array}\right) \begin{pmatrix} \hat{\xi}_k \\ \hline d \\ \hdashline \hat{y}_k \end{pmatrix}, \qquad \tilde{w}_k \in \tilde{F}(\tilde{z}_k)$$

and hence to

$$\begin{pmatrix} \hat{x}_{k+1}^N \\ \hline \tilde{z}_k \\ \hdashline \hat{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A & B_1 & B_2 \\ \hline C_1 & D_1 & D_{12} \\ \hdashline C_2 & D_{21} & D_2 \end{array}\right) \begin{pmatrix} \hat{x}_k^N \\ \hline \tilde{w}_k \\ \hdashline \hat{u}_k \end{pmatrix}, \qquad \begin{pmatrix} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{pmatrix} = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \hat{\xi}_k \\ \hline \hat{y}_k \end{pmatrix}, \qquad \tilde{w}_k \in \tilde{F}(\tilde{z}_k). \tag{160}$$

By Condition 1, we conclude $\lim_{k\to\infty} \hat{x}_k^N = 0$ and $\lim_{k\to\infty} \hat{\xi}_k = 0$.

Note that (160) is a description of $\tilde{F} \star (P \star K)$. Again with $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$ representing $G = P \star K$, we obtain a trajectory of

$$\hat{x}_k = \begin{pmatrix} \hat{x}_k^N \\ \hat{\xi}_k \end{pmatrix}, \qquad \begin{pmatrix} \hat{x}_{k+1} \\ \hline \tilde{z}_k \end{pmatrix} = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \begin{pmatrix} \hat{x}_k \\ \hline \tilde{w}_k \end{pmatrix}, \qquad \tilde{w}_k \in \tilde{F}(\tilde{z}_k) \tag{161}$$

which satisfies $\lim_{k\to\infty} \hat{x}_k = 0$. Since (161) is well-posed (because $\tilde{F} \star G = \tilde{F} \star (P \star K)$ is) and has the fixed-point $(0,0,0)$ (since $0 \in \tilde{F}(0)$), Proposition 2.1 implies $\lim_{k\to\infty}(\hat{x}_k, \tilde{w}_k, \tilde{z}_k) = 0$. The definition of the error signals $(\tilde{w}, \tilde{z})$ in (40) lets us conclude that $\lim_{k\to\infty} z_k = \mathbf{1}_s \otimes \beta^*$ and $\lim_{k\to\infty} w_k = N\hat{w}^*$.

On the one hand, we infer $\lim_{k\to\infty}(\hat{x}_k^N, \hat{\xi}_k) = 0$, and hence $\lim_{k\to\infty}(x_k^N, \xi_k) = (\Pi d, \Theta d)$ due to (158). On the other hand, well-posedness of $P \star K$ permits to infer $\lim_{k\to\infty}(\hat{u}_k, \hat{y}_k) = 0$ (Section 2.2) which in turn gives $\lim_{k\to\infty}(u_k, y_k) = (\Gamma d, \Phi d)$ by (158). Since $d = \mathrm{col}(-\beta^*, \hat{w}^*)$, this concludes the proof.

D.3 Coordinate-Independence of Consensus Matrix

Proposition D.1. *Feasibility of the regulator equation in* (47) *is independent of the choice of the consensus matrix N from Definition 7.*

Proof. Let $(\Pi, \Theta, \Gamma, \Phi)$ be a solution to (47) with respect to N. If N' is another consensus matrix, there exists an invertible $W \in \mathbb{R}^{(s-1)c \times (s-1)c}$ such that $N' = NW$. The regulator equations in (47a) reads with $N = N'W^{-1}$ and $\hat{W} = \mathrm{blkdiag}(I_c, W)$ as

$$\left(\begin{array}{c|cc:c} A & 0 & B_1 N' W^{-1} & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N' W^{-1} & D_{12} \\ \hdashline C_2 & 0 & D_{21} N' W^{-1} & D_2 \end{array}\right) \begin{pmatrix} \Pi \\ \hline I \\ \hdashline \Gamma \end{pmatrix} = \left(\begin{array}{c|cc:c} A & 0 & B_1 N' & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N' & D_{12} \\ \hdashline C_2 & 0 & D_{21} N' & D_2 \end{array}\right) \begin{pmatrix} \Pi \\ \hline \hat{W}^{-1} \\ \hdashline \Gamma \end{pmatrix} = \begin{pmatrix} \Pi \\ \hline 0 \\ \hdashline \Phi \end{pmatrix}. \tag{162}$$

Right-multiplying all terms in the regulator equations by $\hat{W}$ leads to

$$\left(\begin{array}{c|cc:c} A & 0 & B_1 N' & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N' & D_{12} \\ \hdashline C_2 & 0 & D_{21} N' & D_2 \end{array}\right) \begin{pmatrix} \Pi\hat{W} \\ \hline I \\ \hdashline \Gamma\hat{W} \end{pmatrix} = \begin{pmatrix} \Pi\hat{W} \\ \hline 0 \\ \hdashline \Phi\hat{W} \end{pmatrix} \qquad \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \begin{pmatrix} \Theta\hat{W} \\ \Phi\hat{W} \end{pmatrix} = \begin{pmatrix} \Theta\hat{W} \\ \Gamma\hat{W} \end{pmatrix}. \tag{163}$$

As a conclusion, if $(\Pi, \Theta, \Gamma, \Phi)$ is a solution to the regulator equations (47) for N, then $(\hat{\Pi}, \hat{\Theta}, \hat{\Gamma}, \hat{\Phi}) := (\Pi\hat{W}, \Theta\hat{W}, \Gamma\hat{W}, \Phi\hat{W})$ is a solution to the regulator equation for N', which establishes the proof. ☐

E Proof of Theorem 4.2: Algorithm Structure

We follow the procedure of [81, Theorem 3.2] to derive structural constraints on the controller K. By assumption, there exists a solution $(\Pi, \Gamma, \Phi, \Theta)$ of (47). The regulator equation in (47b) can be expressed as

$$\begin{pmatrix} A_c - I \\ C_c \end{pmatrix} \Theta = \begin{pmatrix} -B_c \Phi \\ \Gamma - D_c \Phi \end{pmatrix}. \tag{164}$$

Because (A_c, B_c, C_c, D_c) is a minimal representation of K, the pair (A_c, C_c) is observable, and thus Θ is uniquely determined by (Γ, Φ) in (164). Assumption 4 ensures that (Π, Γ, Φ) solving (47a) is unique, from which it follows that $(\Pi, \Gamma, \Phi, \Theta)$ solving (47) is unique. We now show a rank property of Θ.

Assumption 5 separates the nullspaces of Φ and Γ as follows.

Lemma E.1. *Under Assumptions 1-5, any solution* (Π, Γ, Φ) *to the regulator equation* (47a) *satisfies* $\mathrm{null}(\Phi) \cap \mathrm{null}(\Gamma) = 0$.

