Coding Rate Optimization for Distributed Average Consensus

Ryan Pilgrim
Defense
Master of Science in Electrical Engineering
North Carolina State University
20 June 2017

Motivation

Motivation

- Big, bad, distributed data
- Pervasive computing and sensor nets

Different communication costs

- Sensor networks and cloud services are two seemingly very different settings
- Can we solve both simultaneously?

Different communication costs

- Sensor networks and cloud services are two seemingly very different settings
- Can we solve both simultaneously?

Design compression strategy!

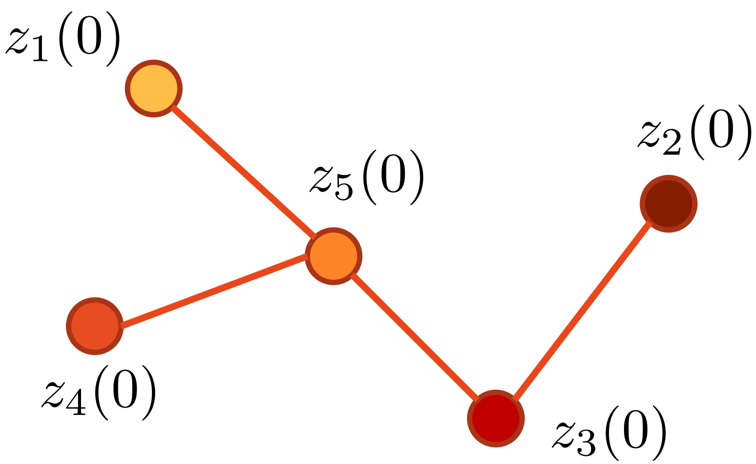
[Yildiz & Scaglione '08;
Su & El Gamal '10;
Yang *et al.* '17;
Zhu & coauthors '16-'17]

Consensus background

Distributed average consensus

[DeGroot ’74; Borkar & Varaiya ’82; Tsitsiklis ’84; Tsitsiklis & Athans ’86]

- Goal: compute sample mean of data distributed throughout network at every node



$$z_i(0)$$

“state” at node i and iteration $t = 0$

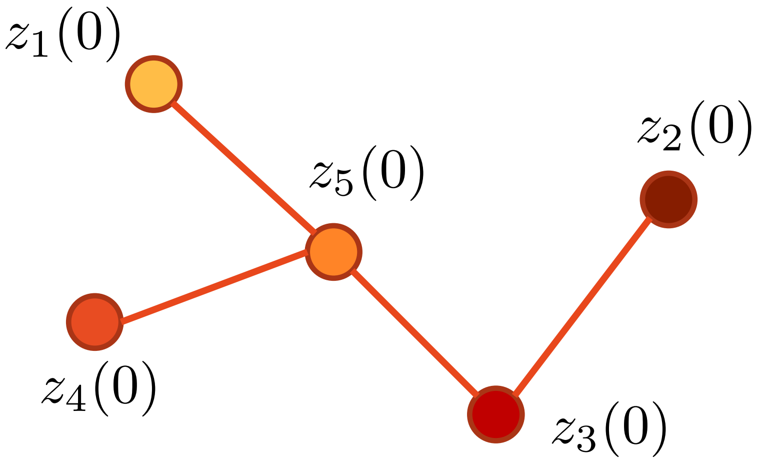
$$m$$

number of nodes

Distributed average consensus

[DeGroot '74; Borkar & Varaiya '82; Tsitsiklis '84; Tsitsiklis & Athans '86]

- Goal: compute sample mean of data distributed throughout network at every node

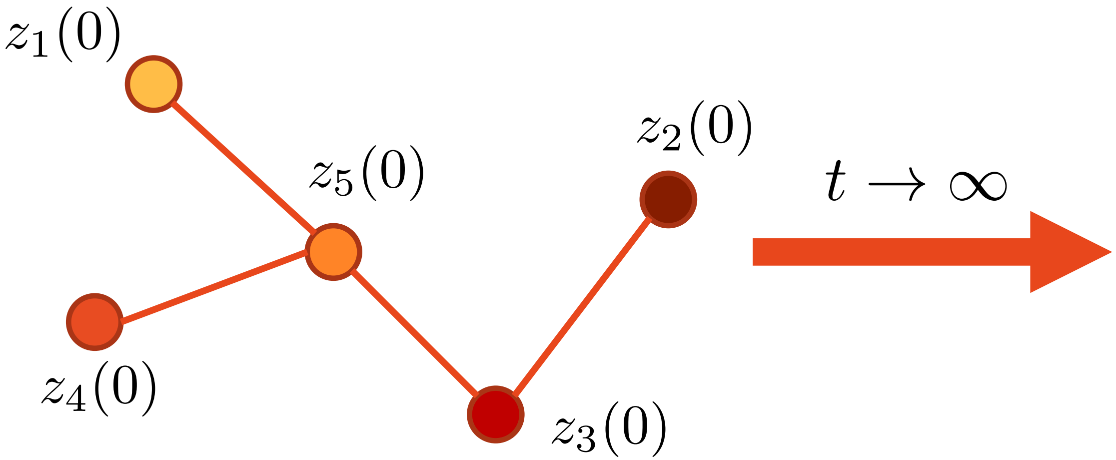


t iteration index

Distributed average consensus

[DeGroot ’74; Borkar & Varaiya ’82; Tsitsiklis ’84; Tsitsiklis & Athans ’86]

- Goal: compute sample mean of data distributed throughout network at every node

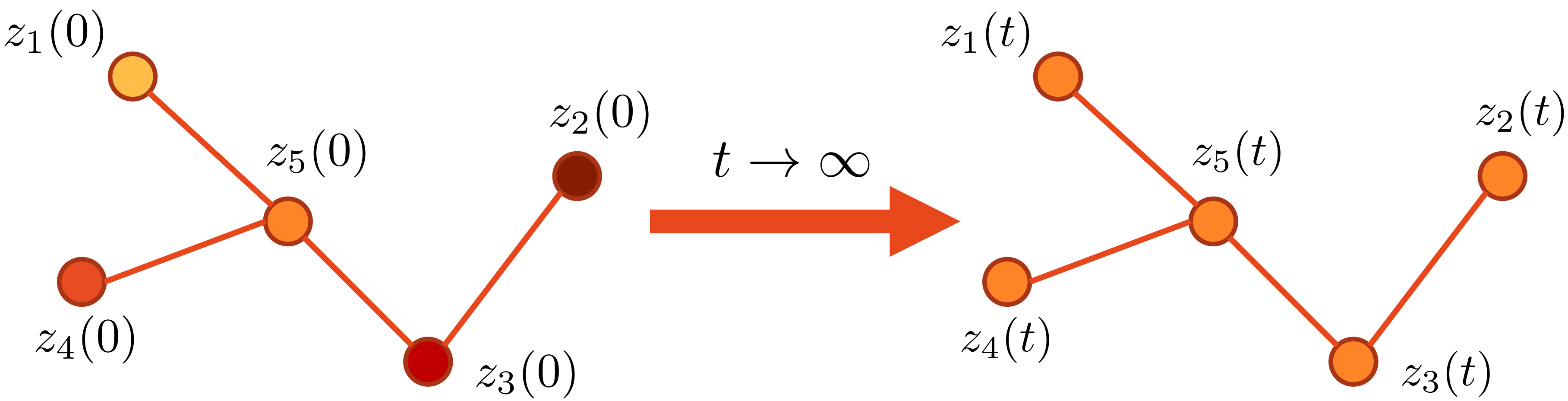


t iteration index

Distributed average consensus

[DeGroot ’74; Borkar & Varaiya ’82; Tsitsiklis ’84; Tsitsiklis & Athans ’86]

- Goal: compute sample mean of data distributed throughout network at every node

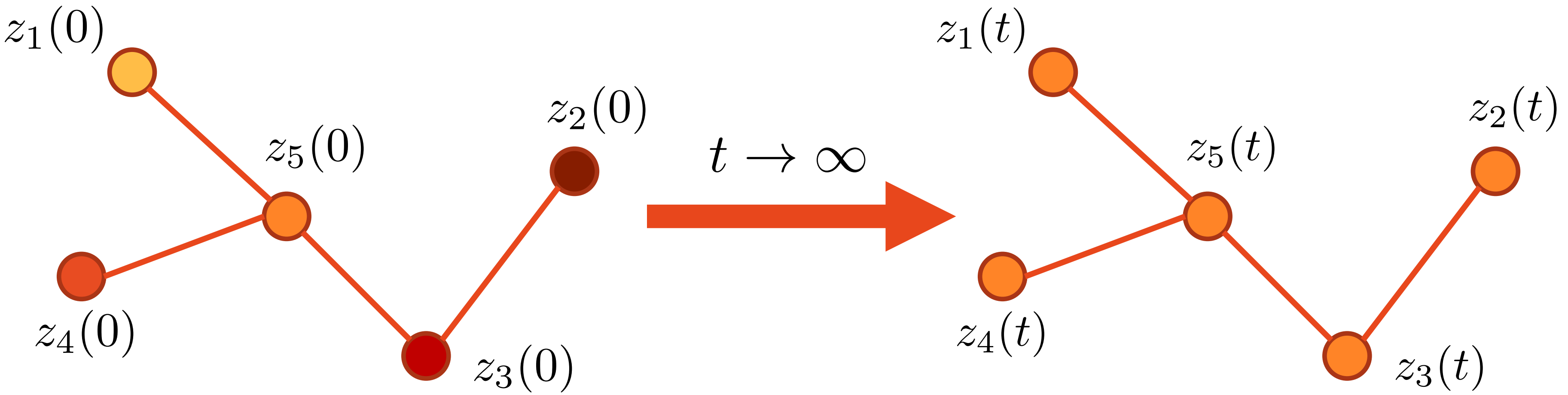


t iteration index

Distributed average consensus

[DeGroot ’74; Borkar & Varaiya ’82; Tsitsiklis ’84; Tsitsiklis & Athans ’86]

