Accelerated Consensus via Min-Sum Splitting

Patrick Rebeschini[1] **and Sekhar Tatikonda**[2]

[1]patrick.rebeschini@stats.ox.ac.uk, University of Oxford [2]sekhar.tatikonda@yale.edu, Yale University

UNIVERSITY OF OXFORD DEPARTMENT OF STATISTICS

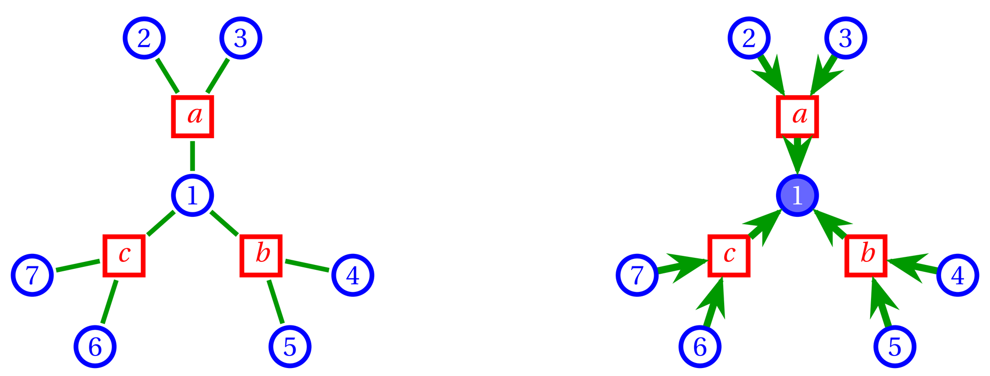
YALE INSTITUTE FOR NETWORK SCIENCE

1. Min-Sum

Min-Sum is a distributed algorithm to optimize a sum of functions.

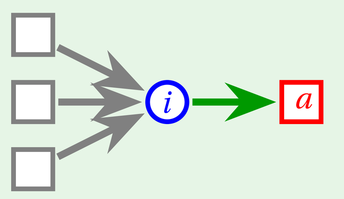
$$\min_x \sum_a f_a(x_{\partial a})$$

The algorithm is exact on trees, it corresponds to dynamic programming.

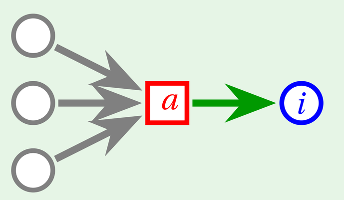


$$\min_x \Big(f_a(x_1,x_2,x_3) + f_b(x_1,x_4,x_5) + f_c(x_1,x_6,x_7) \Big)$$
$$= \min_{x_1} \Big(\underbrace{\min_{x_2,x_3} f_a(x_1,x_2,x_3)}_{\mu_{a\to 1}(x_1)} + \underbrace{\min_{x_4,x_5} f_b(x_1,x_4,x_5)}_{\mu_{b\to 1}(x_1)} + \underbrace{\min_{x_6,x_7} f_c(x_1,x_6,x_7)}_{\mu_{c\to 1}(x_1)} \Big)$$

Min-Sum



$$\mu^{(t)}_{i\to a}(x_i) = f_i(x_i) + \sum_{b\in\partial i\setminus a} \mu^{(t-1)}_{b\to i}(x_i)$$



$$\mu^{(t)}_{a\to i}(x_i) = \min_{x_{\partial a\setminus i}} \Big(\sum_{j\in\partial a\setminus i} \mu^{(t-1)}_{j\to a}(x_j) + f_a(x_{\partial a\setminus i}, x_i) \Big)$$

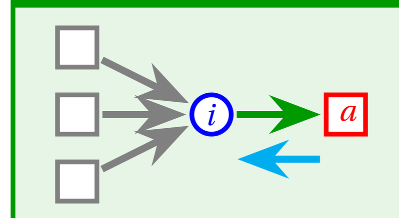
- Messages can be exchanged on *any* graphs, even with loops.
- In general, convergence and correctness are not guaranteed.