Proof. Applying the signal transformation $\hat{x}_k^N = x_k^N - \Pi d$ to the system (44) yields

$$\begin{pmatrix} \hat{x}_{k+1}^N \\ d \\ \hline \tilde{y}_k \end{pmatrix} = \left[\left(\begin{array}{c:c} A & -B_2\Gamma \\ \hdashline 0 & I_{sc} \\ \hline C_2 & \Phi - D_2\Gamma \end{array} \right) + \begin{pmatrix} B_1 \\ \hdashline 0 \\ \hdashline D_{21} \end{pmatrix} \mathbf{m}E(\mathbf{m})^{-1} \left(\begin{array}{c:c} C_1 & -D_{12}\Gamma \end{array} \right) \right] \begin{pmatrix} \hat{x}_k^N \\ d \end{pmatrix}, \tag{165}$$

which is detectable by by Assumption 5. Hence,

$$\begin{pmatrix} A + B_1\mathbf{m}E(\mathbf{m})^{-1}C_1 - I & -(B_2 + B_1\mathbf{m}E(\mathbf{m})^{-1}D_{12})\Gamma \\ 0 & 0_{sc} \\ \hline C_2 + D_{21}\mathbf{m}E(\mathbf{m})^{-1}C_1 & \Phi - (D_{22} + D_{21}\mathbf{m}E(\mathbf{m})^{-1}D_{12})\Gamma \end{pmatrix} \tag{166}$$

has full column rank, which implies that

$$\mathrm{null}(\Phi) \cap \mathrm{null}\left[\begin{pmatrix} B_2 + B_1\mathbf{m}E(\mathbf{m})^{-1}D_{12} \\ D_{22} + D_{21}\mathbf{m}E(\mathbf{m})^{-1}D_{12} \end{pmatrix} \Gamma \right] = \{0\}. \tag{167}$$

Since $\mathrm{null}(\Gamma) \subseteq \mathrm{null}(M\Gamma)$ for any matrix M, we conclude $\mathrm{null}(\Phi) \cap \mathrm{null}(\Gamma) = \{0\}$. ☐

Lemma E.2. *Under Assumptions 1-5, for any solution* $(\Pi, \Gamma, \Phi, \Theta)$ *of the regulator equations* (47) *and any invertible matrix* $R = (R_1 \ R_2)$ *with* $ran(R_1) = null(\Phi)$*, the matrices* $(\Theta_1 \ \Theta_2) := \Theta R$ *and* $(\Gamma_1 \ \Gamma_2) := \Theta R$ *satisfy* $rank(\Theta_1) = rank(\Gamma_1) = dim(null(\Phi))$.

Proof. We prove this Lemma by contradiction.

Γ_1 rank condition: If $v \neq 0$ is a vector such that $\Gamma_1 v = 0$, then $\Gamma R_1 v = 0$. Because $\mathrm{ran}(R_1) = \mathrm{null}(\Phi)$, it holds that $\Phi R_1 v = 0$, which implies that $R_1 v \in \mathrm{null}(\Phi) \cap \mathrm{null}(\Gamma)$. By Lemma E.1, we have $\mathrm{null}(\Phi) \cap \mathrm{null}(\Gamma) = \{0\}$, so $R_1 v = 0$. Since R_1 has full column rank, $R_1 v = 0$ implies $v = 0$, which contradicts $v \neq 0$. Consequently, Γ_1 has full column rank, which implies $\mathrm{rank}(\Gamma_1) = \dim(\mathrm{null}(\Phi))$.

Θ_1 rank condition: We now post-multiply both sides of the bottom equation in (47b) to yield

$$\Gamma R_1 = C_c\Theta R_1 + D_c\Phi R_1 \qquad \text{which reads} \quad \Gamma_1 = C_c\Theta_1 + 0 = C_c\Theta_1. \tag{168}$$

Hence $\dim(\mathrm{null}(\Phi)) = \mathrm{rank}(\Gamma_1) = \mathrm{rank}(C_c\Theta_1) = \mathrm{rank}(\Theta_1) - \dim(\mathrm{ran}(\Theta_1) \cap \mathrm{null}(C_c)) \leq \mathrm{rank}(\Theta_1)$. Because Θ_1 has $\dim(\mathrm{null}(\Phi))$ columns, we infer that Θ_1 has full column rank.

☐

We now proceed to isolate structural properties of controllers K in convergent optimization algorithms $F \star (P \star K)$. By Lemmas E.2 and E.1, there exists an invertible matrix Q such that

$$Q^{-1}\Theta = \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \end{pmatrix}. \tag{169}$$

Applying Q as a similarity transformation to the representation of K in (68) leads to

$$\begin{pmatrix} Q^{-1}A_cQ & Q^{-1}B_c \\ C_cQ & D_c \end{pmatrix} \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \\ \hdashline 0 & \Phi_2 \end{pmatrix} = \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \\ \hdashline \Gamma_1 & \Gamma_2 \end{pmatrix}. \tag{170}$$

Now we partition the matrices of the transformed controller representation according to (68) to get

$$\left(\begin{array}{cc|c} A_c{}'_{11} & A_c{}'_{12} & B_c{}'_1 \\ A_c{}'_{21} & A_c{}'_{22} & B_c{}'_2 \\ \hline C_c{}'_1 & C_c{}'_2 & D_c{}' \end{array}\right) := \left(\begin{array}{c|c} Q^{-1}A_cQ & Q^{-1}B_c \\ \hline C_cQ & D_c \end{array}\right). \tag{171}$$

Substitution of (171) into (170) leads to the following structural constraints:

$$\begin{pmatrix} A_c{}'_{11} \\ A_c{}'_{21} \\ \hline C_c{}'_1 \end{pmatrix} = \begin{pmatrix} I_r \\ 0 \\ \hline -\Gamma_1 \end{pmatrix}, \qquad \begin{pmatrix} A_c{}'_{12} & B_c{}'_1 \\ A_c{}'_{22} & B_c{}'_2 \\ \hline C_c{}'_2 & D_c{}' \end{pmatrix} \begin{pmatrix} \Theta_{22} \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \Theta_{22} \\ \Gamma_1\Theta_{12} + \Gamma_2 \end{pmatrix}. \tag{172}$$

The controller structure in (172) can be recognized as a cascade between an internal model subsystem $\Sigma_{\min}$ with internal dynamics I_r and a subcontroller Σ_{core} described by

$$\Sigma_{\min}: \quad \left[\begin{array}{c|cc} I_r & I & 0 \\ \hline -\Gamma_1 & 0 & I \end{array}\right], \qquad \text{and} \qquad \Sigma_{\text{core}}: \quad \left[\begin{array}{c|c} A_c{}'_{22} & B_c{}'_2 \\ \hline A_c{}'_{12} & B_c{}'_1 \\ C_c{}'_2 & D_c{}' \end{array}\right] = \left[\begin{array}{c|c} A_{\text{core}} & B_{\text{core}} \\ \hline C_{\text{core}1} & D_{\text{core}1} \\ C_{\text{core}2} & D_{\text{core}2} \end{array}\right]. \tag{173}$$

Then (172) arises from (172) after relabeling the matrices describing K according to those of the core subcontroller Σ_{core}.

F Connections to the Internal Model Principle

The necessary presence of the internal model' $\Sigma_{\min}$ as a factor of K is an incarnation of the celebrated 'internal model principle' [15] for algorithms. The state-matrix I_r with $r \leq cs$ of the internal model $\Sigma_{\min}$ is a submatrix of the signal generator I_{sc} used to describe the constant exogenous disturbances as $d_{k+1} = I_{cs}d_k$ in accordance with [81].

F.1 Nominal and Robust Regulation

The work in [15] performs regulation of a measured output of $e = y$. This constraint requires that the entire signal generator I_{cs} must be present in an internal model in order for a controller to reject constant disturbances. The less restrictive condition of 'readability' [15, 18] involves the existence of a wide matrix E_y such that $e = E_y y$ and also ensures that the full matrix I_{cs} must appear in any regulator. Our approach builds on the results about nominal regulation in [81] without the assumption of readability. As a result, only a portion of the signal generator I_{cs} is required to be contained in the internal model $\Sigma_{\min}$.

