Analysis and Design of Algorithms for Dynamic Average Consensus and Convex Optimization

Bryan R. Van Scoy

Northwestern University

May 2, 2017

Overview

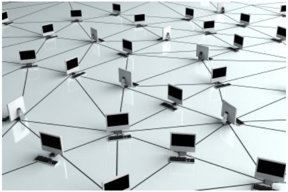
algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Overview

algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Centralized

Distributed

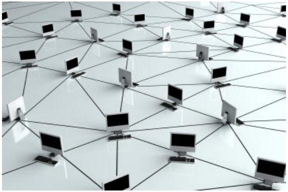


Overview

algorithm - a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer.

Centralized

Distributed



Analysis prove properties of the algorithm

Design develop new algorithms to meet specific criteria

Outline

Problem setup

Consider the optimization problem

$$\min_{x \in \mathbb{R}^n} \; f(x)$$

where $f \in \mathcal{S}_{m,L}$.

Definition (function class)

Let $\mathcal{S}_{m,L}$ be the set of functions $f : \mathbb{R}^n \to \mathbb{R}$ that are

- *continuously differentiable,*
- *strongly convex with parameter m, and*
- *have Lipschitz gradients with parameter L.*

Furthermore, $\kappa := L/m$ is called the condition number of $f \in \mathcal{S}_{m,L}$.

Gradient-based methods

$$\xi_{k+1} = (1+\beta)\xi_k - \beta\xi_{k-1} - \alpha\nabla f(y_k), \quad \xi_0, \xi_{-1} \in \mathbb{R}^n$$
$$y_k = (1+\gamma)\xi_k - \gamma\xi_{k-1}$$
$$x_k = (1+\delta)\xi_k - \delta\xi_{k-1}$$

Method	Parameters $(\alpha, \beta, \gamma, \delta)$
Gradient descent	$(\alpha, 0, 0, 0)$
Heavy-ball method (Polyak, 1964)	$(\alpha, \beta, 0, 0)$
Nesterov's accelerated gradient descent (Nesterov, 2004)	$(\alpha, \beta, \beta, 0)$
Algorithm in (Lessard, Recht, and Packard, 2016)	$(\alpha, \beta, \gamma, 0)$

Definition (Triple momentum method)

$$(\alpha, \beta, \gamma, \delta) = \left(\frac{1+\rho}{L}, \ \frac{\rho^2}{2-\rho}, \ \frac{\rho^2}{(1+\rho)(2-\rho)}, \ \frac{\rho^2}{1-\rho^2}\right)$$

where $\rho = 1 - 1/\sqrt{\kappa}$

Results

Theorem (Triple momentum method)

Let $f \in \mathcal{S}_{m,L}$ with $0 = \nabla f(x_\star)$. For any initial condition $\xi_0, \xi_{-1} \in \mathbb{R}^n$, the TM method produces iterates which satisfy

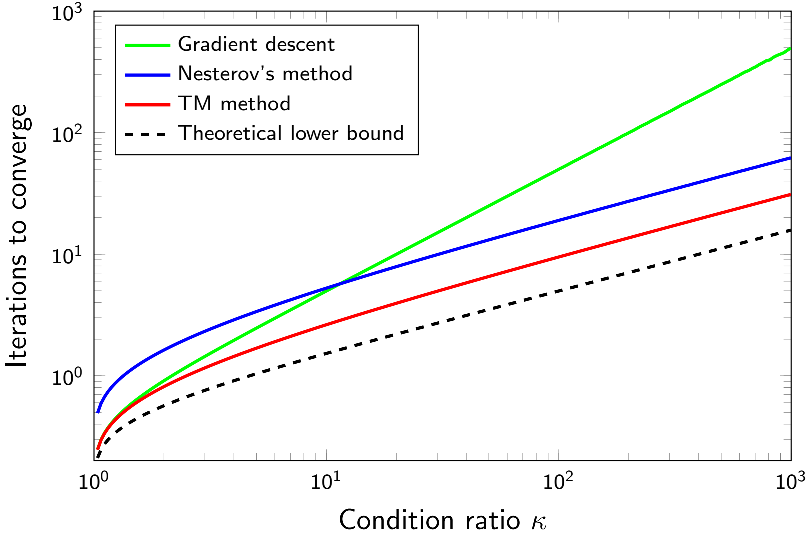
$$\|x_k - x_\star\| \leq \left(1 - \frac{1}{\sqrt{\kappa}}\right)^{k-1} \|x_1 - x_\star\|, \quad \forall k \geq 1.$$

Theorem (Gradient descent)

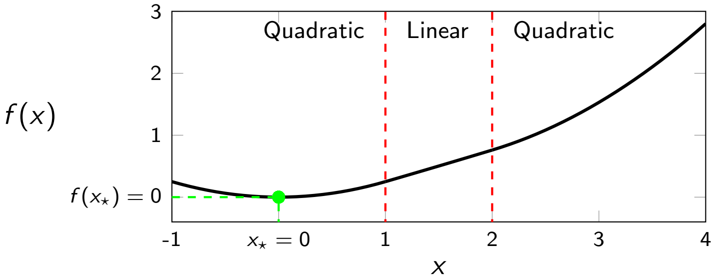
Let $f_k \in \mathcal{S}_{m,L}$ with $0 = \nabla f_k(x_\star)$. For any initial condition $\xi_0 \in \mathbb{R}^n$, the gradient descent method with $\alpha = 2/(L+m)$ produces iterates which satisfy

$$\|x_k - x_\star\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^{k} \|x_0 - x_\star\|, \quad \forall k \geq 0.$$

