Structure, Analysis, and Synthesis of Optimization Algorithms

Jared Miller
Carsten Scherer
Fabian Jakob
Andrea Iannelli

GAMM, S20.4
March 18, 2026

Composite Optimization

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

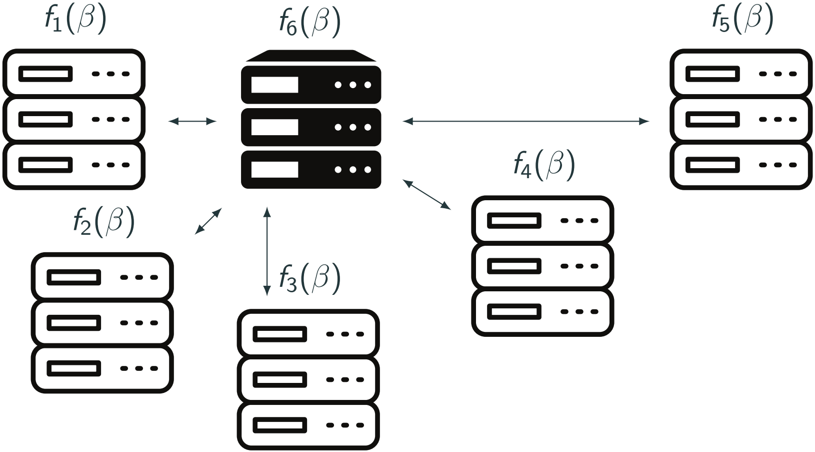
$$\beta^* \in \operatorname{argmin}_{\beta \in \mathbb{R}^c} \sum_{i=1}^{s} f_i(\beta)$$

Wide range of applications:

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

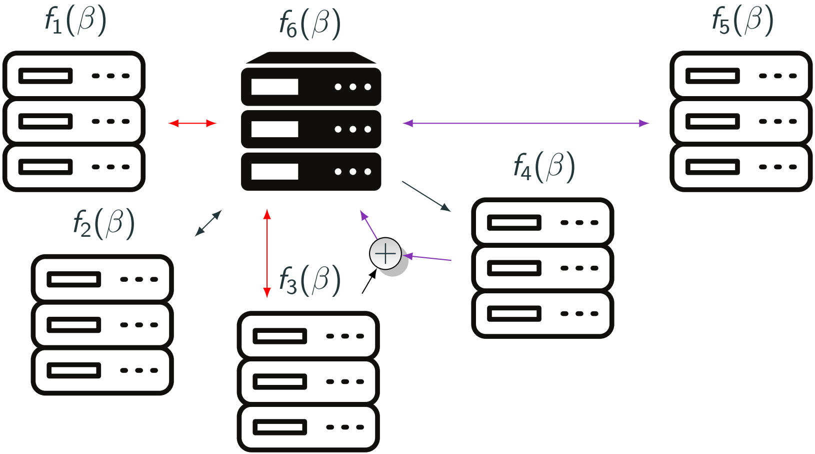
Optimization Setup

Assimilation/regression: each server might have different data



Optimization over Dynamical Networks

Memory, cross-talk, time delays, other nightmares



Main Ideas

When do networked optimization algorithms converge?

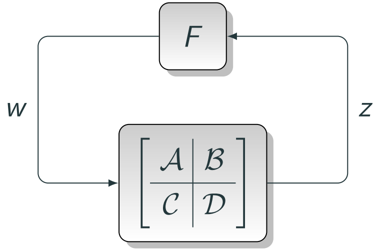
Answer by continuing a link of Optimization and Control

	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 $F(z) = \sum_{i=1}^{s} \partial f_i(z^i)$:



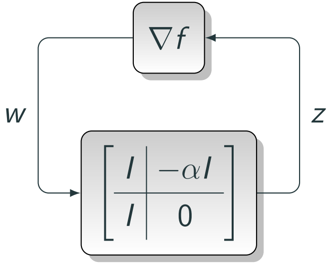
$$\begin{pmatrix} x_{k+1} \\ \hline z_k \end{pmatrix} = \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & \mathcal{D} \end{array}\right) \begin{pmatrix} x_k \\ \hline w_k \end{pmatrix}, \quad w_k \in F(z_k)$$