The Regulator Equation in (47) and the structural property in Theorem 4.2 are based on nominal regulation (Appendix C) in which the solution $(\Pi, \Theta, \Gamma, \Phi)$ is independent from the uncertainty in $\mathcal{O}_{m,L}$. Replacing the test quadratics by more general quadratic functions resulting in $F_i(\beta) = \partial f_i(\beta) = \Delta_i\beta + b_i$ leads to $P^{\Delta} = \Delta \star P$ with $\Delta = \operatorname{diag}(\Delta_1, \ldots, \Delta_s)$ instead of $P^{\mathbf{m}}$ in Section 3.2. The resulting structured robust regulation problem cannot be handled with existing results in [81] or [82, Section III.H], since the readability condition clearly fails for the system (41b).

F.2 Structural Stability

The design of uncertainty-independent controllers that achieve output regulation in the context of arbitrary perturbations in the plant parameters is classically referred to as the 'robust regulation problem' [16], and has been studied in [97, 98, 99, 80]. Performance optimization under robust regulation under a readability assumption is documented in [82, Section III.I]. For a signal generator I_{sc}, any robust regulator must contain an internal model involving I_{sc} ([97, Theorem 1] for the readable setting and [80, Theorem 2.8.2] for the non-readable setting). Because our optimization setting for quadratic functions involves a plant with structured uncertainty instead of arbitrary elementwise perturbations in the entries of a plant's realization, achieving classical robust regulation is a much stronger requirement. Our internal model structure does not achieve robust regulation in the classical sense, but achieves robust regulation for structured uncertainties, even if thy are nonlinear and belong to the class $\mathcal{O}_{m,L}$.

F.3 Frequency Domain Arguments

The work in [52, 54, 58] use frequency domain interpolation constraints to characterize structural properties of optimization algorithms. To any transfer function $G(\mathbf{z})$, there exists a left coprime factorization as $G(\mathbf{z}) = \tilde{M}(\mathbf{z})^{-1}\tilde{N}(\mathbf{z})$ in which $\tilde{M}(\mathbf{z})$ is square, and both $\tilde{M}(\mathbf{z})$ and $\tilde{N}(\mathbf{z})$ are stable [51]. If the system G in an algorithmic interconnection (15) is an optimization algorithm, then it must structurally satisfy Consensus and Zero-Inclusion properties. The Consensus property $N^\top z^* = 0$ may be expressed as the presence of a blocking zero in G as $\tilde{N}(1)N = 0$, and Optimality $\mathbf{1}^\top w^* = 0$ is posed as the presence of a pole $\tilde{M}(1)(\mathbf{1}_s \otimes I_c) = 0$. The Consensus and Zero-Inclusion properties are tangential interpolation constraints [55] on the transfer function matrix of G. Information structures in $\mathcal{L}_{\text{info}}$ about resolvent evaluation/explicitness/dependencies may be implemented as bitangental interpolation constraints. Nevanlinna-Pick interpolation in [52, 54, 58] is used to solve the resultant algorithm synthesis problems in the setting of direct interconnections and for fixed stability filter coefficients. The structure of $\tilde{N}(1)N = 0$ and $\tilde{M}(1)(\mathbf{1}_s \otimes I_c) = 0$ in an optimization algorithm is referred to in [52, 54, 58] as an 'internal model principle.' The work in [52, 54, 58] performs transfer-function-based factorizations to parameterize algorithms that meet these interpolation conditions. Our work instead focuses on state-space factorization of algorithms, solutions of regulator equations, explicit internal models, the broader networked setting, and structured control design methodologies.

G Structured Synthesis Implementation

In the structured control framework, we utilize a minimal internal model Σ_{min} and restrict the core subcontroller as $(\Sigma_{\text{core}})_{ss} \in \mathcal{L}_{\text{core}}(\Theta)$.

The optimization variables of the structured control synthesis Problem 4 are $(\Theta_{12}, \Theta_{22}, (\Sigma_{\text{core}})_{ss}, \boldsymbol{\lambda})$. The goal is to choose a controller $(\Sigma_{\text{core}})_{ss}$ such that the ρ-weighted plant $\bar{\Sigma}_{\text{core}}$ formed by $\hat{G}^{\text{min}}(\boldsymbol{\lambda}) = \Psi(\boldsymbol{\lambda})(\Sigma_{m,L} \star \bar{P}_N \star (\bar{\Sigma}_{\text{min}}\Sigma_{\text{core}}))$ satisfies the LMI in (89) at fixed ρ. This feasibility can be certified by using the tool `hinfstruct` for structured H_∞-norm optimization. A feasible solution of the `hinfstruct` method can then be used to find controller parameters at fixed ρ [93]. H_∞ methods are classical in control theory, refer to [51] for more detail on H_∞ techniques.

The LMI constraint (item 1) of Problem 4 can be replaced by the following requirement [92, Lemma A.2]:

1. The plant $\left[\begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 1 \end{pmatrix} \otimes I_{sc}\right] \star (\Psi(\boldsymbol{\lambda})(\Sigma_{m,L} \star \bar{P}_N \star (\bar{\Sigma}_{\text{min}}\Sigma_{\text{core}}))$ has an H_∞ norm less than 1.

For a fixed matrix Θ_{22}, the constraint set $\mathcal{L}_{\text{core}}(\Theta)$ can be expressed as

$$\mathcal{L}_{\text{core}}(\Theta): \begin{pmatrix} A_{\text{core}} & B_{\text{core}} \\ C_{\text{core}1} & D_{\text{core}1} \\ C_{\text{core}2} & D_{\text{core}2} \end{pmatrix} \begin{pmatrix} \Theta_{22} \\ \Phi_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -\Gamma_1 \end{pmatrix} \Theta_{12} = \begin{pmatrix} \Theta_{22} \\ 0 \\ \Gamma_2 \end{pmatrix}, \tag{174}$$

which defines a subspace in $(A_{\text{core}}, B_{\text{core}}, C_{\text{core}1}, C_{\text{core}2}, D_{\text{core}1}, D_{\text{core}2}, \Theta_{11})$. The constraint $D_{\text{core}2} \in \mathcal{L}_{\text{info}}$ induces a further constraint.

The matrix Θ_{22} used to define the constraint $\mathcal{L}_{\text{core}}(\Theta)$ can be treated as a hyperparameter in optimization. For a fixed Θ_{22}, let $J(\Theta_{22})$ denote the minimal exponential convergence rate ρ obtained by solving Problem 4 when appropriately constraining $(\Theta_{12}, (\Sigma_{\text{core}})_{ss}, \lambda)$ according to items 1 and 2 of Problem 4. The nonconvex fixed-structure H_∞ optimization tool `hinfstruct` [93] can be used to perform this optimization. A higher level optimization algorithm can search over $\Theta_{22} \in \mathbb{R}^{n_\xi \times sc-r}$ to minimize $J(\Theta_{22})$, because by default `hinfstruct` can only impose box constraints on the parameters. In the case of static subcontrollers Σ_{core} ($A_{\text{core}} = [\cdot]$), optimization over Θ_{22} is unnecessary because $\Theta_{22} = [\cdot]$.