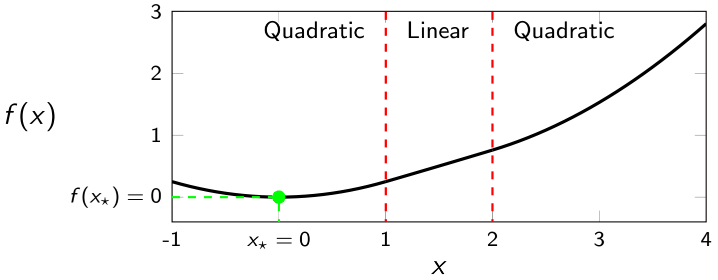
Iterations to converge



Piecewise quadratic objective function



Piecewise quadratic objective function



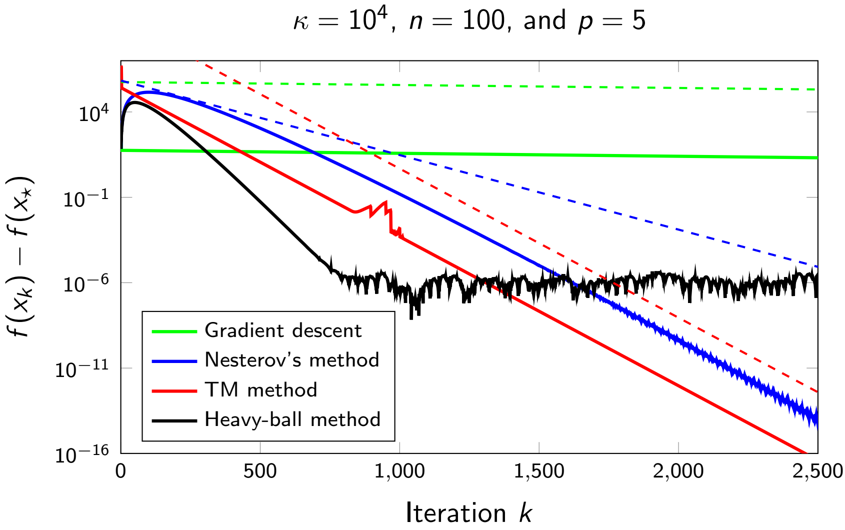
Multidimensional piecewise quadratic:

$$f(x) = \sum_{i=1}^{p} g(a_i^T x - b_i) + \frac{m}{2}\|x\|^2, \quad x \in \mathbb{R}^n$$

where

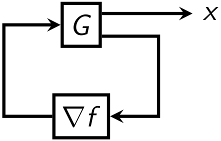
$$g(y) = \begin{cases} \frac{1}{2}y^2, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

Simulations



Proof method

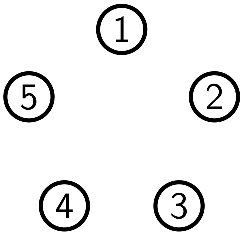
- We can use integral quadratic constraints (IQCs) from robust control theory. The gradient is characterized by constraints which its input and output must satisfy.



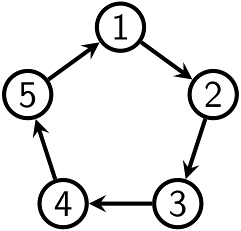
- We also have a simple convergence proof which does not rely on control theory.

Outline

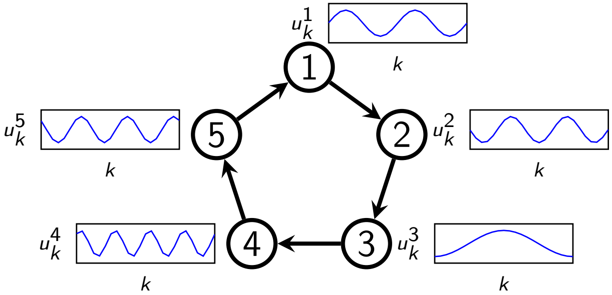
Dynamic Average Consensus (DAC)



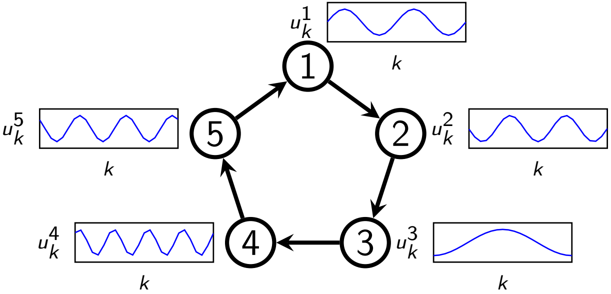
Dynamic Average Consensus (DAC)

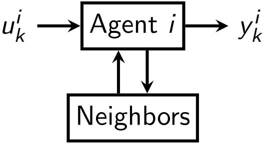


Dynamic Average Consensus (DAC)



Dynamic Average Consensus (DAC)





Error of agent i at time k =

$$\underbrace{y_k^i}_{\text{Local output}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n} u_k^i}_{\text{Global average}}$$

Applications of dynamic average consensus

- distributed multi-agent coordination
  - (Yang, Freeman, and Lynch, 2008)
- distributed environmental monitoring
  - (Lynch, Schwartz, Yang, and Freeman, 2008)
- distributed Kalman filtering
  - (Bai, Freeman, and Lynch, 2011)
- distributed Kriged Kalman filtering
  - (Cortés, 2009)
- distributed dynamic merging of feature-based maps
  - (Aragüés, Cortés, and Sagüés, 2012)
- distributed optimization
  - (Qu and Li, 2016)

Communication graph

To diffuse information among the agents, we use the graph Laplacian

$$[Lx]_i = \sum_{j \in \mathcal{N}_{\text{in}}(i)} a_{ij} \, (x_i - x_j)$$

where $\mathcal{N}_{\text{in}}(i)$ are the agents from which agent i receives information and a_{ij} are the edge weights.

Graph properties:

connected	$\iff$	directed link between any two nodes
undirected	$\iff$	$a_{ij} = a_{ji}$
balanced	$\iff$	$\sum_{j \in \mathcal{N}_{\text{in}}(i)} a_{ij} = \sum_{j \in \mathcal{N}_{\text{out}}(i)} a_{ij}$

Problem

Problem (Distributed algorithm design)

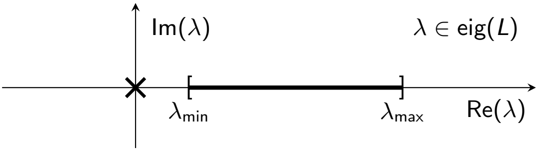
Given:

1. *assumptions on the communication among agents*
2. *assumptions on the input signals*
3. *desired properties of the algorithm*

$\Longrightarrow$ *distributed algorithm (estimator)*

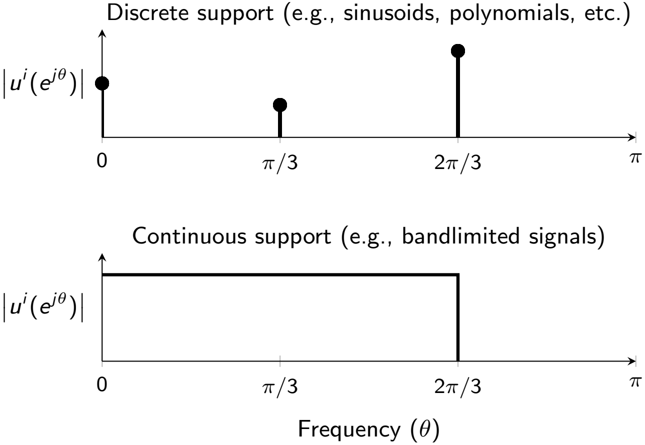
Assumptions on the communication graph

- constant or time-varying
- balanced, undirected, or directed
- randomly generated from a given distribution
- drops packets independently with given probability
- known upper bound on the number of agents
- eigenvalues of the Laplacian matrix are in a known region



Assumptions on the signals

We consider two class of inputs signals depending on the support of the frequency spectrum.