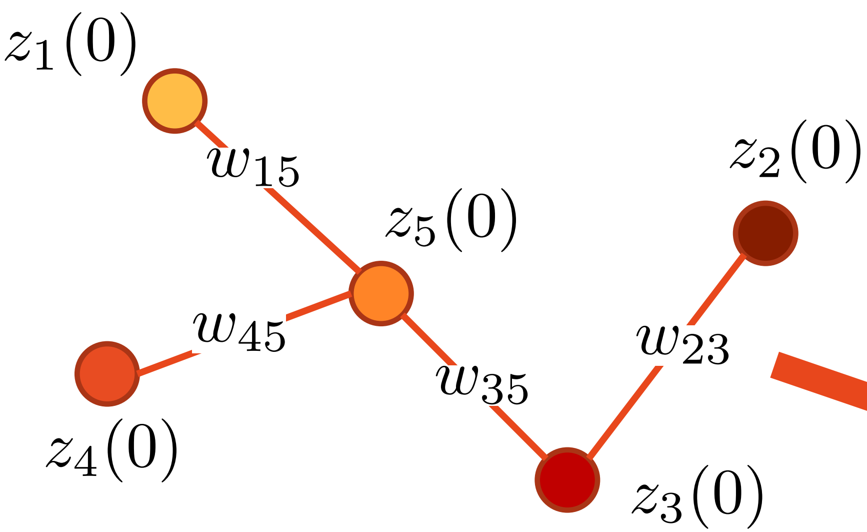
- Goal: compute sample mean of data distributed throughout network at every node

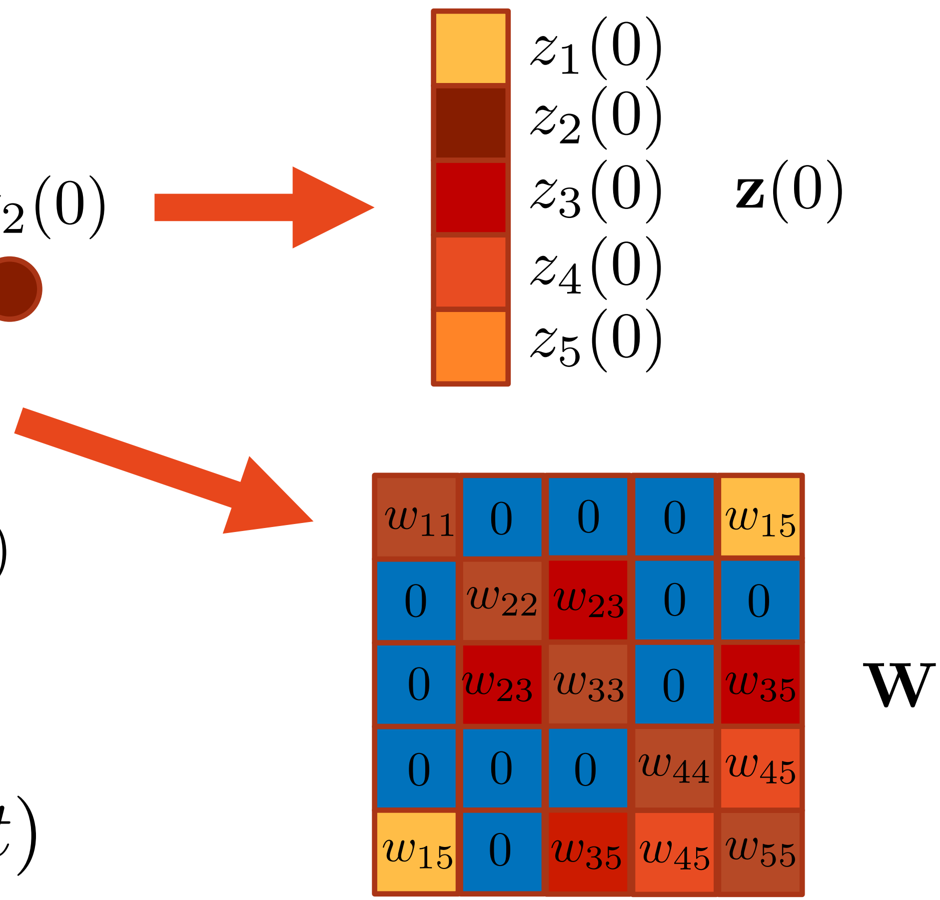


$$\lim_{t\to\infty} z_i(t) = \frac{1}{m}\sum_{i}^{m} z_i(0), \quad \forall i$$

Distributed average consensus

[DeGroot ’74; Borkar & Varaiya ’82; Tsitsiklis ’84; Tsitsiklis & Athans ’86]



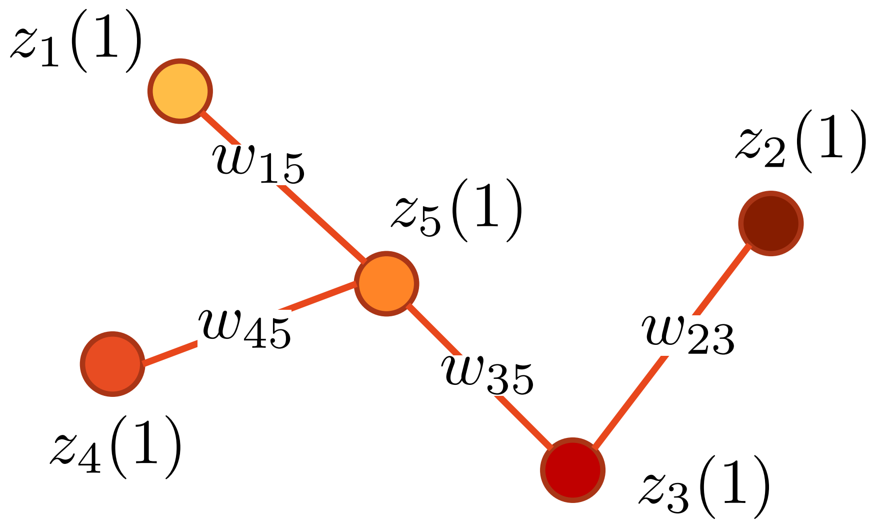


$$[\mathbf{W}]_{ij} = w_{ij}$$

$$[\mathbf{z}(t)]_i = z_i(t)$$

Distributed average consensus

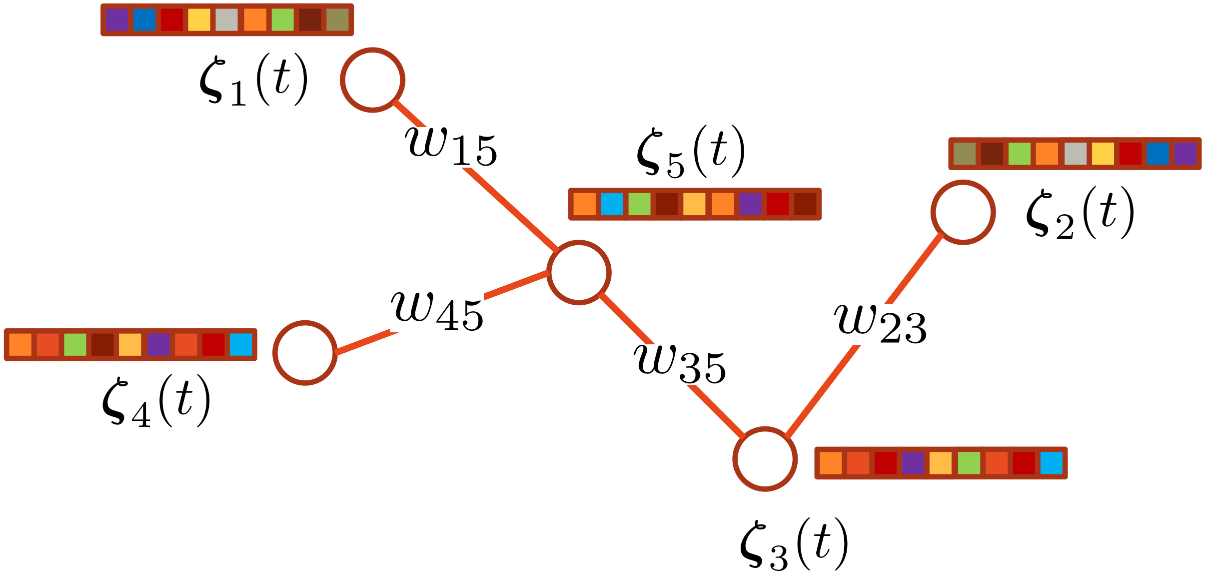
[DeGroot ’74; Borkar & Varaiya ’82; Tsitsiklis ’84; Tsitsiklis & Athans ’86]



$$\mathbf{z}(t+1) = \mathbf{W}\mathbf{z}(t)$$

Vector extension to consensus

[Huang & Hua '11]

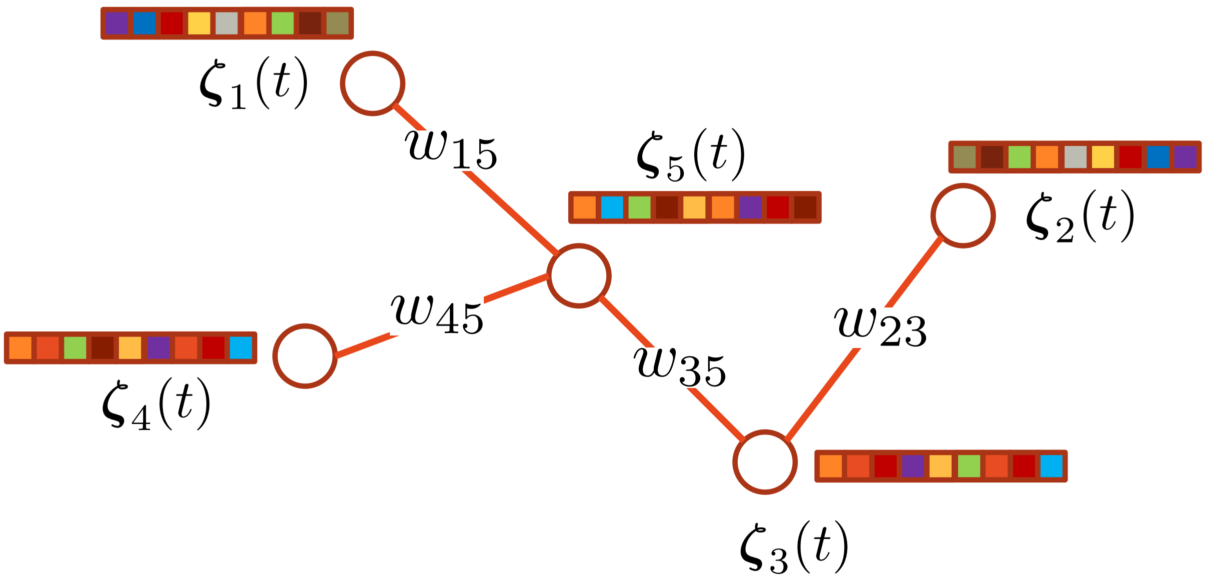


Vector extension to consensus

[Huang & Hua '11]

