Structure, Analysis, and Synthesis of Optimization Algorithms

Jared Miller
Carsten Scherer
Fabian Jakob
Andrea Iannelli
KTH
March 27, 2026

Composite Optimization

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

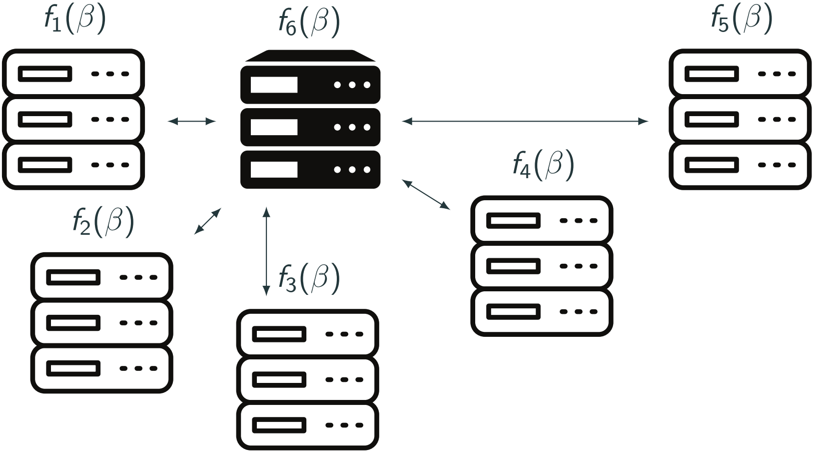
$$\beta^* \in \operatorname{argmin}_{\beta \in \mathbb{R}^c} \sum_{i=1}^{s} f_i(\beta)$$

Wide range of applications:

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

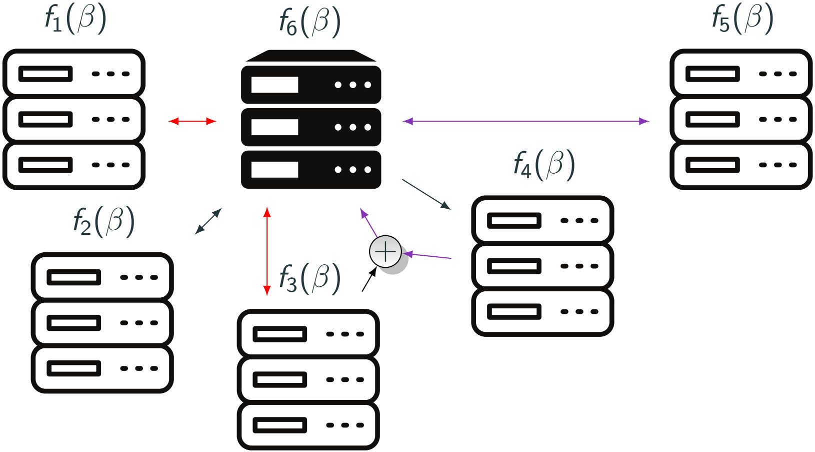
Optimization Setup

Assimilation/regression: 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

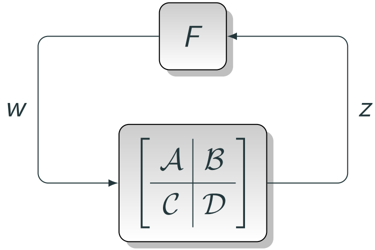
Accomplish by continuing a link of Optimization and Control

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

Background: Algorithm Setting

Algorithmic Interconnection

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



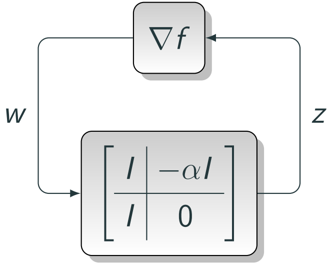
$$\begin{pmatrix} x_{k+1} \\ \hline z_k \end{pmatrix} = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \begin{pmatrix} x_k \\ \hline w_k \end{pmatrix}, \quad w_k \in F(z_k)$$

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

Examples of Interconnections: Gradient/Prox

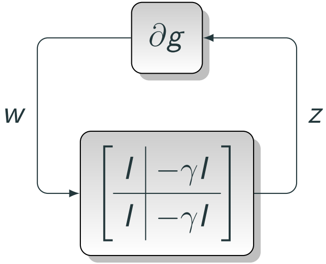
Gradient Descent

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



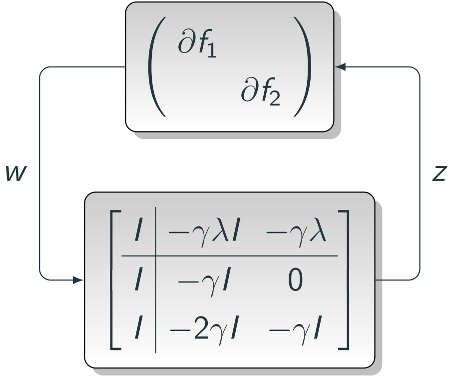
Proximal Point Method

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



Example of Interconnection: Douglas-Rachford

Douglas-Rachford Algorithm[1] with parameters (γ, λ)



[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.