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

Examples of Interconnections: GD/PPM

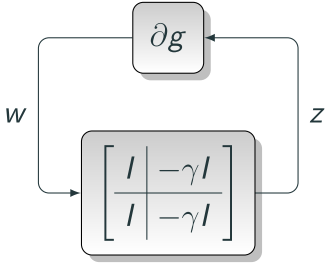
Gradient Descent

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



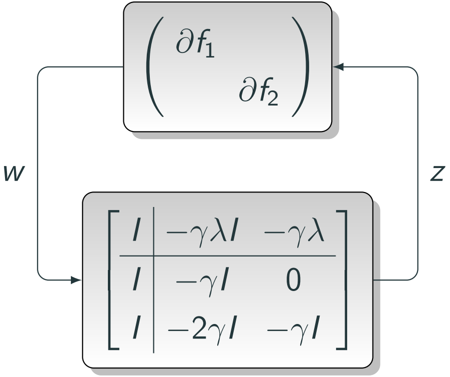
Proximal Point Method

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



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

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

Original Image X

Noisy Image Y

Denoised β^*

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

Algorithms over Networks

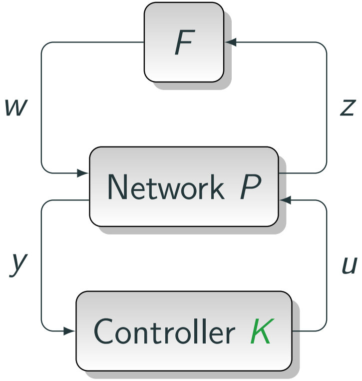
F 'sees' system $G = P \star K$

Network P:

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

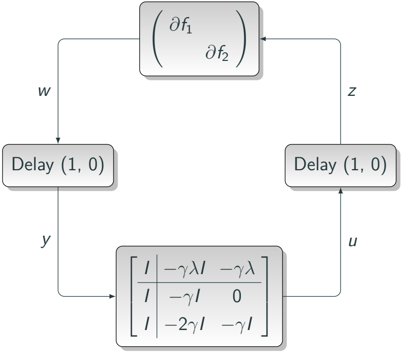
Controller K:

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



Examples of Interconnections: Delayed DR

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



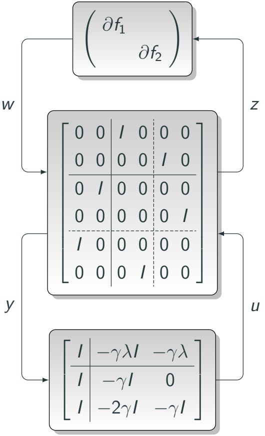
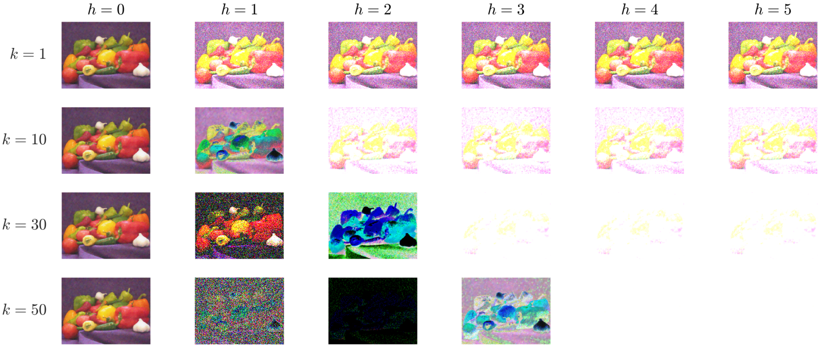