- Each node has vector-valued state $\zeta_i(t) \in \mathbb{R}^L \;\; \forall i$

$$t = 0$$

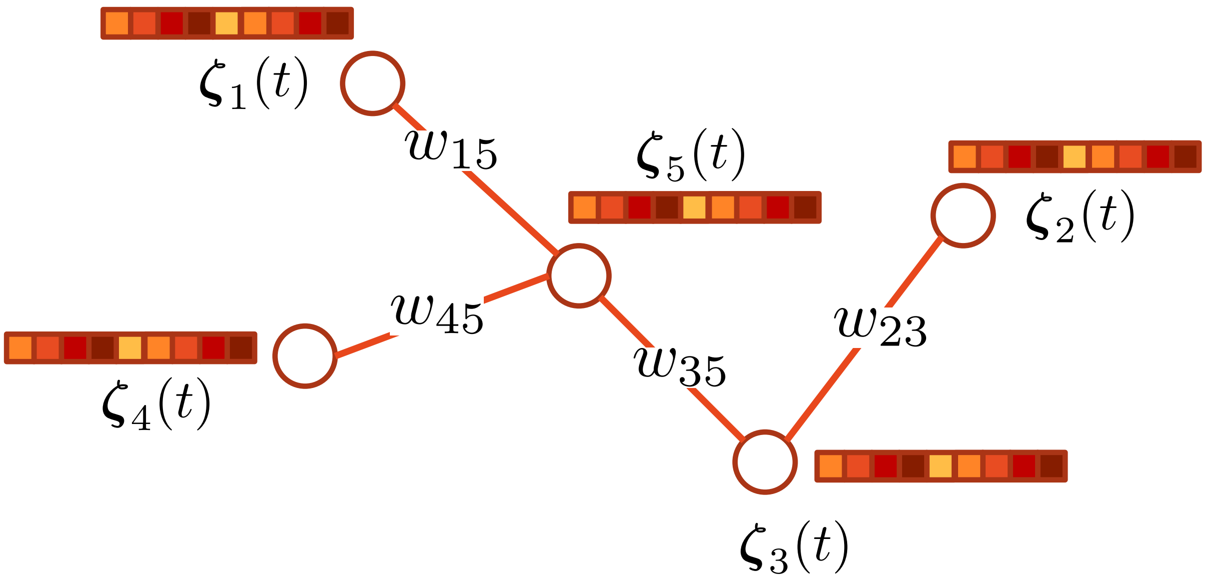


Vector extension to consensus

[Huang & Hua '11]

- Each node has vector-valued state $\zeta_i(t) \in \mathbb{R}^L \;\; \forall i$

$$t \rightarrow \infty$$

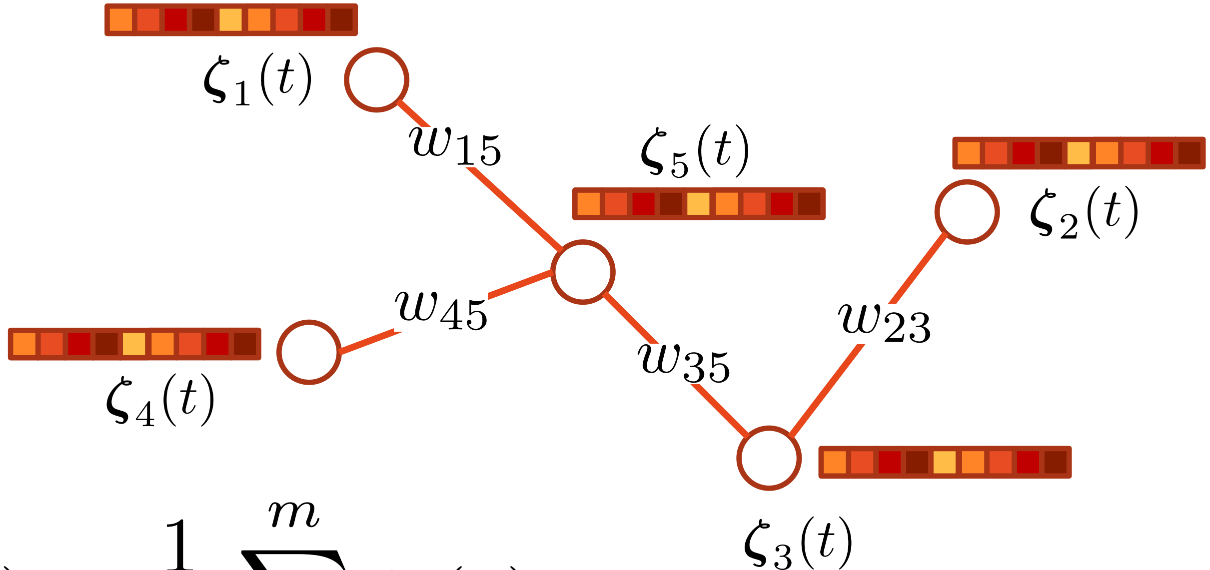


Vector extension to consensus

[Huang & Hua '11]

- Each node has vector-valued state $\zeta_i(t) \in \mathbb{R}^L \ \forall i$

$$t \to \infty$$



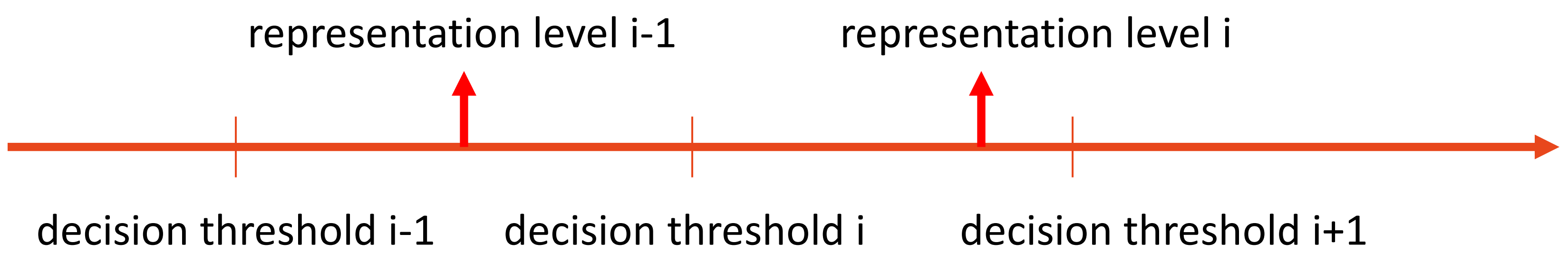
$$\lim_{t \to \infty} \zeta_i(t) = \frac{1}{m} \sum_{i=1}^{m} \zeta_i(0), \quad \forall i$$

Compression background

Quantization

[Gersho & Gray '91]

- Quantizer assigns inputs (states) to representation levels based on partition



- Average distortion $D = \mathbb{E}\left[(z-\hat{z})^2\right]$

- Vector quantization (VQ) = multidim. quantization of vector; achieves lower *D*

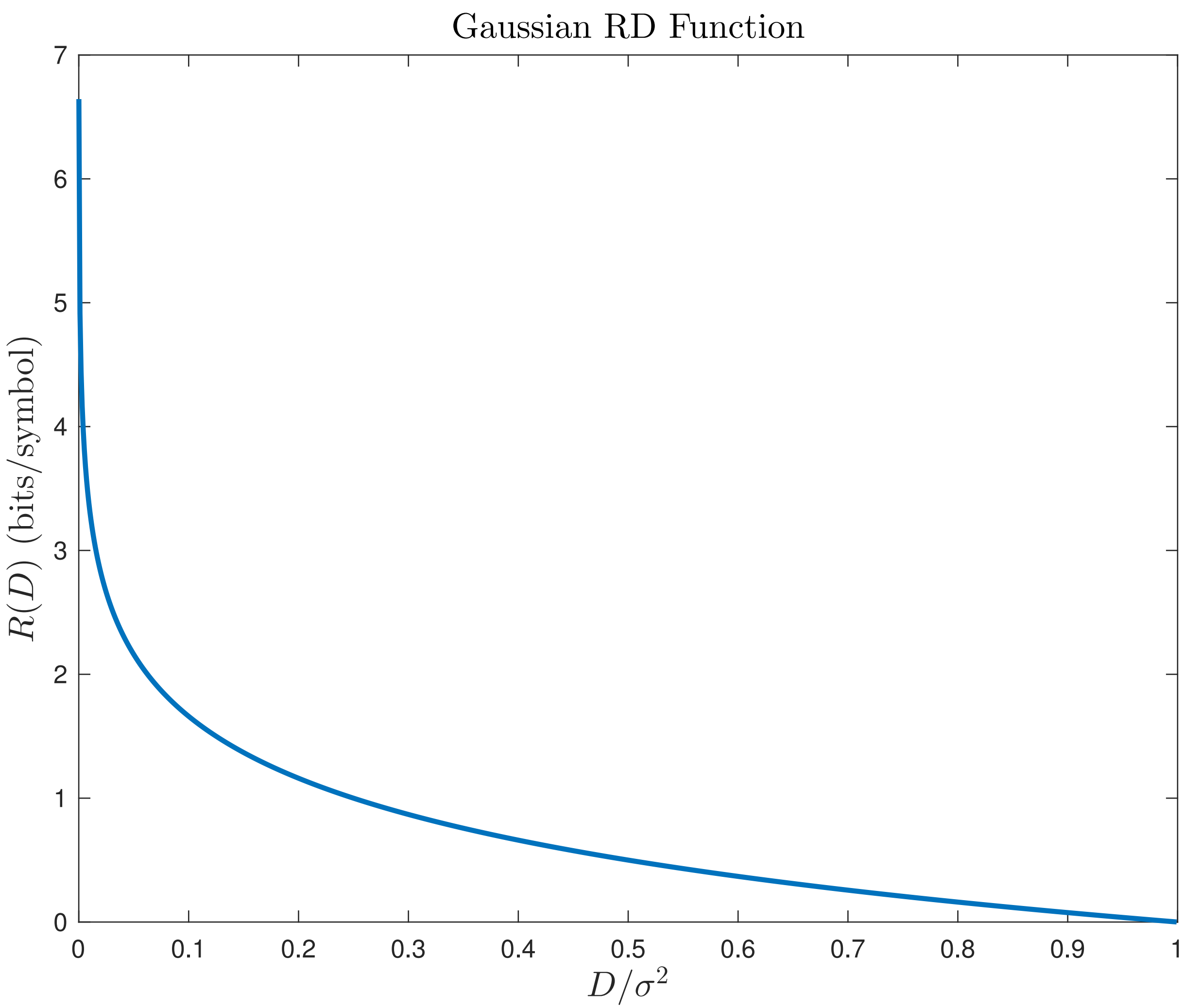
Coding

[Gersho & Gray '91; Cover & Thomas '06]