H Proof of Theorem 5.2: Full-Order and Minimal Internal Models

We show how a full-order subcontroller Σ_f can be factorized into a unique core subcontroller Σ_{core}, and how a core subcontroller Σ_{core} can be lifted into a possibly non-unique full-order subcontroller Σ_f.

From Σ_f to Σ_{core}: Given the basis change matrix R from (67), we perform a coordinate change of K with blkdiag(R, I) as

$$K' := \left[\begin{array}{cc|c} I_{cs} - R^{-1} D_{f1} \Phi R & R^{-1} C_{f1} & R^{-1} D_{f1} \\ -B_f \Phi R & A_f & B_f \\ \hline -\Gamma R - D_{f2} \Phi R & C_{f2} & D_{f2} \end{array}\right] = \left[\begin{array}{cc:c|c} I_r & -D_{f1}^1 \Phi_2 & C_{f1}^1 & D_{f1}^1 \\ 0 & I_{cs-r} - D_{f1}^2 \Phi_2 & C_{f1}^2 & D_{f2}^1 \\ \hdashline 0 & -B_f \Phi_2 & A_f & B_f \\ \hline -\Gamma_1 & -\Gamma_2 - D_{f2}\Phi_2 & C_{f2} & D_{f2} \end{array}\right], \tag{175}$$

under the definitions

$$\begin{pmatrix} C_{f1}^1 \\ C_{f1}^2 \end{pmatrix}, := R^{-1} C_{f1} \qquad \begin{pmatrix} D_{f1}^1 \\ D_{f1}^2 \end{pmatrix}, := R^{-1} D_{f2}. \tag{176}$$

The expression in (175) can be recognized as a cascade between Σ_{core} and $\Sigma_{\min}$, in which Σ_{core} can be represented by

$$\Sigma_{\text{core}} = \left[\begin{array}{c:c|c} I_{cs-r} - D_{f1}^2 \Phi_2 & C_{f1}^2 & D_{f2}^2 \\ \hdashline -B_f \Phi_2 & A_f & B_f \\ \hline -D_{f1}^1 \Phi_2 & C_{f1}^1 & D_{f1}^1 \\ \hdashline -\Gamma_2 - D_{f2}\Phi_2 & C_{f2} & D_{f2} \end{array}\right] = \left[\begin{array}{c|c:ccc} I_{cs-r} & 0 & 0 & I & 0 \\ 0 & 0 & I & 0 & 0 \\ -\Gamma_2 & 0 & 0 & 0 & I \\ \hdashline -\Phi_2 & I & 0 & 0 & 0 \end{array}\right] \star \left[\begin{array}{c|c} A_f & B_f \\ \hline C_{f1}^1 & D_{f1}^1 \\ C_{f1}^2 & D_{f2}^2 \\ C_{f2} & D_{f2} \end{array}\right]. \tag{177}$$

From Σ_{core} to Σ_f: This correspondence uses the method of [81, Theorem 6]. Since Φ_2 has full column rank $sc - r$, it has a left-inverse and there exists a matrix W such that $M := I_{sc-r} - W\Phi_2$ is Schur. Next, the system Σ_{core} is enriched by unobservable modes in M, forming the nonminimal representations

$$\Sigma_{\text{core}f} : \left(\begin{array}{cc|c} A_{\text{core}} & 0 & B_{\text{core}} \\ 0 & M & W \\ \hline C_{\text{core}1} & 0 & D_{\text{core}1} \\ \hdashline C_{\text{core}2} & 0 & D_{\text{core}2} \end{array}\right), \qquad K_f = \Sigma_{\min}\Sigma_{\text{core}f} : \left(\begin{array}{c:cc|c} I_r & C_{\text{core}1} & 0 & D_{\text{core}1} \\ \hdashline 0 & A_{\text{core}} & 0 & B_{\text{core}} \\ 0 & 0 & M & W \\ \hline -\Gamma_1 & C_{\text{core}2} & 0 & D_{\text{core}2} \end{array}\right). \tag{178}$$

The systems $\Sigma_{\text{core}\,f}$ and Σ_{core} are related as $\Sigma_{\text{core}\,f}(\mathbf{z}) = \Sigma_{\text{core}}(\mathbf{z})$ because the modes in M are unobservable. The same relation holds for $\Sigma_{\min}\Sigma_{\text{core}\,f}$ and $\Sigma_{\min}\Sigma_{\text{core}}$. Moreover, K_f represented by (178) satisfies the regulator equation (47b) as

$$\left(\begin{array}{c:cc|c} I_r & C_{\text{core}1} & 0 & D_{\text{core}1} \\ \hdashline 0 & A_{\text{core}} & 0 & B_{\text{core}} \\ 0 & 0 & M & W \\ \hline -\Gamma_1 & C_{\text{core}2} & 0 & D_{\text{core}2} \end{array}\right) \left(\begin{array}{cc} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \\ 0 & I_{sc-r} \\ \hdashline 0 & \Phi_2 \end{array}\right) = \left(\begin{array}{cc} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \\ 0 & I_{sc-r} \\ \hdashline \Gamma_1 & \Gamma_2 \end{array}\right), \tag{179}$$

since $M + W\Phi_2 = I_{sc-r}$. We now define the matrix T_f as

$$T_f := \begin{pmatrix} I_r & 0 & -\Theta_{12} \\ 0 & I_{n_\xi} & -\Theta_{22} \\ 0 & 0 & -I_{sc-r} \end{pmatrix}, \qquad \text{resulting in } T_f \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \\ 0 & I_{sc-r} \end{pmatrix} = \begin{pmatrix} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \end{pmatrix}. \tag{180}$$

Applying T_f to the regulator equation in (179) as a state-coordinate change in $\Sigma_{\min}\Sigma_{\text{core}\,f}$ produces

$$\begin{pmatrix} T_f & 0 \\ 0 & I \end{pmatrix} \left(\begin{array}{c:cc|c} I_r & C_{\text{core}1} & 0 & D_{\text{core}1} \\ \hdashline 0 & A_{\text{core}} & 0 & B_{\text{core}} \\ 0 & 0 & M & W \\ \hline -\Gamma_1 & C_{\text{core}2} & 0 & D_{\text{core}2} \end{array}\right) \begin{pmatrix} T_f & 0 \\ 0 & I \end{pmatrix} \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline 0 & \Phi_2 \end{array}\right) = \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline \Gamma_1 & \Gamma_2 \end{array}\right), \tag{181a}$$

$$\left(\begin{array}{c:cc|c} I_r & C_{\text{core}1} & -\Theta_{12}M & D_{\text{core}1} - \Theta_{12}W \\ \hdashline 0 & A_{\text{core}} & -\Theta_{22}M & B_{\text{core}} - \Theta_{22}W \\ 0 & 0 & -M & -W \\ \hline -\Gamma_1 & C_{\text{core}2} & 0 & D_{\text{core}2} \end{array}\right) \begin{pmatrix} T_f & 0 \\ 0 & I \end{pmatrix} \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline 0 & \Phi_2 \end{array}\right) = \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline \Gamma_1 & \Gamma_2 \end{array}\right), \tag{181b}$$

$$\left(\begin{array}{c:cc|c} I_r & C_{\text{core}1} & -\Theta_{12} - C_{\text{core}1}\Theta_{22} + \Theta_{12}M & D_{\text{core}1} - \Theta_{12}W \\ \hdashline 0 & A_{\text{core}} & (M - A_{\text{core}})\Theta_{22} & B_{\text{core}} - \Theta_{22}W \\ 0 & 0 & M & -W \\ \hline -\Gamma_1 & C_{\text{core}2} & \Gamma_1\Theta_{12} - C_{\text{core}2}\Theta_{22} & D_{\text{core}2} \end{array}\right) \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline 0 & \Phi_2 \end{array}\right) = \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline \Gamma_1 & \Gamma_2 \end{array}\right). \tag{181c}$$

