Structure, Analysis, and Synthesis of Optimization Algorithms

Jared Miller
Carsten Scherer
Fabian Jakob
Andrea Iannelli

Oxford Control

June 9, 2026

Composite Optimization

Want to minimize sum of functions: find β^* with

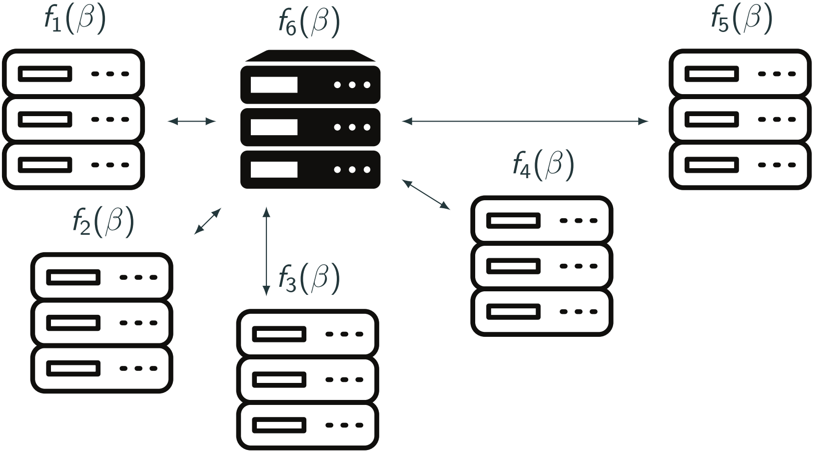
$$\beta^* \in \operatorname{argmin}_{\beta \in \mathbb{R}^c} \sum_{i=1}^{s} f_i(\beta)$$

Wide range of applications:

Regression	Recommender Systems
Bioinformatics	Energy Management
Image Denoising	Robust/Model Predictive Control

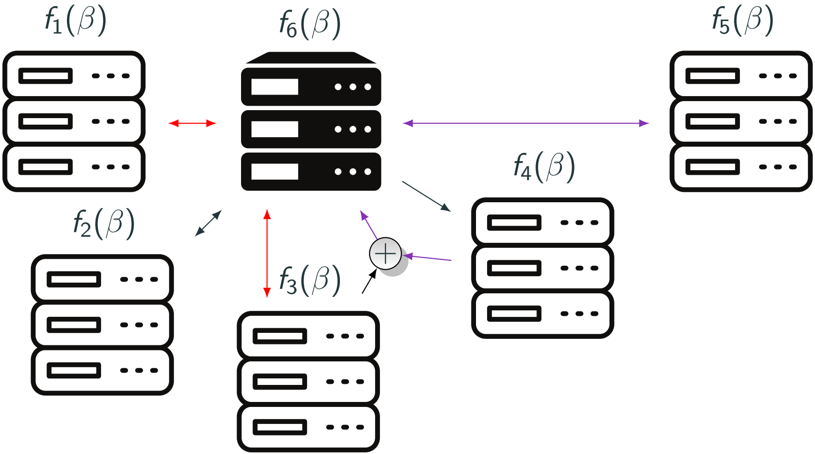
Optimization Setup

Federated Learning: each server might have different data



Optimization over Dynamical Networks

Memory, cross-talk, time delays, other nightmares



Main Ideas

Perform automated, disciplined design of fast algorithms

Restrict search using necessary convergence conditions

Accomplish by continuing a link of Optimization and Control

	Optimization	**Control**
1)	Convergence	Regulation Theory
2)	Structure	Internal Model Principle
3)	Synthesis	Robust Control

Background: Algorithm Setting

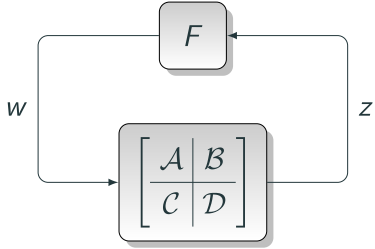
Algorithmic Interconnection

System G connected to map $F(z) = \sum_{i=1}^{s} \partial f_i(z^i)$:

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

$F \star G$ well-posed: $(F^{-1} - \mathcal{D})^{-1}$ continuous, globally defined

Block Diagram Representation



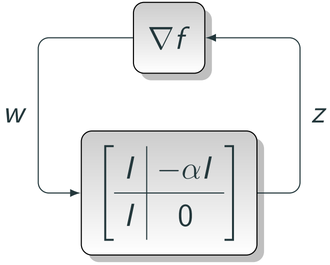
abbreviates the interconnection $F \star G$

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

Interconnections Examples: GD/PPM

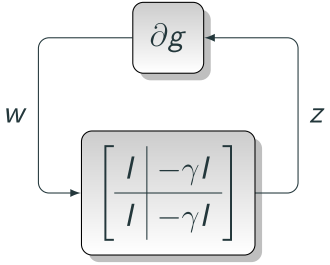
Gradient Descent

$$x_{k+1} = x_k - \alpha \nabla f(x_k)$$



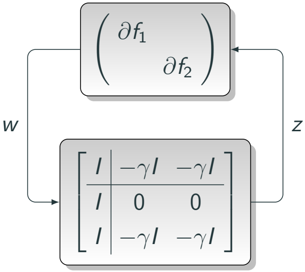
Proximal Point Method

$$x_{k+1} = (I + \gamma \partial g)^{-1}(x_k)$$

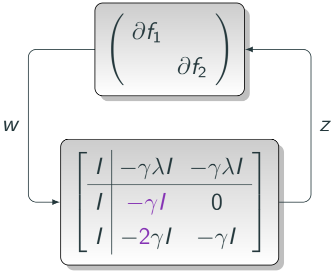


Interconnection Examples: PGD/Douglas-Rachford

Projected Gradient Descent



Douglas-Rachford[1]



[1]Douglas, J., & Rachford, H. H. (1956). On the numerical solution of heat conduction problems in two and three space variables. Transactions of the American mathematical Society, 82(2), 421-439.

Fixed Points and Convergence

Fixed point (x^*, w^*, z^*) of algorithmic interconnection:

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

[2]Upadhyaya, Manu, et al. "Automated tight Lyapunov analysis for first-order methods." Mathematical Programming 209.1 (2025): 133-170.

Fixed Points and Convergence

Fixed point (x^*, w^*, z^*) of algorithmic interconnection:

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

Convergent: (x^*, w^*, z^*) exists with[2]

$$\begin{aligned} \text{Consensus:} \quad & z^{1*} = z^{2*} = \ \dots \ = z^{s*} \\ \text{Optimality:} \quad & \textstyle\sum_{i=1}^{s} w^{i*} = 0 \\ \text{Attractivity:} \quad & \forall x_0 : (x, w, z) \to (x^*(x_0), w^*(x_0), z^*(x_0)) \end{aligned}$$

[2] Upadhyaya, Manu, et al. "Automated tight Lyapunov analysis for first-order methods." Mathematical Programming 209.1 (2025): 133-170.

Image Restoration

Total-variation denoising of multi-color images

$$\beta^* \in \underset{\beta\in\mathbb{R}^{n_x n_y n_c}}{\operatorname{argmin}} \quad \underbrace{\frac{1}{2}\|\mathrm{vec}(Y)-\beta\|_2^2}_{f_1(\beta)} + \underbrace{\mu_{\mathrm{TV}}\mathrm{TV}(\beta)+\chi_{[0,255]}(\beta)}_{f_2(\beta)}$$

Original Image X

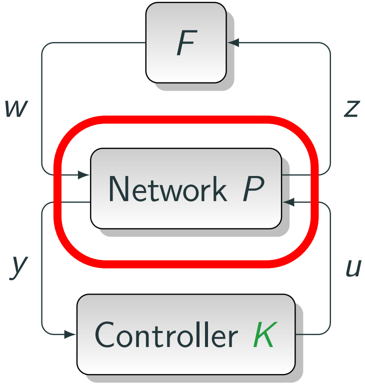
Noisy Image Y

Denoised β^*

Solved by Douglas-Rachford: $(\gamma,\lambda)=(2,1)$, $\mu_{\mathrm{TV}}=0.15$

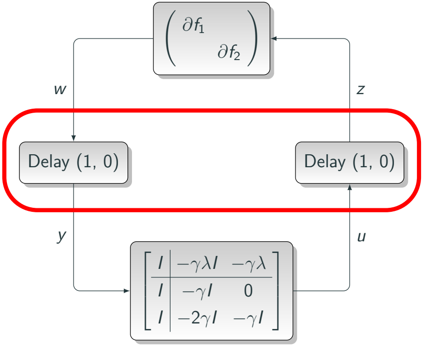
Algorithms over Networks

Map F may not be accessible (e.g. time-delays, cross-talk)