Image Restoration

Total-variation denoising of multi-color images

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

Original Image X

Noisy Image Y

Denoised β^*

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

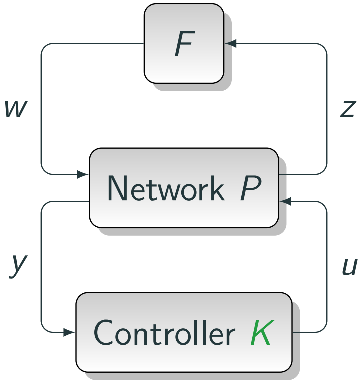
Algorithms over Networks

Network P:

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

Controller K:

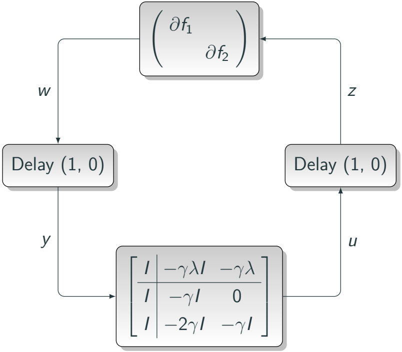
$$\left(\begin{array}{c} \xi_{k+1} \\ \hline u_k \end{array}\right) = \left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \left(\begin{array}{c} \xi_k \\ \hline y_k \end{array}\right)$$



F 'sees' system $G = P \star K$

Examples of Interconnections: Delayed DR

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



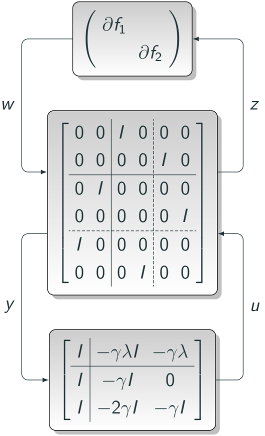
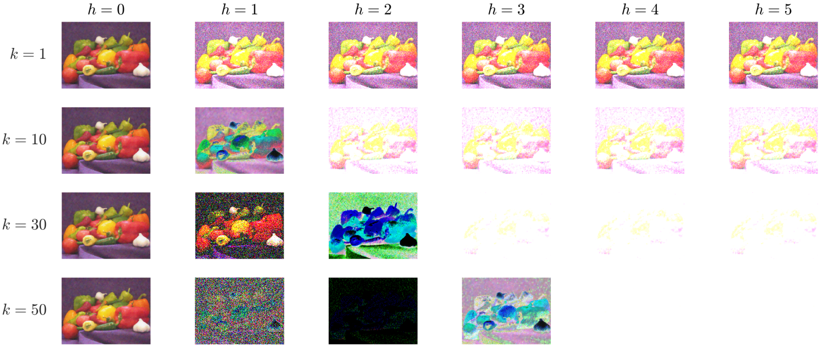


Image Restoration: Douglas-Rachford with Delays

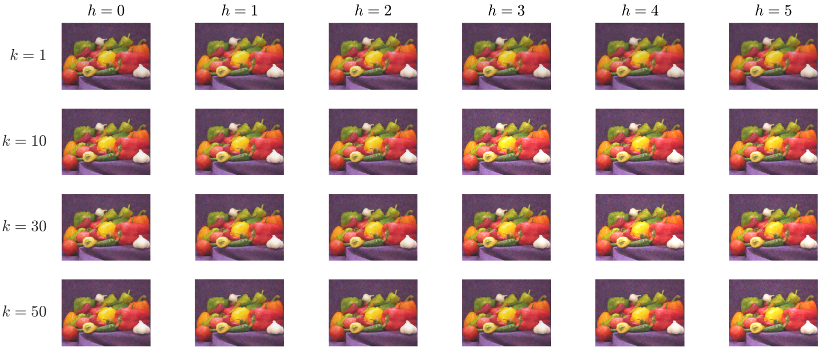
Delays can break convergent algorithms



h-step delay on viewing image data Y

Image Restoration: Synthesized with Delays

Our synthesis returns algorithms that converge under delays



h-step delay on viewing image data Y

Fixed Points and Convergence

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

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

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

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

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

Operator Classes

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

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

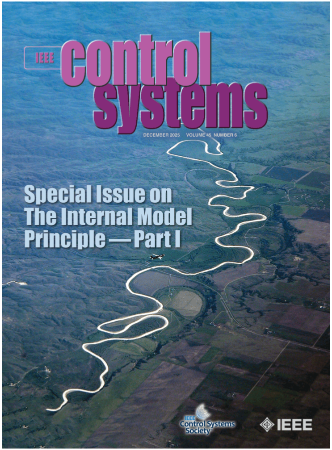
Example class: each f_i is m_i-convex, L_i smooth

Algorithm Synthesis:

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

Rejection of Disturbances

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

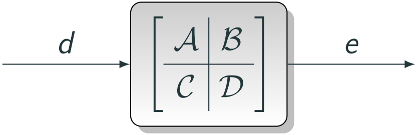


[3] Francis, Bruce A., and Walter Murray Wonham. "The internal model principle of control theory." Automatica 12.5 (1976): 457-465.