- Encode representation levels into bits
- Rate R (per symbol) defined as average # bits/source element
- Fixed rate: same # bits for each representation level
- Entropy H = best possible R
- Entropy coding allows R to approach H

Rate-distortion (RD) theory

[Berger ’71; Cover & Thomas ’91]



Quantized consensus

- Quantization error issues [Frasca *et al.* ’08]

$$z_i(t+1) = z_i(t) + \sum_{j=1}^{m} w_{ij} \left(Q(z_j(t)) - Q(z_i(t))\right)$$

$$\Updownarrow$$

$$\mathbf{z}(t+1) = \mathbf{z}(t) + (\mathbf{W} - \mathbf{I})\, Q(\mathbf{z}(t))$$

- Careful state update design needed for convergence

Prior art

- Predictive, differential, and Wyner-Ziv lattice coding [Yildiz & Scaglione ’08]
- Convergent differential coding approach [Thanou *et al.* ’13]
- Information theory (RD analyses) [Yang *et al.* ’17; Su & El Gamal ’10]

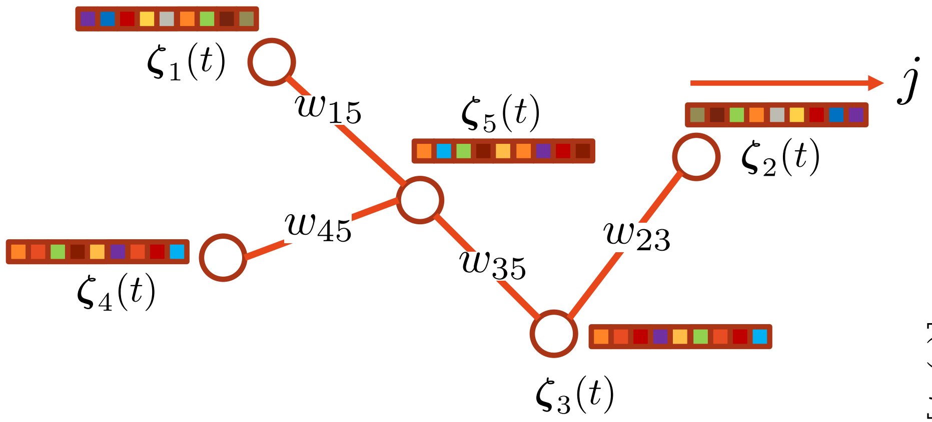
Contributions

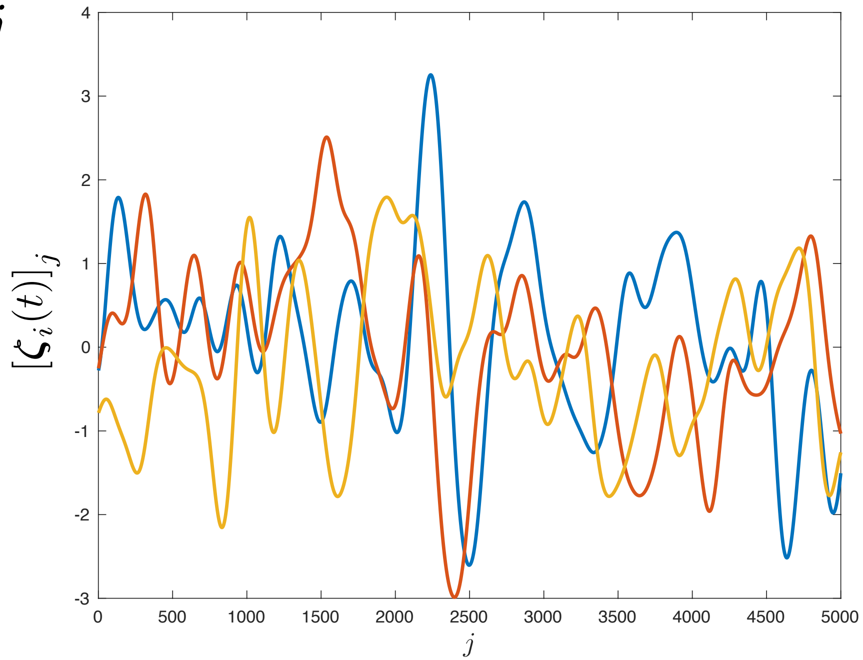
Highlights

- Framework for solving comm. optimization with mean square error (MSE) constraint
- Useful for both long-block VQ and scalar quantization (with/without entropy coding)
- Heuristic search for fixed-rate coding
- Numerical results: impact of topology on rates

Assumptions I

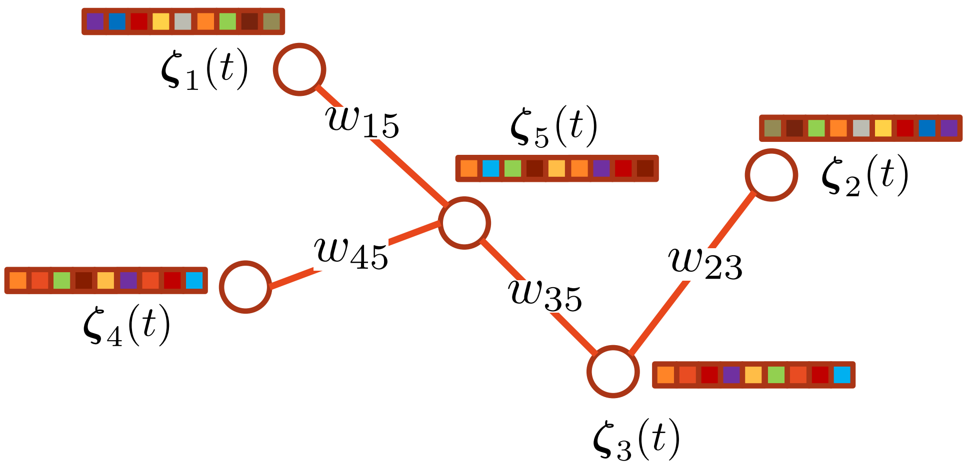
- Multivariate Gaussian distribution
  - Assumed stationary





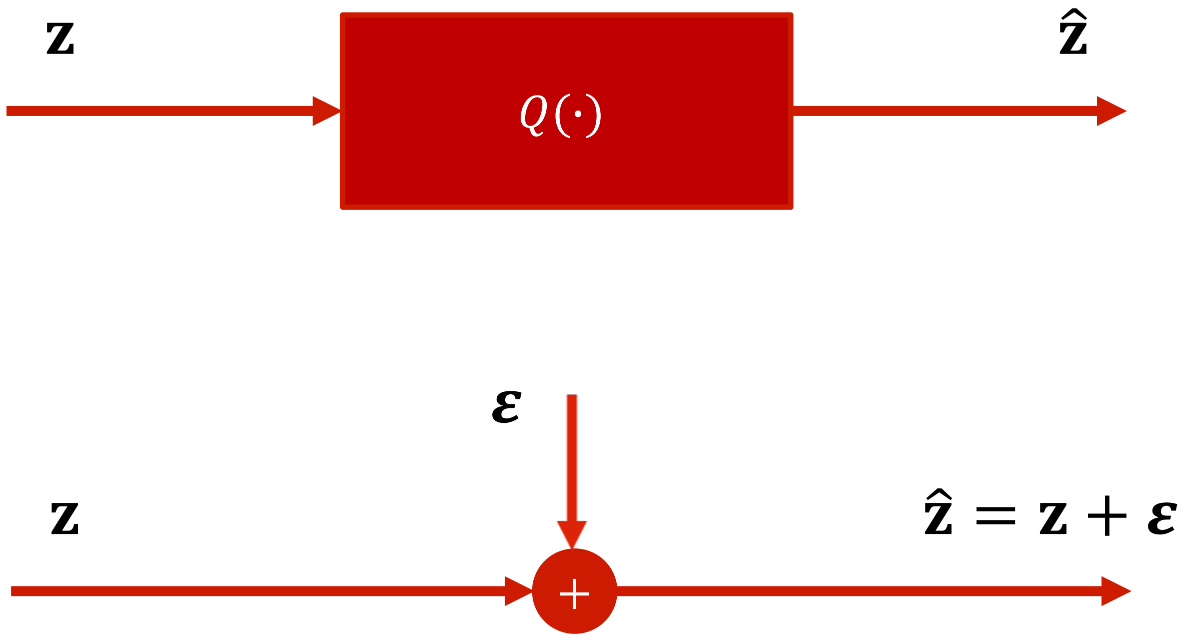
Assumptions II

- Broadcast communications
- Static network topology



Assumptions III

- Quantization error modeled as additive noise



Sneak peek - technical approach

- Rate-distortion model given assumptions
- Statistical model
- Optimization problem classification (proof)
- Heuristic for fixed rate coding

Operational rate-distortion

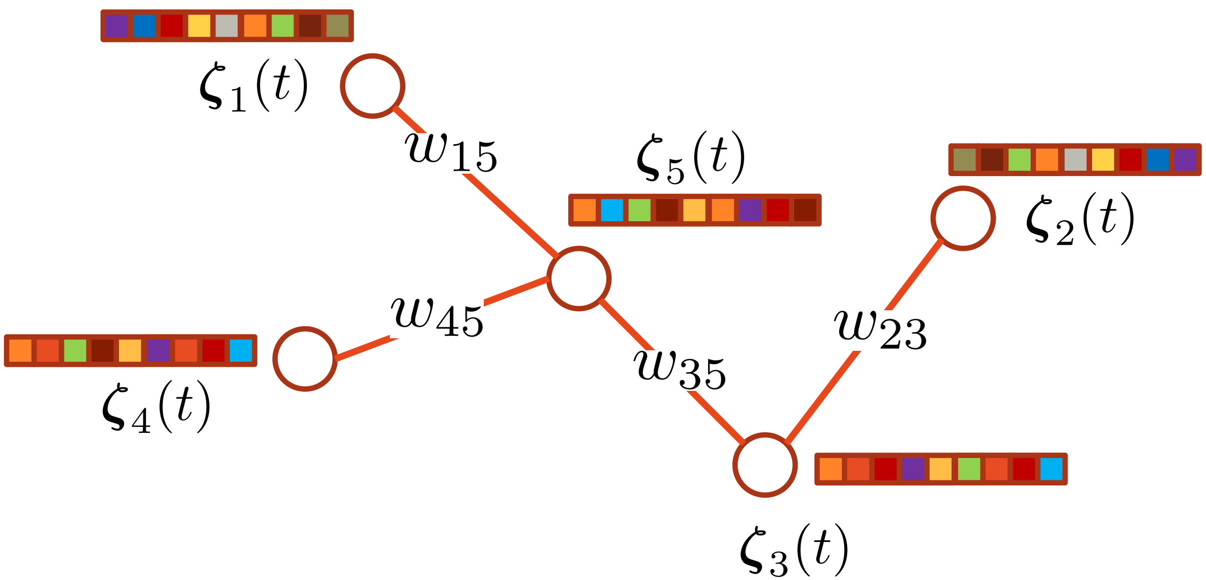
- Quantizers of interest have form (high rate):

$$R(D) \approx \begin{cases} \frac{1}{2}\log_2\left(\frac{\sigma^2}{D}\right) + R_c, & \frac{D}{\sigma^2} \in (0, \text{const.}] \\ 0, & \text{otherwise} \end{cases}$$