Examples of Interconnections: Delayed DR

Douglas-Rachford with 1-step delay before/after viewing image



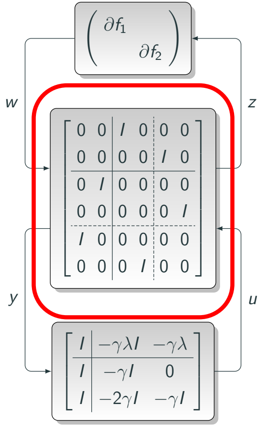
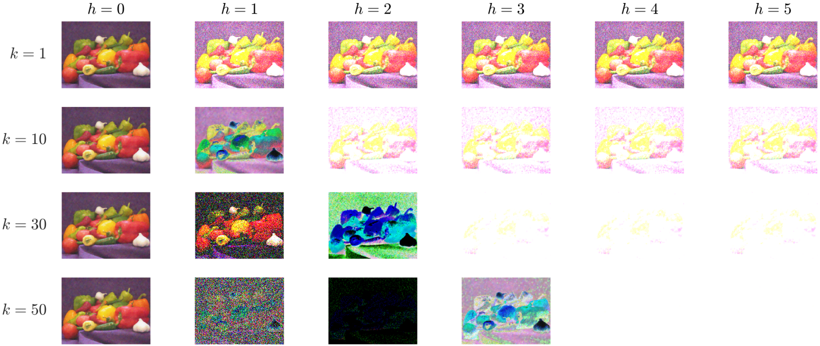


Image Restoration: Douglas-Rachford with Delays

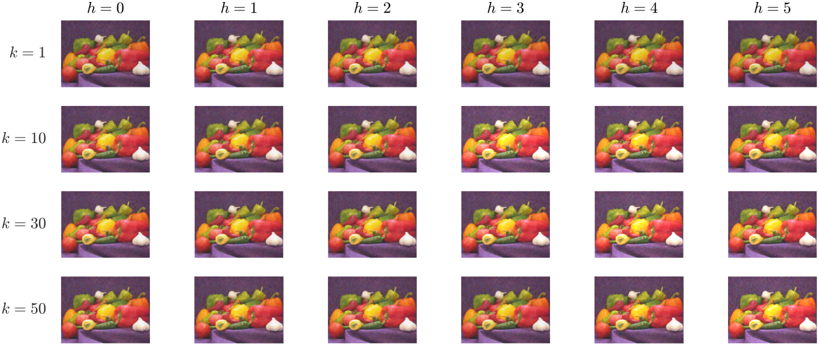
Delays can break convergent algorithms



h-step delay on viewing image data Y

Image Restoration: Synthesized with Delays

Want algorithms that converge with network dynamics



h-step delay on viewing image data Y

Operator Classes

Ensure convergence over *class* of F, not just specific F

$\mathcal{O}$: desired class of operators F having unique (β^*, w^*) with

$$z^* = \mathbf{1}_s \otimes \beta^*, \qquad w^* \in F(z^*), \qquad \textstyle\sum_{i=1}^s w^{i*} = 0$$

Example: each f_i is m_i-convex, L_i smooth, one is nonsmooth

Operator Classes

Ensure convergence over *class* of F, not just specific F

$\mathcal{O}$: desired class of operators F having unique (β^*, w^*) with

$$z^* = \mathbf{1}_s \otimes \beta^*, \qquad w^* \in F(z^*), \qquad \textstyle\sum_{i=1}^s w^{i*} = 0$$

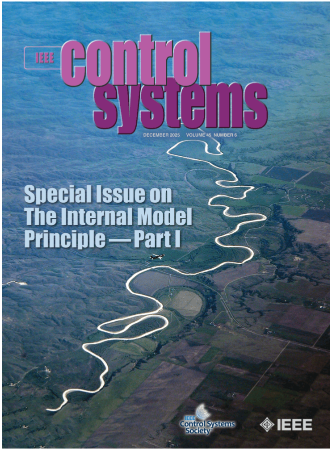
Example: each f_i is m_i-convex, L_i smooth, one is nonsmooth

Algorithm Synthesis:

Design K such that $F \star (P \star K)$ is convergent for all $F \in \mathcal{O}$

Optimization as Disturbance Rejection

Send **unknown** errors $(w - w^*, z - z^*)$ to 0 using Regulation [3]

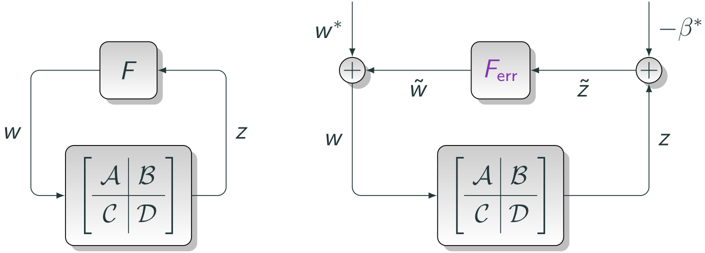


[3] Francis, Bruce A., and Walter Murray Wonham. "The internal model principle of control theory." Automatica 12.5 (1976): 457-465.

Error System

Robust control theory requires convergence to 0, not to β^*

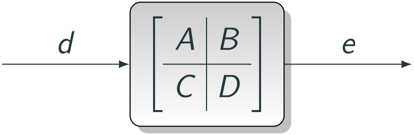
Error map $F_{\text{err}}(\tilde{z}) := F(\tilde{z} + \mathbf{1}_s \otimes \beta^*) - w^*$ satisfies $0 \in F_{\text{err}}(0)$



When F affine, this is a linear system disturbed by (β^*, w^*)

Linear Regulation (Constant Disturbance)

Rejection of constant disturbances d (e.g. setpoint tracking)



Want to satisfy two properties:

1. **Stability:** $x \to 0$ when $d = 0$,
2. **Output Regulation:** $e \to 0$ even when $d \neq 0$

Disturbance Rejection (Constant Disturbance)

Necc. and suff.[4] for disturbance rejection in linear systems:

1. **Stability:** A is Schur ($\rho(A) < 1$)
2. **Regulator Equation:** Linear equation is solvable

$$\exists \mathbf{X} : \qquad \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \mathbf{X}^* \\ I \end{pmatrix} = \begin{pmatrix} \mathbf{X}^* \\ 0 \end{pmatrix}$$

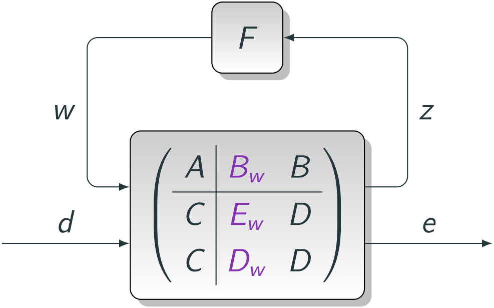
Asymptotic properties: $x \to x^* = \mathbf{X}d,\ e \to 0$ for all (x_0, d)

[4] Francis, Bruce A. "The linear multivariable regulator problem." SIAM Journal on Control and Optimization 15.3 (1977): 486-505. Includes non-constant persistent disturbances and the generalization to controller synthesis.

Disturbance Rejection through Maps

$\mathcal{O}^0$: zero-centered operators $F \in \mathcal{O}$, $0 \in F(0)$

Disturbance rejection: want $e \to 0$ for all (x_0, d, F)



Regulation through Maps

If linear map $F(z) = \Delta z$ is applied (quadratic objective), then

Stability: $A + B_w \Delta (I - E_w \Delta)^{-1} C$ is Schur

Regulation through Maps

If linear map $F(z) = \Delta z$ is applied (quadratic objective), then

Stability: $A + B_w \Delta (I - E_w \Delta)^{-1} C$ is Schur

Regulator Equation: **Same** as before

$$\exists \mathbf{X} : \qquad \begin{pmatrix} A - I & B \\ C & D \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Regulation through Maps

If linear map $F(z) = \Delta z$ is applied (quadratic objective), then

Stability: $A + B_w \Delta (I - E_w \Delta)^{-1} C$ is Schur

Regulator Equation: **Same** as before

$$\exists \mathbf{X} : \qquad \begin{pmatrix} A - I & B \\ C & D \end{pmatrix} \begin{pmatrix} \mathbf{X} \\ I \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Cancellation of Δ contribution by regulator equation

$$\left(\frac{x_{k+1}}{e_k} \right) = \left[\left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) + \text{Something}(\Delta) \left(\begin{array}{c|c} C & D \end{array} \right) \right] \left(\frac{x_k}{d} \right)$$

Extends to nonlinearities $F \in \mathcal{O}^0$ if well-posed

Convergence of Algorithms

Convergence Conditions

Regulation: $e := z - z^* \to 0$ in the presence $d := (\beta^*, w^*)$

Convergent algorithm for unique (β^*, w^*) must satisfy:

1.	**Robust Stability**	Dynamics (difficult)
2.	**Regulator Equation**	Linear Algebra (easy)

Stability tests used for Dynamics in Optimization Literature:

- Lyapunov
- Dissipativity
- Interpolation Conditions
- Integral Quadratic Constraints

Convergence: No Network Dynamics

Robust Stability: $x \to 0$ for all x_0 and $F \in \mathcal{O}^0$

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

Regulator Equation [5]: with consensus N, $(\mathbf{1}_s \otimes I)^\top N = 0$

$$\exists \mathbf{X} : \left(\begin{array}{c:cc} \mathcal{A} - I & 0 & \mathcal{B}N \\ \mathcal{C} & \mathbf{1}_s \otimes I_c & \mathcal{D}N \end{array}\right) \left(\begin{array}{c} \mathbf{X} \\ \hdashline I \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right),$$

[5] Prop 1., Range Space condition from Upadhyaya, Manu, et al. "Automated tight Lyapunov analysis for first-order methods." Mathematical Programming 209.1 (2025): 133-170.