Background: Regulation Theory

Regulation Framework (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:** $\mathcal{A}$ is Schur
2. **Regulator Equation:** Linear equation is solvable

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

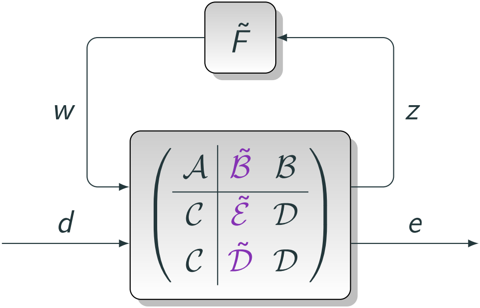
Asymptotic properties: $x \to \Upsilon d, \; e \to 0$ for all (x_0, d)

[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 $\tilde{F} \in \mathcal{O}$, $0 \in \tilde{F}(0)$

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



Regulation through Maps

If linear map $\tilde{F}(z) = \Delta z$ is applied, then

Stability: $\mathcal{A} + \tilde{\mathcal{B}}\Delta(I - \tilde{\mathcal{E}}\Delta)^{-1}\mathcal{C}$ is Schur

Regulator Equation: **Same** as before

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

Cancellation of Δ contribution by regulator equation

$$\begin{pmatrix} x_{k+1} \\ \hline e_k \end{pmatrix} = \left[\begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{pmatrix} + \begin{pmatrix} \tilde{\mathcal{B}} \\ \hline \tilde{\mathcal{D}} \end{pmatrix} \Delta(I - \tilde{\mathcal{E}}\Delta)^{-1} \begin{pmatrix} \mathcal{C} & \mathcal{D} \end{pmatrix} \right] \begin{pmatrix} x_k \\ \hline d \end{pmatrix}$$

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

Convergence of Algorithms

Convergence Conditions

If $F \star (P \star K)$ is convergent for all $F \in \mathcal{O}$, then it satisfies

1. **Robust Stability**
2. **Regulator Equation**

Use consensus parameterization $d = (-\beta^*, \hat{w}^*),\ w^* = N\hat{w}^*$:

$$N = \begin{pmatrix} I_{s-1} \\ \mathbf{1}_s^\top \end{pmatrix} \otimes I_c, \qquad \text{at } s = 3\text{: } N = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \otimes I_c$$

Convergence: No Network Dynamics

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

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

Regulator Equation [5]:

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

[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

Robust Stability: $x = (x^N, \xi) \to (0,0)$ for all $\tilde{F} \in \mathcal{O}^0$

Regulator Equation: Matrices $(\Pi, \Theta, \Gamma, \Phi)$ exist with

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

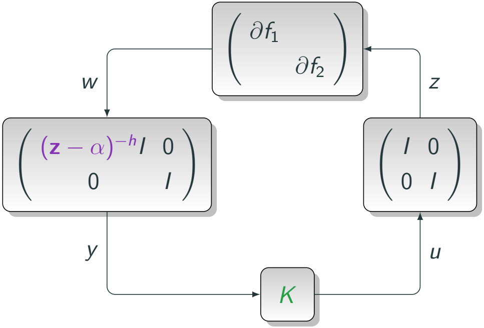
$$\left(\begin{array}{c|c} A_c & B_c \\ \hline C_c & D_c \end{array}\right) \left(\begin{array}{c} \Theta \\ \hline \Phi \end{array}\right) = \left(\begin{array}{c} \Theta \\ \hline \Gamma \end{array}\right)$$

Top: P only. Bottom: K only. Neither involves $\mathcal{O}$.

Top equation infeasible: no convergent algorithm exists

Channel Memory: Network

Network P_h^α: pole at α of order $h > 0$ on link $w_1 \to y_1$



Regulator Equation unsolvable at $\alpha = 1$ (integrator/ramp)

Channel Memory: Regulation

Controllers that satisfies Regulator Equation if $\alpha \neq 1$

$$K_h^\alpha[b]: \quad \left(\begin{array}{c|cc} I & (1-\alpha)^h b_0 I & b_0 I \\ \hline I & (1-\alpha)^h b_2 I & 0 \\ I & (1-\alpha)^h (b_1 + b_2) I & b_1 I \end{array}\right)$$

Douglas Rachford[6]: $(b_0, b_1, b_2) = (-\gamma\lambda, -\lambda, -\lambda)$ with $\alpha = 0$

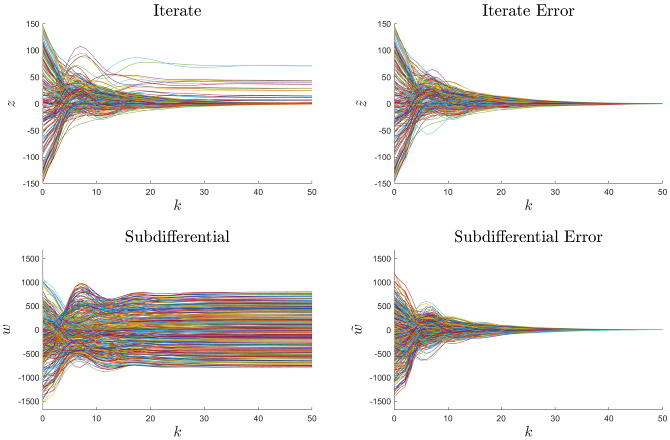
Robust Stability depends on $(b, \mathcal{O}, P_h^\alpha)$

---

[6] 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.