2. Min-Sum Splitting

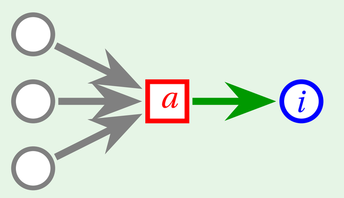
= Min-Sum applied to a reparametrization of the objective function.[4]

$$\min_x \sum_a \sum_{k=1}^{\Gamma_a} \frac{f_{a,k}(x_{\partial a})}{\Gamma_a} \qquad f_{a,k} := f_a$$

Min-Sum Splitting



$$\mu^{(t)}_{i\to a}(x_i) = f_i(x_i) + \sum_{b\in\partial i\setminus a} \Gamma_b \mu^{(t-1)}_{b\to i}(x_i) + (\Gamma_a - 1)\mu^{(t-1)}_{a\to i}(x_i)$$

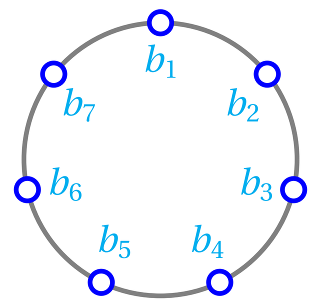


$$\mu^{(t)}_{a\to i}(x_i) = \min_{x_{\partial a\setminus i}} \Big(\sum_{j\in\partial a\setminus i} \mu^{(t-1)}_{j\to a}(x_j) + \frac{f_a(x_{\partial a\setminus i}, x_i)}{\Gamma_a} \Big)$$

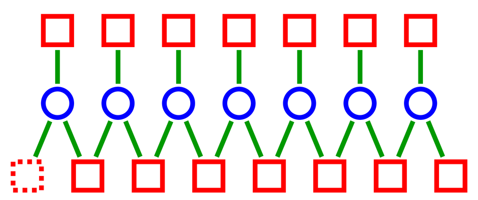
Q. Can we tune directionality to get convergence, correctness, and possibly faster convergence rate?

3. Consensus: Network Averaging

Consensus is a fundamental primitive in distributed optimization.



$$\min. \sum_{v\in V} (x_v - b_v)^2$$
$$\text{sub. to } x_v = x_w, \{v,w\}\in E$$



- Classical algorithms are linear systems: $x^{(0)} = b$ and $x^{(t)} = W x^{(t-1)}$.
- Necessary and sufficient[6]: $W\mathbf{1} = \mathbf{1}, \mathbf{1}^T W = \mathbf{1}^T, \lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$.
- Common choice is Metropolis-Hastings: $W^{MH}_{ij} = \begin{cases} 1/(2d_{\max}) & \text{if } \{i,j\}\in E \\ 1-d_i/(2d_{\max}) & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$
- Rate of convergence is controlled by $\rho(W - \mathbf{1}\mathbf{1}^T/n)$.
- $\min\{\rho(W-\mathbf{1}\mathbf{1}^T/n) : W$ symmetrical$\}$ is a convex problem (SDP).
- **Optimal matrix yields slow rate** $O(D^2)$, achieved by W^{MH}.
- Lower-bound: $\Omega(D)$, where D is graph diameter.
- To get fast rates, two approaches have been developed independently: **Lifted Markov chains**[5] and **multi-step gradient methods**[1].

Q. Can we get fast rates with Min-Sum?

4. Min-Sum Splitting for Consensus

Min-Sum does **not** converge.[3] **Min-Sum Splitting does converge.**

$$\hat{h}_{(w,v)} := b_w \qquad \widehat{K}_{(w,v)(z,u)} := \begin{cases} \Gamma_{zw} & \text{if } u=w, z\in\mathcal{N}(w)\setminus\{v\} \\ \Gamma_{vw}-1 & \text{if } u=w, z=v \\ 0 & \text{otherwise} \end{cases}$$

ALGORITHM 1. Min-Sum Splitting for Consensus

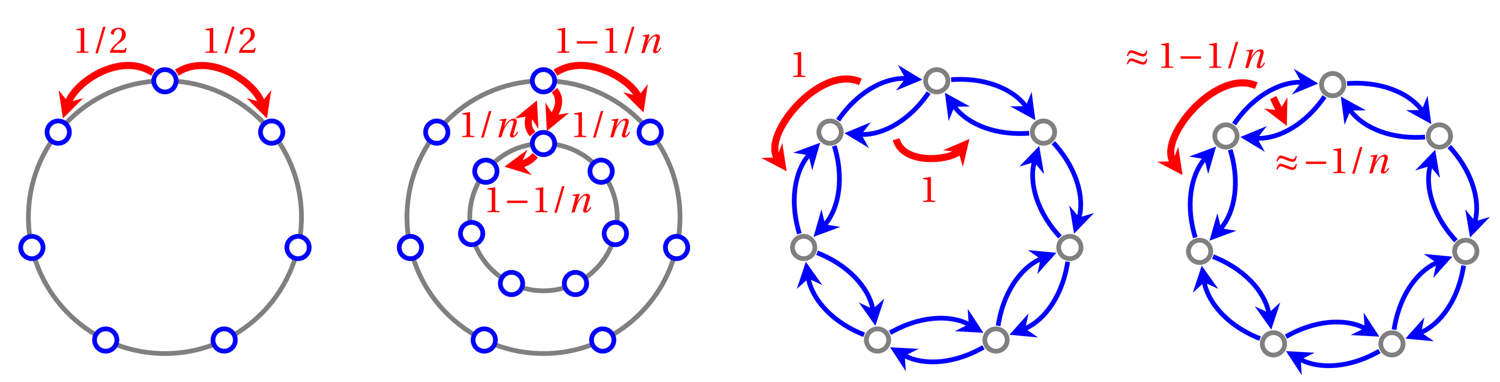
Input: Initial messages $R^{(0)}_{(v,w)}, r^{(0)}_{(v,w)}$; symmetric $\Gamma\in\mathbb{R}^{V\times V}$.