Image Restoration: Douglas-Rachford with Delays

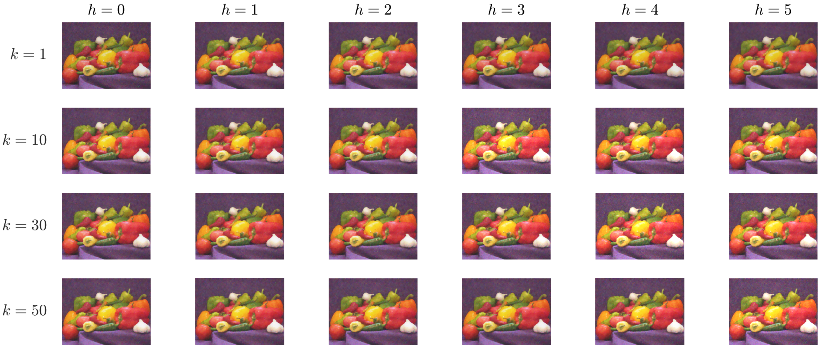
Delays can break convergent algorithms



h-step delay on viewing image data Y

Image Restoration: Synthesized with Delays

Our synthesis returns algorithms that converge under delays



Convergence and Structure

Fixed Points and Convergence

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

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

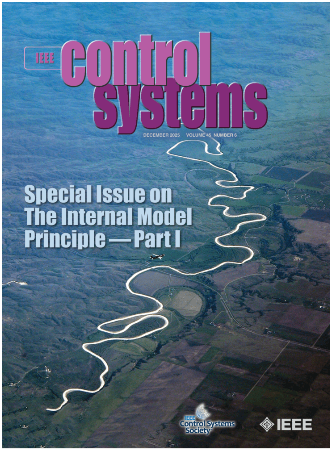
Convergent: (x^*, w^*, z^*) exists with[2]

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

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

Rejection of Disturbances

Convergence: errors $(w - w^*, z - z^*)$ driven to 0 [3]



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

Operator Classes

$\mathcal{O}$: maps F having unique (β^*, w^*) with

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

Example requirements in $\mathcal{O}$:

- f_i is proper, convex, closed
- f_i is m_i-convex, L_i smooth

Algorithm Synthesis:

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

Convergence Conditions

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

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

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

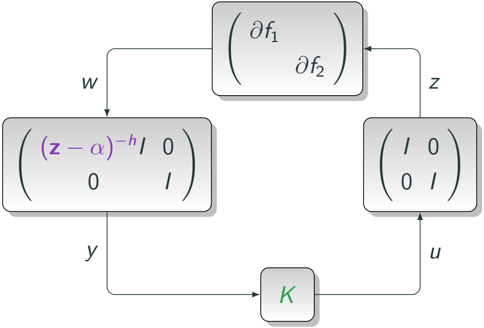
$$\left(\begin{array}{c|cc:c} A - I & 0 & B_1 N & B_2 \\ \hline C_1 & \mathbf{1}_s \otimes I_c & D_1 N & D_{12} \\ \hdashline C_2 & 0 & D_{21} N & D_2 \end{array}\right) \left(\begin{array}{c} \Pi \\ \hline I \\ \hdashline \Gamma \end{array}\right) = \left(\begin{array}{c} 0 \\ \hline 0 \\ \hdashline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

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

$N \in \mathbb{R}^{sc \times (s-1)c}$: full-rank consensus matrix, $(\mathbf{1}_s \otimes I_c)N = 0$

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 satisfy Regulator Equation if $\alpha \neq 1$

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

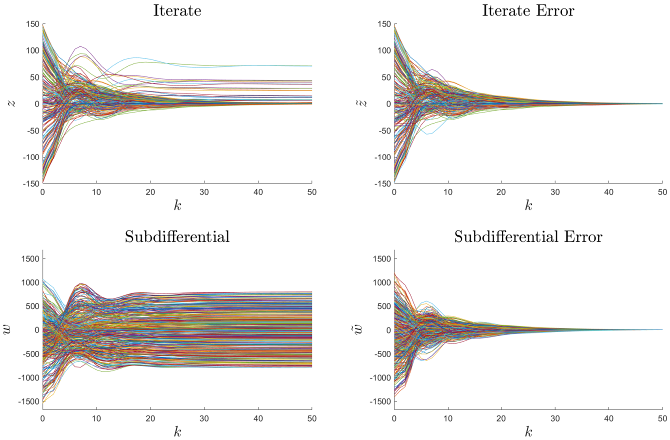
Douglas Rachford[4]: $(b_0, b_1, b_2) = (-\gamma\lambda, -\lambda, -\lambda)$ with $\alpha = 0$

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

[4] 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.