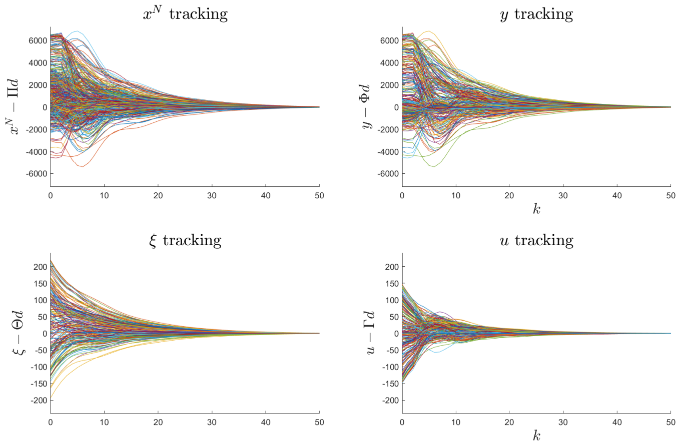
Channel Memory: Convergence

LASSO solved using $P_3^{0.5} \star K_3^{0.5}[-0.04, -0.2, -0.1]$



Channel Memory: Tracking

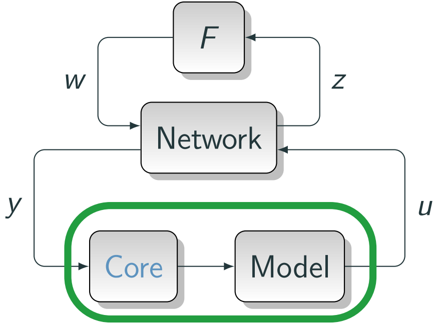
Asymptotic tracking of $(x^N, \xi, y, u) \to (\Pi d, \Theta d, \Phi d, \Gamma d)$



Structure of Algorithms

Structural Factorization

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

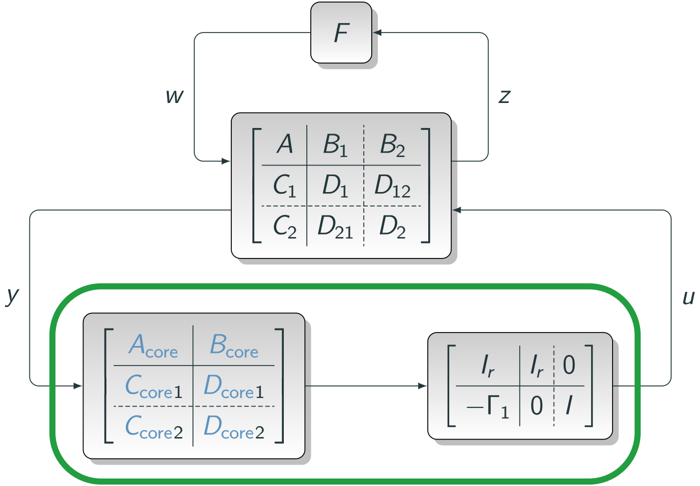


Model depends only on network P, not on K or $\mathcal{O}$

Based on the Internal Model Principle of control theory[7]

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

Factorization: General Setting



Size of model is $r := \dim(\text{null}(\Phi))$

Subspace Restriction

Core not free[8], entries must satisfy constraint $\mathcal{L}_{\text{core}}(\Theta)$:

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

Affine set $\mathcal{L}_{\text{core}}(\Theta)$ parameterized by $\Theta = \begin{pmatrix} -I_r & \Theta_{12} \\ 0 & \Theta_{22} \end{pmatrix}$

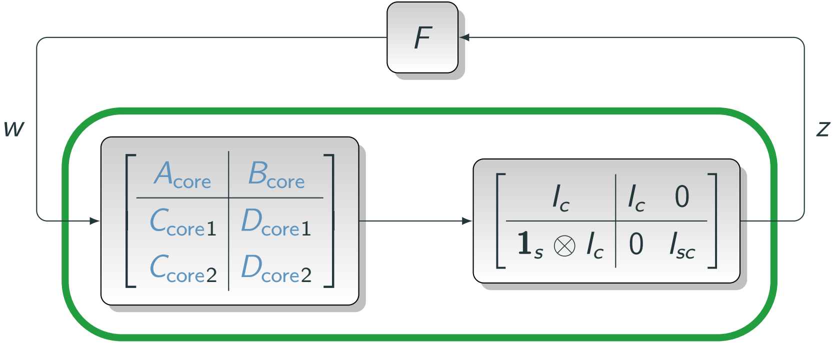
Only search over elements of $\mathcal{L}_{\text{core}}(\Theta)$ in design