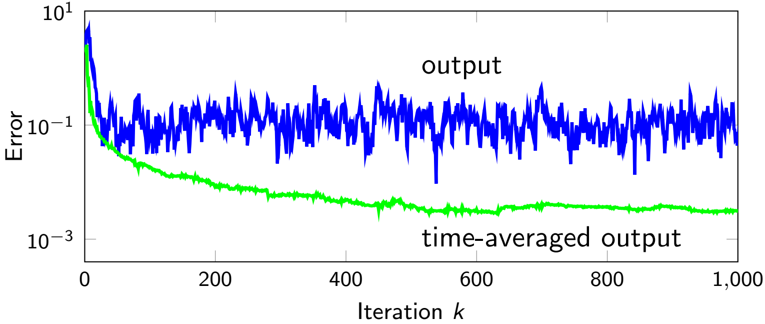


Estimator properties

- **Scalable:** The number of variables and computations on each agent does not scale with the number of agents.
- **Exact:** The error converges to zero.
- **Internally stable:** The internal states are bounded.
- **Time invariant:** The dynamics do not change with time.
- **Robust to initial conditions:** The steady-state output does not depend on the initial condition.

Asymptotic mean ergodicity

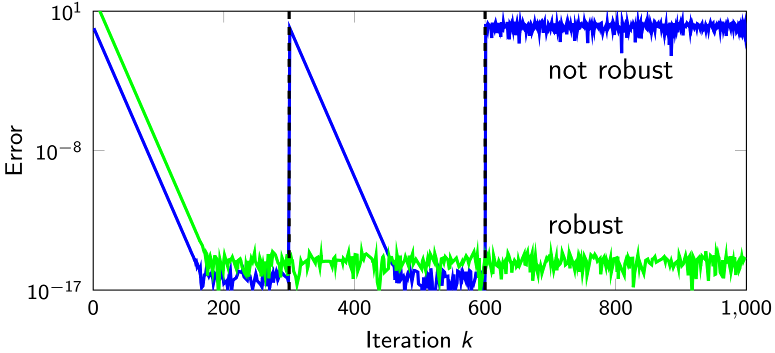
- **Ergodic:** The time average of the output process converges to its statistical average (if the graph is connected and balanced on average and L_k is i.i.d.).



Bryan Van Scoy, Randy A. Freeman, and Kevin M. Lynch (June 2014). "Asymptotic mean ergodicity of average consensus estimators". In: *Proc. of the 2014 Amer. Control Conf.* Pp. 4696–4701

Robust to changes in the graph

- **Robust to changes in the graph:** The steady-state error using a time-varying sequence of graphs is no worse than when using the "worst-case" constant graph.



Outline

1. Convex Optimization
2. Dynamic Average Consensus
   - Assumptions on the communication
   - Assumptions on the signals
   - Properties of estimators
3. Estimators for Signals with Discrete Frequency Spectrum
   - Static estimator
   - Proportional estimator
   - Proportional-integral estimator
   - Summary
4. Estimators for Signals with Continuous Frequency Spectrum
   - Feedback estimators
   - Feedforward estimators
5. Conclusion

Problem

Assumption (Graph)

Assume the graph is constant, connected, undirected, and has nonzero Laplacian matrix eigenvalues in $[\lambda_{\min}, \lambda_{\max}]$.

Assumption (Signals)

Assume the input signals have a known model, i.e., $d(z)$ *is known where* $u_i(z) = n_i(z)/d(z)$.

Desired properties:

1. one-hop communication
2. scalable
3. exact
4. internally stable
5. time-invariant
6. robust to ICs
7. ergodic
8. fast convergence

Problem

Assumption (Graph)

Assume the graph is constant, connected, undirected, and has nonzero Laplacian matrix eigenvalues in $[\lambda_{\text{min}}, \lambda_{\text{max}}]$.

Assumption (Signals)

Assume the input signals have a known model, i.e., $d(z)$ *is known where* $u_i(z) = n_i(z)/d(z)$.

Desired properties:

1. one-hop communication
2. scalable
3. exact
4. internally stable
5. time-invariant
6. robust to ICs
7. ergodic
8. fast convergence

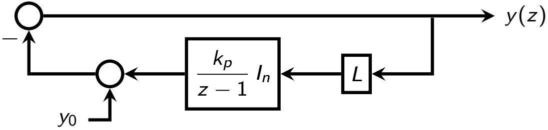
Static estimator

The static estimator is implemented on agent i using

$$y_{k+1}^i = y_k^i - k_p \sum_{j \in \mathcal{N}_i} a_{ij} \, (y_k^i - y_k^j), \quad y_0^i = u^i$$

where y_k is the estimate of the average at time k. Using the Laplacian matrix,

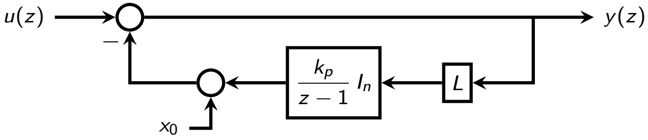
$$y_{k+1} = (I - k_p L) y_k, \quad y_0 = u.$$



John Tsitsiklis (Nov. 1984). "Problems in decentralized decision making and computation". PhD thesis. Massachusetts Institute of Technology

Proportional estimator

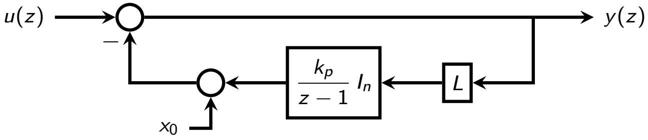
Internally stable but not robust to initial conditions:



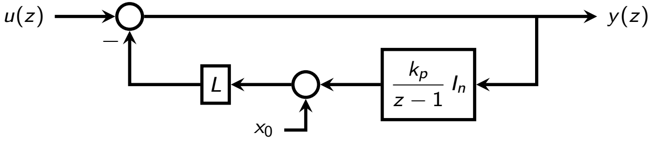
R.A. Freeman, Peng Yang, and K.M. Lynch (2006). "Stability and Convergence Properties of Dynamic Average Consensus Estimators". In: *Proc. of the 45th IEEE Conf. on Decision and Control*, pp. 338–343

Proportional estimator

Internally stable but not robust to initial conditions:



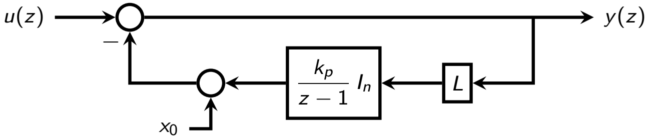
Robust to initial conditions but not internally stable:



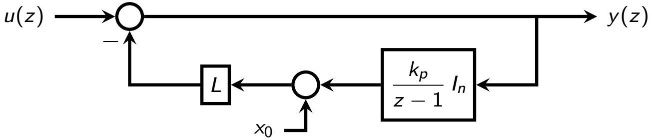
R.A. Freeman, Peng Yang, and K.M. Lynch (2006). "Stability and Convergence Properties of Dynamic Average Consensus Estimators". In: *Proc. of the 45th IEEE Conf. on Decision and Control*, pp. 338–343

Proportional estimator

Internally stable but not robust to initial conditions:



Robust to initial conditions but not internally stable:

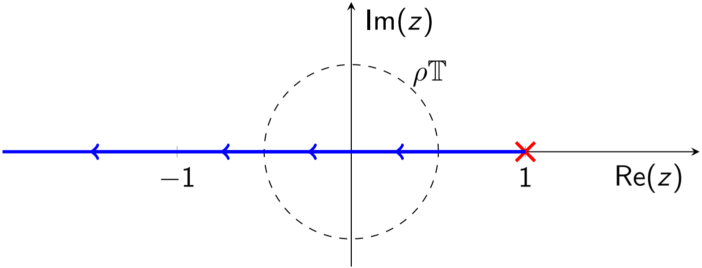


- How to choose k_p to optimize the convergence rate ρ?
- How to get both internal stability and robustness to ICs?

R.A. Freeman, Peng Yang, and K.M. Lynch (2006). "Stability and Convergence Properties of Dynamic Average Consensus Estimators". In: *Proc. of the 45th IEEE Conf. on Decision and Control*, pp. 338–343

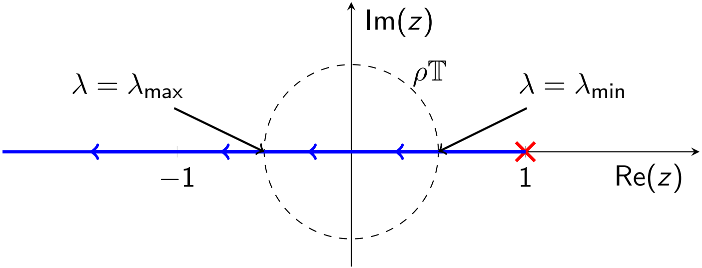
Optimizing the convergence rate

Characteristic equation: $\quad 0 = 1 + \lambda \dfrac{k_p}{z-1}, \quad \lambda \in \mathrm{eig}(L)$



Optimizing the convergence rate

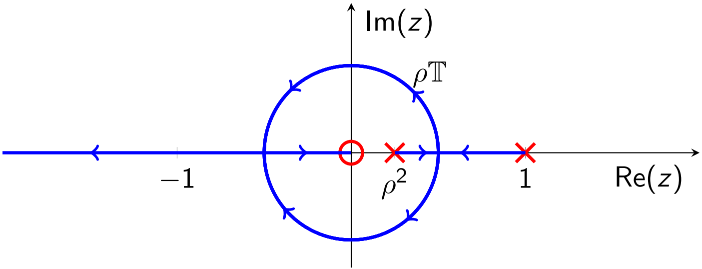
Characteristic equation: $\quad 0 = 1 + \lambda \frac{k_p}{z-1}, \quad \lambda \in \mathrm{eig}(L)$



$$k_p = \frac{2}{\lambda_{\max} + \lambda_{\min}} \qquad \rho = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}$$

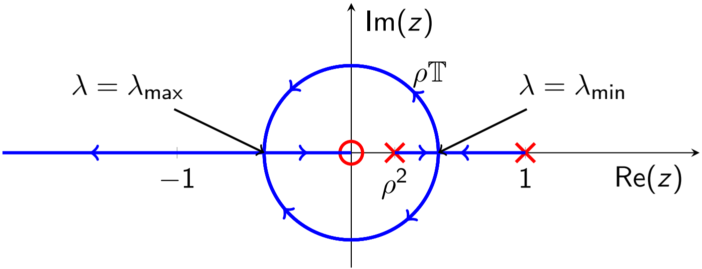
Optimizing the convergence rate

Characteristic equation: $$0 = 1 + \lambda \frac{k_p\, z}{(z-\rho^2)(z-1)}, \quad \lambda \in \mathrm{eig}(L)$$



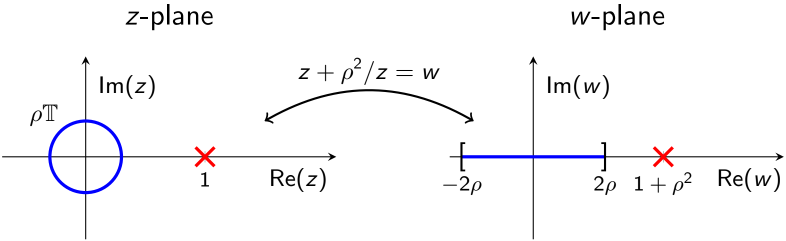
Optimizing the convergence rate

Characteristic equation: $0 = 1 + \lambda \frac{k_p\, z}{(z-\rho^2)(z-1)}, \quad \lambda \in \mathrm{eig}(L)$

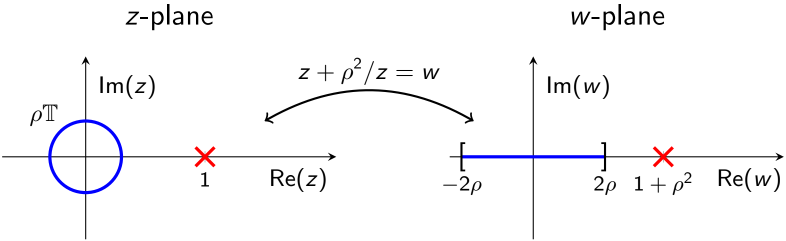


$$k_p = \frac{4}{(\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}})^2} \qquad \rho = \frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}$$

Palindromic transformation



Palindromic transformation



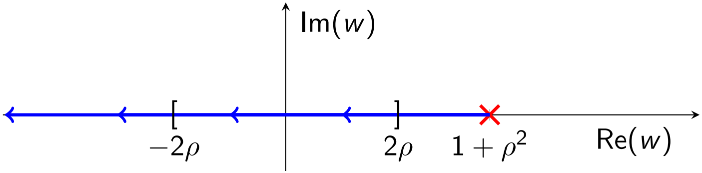
$$z + \rho^2/z = w$$

$$\frac{k_p\, z}{(z-\rho^2)(z-1)} \quad \longleftrightarrow \quad \frac{k_p}{w-(1+\rho^2)}$$