Expanding $M := I_{sc-r} - W\Phi_2$ in (181c) and applying the structural constraints in $\mathcal{L}_{\text{core}}(\Theta)$ via (71) yields

$$\left(\begin{array}{c:cc|c} I_r & C_{\text{core}1} & D_{\text{core}1}\Phi_2 - \Theta_{12}W\Phi_2 & D_{\text{core}1} - \Theta_{12}W \\ \hdashline 0 & A_{\text{core}} & B_{\text{core}}\Phi_2 - \Theta_{22}W\Phi_2 & B_{\text{core}} - \Theta_{22}W \\ 0 & 0 & I_{sc-r} - W\Phi_2 & -W \\ \hline -\Gamma_1 & C_{\text{core}2} & -\Gamma_2 + D_{\text{core}2}\Phi_2 & D_{\text{core}2} \end{array}\right) \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline 0 & \Phi_2 \end{array}\right) = \left(\begin{array}{cc} -I_r & 0 \\ 0 & 0 \\ 0 & -I_{sc-r} \\ \hdashline \Gamma_1 & \Gamma_2 \end{array}\right). \tag{181d}$$

Permuting the second and third rows and columns in the controller representation in (181d) leads to the system

$$K_f = \left[\begin{array}{cc:c|c} I_r & D_{\text{core}1}\Phi_2 - \Theta_{12}W\Phi_2 & C_{\text{core}1} & D_{\text{core}1} - \Theta_{12}W \\ 0 & I_{sc-r} - W\Phi_2 & 0 & -W \\ \hdashline 0 & B_{\text{core}}\Phi_2 - \Theta_{22}W\Phi_2 & A_{\text{core}} & B_{\text{core}} - \Theta_{22}W \\ \hline -\Gamma_1 & -\Gamma_2 + D_{\text{core}2}\Phi_2 & C_{\text{core}2} & D_{\text{core}2} \end{array}\right], \tag{182}$$

which can be recognized as the star product

$$K_f = \left[\begin{array}{cc|c:cc:c} I_r & 0 & 0 & I_r & 0 & 0 \\ 0 & I_{sc-r} & 0 & 0 & I_{sc-r} & 0 \\ \hline -\Gamma_1 & -\Gamma_2 & 0 & 0 & 0 & I \\ \hdashline 0 & \Phi_2 & I & 0 & 0 & 0 \end{array}\right] \star \left[\begin{array}{c|c} A_{\text{core}} & B_{\text{core}} - \Theta_{22} W \\ \hline C_{\text{core}1} & D_{\text{core}1} - \Theta_{12} W \\ 0 & -W \\ \hdashline C_{\text{core}2} & D_{\text{core}2} \end{array}\right]. \tag{183}$$

By exploiting (67), we obtain

$$K_f = \left[\begin{array}{c|c:c:c} I & 0 & I & 0 \\ \hline -\Gamma R & 0 & 0 & I \\ \Phi R & I & 0 & 0 \end{array}\right] \star \left[\begin{array}{c|c} A_{\text{core}} & B_{\text{core}} - \Theta_{22} W \\ \hline C_{\text{core}1} & D_{\text{core}1} - \Theta_{12} W \\ 0 & -W \\ \hdashline C_{\text{core}2} & D_{\text{core}2} \end{array}\right] \tag{184}$$

$$= \left[\begin{array}{c|c:c:c} I & 0 & I & 0 \\ \hline -\Gamma & 0 & 0 & I \\ \Phi & I & 0 & 0 \end{array}\right] \star \left[\begin{array}{c|c} A_{\text{core}} & B_{\text{core}} - \Theta_{22} W \\ \hline R\begin{pmatrix} C_{\text{core}1} \\ 0 \end{pmatrix} & R\begin{pmatrix} D_{\text{core}1} - \Theta_{12} W \\ -W \end{pmatrix} \\ C_{\text{core}2} & D_{\text{core}2} \end{array}\right] = \Sigma_{\text{full}} \star \Sigma_f, \tag{185}$$

which allows us to extract the full-order subcontroller Σ_f as desired. The representation of $K_f = \Sigma_{\text{full}} \star \Sigma_f$ is generally nonminimal, even if Σ_{core} is described by a minimal representation. If the given representation of Σ_{core} has order n_{core}, then the representation of $K = \Sigma_{\min}\Sigma_{\text{core}}$ as constructed by the star product (115) has $n_{\text{core}} + r$ states. The representation of $K_f = \Sigma_{\text{full}} \star \Sigma_f$ through the same star product operation has $n_{\text{core}} + sc$ states, even though the transfer functions of K and K_f are the same.

I Full-Order Synthesis of Optimization Algorithms

The ρ-weighting of the full-order internal model Σ_{full} is

$$\bar{\Sigma}_{\text{full}} : \quad \left(\begin{array}{c|c:cc} \rho^{-1} I_{sc} & 0 & \rho^{-1} I_{sc} & 0 \\ \hline -\Gamma & 0 & 0 & I_{n_u} \\ \hdashline \Phi & I_{n_y} & 0 & 0 \end{array}\right). \tag{186}$$

Then the interconnection $\bar{P} \star \bar{\Sigma}_{\text{full}}$ has the representation

$$\bar{P} \star \bar{\Sigma}_{\text{full}} : \quad \left(\begin{array}{c:c|c:cc} \rho^{-1} A & -\rho^{-1} B_2 \Gamma & \rho^{-1} B_1 & 0 & \rho^{-1} B_2 \\ \hdashline 0 & \rho^{-1} I_{cs} & 0 & \rho^{-1} I_{cs} & 0 \\ \hline C_1 & -D_{12}\Gamma & D_1 & 0 & D_{12} \\ C_2 & -D_{22}\Phi & D_{21} & 0 & D_2 \end{array}\right). \tag{187}$$

This permits us to compute a state-space description of $\hat{P}(\lambda) = \Psi(\lambda)(\Sigma_{m,L} \star \bar{P} \star \bar{\Sigma}_{\text{full}})$ denoted as

$$\hat{P}(\lambda) : \quad \begin{pmatrix} \eta_{k+1} \\ \hline \bar{r}_k \\ \hdashline \bar{\bar{y}}_k \end{pmatrix} = \left(\begin{array}{c|c:c} \hat{A}(\lambda) & \hat{B}_1(\lambda) & \hat{B}_2(\lambda) \\ \hline \hat{C}_1(\lambda) & \hat{D}_1(\lambda) & \hat{D}_{12}(\lambda) \\ \hdashline \hat{C}_2(\lambda) & \hat{D}_{21}(\lambda) & \hat{D}_2(\lambda) \end{array}\right) \begin{pmatrix} \eta_k \\ \hline \bar{q}_k \\ \hdashline \bar{\bar{u}}_k \end{pmatrix}. \tag{188}$$