---

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

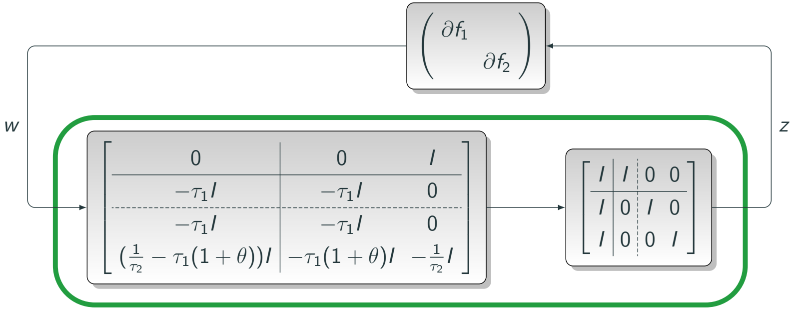
Factorization for No Network Dynamics

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



Factorization Example: Chambolle-Pock

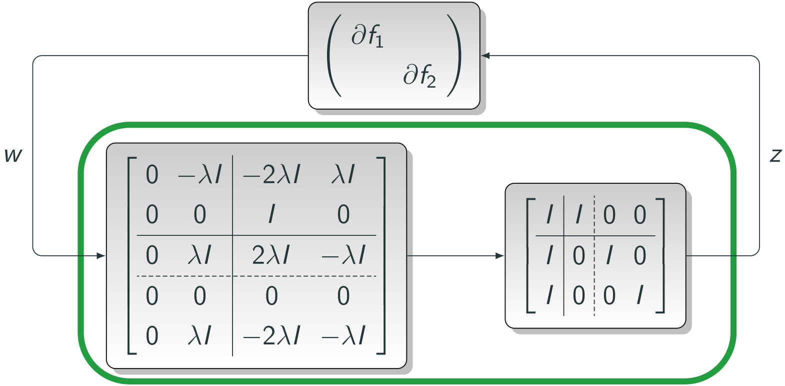
Chambolle-Pock [9] with parameters (τ_1, τ_2, θ)



[9] 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: FRBS

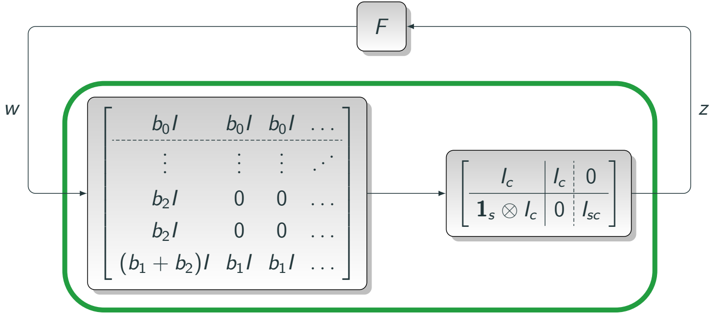
Forward-Reflected-Backward Splitting [10] with param. λ



[10] Malitsky, Yura, and Tam, Matthew K. "A forward-backward splitting method for monotone inclusions without cocoercivity." SIAM Journal on Optimization 30.2 (2020): 1451-1472.

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:	$(-\gamma, -\gamma, 0)$

Analysis and Synthesis

Exponential Convergence Rates

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

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

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

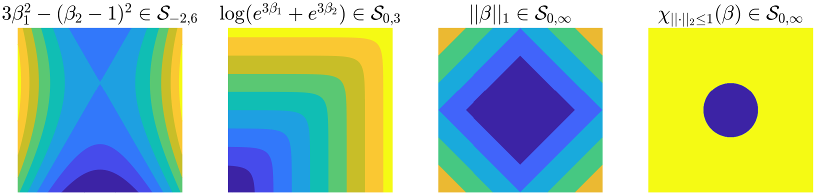
Algorithm Analysis:

Verify worst-case rate ρ over all $F \in \mathcal{O},\ x_0 \in \mathbb{R}^n$

Properties of Functions

Function class $\mathcal{S}_{m,L}$ described by $-\infty < m \leq L \leq \infty$

$\mathcal{S}_{m,L}$: $f - \frac{m}{2}\|\cdot\|_2^2$ p.c.c. Also $\frac{L}{2}\|\cdot\|_2^2 - f$ p.c.c. if $L < \infty$



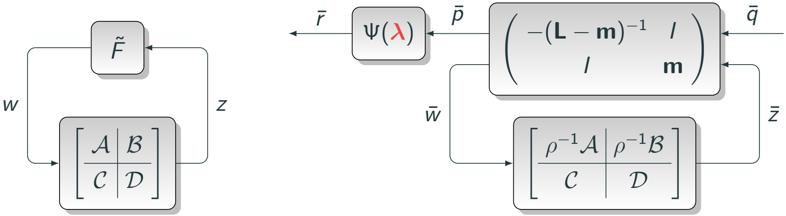
Contours of functions in $\mathcal{S}_{m,L}$

- $\mathcal{S}_{m,\infty}$: m-convex (strong if $m > 0$) with convex domain
- $\mathcal{S}_{m,L}$: L-smooth, differentiable, globally defined if $L < \infty$