Optimizing the convergence rate

Original system: $0 = 1 + \lambda \dfrac{k_p\, z}{(z-\rho^2)(z-1)}, \quad \lambda \in \operatorname{eig}(L)$

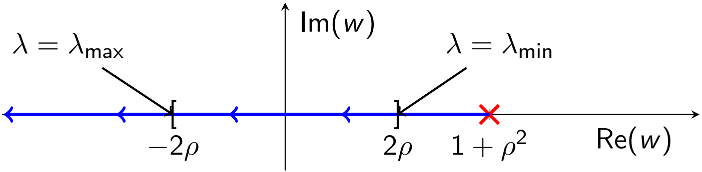
Palindromic system: $0 = 1 + \lambda \dfrac{k_p}{w-(1+\rho^2)}$



Optimizing the convergence rate

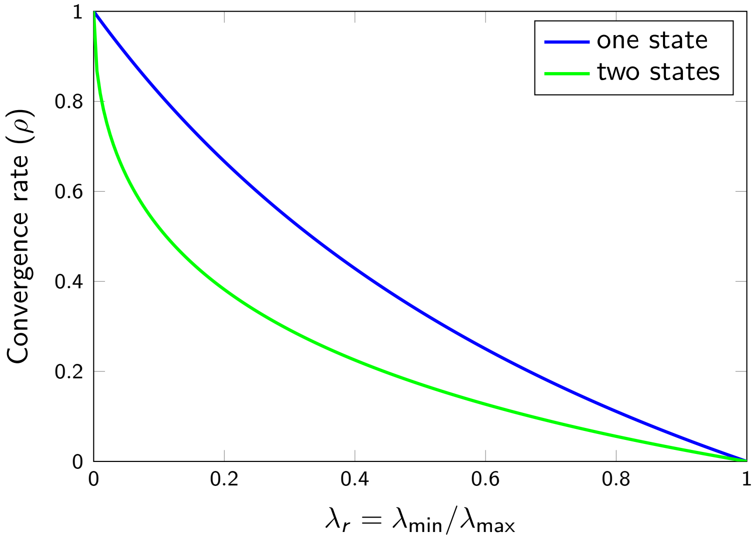
Original system: $0 = 1 + \lambda \dfrac{k_p\, z}{(z-\rho^2)(z-1)}, \quad \lambda \in \mathrm{eig}(L)$

Palindromic system: $0 = 1 + \lambda \dfrac{k_p}{w - (1+\rho^2)}$



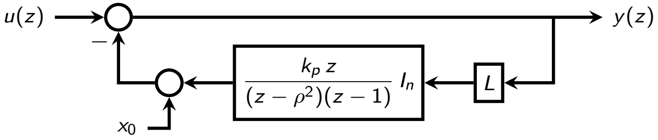
$$k_p = \frac{4}{(\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}})^2} \qquad \rho = \frac{\sqrt{\lambda_{\max}} - \sqrt{\lambda_{\min}}}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}$$

P estimator: Convergence rate

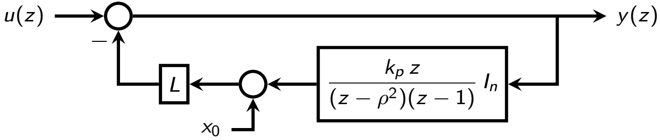


P estimator: Accelerated versions

Internally stable but not robust to initial conditions:

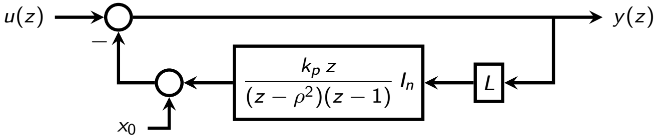


Robust to initial conditions but not internally stable:

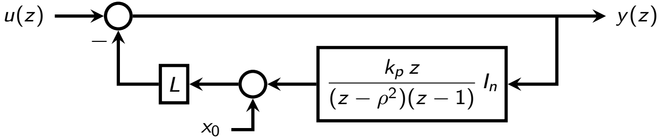


P estimator: Accelerated versions

Internally stable but not robust to initial conditions:



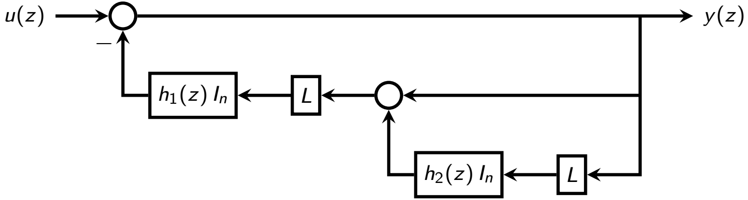
Robust to initial conditions but not internally stable:



- How to get both internal stability and robustness to ICs?

Proportional-integral (PI) estimator

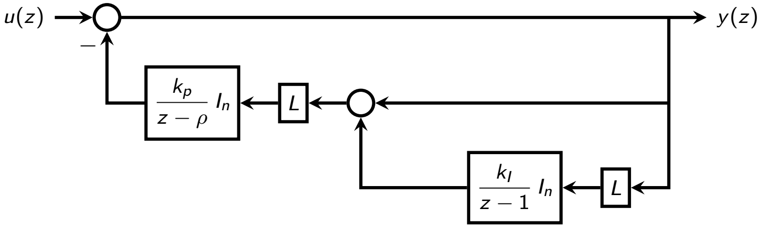
To obtain both robustness to initial conditions and internal stability, we can use two Laplacian blocks.



- Want $h_2(z)$ to have a pole at $z = 1$ to be exact.
- Want $h_1(z)$ and $h_2(z)$ to be strictly proper for one-hop communication.

R.A. Freeman, Peng Yang, and K.M. Lynch (2006). "Stability and Convergence Properties of Dynamic Average Consensus Estimators". In: *Proc. of the 45th IEEE Conf. on Decision and Control*, pp. 338–343

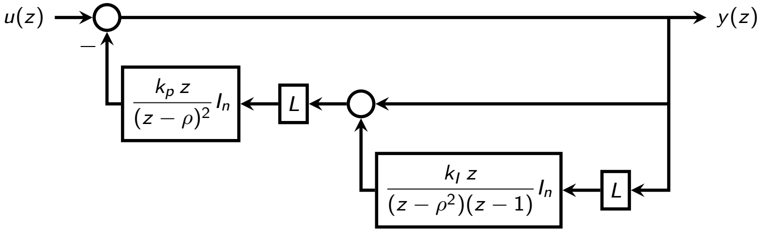
PI estimator: Two states



- Each agent has two internal state variables and transmits two variables per iteration.
- We have closed-form expressions for k_p, k_I, and ρ in terms of λ_{min} and λ_{max}.
- This estimator has all the desired properties, except the convergence rate is slow.