Rate-distortion model

- In general,

$$R_i(t) \neq R_j(s) \iff D_i(t) \neq D_j(t)$$



$$\mathbf{D} := [D_1(0) \cdots D_m(0) \cdots D_1(T-1) \cdots D_m(T-1)]^\top$$

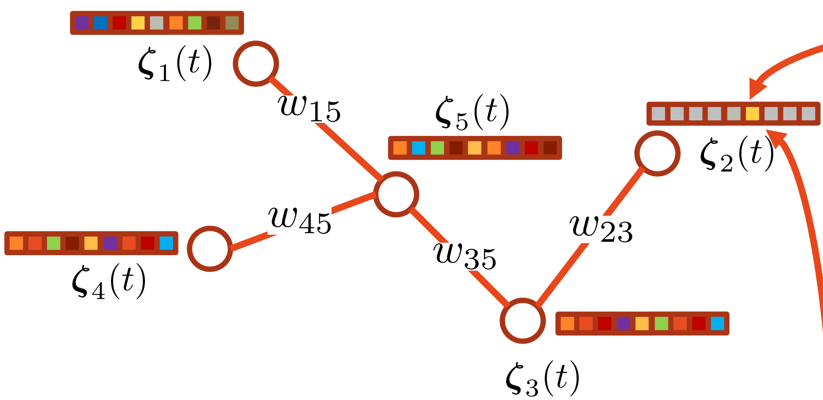
State evolution

- Gaussian → {mean, covariance} sufficient stats

- Used to compute optimization model parameters (marginal variances and MSE values)
  - Marginal variance suffices for fixed-rate and entropy coding [Widrow & Kollar ’08; Gersho Gray ’91]

Necessary statistics for optimization

- Need variance, MSE (use marginals)
- Node index i, element index j



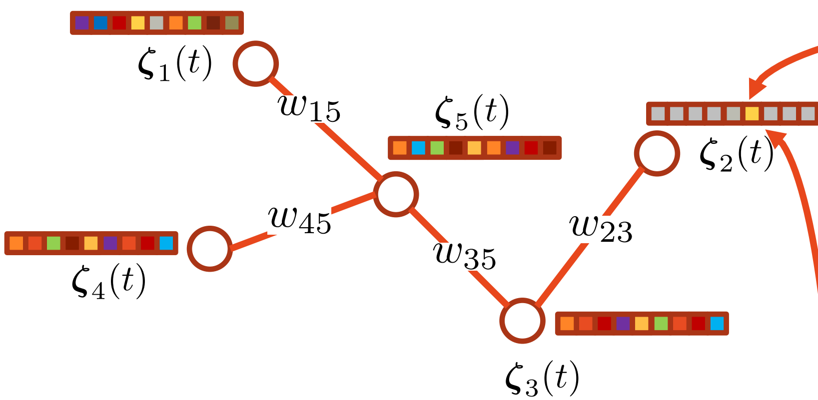
$$\nu_i(\mathbf{D}, t) = \mathrm{var}\left([\zeta_i(t)]_j \right)$$

$$\mathrm{MSE}(\mathbf{D}, t) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{MSE}_i(\mathbf{D}, t)$$

$$\mathrm{MSE}_i(\mathbf{D}, t) = \mathbb{E}\left[\left([\zeta_i(t)]_j - \frac{1}{m} \sum_{i=1}^{m} [\zeta_i(0)]_j \right)^2 \right]$$

Necessary statistics for optimization

- Need variance, MSE (use marginals)
- Node index i, element index j



$$\nu_i(\mathbf{D}, t) = \mathrm{var}\left(\left[\zeta_i(t)\right]_j\right)$$

$$\mathrm{MSE}(\mathbf{D}, t) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{MSE}_i(\mathbf{D}, t)$$

$$\mathrm{MSE}_i(\mathbf{D}, t) = \mathbb{E}\left[\left(\left[\zeta_i(t)\right]_j - \frac{1}{m}\sum_{i=1}^{m}\left[\zeta_i(0)\right]_j\right)^2\right]$$

Cost function

- Goal: minimize total coding rate used

$$R_{\text{agg}} = \sum_{t=0}^{T-1} \sum_{i=1}^{m} R_i(\mathbf{D}, t)$$

Cost function

- Goal: minimize total coding rate used

$$R_{\text{agg}} = \sum_{t=0}^{T-1} \sum_{i=1}^{m} R_i(\mathbf{D}, t)$$

- “Aggregate rate”

Cost function

- Want to minimize (neglecting constraints for now)

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} R_i(\mathbf{D}, t)$$

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \begin{cases} \frac{1}{2}\log_2\left(\frac{\nu_i(\mathbf{D},t)}{D_i(t)}\right) + R_c, & \frac{D_i(t)}{\nu_i(\mathbf{D},t)} \in (0, \text{const.}] \\ 0, & \text{otherwise} \end{cases}$$

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \frac{1}{2}\log_2\left(\max\left\{\frac{\nu_i(\mathbf{D},t)}{D_i(t)}, k\right\}\right) + R_c$$

Cost function

- Want to minimize (neglecting constraints for now)

$$\sum_{t=0}^{T-1}\sum_{i=1}^{m} R_i(\mathbf{D},t)$$

$$\sum_{t=0}^{T-1}\sum_{i=1}^{m} \begin{cases} \frac{1}{2}\log_2\left(\frac{\nu_i(\mathbf{D},t)}{D_i(t)}\right) + R_c, & \frac{D_i(t)}{\nu_i(\mathbf{D},t)} \in (0, \text{const.}] \\ 0, & \text{otherwise} \end{cases}$$

$$\sum_{t=0}^{T-1}\sum_{i=1}^{m} \frac{1}{2}\log_2\left(\max\left\{\frac{\nu_i(\mathbf{D},t)}{D_i(t)}, k\right\}\right) + R_c$$

Cost function

- Want to minimize (neglecting constraints for now)

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} R_i(\mathbf{D}, t)$$

↓

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \begin{cases} \frac{1}{2} \log_2 \left(\frac{\nu_i(\mathbf{D},t)}{D_i(t)} \right) + R_c, & \frac{D_i(t)}{\nu_i(\mathbf{D},t)} \in (0, \text{const.}] \\ 0, & \text{otherwise} \end{cases}$$

↓

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(\max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right) + R_c$$

Key theoretical contribution:

Generalized geometric programming (GGP)

Cost function transformation

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(\max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right) + R_c$$

$$\log_2 \left(\prod_{t=0}^{T-1} \prod_{i=1}^{m} \max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right)$$

$$\prod_{t=0}^{T-1} \prod_{i=1}^{m} \max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\}$$

Cost function transformation

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(\max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right) + R_c$$

Eliminate constants

$$\log_2 \left(\prod_{t=0}^{T-1} \prod_{i=1}^{m} \max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right)$$

$$\prod_{t=0}^{T-1} \prod_{i=1}^{m} \max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\}$$

Cost function transformation

$$\sum_{t=0}^{T-1} \sum_{i=1}^{m} \frac{1}{2} \log_2 \left(\max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right) + R_c$$

$$\log_2 \left(\prod_{t=0}^{T-1} \prod_{i=1}^{m} \max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\} \right)$$

↓ $\exp(\cdot)$

$$\prod_{t=0}^{T-1} \prod_{i=1}^{m} \max \left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\}$$

Optimization problem

$$\begin{aligned} \underset{\mathbf{D}}{\text{minimize}} \quad & \prod_{t=0}^{T-1} \prod_{i=1}^{m} \max\left\{ \frac{\nu_i(\mathbf{D}, t)}{D_i(t)}, k \right\}, \\ \text{subject to} \quad & \text{MSE}(\mathbf{D}, T) \leq \text{MSE}^*, \\ & D_i(t) > 0, \quad \forall i, t \end{aligned}$$

Optimization is GGP

[Boyd & Vandenberghe, '04; Boyd *et al.* '07]

- Monomials
$$f(x) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \quad c > 0, x_i > 0 \ \forall i, a_i \in \mathbb{R} \ \forall i$$

- Posynomials
$$f(x) = \sum_{i=1}^{k} b_i g_i(x_1, \ldots, x_n), \ b_i > 0 \forall i$$

- Generalized posynomials formed by $+$, $\times$, $\div$,* and $\max(\cdot)$

* only generalized posynomial by monomial division

Optimization is GGP

- Generalized geometric programs

$$\begin{aligned} \underset{x_1,\ldots,x_n}{\text{minimize}} \quad & C(x_1,\ldots,x_n), \\ \text{subject to} \quad & f_i(x_1,\ldots,x_n) \leq 1, \quad \forall i, \\ & g_i(x_1,\ldots,x_n) = 1, \quad \forall i, \\ & x_i > 0, \quad \forall i \end{aligned}$$

- C, f_i's: generalized posynomials
- g_i's: monomials

Optimization is GGP

- Generalized posynomials:

$$\nu_i(\mathbf{D}, t) = \mathrm{var}\left(\left[\zeta_i(t)\right]_j\right)$$

$$\mathrm{MSE}_i(\mathbf{D}, t) = \mathbb{E}\left[\left(\left[\zeta_i(t)\right]_j - \frac{1}{m}\sum_{i=1}^{m}\left[\zeta_i(0)\right]_j\right)^2\right]$$

$$\mathrm{MSE}(\mathbf{D}, t) = \frac{1}{m}\sum_{i=1}^{m}\mathrm{MSE}_i(\mathbf{D}, t)$$

Proof sketch

- Recall generalized posynomials formed by $+, \times, \div, \max(\cdot)$

$$(1)\ \nu_i(\mathbf{D},t)\ \text{posy.} \Rightarrow \frac{\nu_i(\mathbf{D},t)}{D_i(t)}\ \text{posy.}$$

$$(2)\ k > 0\ \Rightarrow k\ \text{mon.} \Rightarrow \max\left\{\frac{\nu_i(\mathbf{D},t)}{D_i(t)}, k\right\}\ \text{gen. posy.}$$

$$(3)\ \prod \text{gen. posy.} = \text{gen. posy}$$

$$\Rightarrow \prod_{t=0}^{T-1}\prod_{i=1}^{m}\max\left\{\frac{\nu_i(\mathbf{D},t)}{D_i(t)}, k\right\}\ \text{gen. posy.}$$

Key algorithmic contribution:
Search heuristic

Entropy and fixed-length coding

- Entropy coding: $R \approx H(z) \in \mathbb{R}$
- Fixed-length coding: $R \in \mathbb{Z}_{>0}$
- Question: How to deal with integer constraint?