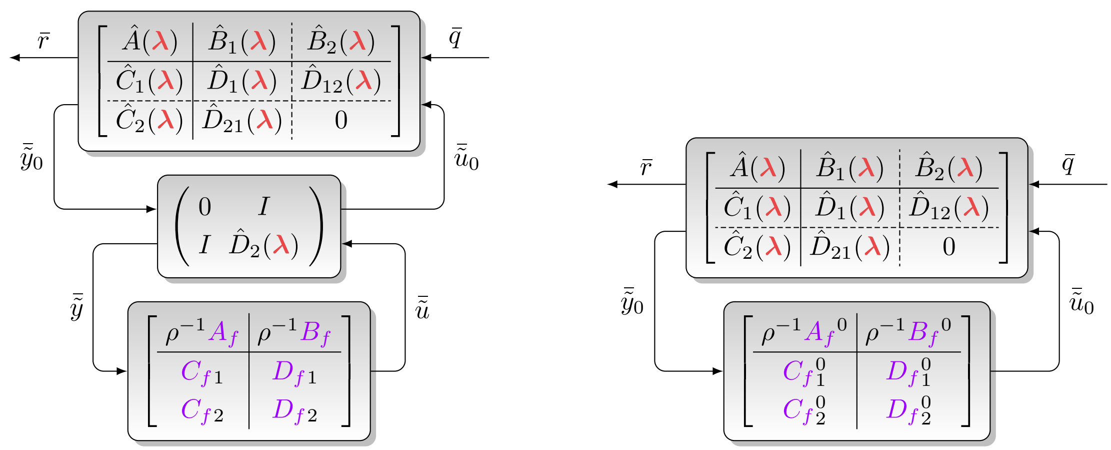
For fixed filter coefficients λ, the goal is to design Σ_{f} such that $\hat{P}(\lambda) \star \bar{\Sigma}_{\text{f}}$ satisfies the properties in Problem 5. To convexify this problem, we follow the nonlinear convexification procedure for controller synthesis in [82]. It only remains to clarify how to handle the fact that $\hat{D}_2(\lambda) = D_2$ does not vanish and that the constraint $D_{f2} \in \mathcal{L}^D$ needs to be satisfied.

The key idea can be visualized as in Figure 27. We first separate the direct feedthrough matrix $\hat{D}_2(\boldsymbol{\lambda})$ from the plant $\hat{P}(\boldsymbol{\lambda})$ using a linear fractional transformation (Figure 27a), thus forming a control synthesis task with respect to a plant Σ_f^0 (Figure 27b). We achieve condition 1 in Problem 5 for the interconnection on the right in Figure 27, by making sure that the designed controller satisfies ${D_f}_2^0 \in \mathcal{L}^D$. Since $\mathcal{L}^D$ is assumed to be convex, this constraint can be taken into account in the transformation approach to controller design as proposed in [82]. In our presented examples with block-sparsity patterns, the set $\mathcal{L}^D$ is representable by a finite number of linear equality constraints in the entries of ${D_f}_2^0$.

If non-zero, the matrix $\hat{D}_2(\boldsymbol{\lambda})$ is then brought back in by choosing the controller on the left in Figure 27 such that both closed-loop systems coincide. This is achieved for the choice

$$\left[\begin{array}{c|c} A_f & B_f \\ \hline C_f & D_f \end{array}\right] = \begin{pmatrix} 0 & I \\ 0 & -\hat{D}_2(\boldsymbol{\lambda}) \end{pmatrix} \star \left[\begin{array}{c|c} {A_f}^0 & {B_f}^0 \\ \hline {C_f}^0 & {D_f}^0 \end{array}\right] = \left[\begin{array}{c|c} {A_f}^0 - {B_f}^0\mathcal{E}^{-1}\hat{D}_2{D_f}^0 & {B_f}^0\mathcal{E}^{-1} \\ \hline (I - {D_f}^0\mathcal{E}^{-1}\hat{D}_2){C_f}^0 & {D_f}^0\mathcal{E}^{-1} \end{array}\right]$$

where $\mathcal{E} = I + \hat{D}_2(\boldsymbol{\lambda}){D_f}^0$. A slight numerical perturbation of $\hat{D}_2(\boldsymbol{\lambda})$ may be required to render $\mathcal{E}$ invertible, and invertibility of $\mathcal{E}$ ensures that the star product is well-posed. Due to Assumption 7, we infer that $D_f = {D_f}^0(I + \hat{D}_2{D_f}^0)^{-1} \in \mathcal{L}_{\text{info}}$ is satisfied. As a conclusion, the constructed controller will satisfy both Conditions 1 and 2 in Problem 5.



(a) Separation of feedthrough

(b) Modified configuration for controller synthesis

Figure 27: Removal of direct feedthrough by interconnection

If the network model and the to-be-designed algorithm all have Kronecker structure, then we perform synthesis for dimension $c = c'$ by performing algorithm design for $c = 1$, and then applying a Kronecker product $\otimes I_{c'}$ to all matrices in the controller representation K.

References

[1] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[2] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *Journal of mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.

[3] E. J. Candes and Y. Plan, "Matrix Completion with Noise," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 925–936, 2010.

[4] B. Kouvaritakis and M. Cannon, "Model Predictive Control: Classical, Robust and Stochastic," *Switzerland: Springer International Publishing*, vol. 38, no. 13-56, p. 7, 2016.

[5] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network Topology and Communication-Computation Tradeoffs in Decentralized Optimization," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.

[6] M. Doostmohammadian, A. Aghasi, M. Pirani, E. Nekouei, H. Zarrabi, R. Keypour, A. I. Rikos, and K. H. Johansson, "Survey of Distributed Algorithms for Resource Allocation over Multi-Agent Systems," *Annual Reviews in Control*, vol. 59, p. 100983, 2025.

[7] S. Wright and J. Nocedal, "Numerical Optimization," *Springer, New York*, vol. 35, no. 67-68, p. 7, 1999.

[8] A. I. Lur'e and V. N. Postnikov, "On the theory of stability of control systems," *Applied mathematics and mechanics*, vol. 8, no. 3, pp. 246–248, 1944.

[9] K. Meyer, "Liapunov functions for the problem of Luré," *Proceedings of the National Academy of Sciences*, vol. 53, no. 3, pp. 501–503, 1965.

[10] J. Wang and N. Elia, "A Control Perspective for Centralized and Distributed Convex Optimization," in *2011 50th IEEE conference on decision and control and European control conference*. IEEE, 2011, pp. 3800–3805.

[11] D. Davis and W. Yin, "A Three-Operator Splitting Scheme and its Optimization Applications," *Set-valued and variational analysis*, vol. 25, no. 4, pp. 829–858, 2017.

[12] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[13] B. Van Scoy and L. Lessard, "A Universal Decomposition for Distributed Optimization Algorithms," *IEEE Control Systems Letters*, vol. 6, pp. 3044–3049, 2022.

[14] M. Upadhyaya, S. Banert, A. B. Taylor, and P. Giselsson, "Automated tight Lyapunov analysis for first-order methods," *Mathematical Programming*, vol. 209, no. 1, pp. 133–170, 2025.

[15] B. A. Francis and W. M. Wonham, "The internal model principle of control theory," *Automatica*, vol. 12, no. 5, pp. 457–465, 1976.

[16] B. A. Francis, "The Linear Multivariable Regulator Problem," *SIAM Journal on Control and Optimization*, vol. 15, no. 3, pp. 486–505, 1977.

[17] W. M. Wonham, "Towards an Abstract Internal Model Principle," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 11, pp. 735–740, 1976.

[18] A. Isidori, L. Marconi, and A. Serrani, *Robust Autonomous Guidance An Internal Model Approach*. Springer Science & Business Media, 2003.