Convergence Conditions (Networked)

Robust Stability: $F_{\text{err}} \star (P \star K)$ has $x \to 0$ for all x_0

Regulator Equation: Matrices $(\mathbf{X}_N^*, \mathbf{X}_c^*, \mathbf{U}^*, \mathbf{Y}^*)$ exist with

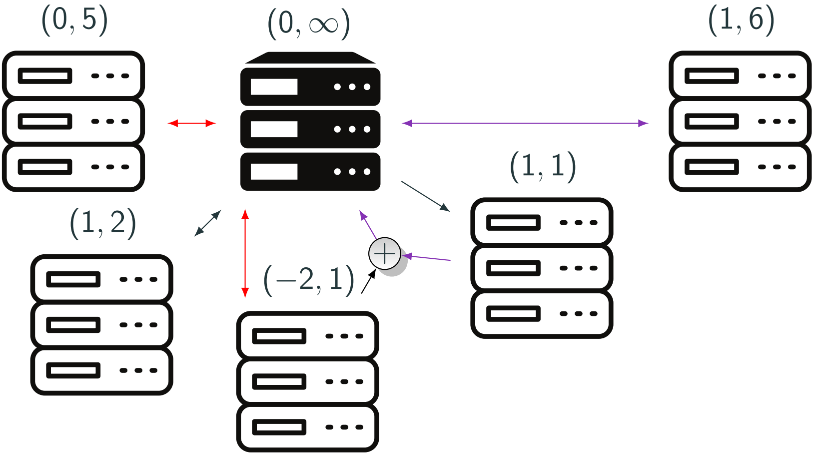
$$\left(\begin{array}{c|cc:c} A & 0 & B_w N & B_u \\ \hline C_z & \mathbf{1}_s \otimes I_c & D_{zw} N & D_{zu} \\ \hdashline C_y & 0 & D_{yw} N & D_{yu} \end{array}\right) \left(\begin{array}{c} \mathbf{X}_N^* \\ \hline I \\ \hdashline \mathbf{U}^* \end{array}\right) = \left(\begin{array}{c} \mathbf{X}_N^* \\ \hline 0 \\ \hdashline \mathbf{Y}^* \end{array}\right),$$

$$\left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \left(\begin{array}{c} \mathbf{X}_c^* \\ \hline \mathbf{Y}^* \end{array}\right) = \left(\begin{array}{c} \mathbf{X}_c^* \\ \hline \mathbf{U}^* \end{array}\right)$$

Top equation infeasible: no convergent algorithm exists

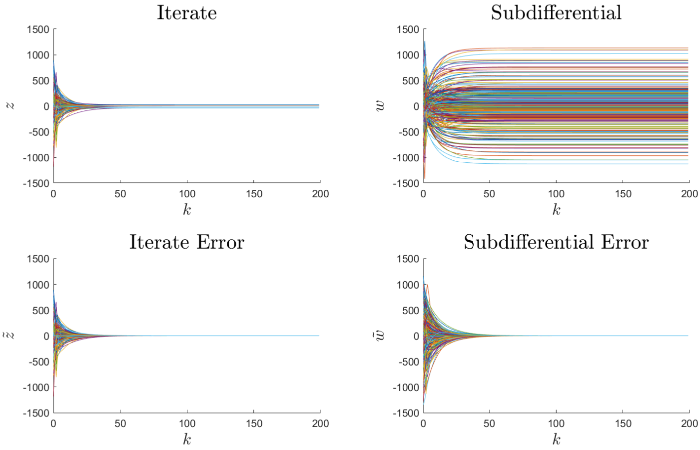
Server Example: With Network Dynamics

Sparse $\beta \in \mathbb{R}^{200}$: $f_{1:5}$ quadratics, $f_6 = \chi_{\|\cdot\|_1 \le 1000}$, $\|\beta^*\|_0 = 27$



Server Example: Convergence

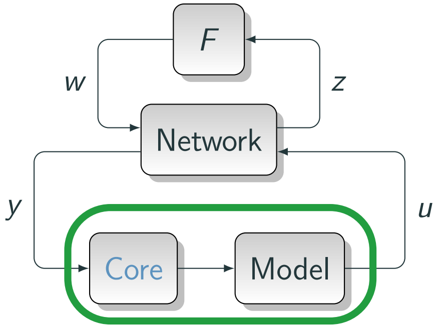
Designed algorithm: order(P) $= 9c$, order(K) $= 21c$



Structure of Algorithms

Structural Factorization

If $F \star (P \star K)$ convergent, then K factors into Core and Model



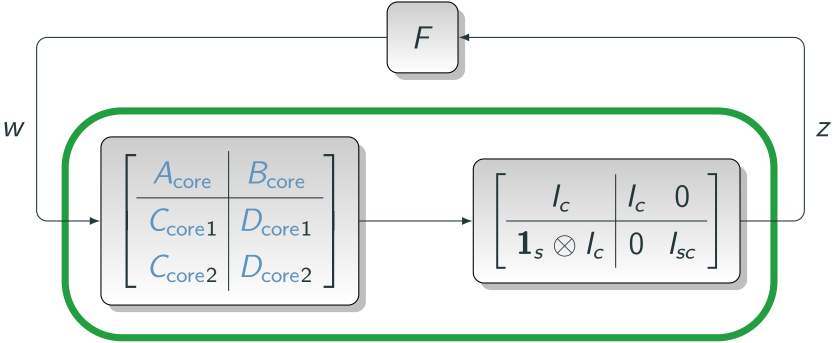
Model depends only on network P, not on K or $\mathcal{O}$

Based on the Internal Model Principle of control theory[6]

[6] Francis, Bruce A., and Walter Murray Wonham. "The internal model principle of control theory." Automatica 12.5 (1976): 457-465.

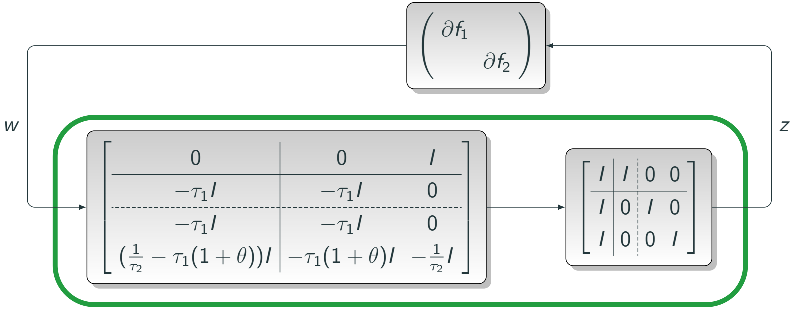
Factorization for No Network Dynamics

Factorization of convergent LTI algorithms for $\sum_{i=1}^{s} m_i > 0$



Factorization Example: Chambolle-Pock

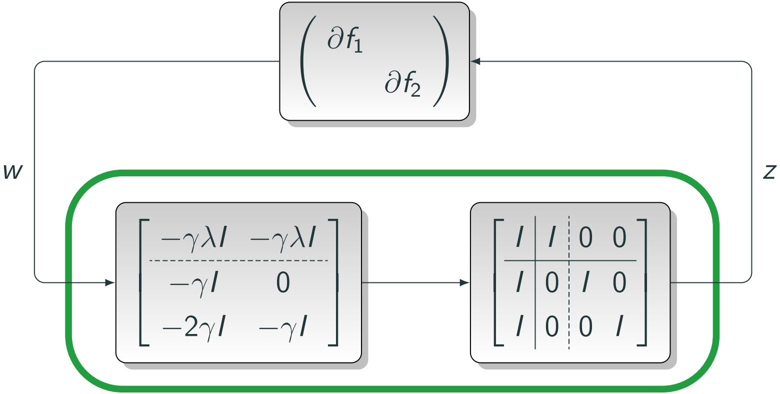
Chambolle-Pock [7] with parameters (τ_1, τ_2, θ)



[7] Chambolle, Antonin, and Pock, Thomas. "A first-order primal-dual algorithm for convex problems with applications to imaging." Journal of Mathematical Imaging and Vision 40.1 (2011): 120-145.