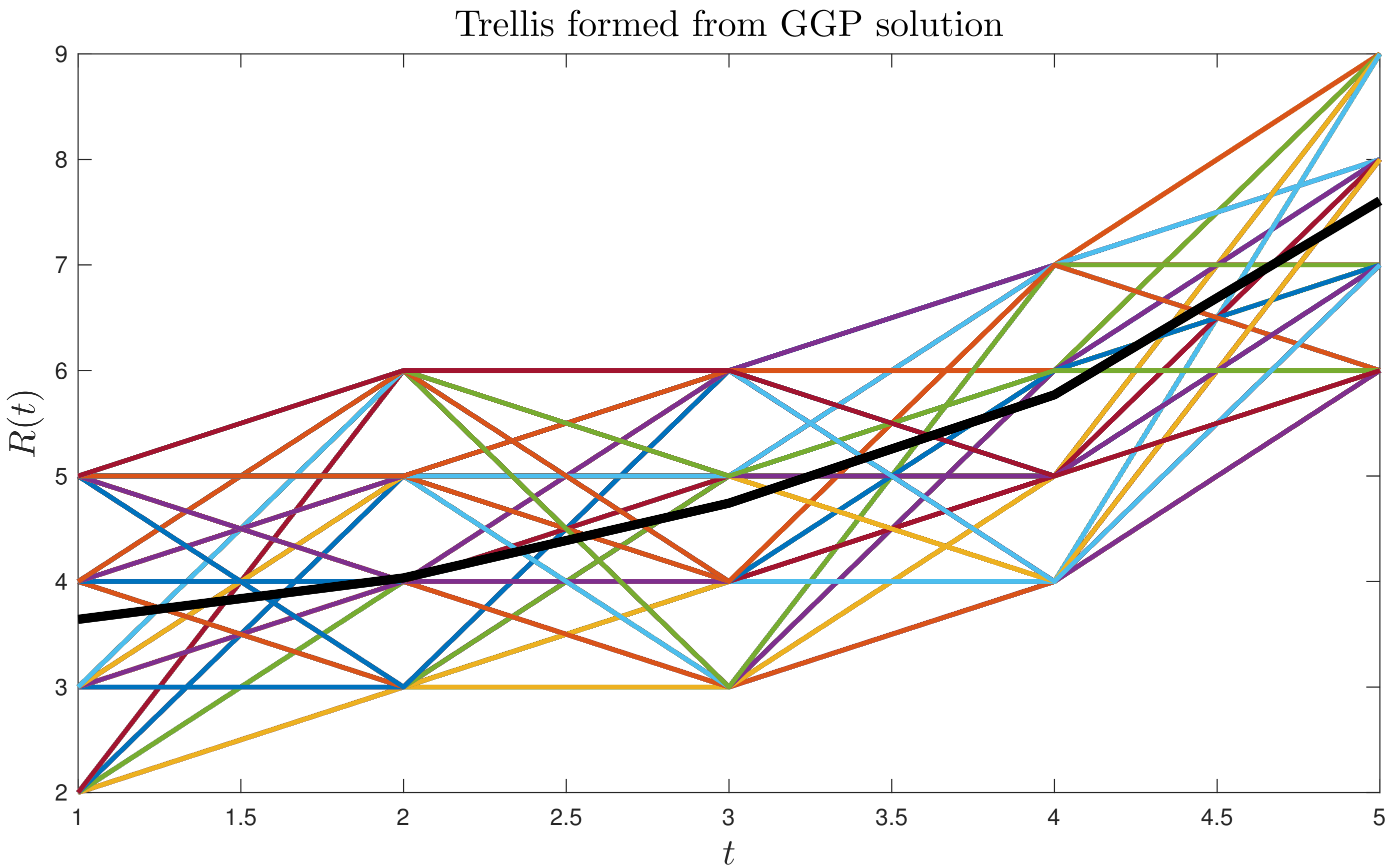
Equal-distortion simplification

$$R_i(t) \neq R_j(s) \Rightarrow \text{search space too large!}$$

$$\begin{aligned} \underset{\mathbf{D}}{\text{minimize}} \quad & \prod_{t=0}^{T-1} \prod_{i=1}^{m} \max\left\{ \frac{\nu_i(\mathbf{D}, t)}{D(t)}, k \right\}, \\ \text{subject to} \quad & \text{MSE}(\mathbf{D}, T) \leq \text{MSE}^*, \\ & D(t) > 0, \quad \forall t \end{aligned}$$