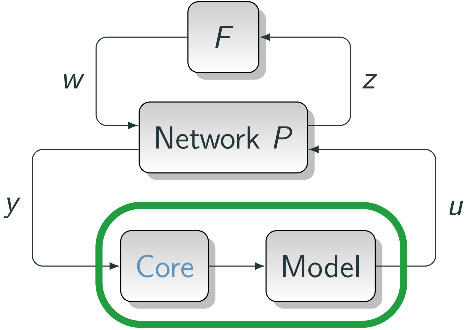
Channel Memory: Convergence

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



Structural Factorization

If $F \star (P \star K)$ convergent, then K factors[5] into Core and Model

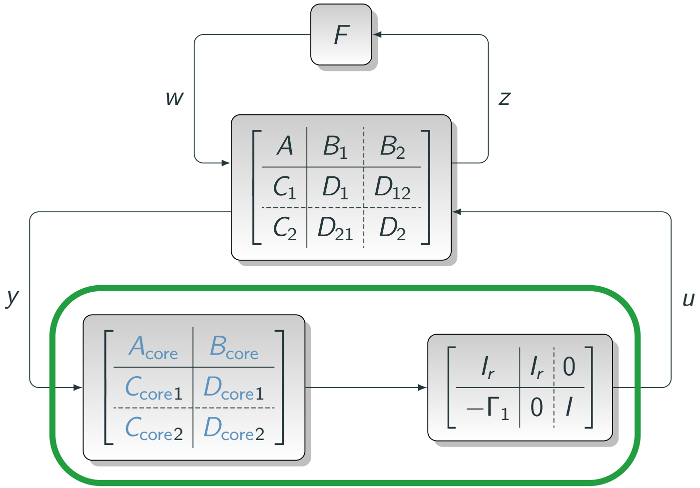


Entries of Core restricted to an affine constraint

Model depends only on network P, not on F

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

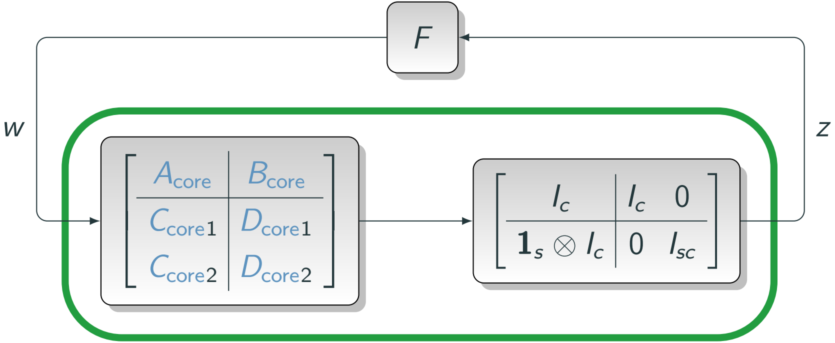
Factorization: General Setting



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

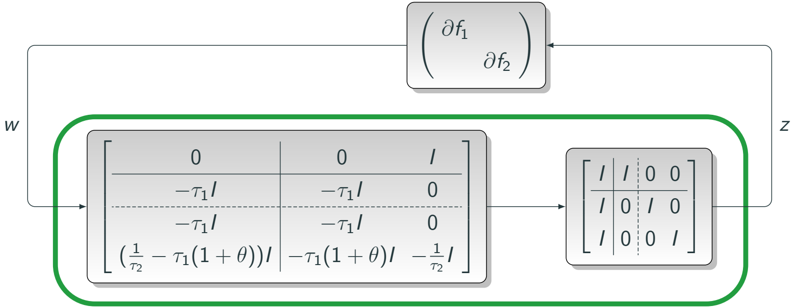
Factorization for No Network Dynamics

Algorithms for strongly convex optimization can be factored



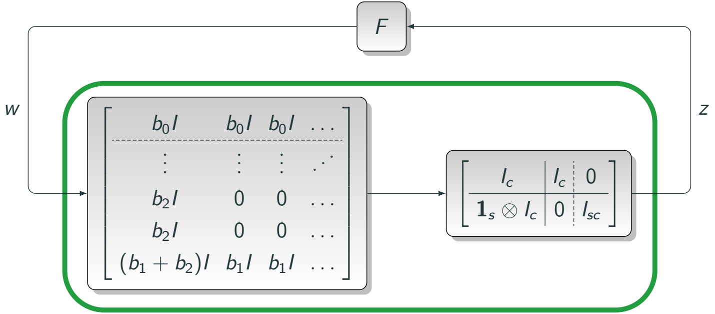
Factorization Example: Chambolle-Pock

Chambolle-Pock with parameters (τ_1, τ_2, θ)