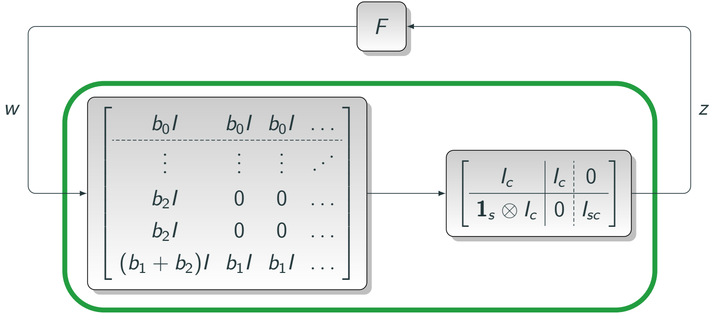
Factorization Example: Douglas-Rachford

Douglas-Rachford with parameters (λ, γ)



Factorization Example: Static Structure

Parameterization $(b_0, b_1, b_2) \in \mathbb{R}^3$ for common algorithms



Algorithm	(b_0, b_1, b_2)
Douglas-Rachford, Davis-Yin:	$(-\gamma\lambda, -\gamma, -\gamma)$
Proximal Point, Forward-Backward:	$(-\gamma, -\gamma, 0)$
Gradient Descent:	$(-\gamma, 0, 0)$

Analysis and Synthesis

Exponential Convergence Rates

Analyze speed of convergence, **Synthesize** fast algorithms

Exponential convergence at rate ρ: $\exists \gamma > 0, \forall\ k \in \mathbb{N},\ x_0 \in \mathbb{R}^n$:

$$\max\{\|x_k - x^*\|_2, \|w_k - w^*\|_2, \|z_k - z^*\|_2\} \leq \gamma \rho^k \|x_0 - x^*\|_2.$$

Exponential Convergence Rates

Analyze speed of convergence, **Synthesize** fast algorithms

Exponential convergence at rate ρ: $\exists \gamma > 0, \forall\ k \in \mathbb{N},\ x_0 \in \mathbb{R}^n$:

$$\max\{\|x_k - x^*\|_2, \|w_k - w^*\|_2, \|z_k - z^*\|_2\} \leq \gamma \rho^k \|x_0 - x^*\|_2.$$

Algorithm Analysis:

Verify worst-case rate ρ over all $F \in \mathcal{O}$

Dissipativity/IQCs: Replace $w_t \in F(z_t)$ by *valid relations*

Convex Functions: Subgradient Inequality

If f is convex, subgradient inequality implies $\forall w \in \partial f(z)$

$$f(z) - f(z') \leq w^\top (z - z')$$

and if $z' = 0,\ f(z') = 0$: $\qquad 0 \leq f(z) \leq w^\top z$

[8] Van der Schaft, A. (2000). L2-gain and passivity techniques in nonlinear control. Berlin, Heidelberg: Springer Berlin Heidelberg.

Convex Functions: Subgradient Inequality

If f is convex, subgradient inequality implies $\forall w \in \partial f(z)$

$$f(z) - f(z') \leq w^\top (z - z')$$

and if $z' = 0,\ f(z') = 0$: $\qquad 0 \leq f(z) \leq w^\top z$

Any sequence $\{(w_k, z_k)\}_{k=-\infty}^{\infty}$ with $w_k \in \partial f(z_k)$ satisfies [8]

$$0 \leq f(z_k) \leq w_k^\top z_k \ \forall k, \quad \text{implying } \textit{passivity } 0 \leq \textstyle\sum_{k=0}^{\infty} w_k^\top z_k$$

[8] Van der Schaft, A. (2000). L2-gain and passivity techniques in nonlinear control. Berlin, Heidelberg: Springer Berlin Heidelberg.

Convex Functions: Zames-Falb

We have static and dynamic relations:

$$\begin{array}{rr} \text{Static} & 0 \leq \sum_{k=0}^{\infty} w_k^\top z_k \\ \text{Dynamic} & \forall k:\ 0 \leq \sum_{k=0}^{\infty} w_k^\top (z_k - z_{k-t}) \end{array}$$

Convex Functions: Zames-Falb

We have static and dynamic relations:

Static	$0 \leq \sum_{k=0}^{\infty} w_k^\top z_k$
Dynamic	$\forall k: \; 0 \leq \sum_{k=0}^{\infty} w_k^\top (z_k - z_{k-t})$

For *any* nonnegative coefficients $\{\mu_t\}_{t=0}^{\infty}$, we add them up:

$$0 \leq \sum_{k=0}^{\infty} \left(\mu_0 w_k^\top z_k\right) + \left(\sum_{t=1}^{\infty} \mu_t w_k^\top (z_k - z_{k-t})\right)$$

Convex Functions: Zames-Falb

We have static and dynamic relations:

Static	$0 \le \sum_{k=0}^{\infty} w_k^\top z_k$
Dynamic	$\forall k:\ 0 \le \sum_{k=0}^{\infty} w_k^\top (z_k - z_{k-t})$

For *any* nonnegative coefficients $\{\mu_t\}_{t=0}^{\infty}$, we add them up:

$$0 \le \sum_{k=0}^{\infty} \left(\mu_0 w_k^\top z_k\right) + \left(\sum_{t=1}^{\infty} \mu_t w_k^\top (z_k - z_{k-t})\right)$$

This forms a *Zames-Falb* filter $\Psi(\mu)$:

$$r_k := \left(\textstyle\sum_{t=0}^{\infty} \mu_t\right) z_k - \textstyle\sum_{t=1}^{\infty} \mu_t z_{k-t} \qquad \textstyle\sum_{k=0}^{\infty} w_k^\top r_k \ge 0$$

Strongly Convex Functions

Strong convexity with parameter $m > 0$:

$$f(z') \geq f(z) + w^\top (z' - z) + \frac{m}{2}\|z' - z\|_2^2, \qquad f - \frac{m}{2}\|\cdot\|_2^2 \text{ is convex}$$

Same Zames-Falb filter $\Psi(\mu)$: [9]

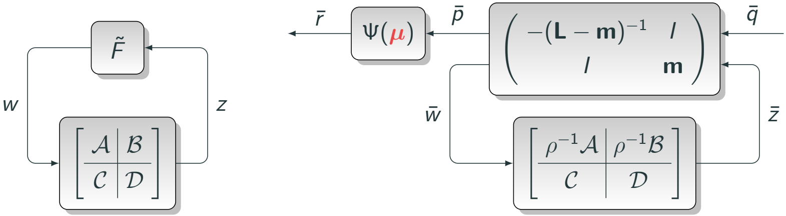
$$r_k := \left(\sum_{t=0}^{\infty} \mu_t\right) z_k - \sum_{t=1}^{\infty} \mu_t z_{k-t} \qquad \sum_{k=0}^{\infty} (w_k - m z_k)^\top r_k \geq 0$$

[9] Scherer, C. W. (2023). Robust exponential stability and invariance guarantees with general dynamic O'Shea-Zames-Falb multipliers. IFAC-PapersOnLine, 56(2), 5799-5804.

Filtering Framework (Analysis)

$\mathcal{O}$: $f_i - \frac{m_i}{2}\|\cdot\|_2^2$ p.c.c, and $\frac{L_i}{2}\|\cdot\|_2^2 - f_i$ p.c.c. if $L_i < \infty$ for all i

Zames-Falb filter[10] : $\Psi(\mu)$: $\quad \bar{r}_k^i = \sum_t \mu_t^i \bar{p}_{k-t}^i$



[10] ρ-weighted generalizations of the subgradient inequality. Scherer, Carsten W., Christian Ebenbauer, and Tobias Holicki. "Optimization algorithm synthesis based on integral quadratic constraints: A tutorial." 2023 62nd IEEE Conference on Decision and Control (CDC). IEEE, 2023.

Analysis Task

Linear Matrix Inequality in $(\boldsymbol{\mu}, \mathcal{X})$ for fixed ρ:

$$\mathcal{X} \succ 0,\ \boldsymbol{\mu}_0^i > 0, \quad \boldsymbol{\mu}_t^i \geq 0,$$

$$[*]^\top \begin{bmatrix} \mathcal{X} & 0 \\ 0 & -\mathcal{X} \end{bmatrix} \begin{bmatrix} \hat{\mathcal{A}}(\boldsymbol{\mu}) & \hat{\mathcal{B}}(\boldsymbol{\mu}) \\ I & 0 \end{bmatrix} + [*]^\top \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} \hat{\mathcal{C}}(\boldsymbol{\mu}) & \hat{\mathcal{D}}(\boldsymbol{\mu}) \\ 0 & I \end{bmatrix} \prec 0$$

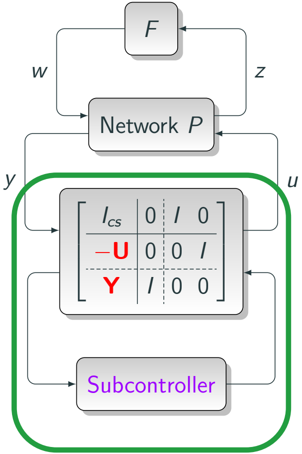
Algorithm ρ-convergent *and well-posed* if (sufficient)

1. Regulator Equation satisfied
2. LMI is feasible (strict negative realness)
3. $\mathcal{D}$ is block-lower-triangular

Full-Order Internal Models and Synthesis

Add a model in the middle

Regulator Equation *always* satisfied

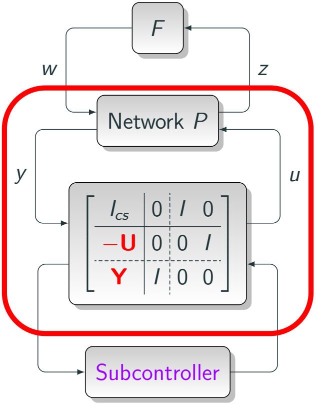


Full-Order Internal Models and Synthesis (cont.)