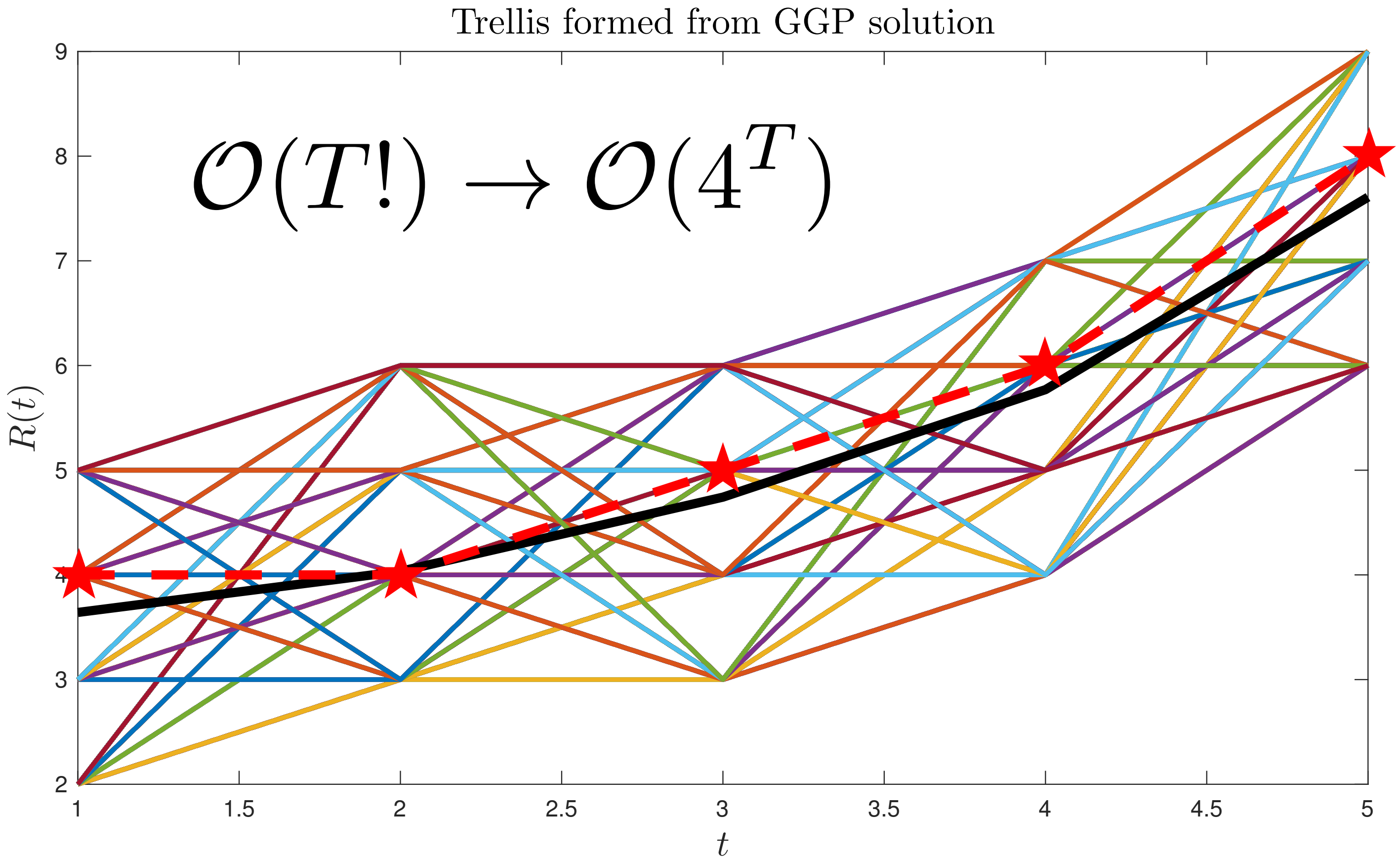
Search heuristic

- Key idea: limit size of search space using GGP solution as a starting point



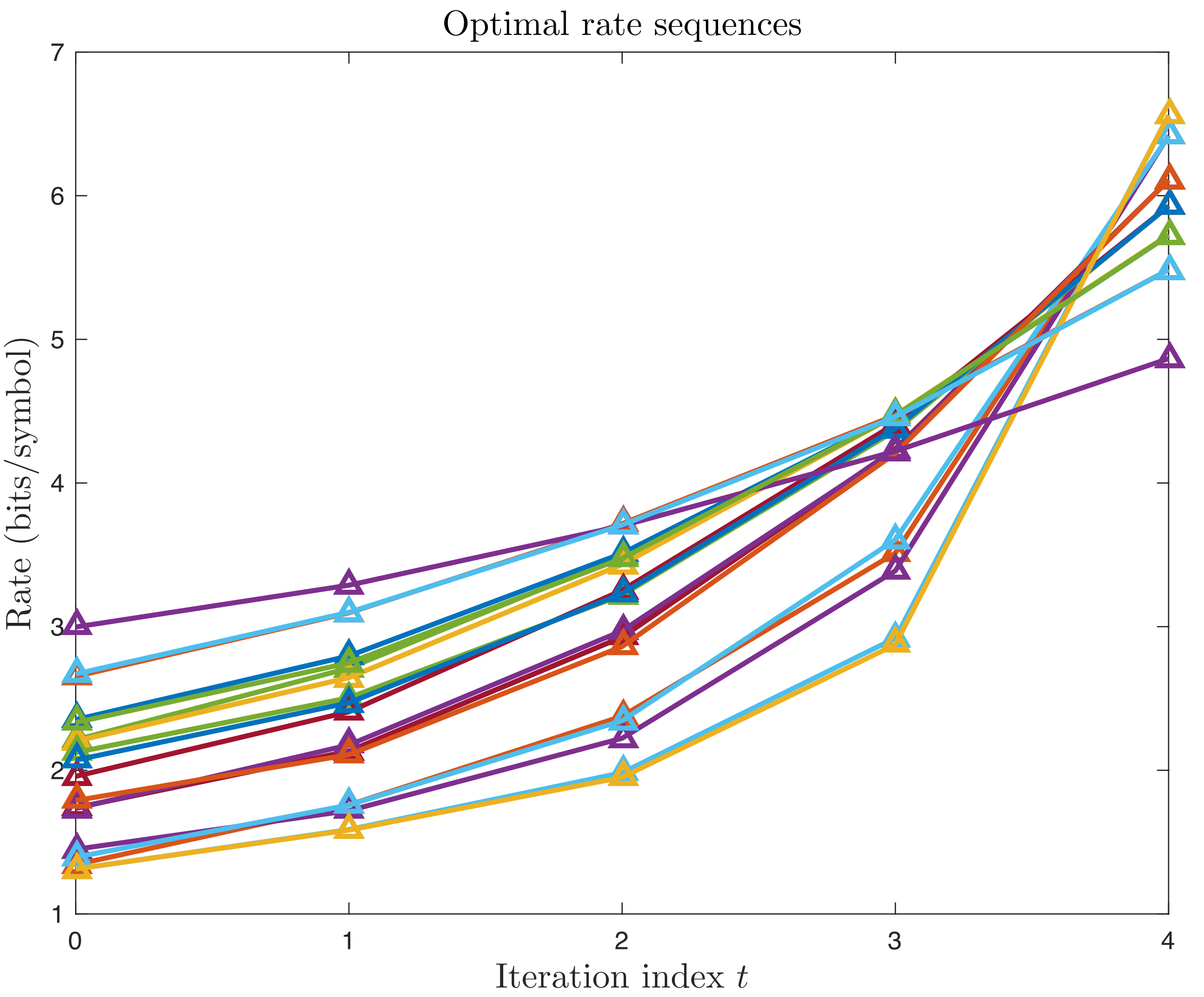
Search heuristic

- Key idea: limit size of search space using GGP solution as a starting point

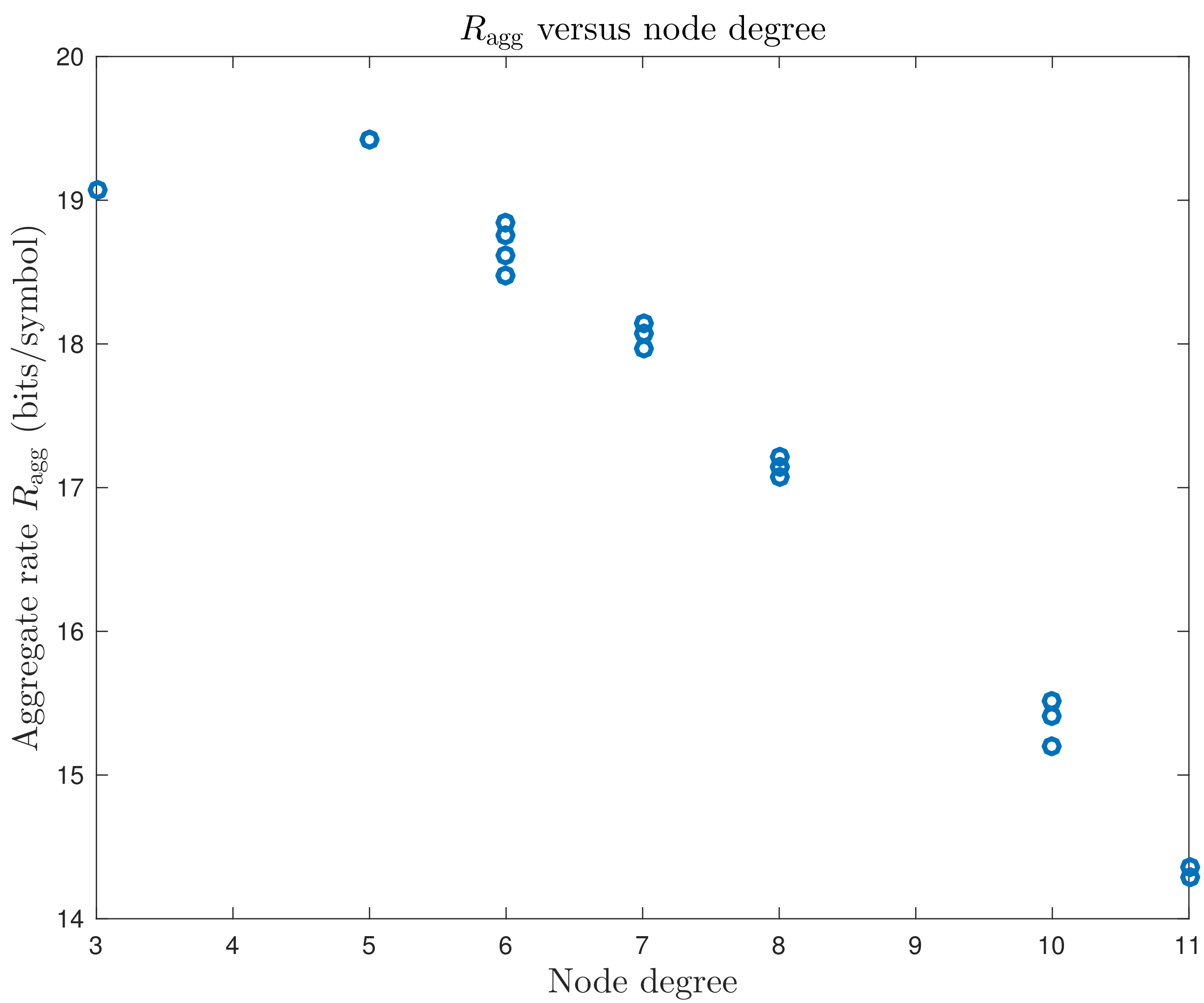


Numerical results

Numerical results – optimal rates



Numerical results – node degree



Lossless case and excess MSE

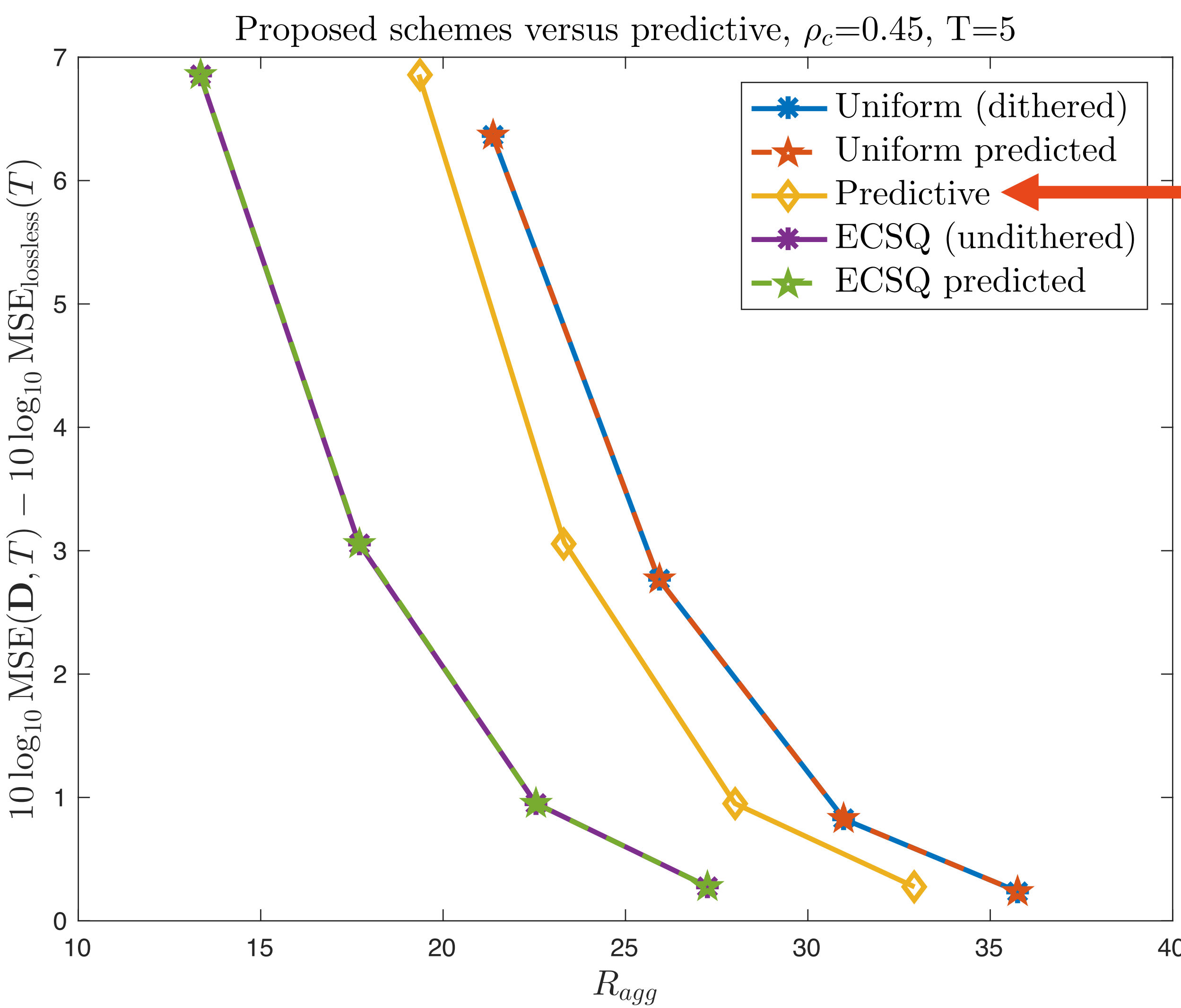
- Lossless case: no distortion (best you can do)

$$\mathrm{MSE}_{\mathrm{lossless}}(t) = \mathrm{MSE}(\mathbf{D}, t)\Big|_{\mathbf{D}=\mathbf{0}}$$