Filtering Framework (Analysis)

Class $\mathcal{O}$: $f_i \in \mathcal{S}_{m_i,L_i}$ for all i

Zames-Falb filter[11] : $\Psi(\boldsymbol{\lambda})$: $\bar{r}^i_k = \sum_\nu \lambda^i_\nu \bar{p}^i_{k-\nu}$



[11] ρ-$\mathcal{S}_{m,L}$ generalizations of the subgradient inequality, anti-passivity constraints. 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 $(\boldsymbol{\lambda}, \mathcal{X})$ for fixed ρ:

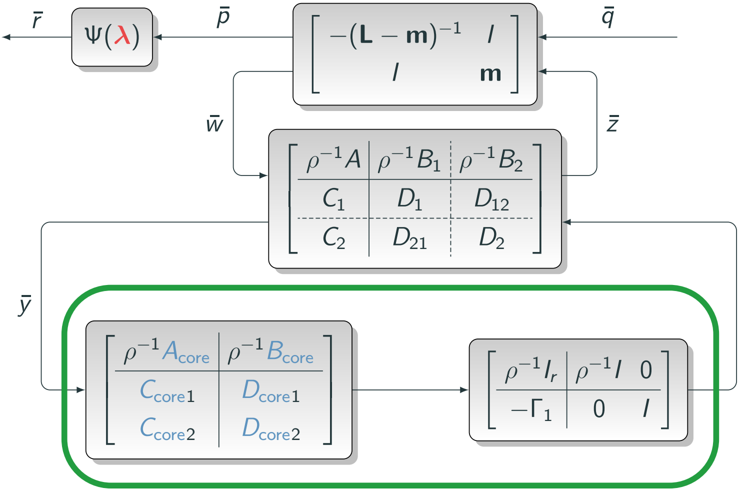
$$\mathcal{X} \succ 0, \ \sum_{\nu=0}^{\nu_{\max}} \rho^{-\nu} \lambda_\nu^i > 0, \quad \forall \nu \geq 1 : \ \lambda_\nu^i \leq 0$$

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

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

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

Nonconvex Structured Control



Nonconvex joint search over $(\Sigma_{\text{core}}, \Theta_{12}, \Theta_{22}, \lambda)$ to minimize ρ

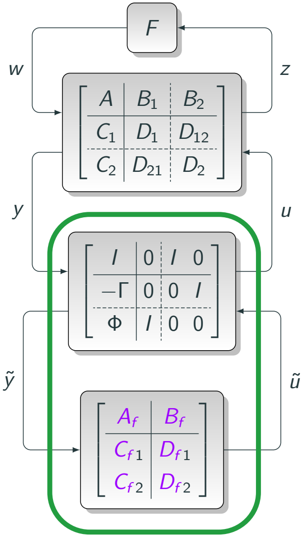
Full-Order Internal Models

Nonconvex constraints on $(A_{\text{core}}, B_{\text{core}}, C_{\text{core}1}, C_{\text{core}2})$ [a]

Use Full-Order internal model (Γ, Φ)

Regulator Equation always satisfied

Can exchange between Σ_f and Σ_{core}



[a] V. Blondel and J. N. Tsitsiklis, "NP-hardness of some linear control design problems," SIAM J. Control Optim., vol. 35, no. 6, pp. 2118–2127, 1997.