Accumulate Network and Model

Robust-control design of Subcontroller

Allows for performance
(e.g. convergence rates, tracking)



Alternating Synthesis

Alternating convex problem

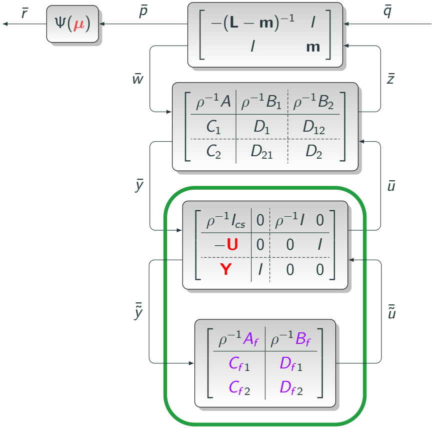
Bisection in ρ: inner steps

Start with $\Psi(\mu) = I_{cs}$

For fixed μ : optimize Σ_f

For fixed Σ_f : optimize μ

Repeat (μ, Σ_f) loop

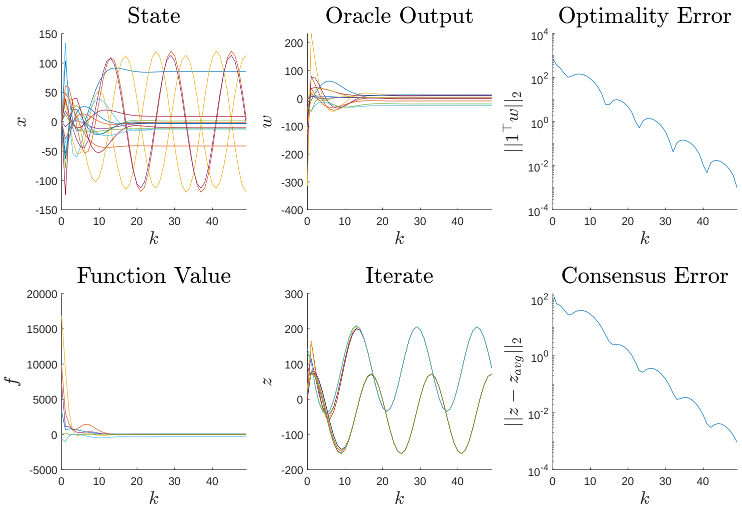


LMI solvable $\bar{q} \to \bar{r}$

Extensions

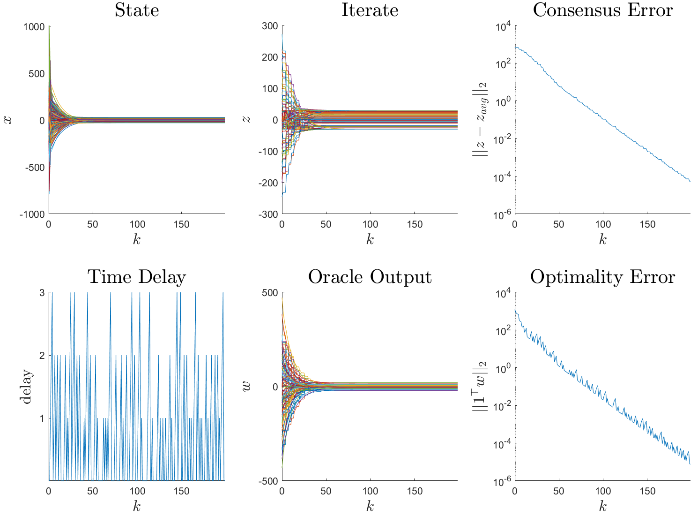
Perfect Tracking

Track the time-varying optimal solution



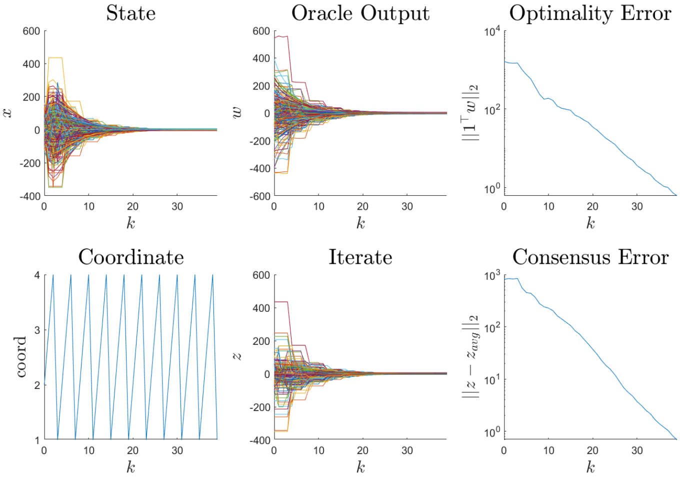
Time-Varying Delay

Time-varying delay before ∂f_1



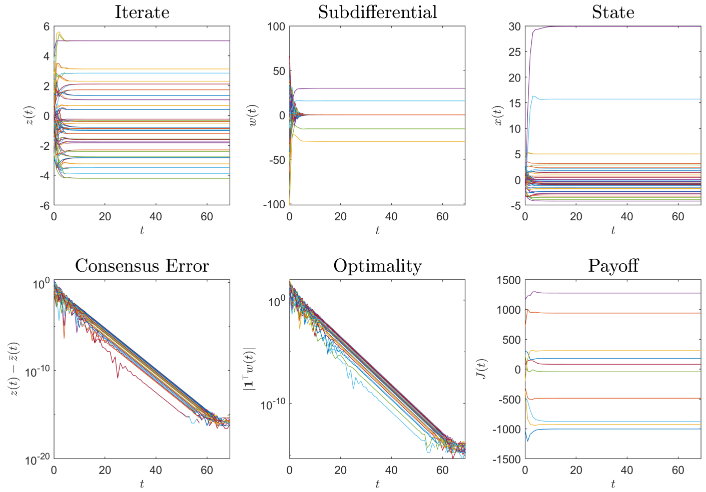
Coordinate Descent

Coordinate descent with 4 blocks



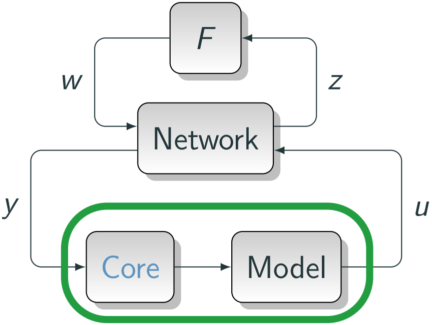
Fixed Point Equations

Constrained Nash-Equilibrium seeking under delays



Take-aways

Conclusion



Convergence

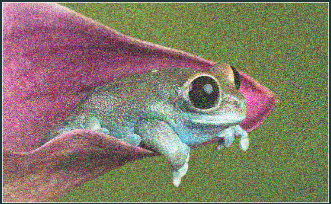
Structure

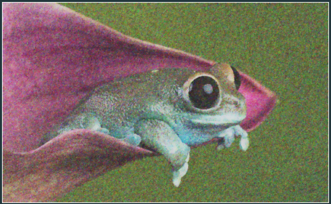
Class of F →

Network P →

→ certified algorithm

Optimize(d) using Structure!