Bryan Van Scoy, Randy A. Freeman, and Kevin M. Lynch (July 2015b). "Optimal worst-case dynamic average consensus". In: *Proc. of the 2015 Amer. Control Conf.* Pp. 5324–5329

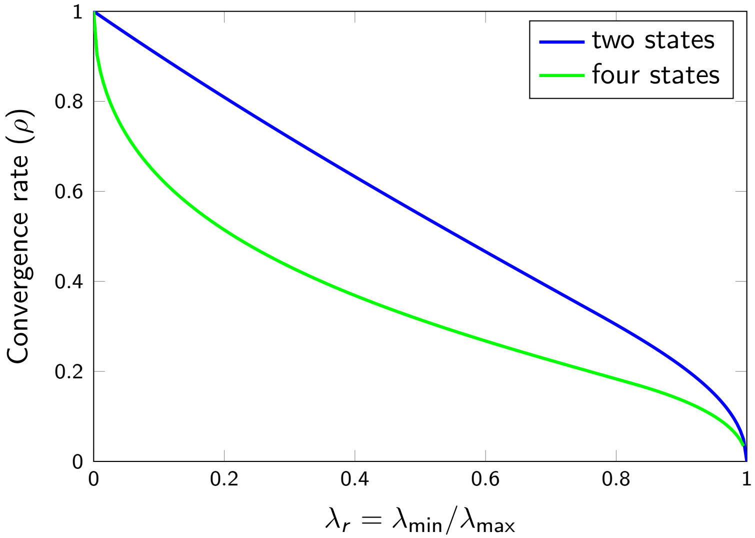
PI estimator: Four states



- Each agent has four internal state variables and transmits two variables per iteration.
- We have closed-form expressions for k_p, k_I, and ρ in terms of $\lambda_{\min}$ and $\lambda_{\max}$.
- This estimator has all the desired properties.

Bryan Van Scoy, Randy A. Freeman, and Kevin M. Lynch (Dec. 2015a). "Design of robust dynamic average consensus estimators". In: *Proc. of the 54th IEEE Conf. on Decision and Control*, pp. 6269–6275

PI estimator: Convergence rate



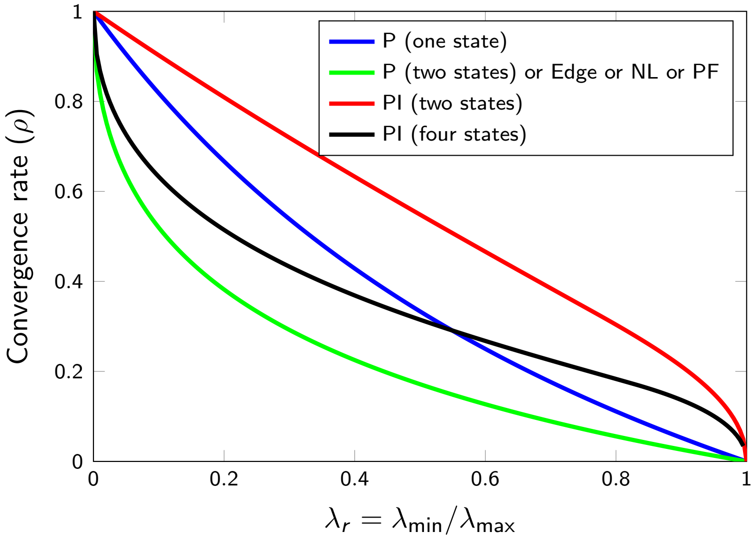
Summary of feedback estimators

	Internal states	Transmission variables	Communication hops	Scalable	Exact	Internally stable	Locally convergent	Globally convergent
P	1 or 2	1	1	✓	✓	✓	✗	✗
	1 or 2	1	1	✓	✓	✗	✓	✓
	1 or 2	1	1	✓	✗	✗	✗	✗
PI	2 or 4	2	1	✓	✓	✓	✓	✓
PF	1 or 2	1	r	✓	✓	✓	✗	✗
	1 or 2	1	r	✓	✓	✓	✓	✓
Edge	$1+\lvert\mathcal{N}_{\text{in}}(i)\rvert$	1	1	✗	✓	✓	✓	✓
NL	2	1	1	✓	✓	✓	✓	✗

Summary of feedback estimators

	Internal states	Transmission variables	Communication hops	Scalable	Exact	Internally stable	Locally convergent	Globally convergent
	1 or 2	1	1	✓	✓	✓	✗	✗
P	1 or 2	1	1	✓	✓	✗	✓	✓
	1 or 2	1	1	✓	✗	✗	✗	✗
PI	2 or 4	2	1	✓	✓	✓	✓	✓
PF	1 or 2	1	r	✓	✓	✓	✗	✗
	1 or 2	1	r	✓	✓	✓	✓	✓
Edge	$1+\lvert\mathcal{N}_{\text{in}}(i)\rvert$	1	1	✗	✓	✓	✓	✓
NL	2	1	1	✓	✓	✓	✓	✗

Summary of feedback estimators



Outline

Problem

Assumption (Graph)

Assume the graph is connected, balanced, and satisfies $\|I - L_k - \mathbf{1}_n \mathbf{1}_n^T / n\|_2 < 1$ *at each iteration.*

Assumption (Signals)

Assume the input signals are bandlimited with known cutoff frequency θ_c*, i.e.,* $|u^i(e^{j\theta})| = 0$ *for all* $\theta \in [\theta_c, \pi]$*.*

Desired properties:

1. one-hop communication
2. scalable
3. internally stable
4. time-invariant
5. robust to ICs
6. robust to changes in the graph
7. small steady-state error

Problem

Assumption (Graph)

Assume the graph is connected, balanced, and satisfies $\|I - L_k - \mathbf{1}_n\mathbf{1}_n^T/n\|_2 < 1$ at each iteration.

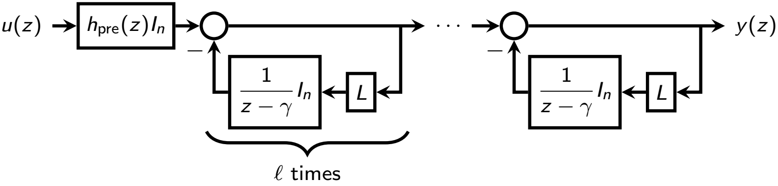
Assumption (Signals)

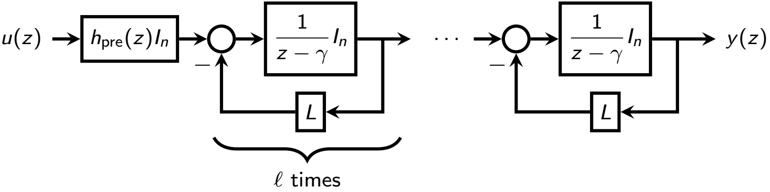
Assume the input signals are bandlimited with known cutoff frequency θ_c, i.e., $|u^i(e^{j\theta})| = 0$ for all $\theta \in [\theta_c, \pi]$.