Alternating Synthesis

Alternating convex problem

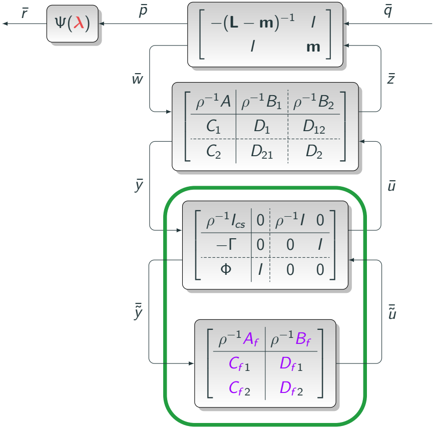
Bisection in ρ: inner steps

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

For fixed λ : optimize Σ_f

For fixed Σ_f : optimize λ

Repeat (λ, Σ_f) loop

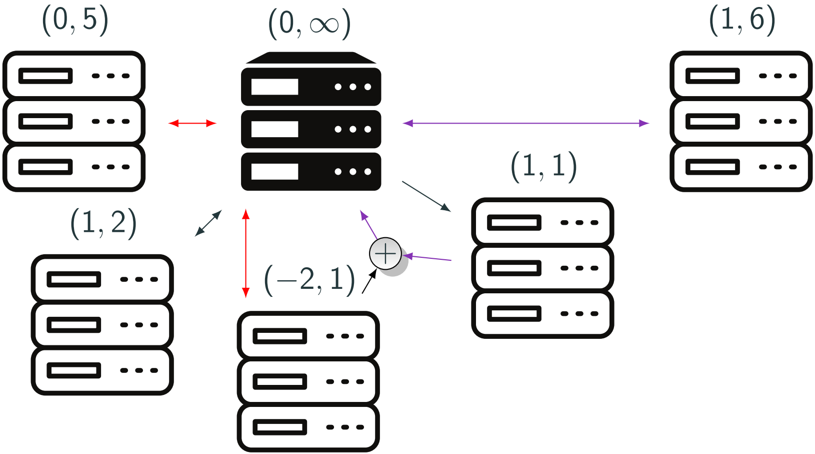


LMI solvable $\bar{q} \to \bar{r}$

Examples

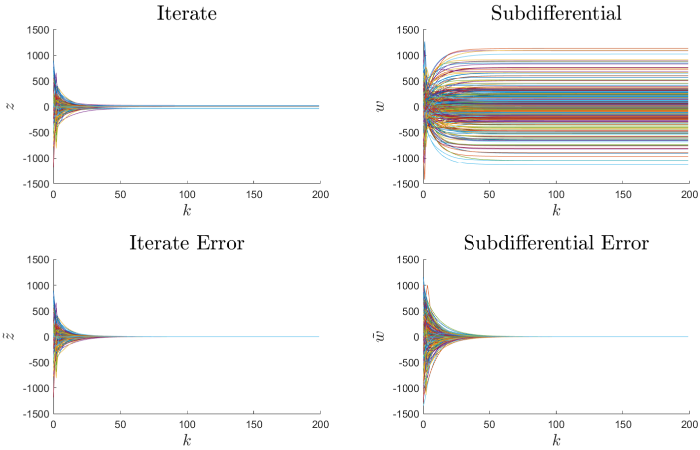
Server Example: With Network Dynamics

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



Server Example: Convergence

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



Quadratic Program over Unstable Network

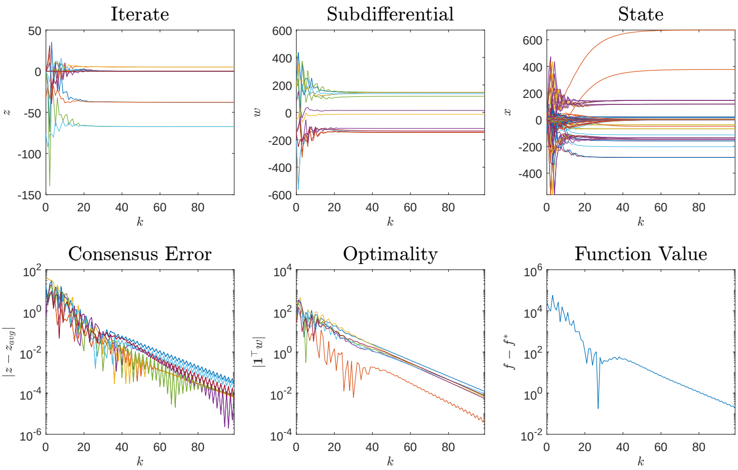
System with $f_1 \in \mathcal{S}_{1,5}$ (quadratic) $f_2 \in \mathcal{S}_{0,\infty}$ (indicator)

$$\beta^* \in \underset{\beta \in \mathbb{R}^c}{\text{argmin}} \, \frac{1}{2}(\beta - \beta_Q)^\top Q(\beta - \beta_Q) \qquad \|\beta\|_1 \leq \tau$$

Unstable dynamics

$$P : \left(\begin{array}{cc|cc} 0 & 0 & \frac{1}{\mathbf{z}+1.1} & 0 \\ 0 & 0 & \frac{1}{\mathbf{z}-0.3} & 2 \\ \hline 1 & \frac{1}{\mathbf{z}} & \frac{2}{\mathbf{z}^2} & \frac{1}{\mathbf{z}-0.9} \\ 1 & 1 - \frac{1}{\mathbf{z}^3} & 0 & 1 \end{array}\right) \otimes I_c$$