Backup Slides

Well-Posedness

Well-posed: $H := (F^{-1} - \mathcal{D})^{-1}$ continuous, globally defined

$$x_{k+1} = \mathcal{A}x_k + \mathcal{B}H(\mathcal{C}x_k), \qquad w_k = H(\mathcal{C}x_k)$$

Unique trajectory $(x_k, w_k, z_k)_{k\in\mathbb{N}}$ exists for all x_0

Well-posed algorithmic interconnections: $F \star G$, $F \star P \star K$

Information Structures

Block sparsity pattern of $\mathcal{D}$: information flow in iteration k

$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$: parallel, w_1, w_2 computed independently

$\begin{pmatrix} 0 & 0 \\ \bullet & 0 \end{pmatrix}$: sequential, z_2 can use knowledge of $w_1 = \nabla f(z_1)$

$\begin{pmatrix} 0 & 0 \\ \bullet & \bullet \end{pmatrix}$: implicit loop, z_2 uses $\nabla f_1(z_1)$ and $\partial f_2(z_2)$

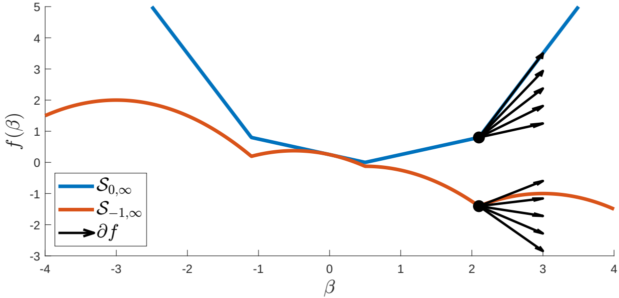
Resolvent-and-gradient evaluation: $\mathcal{D}$ block-lower-triangular

Subdifferentials of Functions (Frechét)

Sufficient for optimality: $0 \in \sum_{i=1}^{s} \partial f_i(\beta^*)$

f p.c.c. : $\partial f(\beta) = \{g \mid \forall \beta' \in \mathbb{R}^c : \ f(\beta') \geq f(\beta) + g^\top(\beta' - \beta)\}$

$f \in \mathcal{S}_{m,L}$: $\partial f(\beta) = \partial(f - \frac{m}{2}\|\cdot\|_2^2)\beta + m\beta$



$\partial f \star G$ well-posed $\forall f \in \mathcal{S}_{m,L}$: $\mathrm{Sym}(\mathcal{D}(I - m\mathcal{D})^{-1} - \frac{1}{L-m}) \prec 0$

Bonus: Regulation Theory (open-loop)

Regulator Equation (constant disturbance)

Apply coordinate transformation $(\hat{x}, \hat{d}) = (x - \Upsilon d, d)$

$$\left(\begin{array}{c} \hat{x}_{k+1} \\ \hat{d}_{k+1} \\ \hline e_k \end{array}\right) = \left(\begin{array}{c:c} \mathcal{A} & (\mathcal{A} - I)\Upsilon + \mathcal{B} \\ \hdashline 0 & I \\ \hline \mathcal{C} & \mathcal{C}\Upsilon + \mathcal{D} \end{array}\right) \left(\begin{array}{c} \hat{x}_k \\ \hdashline \hat{d}_k \end{array}\right) = \left(\begin{array}{c:c} \mathcal{A} & 0 \\ \hdashline 0 & I \\ \hline \mathcal{C} & 0 \end{array}\right) \left(\begin{array}{c} \hat{x}_k \\ \hdashline \hat{d}_k \end{array}\right)$$

$\lim_{k\to 0} e_k = \lim_{k\to 0} \mathcal{C}\hat{x}_k + (\mathcal{C}\Upsilon + \mathcal{D})d_k = (\mathcal{C}\Upsilon + \mathcal{D})d_0$

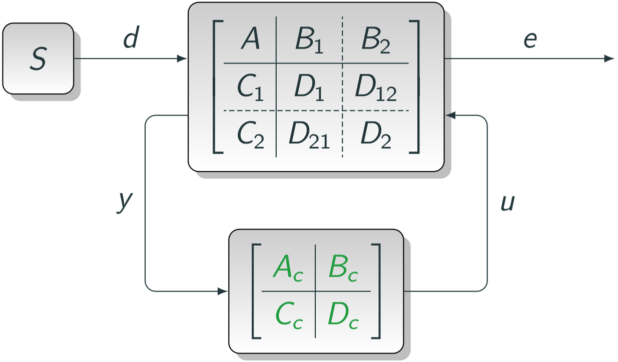
Persistent error if $d \neq 0$ and $(\mathcal{C}\Upsilon + \mathcal{D}) \neq 0$

Stability + Regulator Equation are therefore necessary and sufficient for disturbance rejection

Regulation Framework (open-loop)

Given plant P, exosystem S, design a controller K such that:

1. $G = P \star K$ is internally stable ($\mathcal{A}$ from $(G)_{ss}$ Schur)
2. G achieves output regulation



Regulator Equation (open-loop)

Output regulation of $G = P \star K$: unique Υ certificate

Open-loop formulation based on matrices $(\Pi, \Gamma, \Phi, \Theta)$

After indexing $\begin{pmatrix} \Pi \\ \Theta \end{pmatrix} := \Upsilon$ from G, define

$$\Gamma := (I - D_c D_{22})^{-1}(C_c \Theta + D_c(C_2 \Pi + D_{21})) \tag{1a}$$

$$\Phi := D_2 \Gamma + C_2 \Pi + D_{21} \tag{1b}$$

$(\Pi, \Gamma, \Phi, \Theta)$ exist by output regulation of $G = P \star K$

Regulator Equation (open-loop)

Perform coordinate transformations like before:

$$\hat{x}^N = x^N - \Pi d, \quad \hat{\xi} = \xi - \Theta d, \quad \hat{u} \quad = u - \Gamma d, \hat{y} = y - \Phi d$$

Output regulation ($d \to$ all) must be 0: Regulator Equation

$$d_{k+1} = S d_k \tag{2}$$

$$\begin{pmatrix} \hat{x}^N_{k+1} \\ \hline e_k \\ \hat{y}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A & A\Pi - \Pi S + B_2\Gamma + B_1 & B_2 \\ \hline C_1 & C_1\Pi + D_{12}\Gamma + D_1 & D_{12} \\ C_2 & C_2\Pi + D_2 - \Phi + D_{21} & D_2 \end{array}\right) \begin{pmatrix} \hat{x}^N_k \\ \hline d_k \\ \hat{u}_k \end{pmatrix}$$

$$\begin{pmatrix} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{pmatrix} = \left(\begin{array}{c|c:c} A_c & A_c\Theta - \Theta S + B_c\Phi & B_c \\ \hline C_c & C_c\Theta + D_c\Phi - \Gamma & D_c \end{array}\right) \begin{pmatrix} \hat{\xi}_k \\ \hline d_k \\ \hat{y}_k \end{pmatrix}$$

Regulator Equation (open-loop)

Output regulation achieved if $\mathcal{A}$ Schur and $(\Pi, \Theta, \Gamma, \Phi)$ satisfy

$$\left(\begin{array}{c|cc} A & B_1 & B_2 \\ \hline C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & D_{22} \end{array}\right)\left(\begin{array}{c} \Pi \\ \hline I \\ \Gamma \end{array}\right) = \left(\begin{array}{c} \Pi S \\ \hline 0 \\ \Phi \end{array}\right), \tag{3}$$

$$\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix}\begin{pmatrix} \Theta \\ \Phi \end{pmatrix} = \begin{pmatrix} \Theta S \\ \Gamma \end{pmatrix}, \tag{4}$$

Asymptotic tracking, $\lim_{k\to\infty}(\hat{x}_k^N, \hat{\xi}, \hat{u}, \hat{y},) = 0$ implies

$$\lim_{k\to\infty} \|x_k^N - \Pi d_k,\ \xi_k - \Theta d_k,\ u_k - \Gamma d_k,\ y_k - \Phi d_k\| = 0. \tag{5}$$

Solvability of (3) **necessary** for perfect tracker to exist

Bonus: Proof of Convergence

Test Quadratic

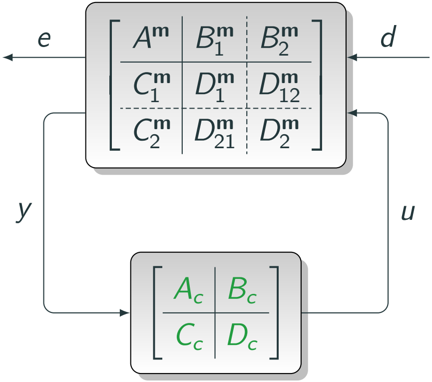
Test quadratic inside $\mathcal{S}_{m_i,L_i}$:
$f_i = m_i\mathbf{q}(z_i) + b_i^\top z_i + b_i^0$

Constant $\mathbf{m} = \operatorname{diag} m \otimes I_c$

$F(z) = \mathbf{m}z + b, \;\; \tilde{F}(e) = \mathbf{m}e$

Linear system $G^{\mathbf{m}} = \mathbf{m} \star P^e$,

$(\mathbf{m}z + b) \star P \star K$ opt. alg. $\Leftrightarrow$
$G^{\mathbf{m}}$ achieves output regulation



Interconnection $G^{\mathbf{m}} \star K$

Assumptions: Convergence

Gives us a chance of assuring convergence in iterates:

A1 $\sum_{i=1}^{s} m_i > 0$ (strong convexity/uniqueness)

A2 $E(\mathbf{m}) = I - D_1\mathbf{m}$ is invertible

A3 $(A^{\mathbf{m}}, B_2^{\mathbf{m}})$ stabilizable, $(A^{\mathbf{m}}, C_2^{\mathbf{m}})$ detectable

A2+A3: $G^{\mathbf{m}} \star K$ must be able to achieve output regulation

Convergence Conditions

Under A1-A3, any $F \star P \star K$ that is an optimization algorithm for all $F \in \mathcal{O}_{m,L}$ must satisfy two conditions:

1. Regulator Equation:
2. Robust Stability

Stability uses $\mathcal{O}^0_{m,L} := \{F \in \mathcal{O}_{m,L} \mid \; f_i(0) = 0, \; 0 \in \partial f_i(0), \forall i\}$

Regulator Equation: verify by linear algebra

Robust Stability: More difficult, requires dynamics

Convergence: Test Regulation

Why these conditions?