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

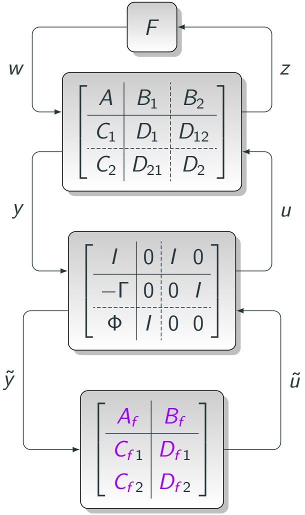
Full-Order Internal Models and Synthesis

Full-Order internal model (Γ, Φ)

Regulator Equation always satisfied

Synthesis [a] using Robust Control

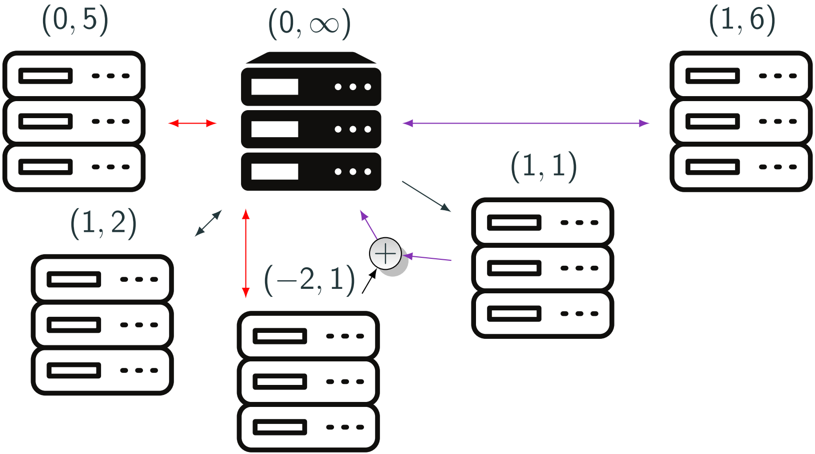
[a] Scherer, C. W., & Ebenbauer, C. (2025). A Tutorial on Convex Design of Optimization Algorithms by Integral Quadratic Constraints. Annual Review of Control, Robotics, and Autonomous Systems, 9.



Synthesis Example: Servers

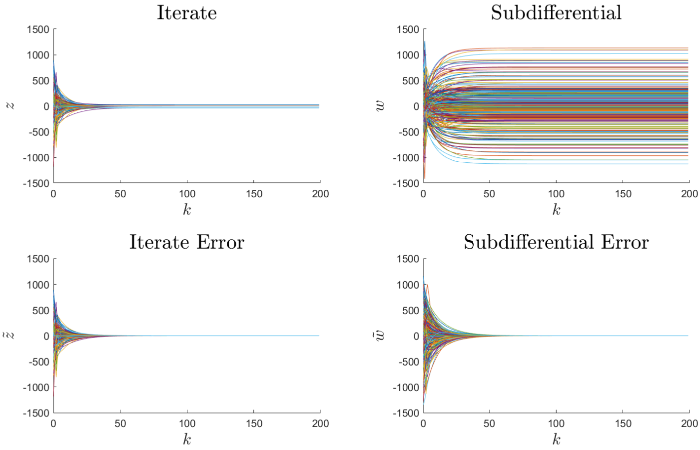
Server Example: With Network Dynamics

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



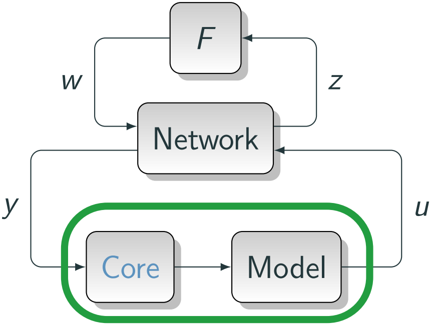
Server Example: Convergence

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



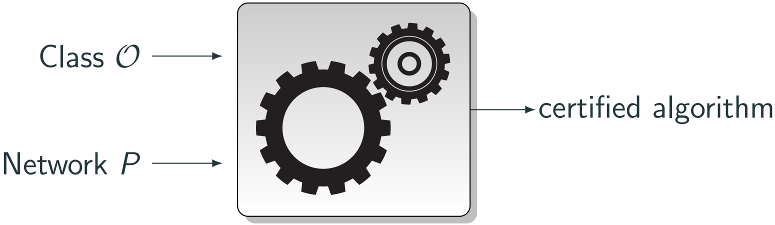
Take-aways

Conclusion



Convergence

Structure



Optimize(d) using Structure!