Desired properties:

1. one-hop communication
2. scalable
3. internally stable
4. time-invariant
5. robust to ICs
6. robust to changes in the graph
7. small steady-state error

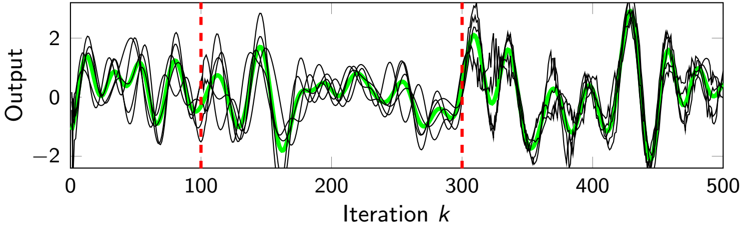
Cascaded feedback estimators

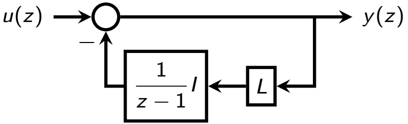




Estimator	**Fig.**	**Parameters**
Freeman, Yang, and Lynch, 2006	top	$h_{\text{pre}}(z) = 1,\ \ell = 1$
Kia, Cortés, and Martínez, 2014	top	$h_{\text{pre}}(z) = 1,\ \ell = 1$
Zhu and Martínez, 2010	bottom	$h_{\text{pre}}(z) = (1 - z^{-1})^{\ell},\ \gamma = 1$
Franceschelli and Gasparri, 2016	bottom	$h_{\text{pre}}(z) = (1 - \gamma)^{\ell}$

Feedback estimator 1

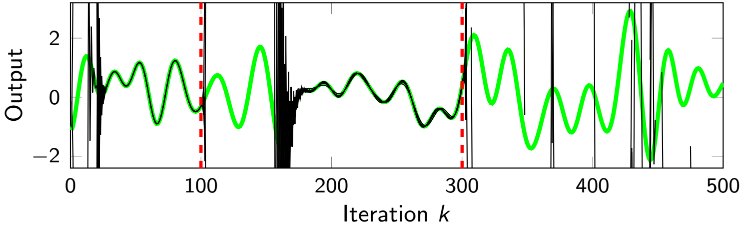


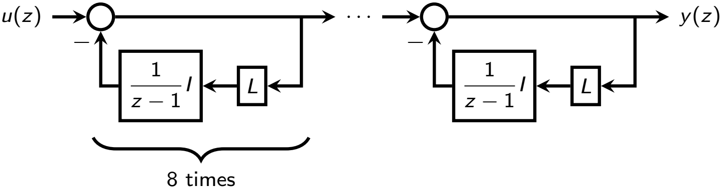


R.A. Freeman, Peng Yang, and K.M. Lynch (2006). "Stability and Convergence Properties of Dynamic Average Consensus Estimators". In: *Proc. of the 45th IEEE Conf. on Decision and Control*, pp. 338–343

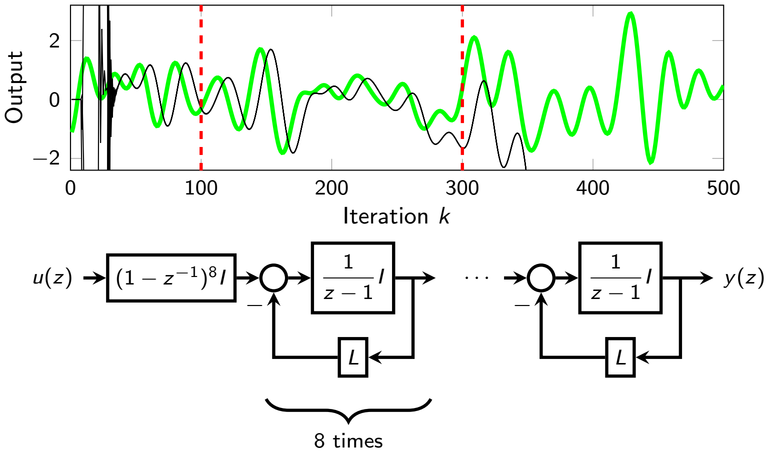
S.S. Kia, J. Cortés, and S. Martínez (2013). "Dynamic Average Consensus under Limited Control Authority and Privacy Requirements". In: *International Journal of Robust and Nonlinear Control*

Feedback estimator 2



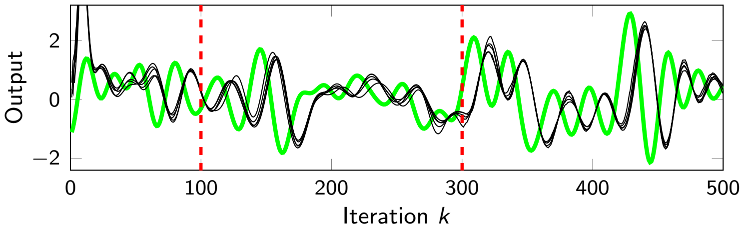


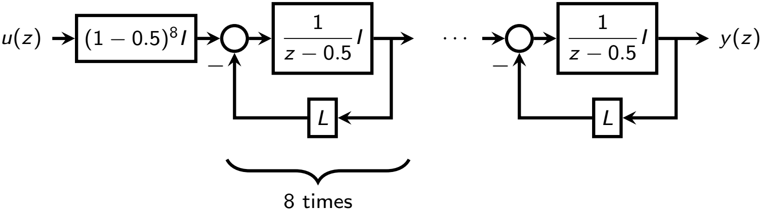
Feedback estimator 3



Minghui Zhu and Sonia Martínez (2010). "Discrete-time dynamic average consensus". In: *Automatica* 46.2, pp. 322 –329

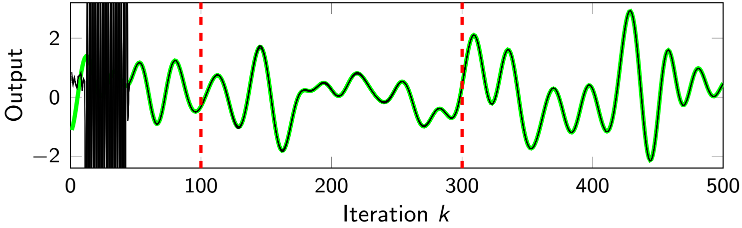
Feedback estimator 4

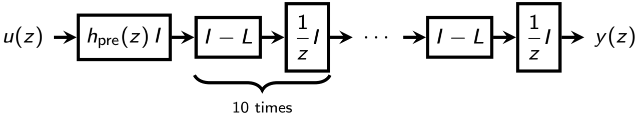




M. Franceschelli and A. Gasparri (Dec. 2016). "Multi-Stage Discrete Time Dynamic Average Consensus". In: *Proc. of the 55nd IEEE Conf. on Decision and Control*