$G^{\mathbf{m}} \star K$ must have output regulation, inside class $\mathcal{O}^0_{m,L}$

Describe state-space matrices of $G^{\mathbf{m}}$ as $\Omega_0 + \Omega_{\mathbf{m}}$:

$$\Omega_0 := \left(\begin{array}{c|cc:c} A & 0 & B_1 N & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_2 & 0 & D_{21} N & D_2 \end{array}\right), \tag{6}$$

$$\Omega_{\mathbf{m}} := \left(\begin{array}{c} B_1 \\ \hline D_1 \\ \hdashline D_{21} \end{array}\right) \mathbf{m} E(\mathbf{m})^{-1} \left(\begin{array}{c|cc:c} C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \end{array}\right)$$

Convergence: Test System Regulation

Exists $(\Pi, \Gamma, \Phi, \Theta)$ solving regulator equation for $G^{\mathbf{m}}$:

$$(\Omega_0 + \Omega_{\mathbf{m}}) \begin{pmatrix} \Pi \\ \hline I \\ \Gamma \end{pmatrix} = \begin{pmatrix} \Pi \\ \hline 0 \\ \Phi \end{pmatrix}, \quad \begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} \Theta \\ \Phi \end{pmatrix} = \begin{pmatrix} \Theta \\ \Gamma \end{pmatrix}. \tag{7}$$

But $\Omega_{\mathbf{m}} \begin{pmatrix} \Pi \\ \hline I \\ \Gamma \end{pmatrix} = 0$, leaving statement $\Omega_0 \begin{pmatrix} \Pi \\ \hline I \\ \Gamma \end{pmatrix} = \begin{pmatrix} \Pi \\ \hline 0 \\ \Phi \end{pmatrix}$

Convergence: Nonlinear Regulation

Perform coordinate transformations

$$\hat{x}^N = x^N - \Pi d, \quad \hat{\xi} = \xi - \Theta d, \quad \hat{y} = y - \Phi d, \quad \hat{u} = u - \Gamma d$$

Decoupling of d by regulator equation:

$$\left(\begin{array}{c} \hat{x}_{k+1}^N \\ \hline \tilde{z}_k \\ \hdashline e_k \\ \hdashline \hat{y}_k \end{array}\right) = \left(\begin{array}{c|c:c:c} A & B_1 & 0 & B_2 \\ \hline C_1 & D_1 & 0 & D_{12} \\ \hdashline C_1 & D_1 & 0 & D_{12} \\ C_2 & D_{21} & 0 & D_2 \end{array}\right) \left(\begin{array}{c} \hat{x}_k^N \\ \hline \tilde{w}_k \\ \hdashline d_k \\ \hdashline \hat{u}_k \end{array}\right) \tag{8}$$

$$\left(\begin{array}{c} \hat{\xi}_{k+1} \\ \hline \hat{u}_k \end{array}\right) = \left(\begin{array}{c|c:c} A_c & 0 & B_c \\ \hline C_c & 0 & D_c \end{array}\right) \left(\begin{array}{c} \hat{\xi}_k \\ \hline d_k \\ \hdashline \hat{y}_k \end{array}\right), \qquad \tilde{w} \in \tilde{F}(\tilde{z}_k). \tag{9}$$

Convergence: Finish

Robust Stability: $\lim_{k\to\infty}(\hat{x}_k^N, \hat{\xi}_k) = (0,0)$ for all $(\hat{x}_0^N, \hat{\xi}_0)$

Now apply well-posedness of $F \star P \star K$ for all $F \in \mathcal{O}_{m,L}$

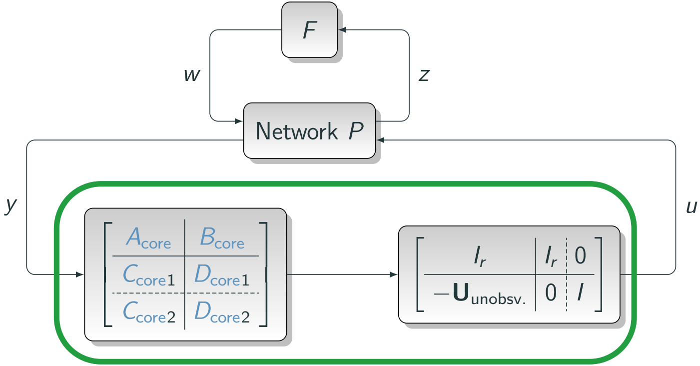
$$\lim_{k\to\infty}(\hat{x}_k^N, \hat{\xi}_k) = (0,0) \implies \lim_{k\to\infty}(\tilde{z}_k, \tilde{w}_k) = (0,0) \tag{10}$$

This ensures output regulation: $\lim_{k\to\infty} e_k = \lim_{k\to\infty} \tilde{z}_k = 0$

Linear regulator theory used for nonlinear problem

Bonus: Proof of Structure

Factorization: General Setting



Core restricted to subspace $\mathcal{L}_{\text{core}}$, size $r := \dim(\text{null}(\mathbf{Y}))$

Assumptions: Structure

New assumptions for common structure of K (nice to have)

A4 $\begin{pmatrix} A - I & B_2 \\ C_1 & D_{12} \end{pmatrix}$ has full column rank (unique Π, Γ, Φ).

A5 $\left(\begin{pmatrix} A^{\mathbf{m}} & B_1^{\mathbf{m}} \\ 0 & I_{cs} \end{pmatrix}, \begin{pmatrix} C_2^{\mathbf{m}} & D_{21}^{\mathbf{m}} \end{pmatrix} \right)$ is detectable.

Direct interconnection: A4 met, A5 requires $\sum_{i=1}^{s} m_i \neq 0$.

Exposure of Internal Model

Internal Model Principle from [11] [12]

With $r = \dim(\text{null}(\Phi))$, recall regulator equation in Θ

$$\begin{pmatrix} A_c & B_c \\ C_c & D_c \end{pmatrix} \begin{pmatrix} \Theta \\ \Phi \end{pmatrix} = \begin{pmatrix} \Theta \\ \Gamma \end{pmatrix}, \tag{11}$$

A5: $\text{null}(\Phi) \cap \text{null}(\Gamma) = \{0\}$, and $C_c(\Theta v) = \Gamma v \quad \forall v \in \text{null}(\Phi)$

Consequence: $\text{rank}(\Theta) \geq r$

---

[11] Francis, Bruce A., and Walter Murray Wonham. "The internal model principle of control theory." Automatica 12.5 (1976): 457-465.

[12] Stoorvogel, Anton A., Ali Saberi, and Peddapullaiah Sannuti. "Performance with regulation constraints." Automatica 36.10 (2000): 1443-1456.

Exposure of Internal Model (cont.)

Choose invertible (P, Q), such that Φ_2 has full column rank

$$\Gamma P = \begin{pmatrix} \Gamma_1 & \Gamma_2 \end{pmatrix}, \quad \Phi P = \begin{pmatrix} 0 & \Phi_2 \end{pmatrix}, \quad Q^{-1}\Theta P = \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \end{pmatrix}.$$

Post-multiply regulator equation by P, transform $(K)_{ss}$ by Q:

$$\left(\begin{array}{cc|c} {A_c}'_{11} & {A_c}'_{12} & {B_c}'_1 \\ {A_c}'_{21} & {A_c}'_{22} & {B_c}'_2 \\ \hline {C_c}'_1 & {C_c}'_2 & {D_c}' \end{array}\right) \begin{pmatrix} -I_r & \Theta_{11} \\ 0 & \Theta_{22} \\ 0 & \Phi_2 \end{pmatrix} = \begin{pmatrix} -I_r & \Theta_{11} \\ 0 & \Theta_{22} \\ \Gamma_1 & \Gamma_2 \end{pmatrix} \qquad (12)$$

Exposure of Internal Model

Line up constraints, label remaining terms as core subsystem

$$\left(\begin{array}{cc|c} I_r & C_{\mathrm{core}1} & D_{\mathrm{core}1} \\ 0 & A_{\mathrm{core}} & B_{\mathrm{core}} \\ \hline -\Gamma_1 & C_{\mathrm{core}2} & D_{\mathrm{core}2} \end{array}\right) \begin{pmatrix} \Theta_{11} \\ \Theta_{22} \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} \Theta_{11} \\ \Theta_{22} \\ \Gamma_2 \end{pmatrix} \tag{13}$$

Recognizable as a cascade involving a system Σ_{core}

Entries of $(\Sigma_{\mathrm{core}})_{ss}$ must satisfy affine constraint

Exposure of Internal Model (Final)