- Excess MSE (EMSE) defined as

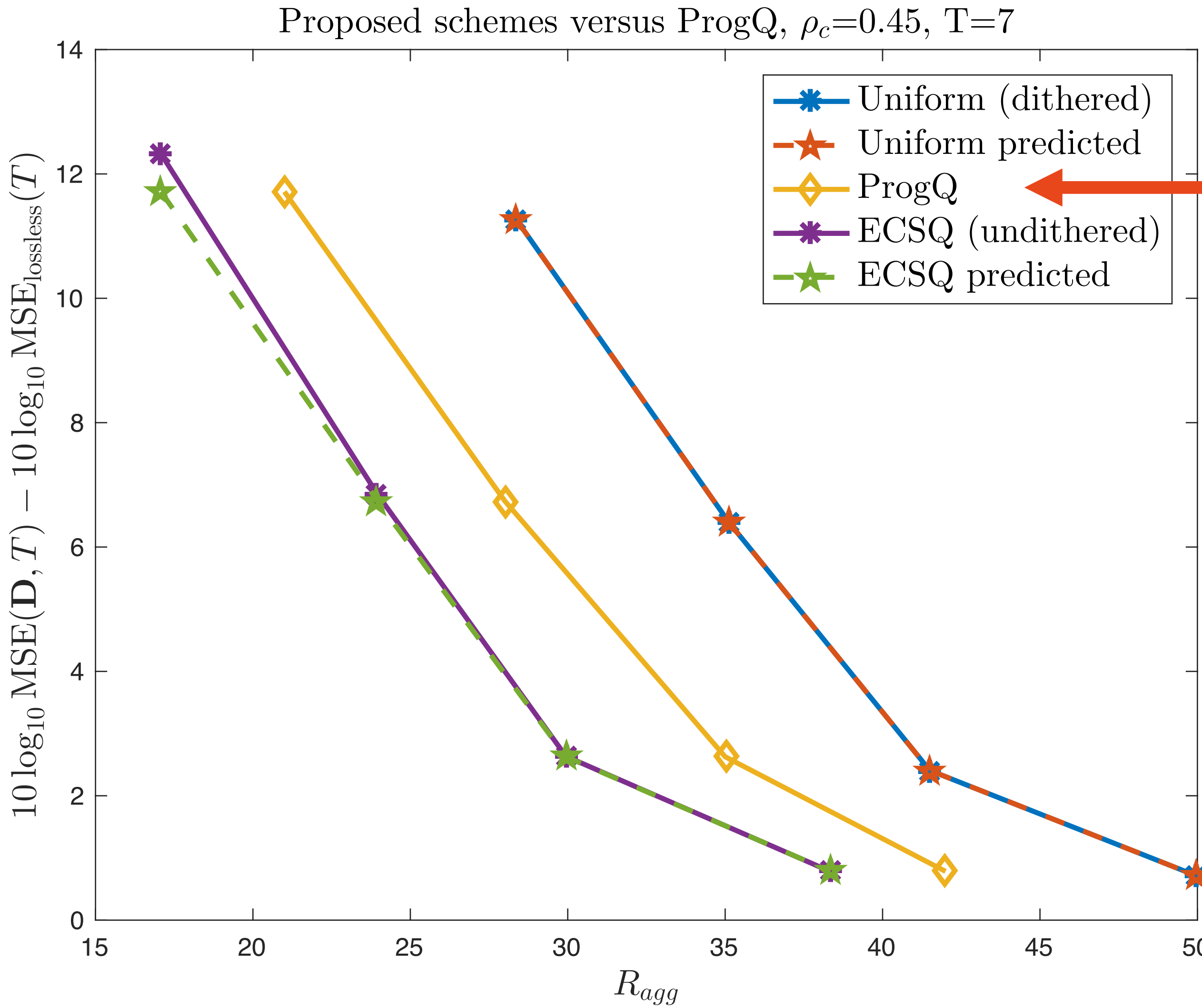
$$\mathrm{EMSE(T)} = 10\log_{10}\mathrm{MSE}(\mathbf{D}, T) - 10\log_{10}\mathrm{MSE}_{\mathrm{lossless}}(T)$$

Numerical results – prior art comparison



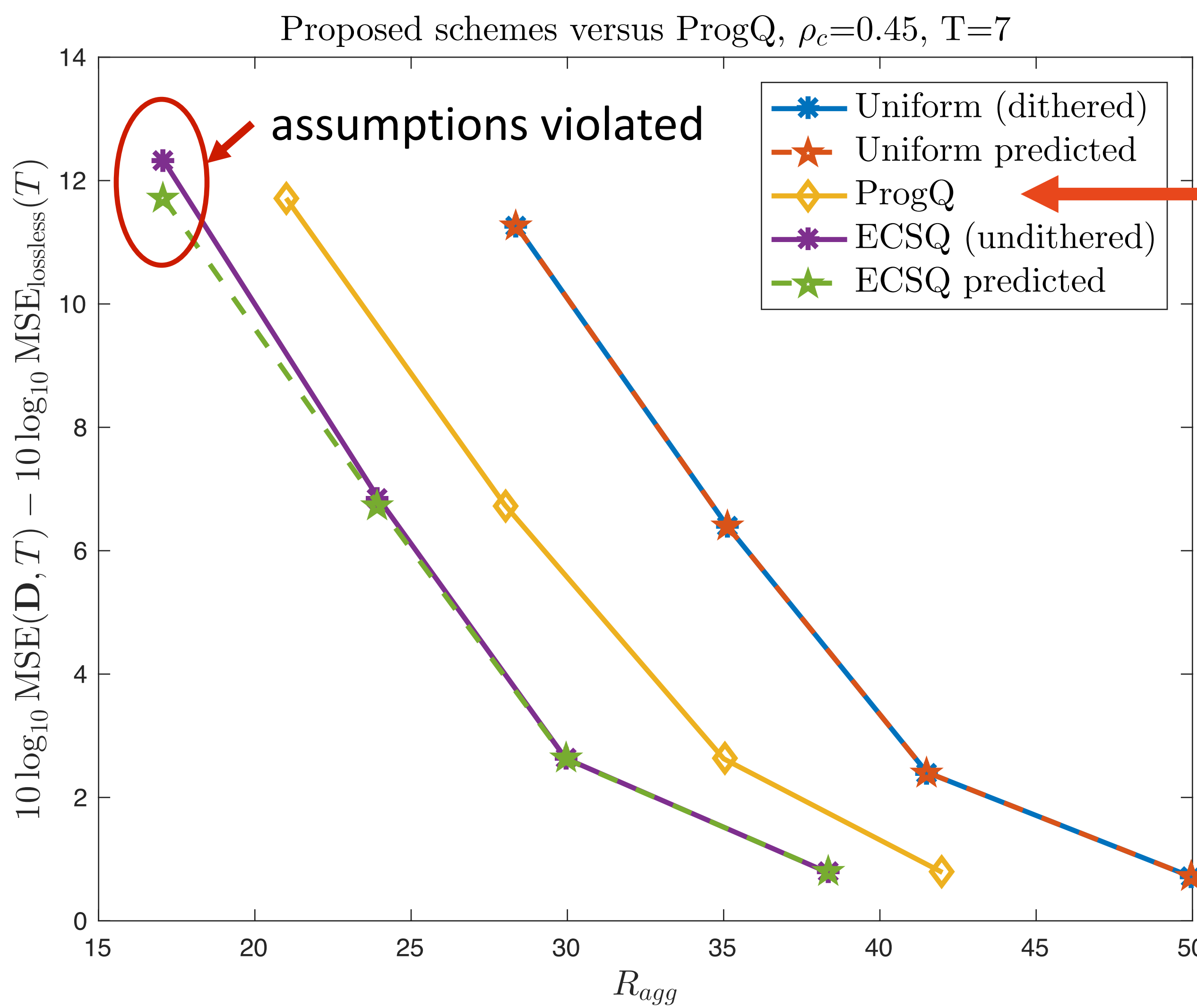
Prior art

Numerical results – prior art comparison



Prior art

Numerical results – prior art comparison



Prior art

Conclusions

Summary

- Proved coding rate optimization in consensus
  - Solvable by GGP (some assumptions)
- Presented search heuristic for fixed-rate coding
- Scalar quantizer simulations & comparison to prior art

Future work

- Incorporate differential/predictive coding: can only do better!
- Application to distributed algorithms such as cloud K-SVD [Raja & Bajwa '16]
- RD for consensus…?

Thank you!

References

1. M. H. DeGroot, “Reaching a consensus,” *J. Amer. Statist. Assoc.*, vol. 69, no. 345, pp. 118–121, 1974.
2. V. Borkar and P. Varaiya, “Asymptotic agreement in distributed estimation,” *IEEE Trans. Autom. Control*, vol. AC-27, no. 3, pp. 650–655, Jun. 1982.
3. J. N. Tsitsiklis, “Problems in decentralized decision making and computation,” PhD thesis, Massachussetts Inst. Technol., Cambridge, MA, Nov. 1984.
4. J. Tsitsiklis, D. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
5. Y. Huang and Y. Hua, “On energy for progressive and consensus estimation in multihop sensor networks,” *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3863–3875, Aug. 2011.
6. A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1993.
7. T. Berger, *Rate Distortion Theory: Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
8. T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley-Interscience, 1991.

References

9. P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus on networks with quantized communication," *Int. J. Robust Nonlinear Control*, vol. 19, no. 16, pp. 1787–1816, Nov. 2008.
10. M. E. Yildiz and A. Scaglione, "Coding with side information for rate-constrained consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3753–3764, Aug. 2008.
11. ——, "Limiting rate behavior and rate allocation strategies for average consensus prob- lems with bounded convergence," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2008, pp. 2717–2720.
12. D. Thanou, E. Kokiopoulou, Y. Pu, and P. Frossard, "Distributed average consensus with quantization refinement," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 194–205, Jan. 2013.
13. Y. Yang, P. Grover, and S. Kar, "Rate distortion for lossy in-network function computation: Information dissipation and sequential reverse water-filling," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, pp. 1–29, May 2017.
14. H.-I. Su and A. El Gamal, "Distributed lossy averaging," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3422–3437, Jul. 2010.
15. H. Raja and W. U. Bajwa, "Cloud K-SVD: A collaborative dictionary learning algorithm for big, distributed data," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 173–188, Jan. 2016.