[19] A. Megretski and A. Rantzer, "System analysis via integral quadratic constraints," *IEEE Transactions on Automatic Control*, vol. 42, no. 6, pp. 819–830, 1997.

[20] V. Choppella, K. Viswanath, and M. Kumar, "Algodynamics: Algorithms as systems," in *2021 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2021, pp. 1–9.

[21] F. Dörfler, Z. He, G. Belgioioso, S. Bolognani, J. Lygeros, and M. Muehlebach, "Towards a Systems Theory of Algorithms," *IEEE Control Systems Letters*, vol. 8, pp. 1198–1210, 2024.

[22] W. I. Zangwill, *Nonlinear Programming: A Unified Approach.* Prentice-Hall, Englewood Cliffs, N.J., 1969.

[23] K. Glover and J. C. Willems, "On the Stability of Numerical Integration Routines for Ordinary Differential Equations," *IMA Journal of Applied Mathematics*, vol. 11, no. 2, pp. 171–180, 1973.

[24] R. W. Brockett, "Dynamical Systems That Sort Lists, Diagonalize Matrices, and Solve Linear Programming Problems," *Linear Algebra and its applications*, vol. 146, pp. 79–91, 1991.

[25] K. Kashima and Y. Yamamoto, "System theory for numerical analysis," *Automatica*, vol. 43, no. 7, pp. 1156–1164, 2007.

[26] F. Santambrogio, "{Euclidean, Metric, and Wasserstein} Gradient Flows: an overview," *Bulletin of Mathematical Sciences*, vol. 7, no. 1, pp. 87–154, 2017.

[27] W. Su, S. Boyd, and E. J. Candes, "A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights," *Journal of Machine Learning Research*, vol. 17, no. 153, pp. 1–43, 2016.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[29] M. Muehlebach and M. I. Jordan, "Accelerated First-Order Optimization under Nonlinear Constraints," *Mathematical Programming*, vol. 215, pp. 407–452, 2026.

[30] V. A. Yakubovich, "Frequency conditions for the absolute stability of control systems with several nonlinear or linear nonstationary blocks," *Avtomatika i telemekhanika*, vol. 6, pp. 5–30, 1967.

[31] J. C. Willems, "Dissipative dynamical systems part I: General theory," *Archive for rational mechanics and analysis*, vol. 45, no. 5, pp. 321–351, 1972.

[32] ——, "Dissipative dynamical systems part II: Linear systems with quadratic supply rates," *Archive for rational mechanics and analysis*, vol. 45, no. 5, pp. 352–393, 1972.

[33] A. Van der Schaft, *L2-Gain and Passivity Techniques in Nonlinear Control.* Springer, 2000.

[34] N. Bastianello, R. Carli, and S. Zampieri, "Internal Model-Based Online Optimization," *IEEE Transactions on Automatic Control*, vol. 69, no. 1, pp. 689–696, 2023.

[35] G. Bianchin and B. Van Scoy, "The internal model principle of time-varying optimization," *IEEE Transactions on Automatic Control*, 2026.

[36] E. K. Ryu, R. Hannah, and W. Yin, "Scaled relative graphs: Nonexpansive operators via 2D Euclidean geometry," *Mathematical Programming*, vol. 194, no. 1, pp. 569–619, 2022.

[37] T. Chaffey, F. Forni, and R. Sepulchre, "Scaled relative graphs for system analysis," in *2021 60th IEEE Conference on Decision and Control (CDC).* IEEE, 2021, pp. 3166–3172.

[38] A. Hauswirth, Z. He, S. Bolognani, G. Hug, and F. Dörfler, "Optimization algorithms as robust feedback controllers," *Annual Reviews in Control*, vol. 57, p. 100941, 2024.

[39] R. M. Murray *et al.*, "Optimization-based control," *California Institute of Technology, CA*, pp. 111–128, 2009.

[40] L. Lessard, B. Recht, and A. Packard, "Analysis and Design of Optimization Algorithms via Integral Quadratic Constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[41] G. Zames and P. Falb, "Stability Conditions for Systems with Monotone and Slope-Restricted Nonlinearities," *SIAM Journal on Control*, vol. 6, no. 1, pp. 89–108, 1968.

[42] W. P. Heath and A. G. Wills, "Zames-Falb multipliers for quadratic programming," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 963–968.

[43] B. Van Scoy, R. A. Freeman, and K. M. Lynch, "The Fastest Known Globally Convergent First-Order Method for Minimizing Strongly Convex Functions," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 49–54, 2018.

[44] A. Taylor and Y. Drori, "An optimal gradient method for smooth strongly convex minimization," *Mathematical Programming*, vol. 199, no. 1, pp. 557–594, 2023.

[45] L. Lessard and P. Seiler, "Direct Synthesis of Iterative Algorithms With Bounds on Achievable Worst-Case Convergence Rate," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 119–125.

[46] G. Zhang, X. Bao, L. Lessard, and R. Grosse, "A Unified Analysis of First-Order Methods for Smooth Games via Integral Quadratic Constraints," *Journal of machine learning research*, vol. 22, no. 103, pp. 1–39, 2021.

[47] F. Jakob and A. Iannelli, "A Linear Parameter-Varying Framework for the Analysis of Time-Varying Optimization Algorithms," *arXiv preprint arXiv:2501.07461*, 2025.

[48] ——, "Online convex optimization and integral quadratic constraints: An automated approach to regret analysis," in *2025 IEEE 64th Conference on Decision and Control (CDC)*. IEEE, 2025, pp. 6298–6303.

[49] A. Karakai, J. Eising, A. Martinelli, and F. Dörfler, "Convergence Analysis of Distributed Optimization: A Dissipativity Framework," *arXiv preprint arXiv:2510.27645*, 2025.

[50] L. Lessard, "The Analysis of Optimization Algorithms, A Dissipativity Approach," *IEEE Control Systems Magazine*, vol. 42, no. 3, pp. 58–72, 2022.

[51] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Prentice hall Upper Saddle River, NJ, 1998, vol. 104.

[52] W. Wu, J. Chen, M. R. Jovanović, and T. T. Georgiou, "Tannenbaum's gain-margin optimization meets Polyak's heavy-ball algorithm," *arXiv preprint arXiv:2409.19882*, 2024.

[53] S. Zhang, W. Wu, Z. Li, J. Chen, and T. T. Georgiou, "Frequency-Domain Analysis of Distributed Optimization: Fundamental Convergence Rate and Optimal Algorithm Synthesis," *IEEE Transactions on Automatic Control*, vol. 69, no. 12, pp. 8539–8554, 2024.

[54] W. Wu, J. Chen, M. R. Jovanović, and T. T. Georgiou, "Frequency-Domain Synthesis of Implicit Algorithms," in *2025 American Control Conference (ACC)*. IEEE, 2025, pp. 2262–2267.

[55] J. Ball, I. Gohberg *et al.*, *Interpolation of Rational Matrix Functions*. Birkhäuser, 2013, vol. 45.

[56] G. Pick, "Über die Beschränkungen analytischer Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden," *Mathematische Annalen*, vol. 77, no. 1, pp. 7–23, 1915.

[57] J. Agler and J. E. McCarthy, *Pick Interpolation and Hilbert Function Spaces*. American Mathematical Society, 2002.

[58] I. K. Ozaslan, W. Wu, J. Chen, T. T. Georgiou, and M. R. Jovanovic, "Automated algorithm design for convex optimization problems with linear equality constraints," *arXiv preprint arXiv:2509.20746*, 2025.