Cascade $K = \Sigma_{\mathsf{min}} \Sigma_{\mathsf{core}}$, constrained as $(\Sigma_{\mathsf{core}})_{ss} \in \mathcal{L}_{\mathsf{core}}(\Theta)$

$$\Sigma_{\mathsf{min}} : \quad \left[\begin{array}{c|c:c} I_r & I_r & 0 \\ \hline -\Gamma_1 & 0 & I_{n_u} \end{array}\right] \quad \Sigma_{\mathsf{core}} : \quad \left[\begin{array}{c|c} A_{\mathsf{core}} & B_{\mathsf{core}} \\ \hline C_{\mathsf{core}1} & D_{\mathsf{core}1} \\ C_{\mathsf{core}2} & D_{\mathsf{core}2} \end{array}\right], \qquad (14)$$

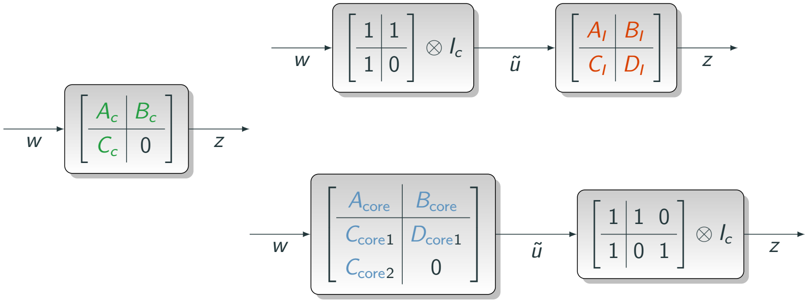
$$\mathcal{L}_{\mathsf{core}}(\Theta) : \begin{pmatrix} A_{\mathsf{core}} & B_{\mathsf{core}} \\ C_{\mathsf{core}1} & D_{\mathsf{core}1} \\ C_{\mathsf{core}2} & D_{\mathsf{core}2} \end{pmatrix} \begin{pmatrix} \Theta_{22} \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \Gamma_1 \Theta_{12} + \Gamma_2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

Σ_{min} has r states, may satisfy $r \ll sc$

Bonus: Algorithm Factorization

Factorization of Gradient Algorithms

Algorithms with $\mathcal{D} = 0,\ s = 1$ have integrator after w [13]

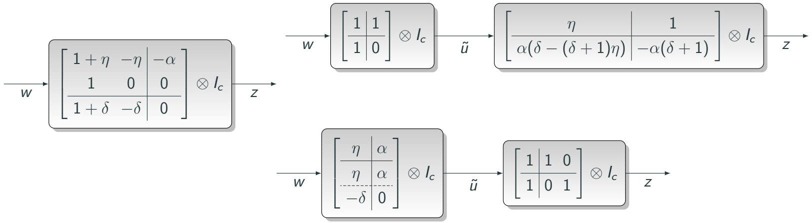


We support $s > 1$ and/or $\mathcal{D} \neq 0$, have an integrator before z

[13] Theorem 2.1: Scherer, Carsten, and Christian Ebenbauer. "Convex Synthesis of Accelerated Gradient Algorithms." SIAM J. Control Optim. 59.6 (2021): 4615-4645.

Factorization of Gradient Algorithms: TMM

Factorization of Triple-Momentum [14]



[14]Van Scoy, Bryan, Randy A. Freeman, and Kevin M. Lynch. "The fastest known globally convergent first-order method for minimizing strongly convex functions." IEEE Control Systems Letters 2.1 (2017): 49-54.

Example: Many Algorithms

Algorithm	Parameters	s	b_0	b_1	b_2
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	γ, λ	2	$-\gamma\lambda$	$-\gamma$	$-\gamma$
Forward-Backward	γ	2	$-\gamma$	0	$-\gamma$
Davis-Yin	γ, λ	3	$-\gamma\lambda$	$-\gamma$	$-\gamma$

Parameters (b_0, b_1, b_2) describing static-Σ_{core} algorithms

Bonus: Analysis and Synthesis

Passivity

ρ-Robust Stability will be verified by Passivity

Static passive operator F: $\forall w, z \mid w \in F(z) : \; w^\top z \geq 0$

Strictly anti-passive linear system: exists positive definite V:

$$V(x_T) - V(x_0) < \textstyle\sum_{k=0}^{T} -u_k^\top y_k \tag{15}$$

Interconnection of passive operator and antipassive system:

$$V(x_T) - V(x_0) < 0 \tag{16}$$

yields bounded solution trajectories

If ρ-weighted system is bounded, then system is ρ-convergent

Exponential Transformation

$F \star G$ is ρ-convergent for all $F \in \mathcal{O}_{m,L}$ if:

1. ρ-Robust Stability[15]: $\bar{x}_k \to 0 \quad \forall \tilde{F} \in \mathcal{O}^0_{m,L},\ \bar{x}_0$

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

2. Regulator Equation: Same as before

[15] Scherer, Carsten W. "Robust Exponential Stability and Invariance Guarantees with General Dynamic O'Shea-Zames-Falb Multipliers." IFAC-PapersOnLine 56.2 (2023): 5799-5804.

Structured Control

Constraints on $(A_{\mathrm{core}}, B_{\mathrm{core}}, C_{\mathrm{core}1}, C_{\mathrm{core}2})$: nonconvex design [16]

Joint search over $(\Sigma_{\mathrm{core}}, \Theta_{12}, \Theta_{22}, \boldsymbol{\lambda})$ to minimize ρ

For fixed Θ_{22}, use *hinfstruct* [17] for H_∞ synthesis

---

[16] 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.

[17] 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. IEEE, 2011, pp. 8205–8210.

Valid Relations

Subgradient inequality $f \in \mathcal{S}_{0,\infty}$, $(x, x') \in \mathrm{dom}(f)$, $g \in \partial f(x)$

$$f(x') \geq f(x) + g^\top (x' - x). \tag{17}$$

Can be generalized to $\mathcal{S}_{m,L}$, ρ-weightings

Define $V(x, g) := f(x) - m\mathbf{q}(x) - \sigma\mathbf{q}(g - mx)$

Given $f \in \mathcal{S}_{m,L}$, $\alpha_i > 0$, $\bar{x}_i \in \mathbb{R}^n, \bar{g}_i \in \alpha_i^{-1}\partial f(\alpha_i \bar{x}_i)$:

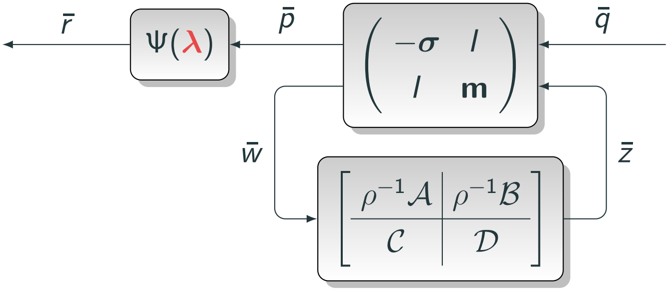
$$\begin{aligned} &V(\alpha_1\bar{x}_1, \alpha_1\bar{g}_1) - V(\alpha_2\bar{x}_2, \alpha_2\bar{g}_2) \leq \\ &\qquad \alpha_1(\bar{g}_1 - m\bar{x}_1)^\top[\alpha_1(\sigma L\bar{x}_1 - \sigma g_1) - \alpha_2(\sigma L\bar{x}_2 - \sigma\bar{g}_2)], \end{aligned} \tag{18}$$

Filtering Framework

One-step delay operator $\mathbf{z}$, filter coefficients λ^i

Zames-Falb filter: $\Psi(\boldsymbol{\lambda}) = \text{blkdiag}(\bar{p}^i_t \rightarrow \sum_{\nu=0}^{\nu_{\max}} \lambda^i_\nu \bar{p}^i_{t-\nu}) \otimes I_c$

$$\sum_{\nu=0}^{\nu_{\max}} \rho^{-\nu} \lambda_\nu > 0 \quad \forall \nu \geq 1: \ \lambda_\nu \leq 0 \tag{19}$$



Zames-Falb Filters

Conic combinations of dissipation relations $\{\mu_\nu\}_{\nu=0}^{\nu_{\max}} \geq 0$

$$\sum_{k=0}^{T-1} \left(\mu_0 \bar{q}_k^\top \bar{p}_k + \sum_{\nu=1}^{\nu_{\max}} \mu_r \bar{q}_k^\top (\bar{p}_k - \rho^\nu \bar{p}_{k-\nu}) \right). \tag{20}$$

Summarized by λ coefficients

$$\sum_{\nu=0}^{\nu_{\max}} \rho^{-\nu} \lambda_\nu > 0 \quad \forall \nu \geq 1 : \ \lambda_\nu \leq 0 \tag{21}$$

Association between λ and ν

$$\lambda_0 = \sum_{\nu=0}^{\nu_{\max}} \mu_n u \qquad\qquad \lambda_\nu = -\rho^\nu \mu_\nu \tag{22}$$