Quadratic Program over Unstable Network

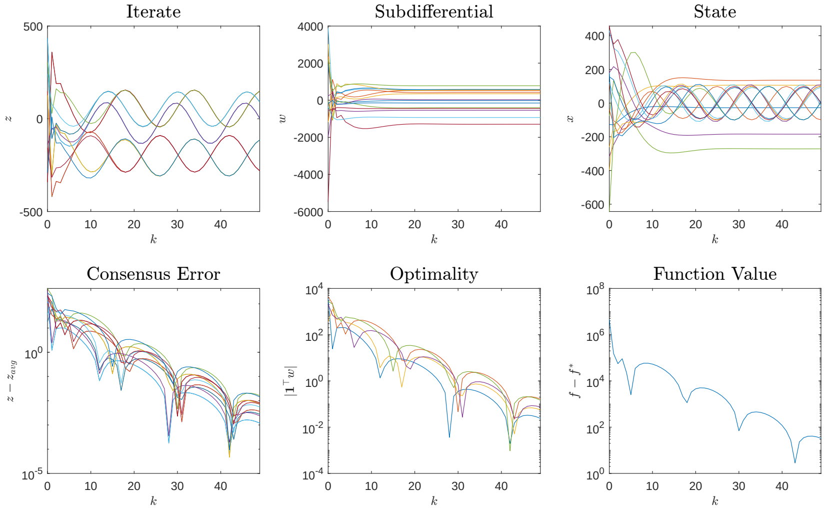


Certified $\rho < 0.9476$ after 4 Synthesis/Analysis iterations

Extensions

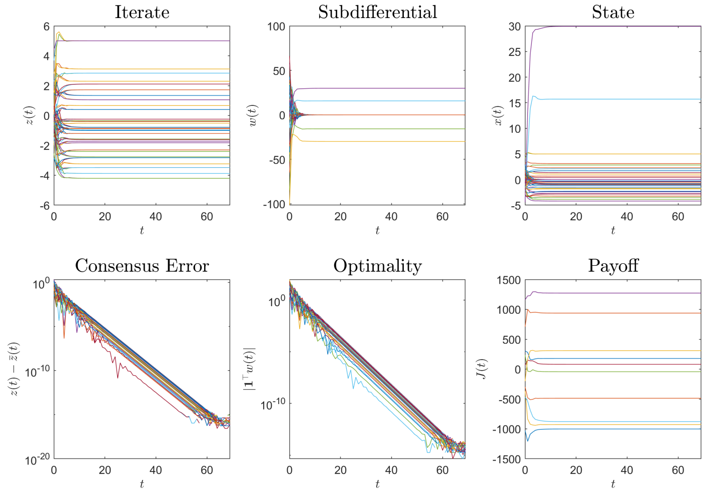
Perfect Tracking

Track the time-varying optimal solution



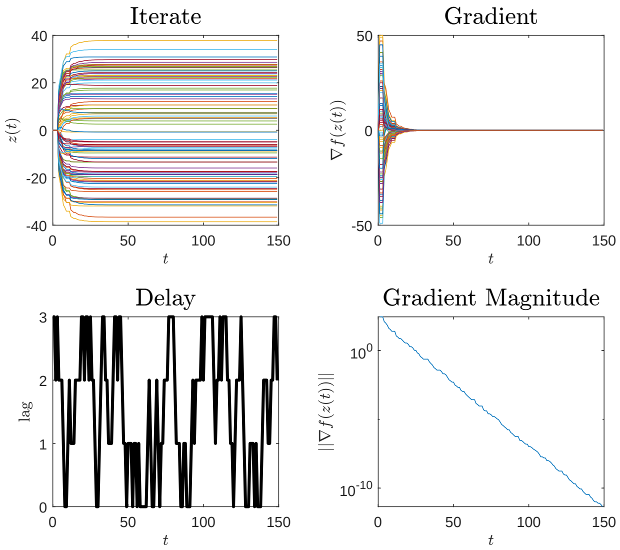
Fixed Point Equations

Constrained Nash-Equilibrium seeking under delays



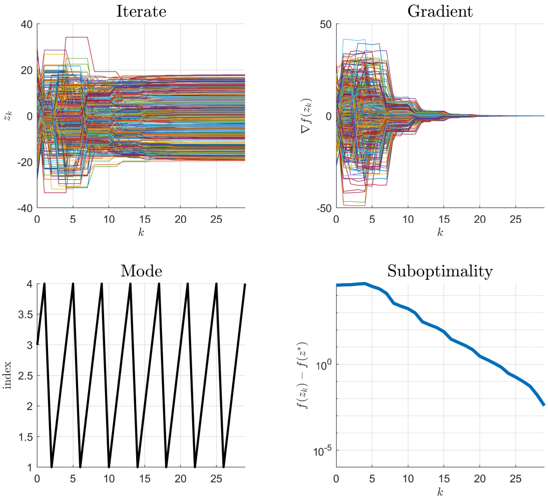
Switching Example: Time-Varying Delay

Time-varying delay between 0 and 3 before ∇f



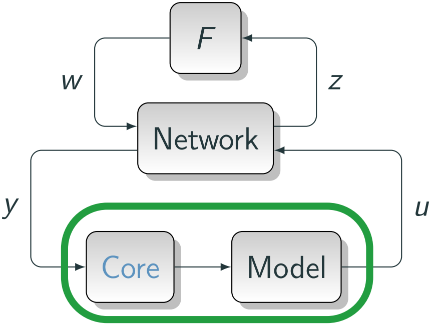
Periodic Example: Coordinate Descent

Coordinate descent with 4 blocks



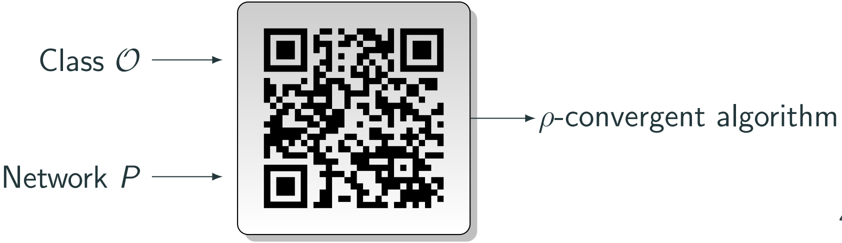
Take-aways

Conclusion



Convergence

Structure



Optimize(d) using Structure!