Feedforward estimator



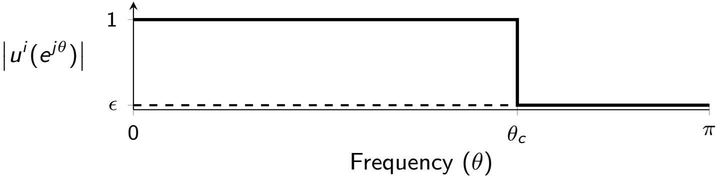


$$h_{\text{pre}}(z) = \left(\frac{3.902z^3 - 5.805z^2 + 3.902z - 1}{z^3}\right)^{10}$$

Bryan Van Scoy, Randy A. Freeman, and Kevin M. Lynch (Dec. 2016). "Feedforward estimators for the distributed average tracking of bandlimited signals in discrete time with switching graph topology". In: *Proc. of the 55th IEEE Conf. on Decision and Control*, pp. 4284–4289

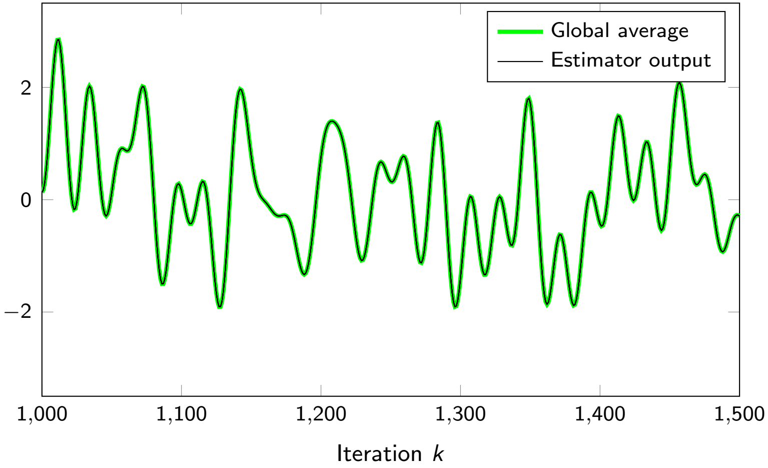
Design of prefilter

- We have two prefilter designs
  - **FIR:** Faster convergence, larger steady-state error
  - **IIR:** Slower convergence, smaller steady-state error
- Must take into account high-frequency components due to numerical error.
- Zero steady-state error if
  - exact arithmetic is used ($\epsilon = 0$),
  - infinite number of stages ($\ell \to \infty$), and
  - graph satisfies the assumptions at each iteration.



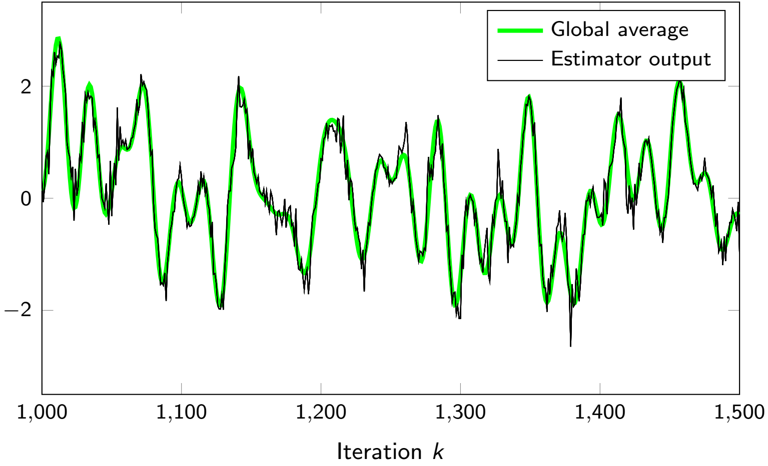
Drop packets with probability 0%

Assumptions: connected=100%, balanced=100%, norm condition=100%



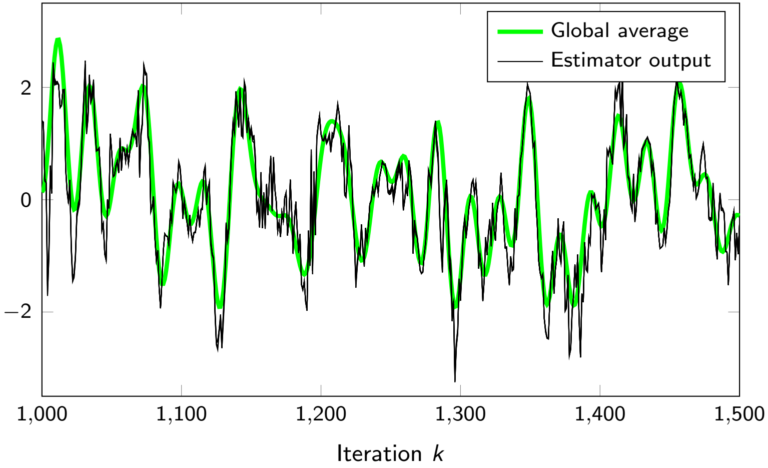
Drop packets with probability 10%

Assumptions: connected=83.5%, balanced=35.7%, norm condition=34.1%



Drop packets with probability 50%

Assumptions: connected=7.4%, balanced=3.3%, norm condition=0.2%



Summary

To summarize, the feedforward estimator

1. uses one-hop discrete-time local broadcast communication,
2. is scalable,
3. is internally stable,
4. is time-invariant,
5. is robust to initial conditions,
6. is robust to changes in the graph,
7. and has bounded steady-state error.

Futhermore, the steady-state error can be made arbitrarily small if the graph satisfies certain properties on average and exact arithmetic is used.

Conclusion

- **Convex optimization** The triple momentum method is the fastest known globally convergent first-order method for the minimization of strongly convex functions.
- **Dynamic average consensus** Estimator design depends on the frequency spectrum of the signals. Both estimators are
  - scalable,
  - time-invariant,
  - internally stable,
  - robust to initial conditions, and
  - use one-hop discrete-time communication.

Discrete: PI-4 estimator

- exact
- ergodic
- fast convergence rate

Continuous: Feedforward estimator

- small steady-state error
- robust to changes in the graph
- convergence rate depends on prefilter