[59] T. Holicki and C. W. Scherer, "Algorithm design and extremum control: Convex synthesis due to plant multiplier commutation," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 3249–3256.

[60] V. Kučera, "Stability of Discrete Linear Feedback Systems," *IFAC Proceedings Volumes*, vol. 8, no. 1, pp. 573–578, 1975.

[61] D. Youla, H. Jabr, and J. Bongiorno, "Modern Wiener-Hopf design of optimal controllers–Part II: The multivariable case," *IEEE Transactions on Automatic Control*, vol. 21, no. 3, pp. 319–338, 1976.

[62] C. Scherer and C. Ebenbauer, "Convex Synthesis of Accelerated Gradient Algorithms," *SIAM Journal on Control and Optimization*, vol. 59, no. 6, pp. 4615–4645, 2021.

[63] C. W. Scherer, C. Ebenbauer, and T. Holicki, "Optimization Algorithm Synthesis based on Integral Quadratic Constraints: A Tutorial," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 2995–3002.

[64] J. Miller, F. Jakob, C. Scherer, and A. Iannelli, "Analysis and Synthesis of Switched Optimization Algorithms," *arXiv preprint arXiv:2510.21490*, 2025.

[65] D. Gramlich, C. Ebenbauer, and C. W. Scherer, "Synthesis of accelerated gradient algorithms for optimization and saddle point problems using Lyapunov functions and LMIs," *Systems & Control Letters*, vol. 165, p. 105271, 2022.

[66] Y. Drori and M. Teboulle, "Performance of first-order methods for smooth convex minimization: a novel approach," *Mathematical Programming*, vol. 145, no. 1, pp. 451–482, 2014.

[67] A. B. Taylor, J. M. Hendrickx, and F. Glineur, "Smooth Strongly Convex Interpolation and Exact Worst-case Performance of First-order Methods," *Mathematical Programming*, vol. 161, no. 1, pp. 307–345, 2017.

[68] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, "Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection," *SIAM Journal on Optimization*, vol. 30, no. 3, pp. 2251–2271, 2020.

[69] S. Das Gupta, B. P. Van Parys, and E. K. Ryu, "Branch-and-Bound Performance Estimation Programming: A Unified Methodology for Constructing Optimal Optimization Methods," *Mathematical Programming*, vol. 204, no. 1, pp. 567–639, 2024.

[70] A. Taylor, B. Van Scoy, and L. Lessard, "Lyapunov Functions for First-Order Methods: Tight Automated Convergence Guarantees," in *International Conference on Machine Learning*. PMLR, 2018, pp. 4897–4906.

[71] B. Van Scoy, J. W. Simpson-Porco, and L. Lessard, "Automated Lyapunov Analysis of Primal-Dual Optimization Algorithms: An Interpolation Approach," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 1306–1311.

[72] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer Science & Business Media, 2009, vol. 317.

[73] C. W. Scherer, "Robust Exponential Stability and Invariance Guarantees with General Dynamic O'Shea-Zames-Falb Multipliers," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 5799–5804, 2023, 22nd IFAC World Congress.

[74] T. Rotaru, F. Glineur, and P. Patrinos, "Exact worst-case convergence rates of gradient descent: a complete analysis for all constant stepsizes over nonconvex and convex functions," *arXiv preprint arXiv:2406.17506*, 2024.

[75] R. A. Freeman, "Noncausal Zames-Falb Multipliers for Tighter Estimates of Exponential Convergence Rates," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 2984–2989.

[76] H. Bauschke and P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Space*. Springer, 01 2011.

[77] M. Upadhyaya, A. B. Taylor, S. Banert, and P. Giselsson, "AutoLyap: A Python package for computer-assisted Lyapunov analyses for first-order methods," *arXiv preprint arXiv:2506.24076*, 2025.

[78] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation II: Applications*. Springer, 2006, vol. 331.

[79] S. Michalowsky, C. Scherer, and C. Ebenbauer, "Robust and structure exploiting optimisation algorithms: an integral quadratic constraint approach," *International Journal of Control*, vol. 94, no. 11, pp. 2956–2979, 2021.

[80] A. Saberi, A. A. Stoorvogel, and P. Sannuti, *Control of Linear Systems with Regulation and Input Constraints*. Springer Science & Business Media, 2000.

[81] A. A. Stoorvogel, A. Saberi, and P. Sannuti, "Performance with regulation constraints," *Automatica*, vol. 36, no. 10, pp. 1443–1456, 2000.

[82] C. Scherer, P. Gahinet, and M. Chilali, "Multiobjective Output-Feedback Control via LMI Optimization ," *IEEE Transactions on Automatic Control*, vol. 42, no. 7, pp. 896–911, 1997.

[83] J. Douglas and H. H. Rachford, "On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables," *Transactions of the American mathematical Society*, vol. 82, no. 2, pp. 421–439, 1956.

[84] E. K. Ryu, "Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting," *Mathematical Programming*, vol. 182, no. 1, pp. 233–273, 2020.

[85] M. Morin, S. Banert, and P. Giselsson, "Frugal Splitting Operators: Representation, Minimal Lifting, and Convergence," *SIAM Journal on Optimization*, vol. 34, no. 2, pp. 1595–1621, 2024.

[86] Y. Malitsky and M. K. Tam, "Resolvent splitting for sums of monotone operators with minimal lifting," *Mathematical Programming*, vol. 201, no. 1, pp. 231–262, 2023.

[87] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*. SIAM, 2000.

[88] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[89] Y. Malitsky and M. K. Tam, "A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity," *SIAM Journal on Optimization*, vol. 30, no. 2, pp. 1451–1472, 2020.

[90] G. B. Passty, "Ergodic convergence to a zero of the sum of monotone operators in Hilbert space," *Journal of Mathematical Analysis and Applications*, vol. 72, no. 2, pp. 383–390, 1979.

[91] V. Blondel and J. N. Tsitsiklis, "NP-hardness of some linear control design problems," *SIAM journal on control and optimization*, vol. 35, no. 6, pp. 2118–2127, 1997.

[92] C. W. Scherer and C. Ebenbauer, "A Tutorial on Convex Design of Optimization Algorithms by Integral Quadratic Constraints," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 9, 2025.

[93] P. Gahinet and P. Apkarian, "Decentralized and fixed-structure H_∞ control in MATLAB," in *2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. IEEE, 2011, pp. 8205–8210.

[94] P. Gahinet and A. Nemirovskii, "General-Purpose LMI solvers with benchmarks," in *Proc. IEEE Conf. Decis. Control*, 1993, pp. 3162–3165 vol.4.

[95] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1, pp. 89–97, 2004.

[96] C. Scherer and S. Weiland, "Linear Matrix Inequalities in Control," *Lecture notes, dutch institute for systems and control, Delft, the Netherlands*, vol. 3, no. 2, pp. 62–74, 2000.

[97] E. J. Davison and A. Goldenberg, "Robust control of a general servomechanism problem: The servo compensator," *Automatica*, vol. 11, no. 5, pp. 461–471, 1975.

[98] C. Desoer and Y. Wang, "On the minimum order of a robust servocompensator," *IEEE Transactions on Automatic Control*, vol. 23, no. 1, pp. 70–73, 2003.

[99] J. Pearson, R. Shields, and P. Staats, "Robust solutions to linear multivariable control problems," *IEEE Transactions on Automatic Control*, vol. 19, no. 5, pp. 508–517, 1974.