for $s\in\{1,\dots,t\}$ **do**

$\hat{R}^{(s)} = \mathbf{1} + \widehat{K}\hat{R}^{(s-1)}$; $\qquad \hat{r}^{(s)} = \hat{h} + \widehat{K}\hat{r}^{(s-1)}$;

Output: $x^{(t)}_v := \frac{b_v + \sum_{w\in\mathcal{N}(v)} \Gamma_{wv}\hat{r}^{(t)}_{wv}}{1+\sum_{w\in\mathcal{N}(v)} \Gamma_{wv}\hat{R}^{(t)}_{wv}}, v\in V$.

KEY: Properly tune the directionality of messages.



Markov chain | Lifted Markov chain | Min-Sum | Min-Sum Splitting

5. Accelerated rate of convergence

Theorem

- Let $W\in\mathbb{R}^{V\times V}$ be symmetric, $W\mathbf{1}=\mathbf{1}$ and $\rho_W := \rho(W-\mathbf{1}\mathbf{1}^T/n) < 1$.
- Let $\Gamma = \gamma W$, with $\gamma = 2/(1+\sqrt{1-\rho_W^2})$.
- Define: $K := \begin{pmatrix} \gamma W & I \\ (1-\gamma)I & 0 \end{pmatrix}$, $K^\infty := \frac{1}{(2-\gamma)n}\begin{pmatrix} \mathbf{1}\mathbf{1}^T & \mathbf{1}\mathbf{1}^T \\ (1-\gamma)\mathbf{1}\mathbf{1}^T & (1-\gamma)\mathbf{1}\mathbf{1}^T \end{pmatrix}$.

Then, $\boxed{\|x^{(t)} - \bar{b}\mathbf{1}\| \le \frac{4\sqrt{2|V|}}{2-\gamma}\|(K-K^\infty)^t\|}$, where $\bar{b} := \frac{1}{|V|}\sum_{v\in V} b_v$.

Asymptotic convergence rate:

$$\rho_K := \rho(K-K^\infty) = \lim_{n\to\infty}\|(K-K^\infty)^n\|^{1/n} = \sqrt{\frac{1-\sqrt{1-\rho_W^2}}{1+\sqrt{1-\rho_W^2}}} < \rho_W < 1,$$

and $\frac{1}{2}\sqrt{1/(1-\rho_W)} \le 1/(1-\rho_K) \le \sqrt{1/(1-\rho_W)}$.

- Same rate as shift-register methods.[2]
- Asymptotic convergence time $O(D\log D)$ for cycles and grids.

KEY: Tune directionality using global information.

6. Contributions

- **Directionality embedded in Belief Propagation protocols can be tuned to yield convergence and accelerated rates.**
- Connection of Min-Sum schemes with lifted Markov chains techniques and **multi-step gradient methods**:

$$\begin{pmatrix} x^{(t)} \\ x^{(t-1)} \end{pmatrix} = K \begin{pmatrix} x^{(t-1)} \\ x^{(t-2)} \end{pmatrix}$$

- New proof technique based on the introduction of an auxiliary process to track the evolution of Min-Sum schemes on the nodes.
- **Quasi-optimal rate** $O(D\log D)$ for the network averaging problem in cycles and grids, improving previous rates for Min-Sum with soft barrier (Consensus Propagation[3]) ($\Theta(D^{2(d-1)/d})$ for $d/2$ dim. grids).

Acknowledgements. This work was partially supported by the NSF under Grant EECS-1609484.

References

1. E. Ghadimi, I. Shames, and M. Johansson. Multi-step gradient methods for networked optimization. *IEEE Transactions on Signal Processing*, 61(21):5417–5429, 2013. 2. Ji Liu, Brian D.O. Anderson, Ming Cao, and A. Stephen Morse. Analysis of accelerated gossip algorithms. *Automatica*, 49(4):873–883, 2013. 3. C. C. Moallemi and B. Van Roy. Consensus propagation. *IEEE Transactions on Information Theory*, 52(11):4753–4766, 2006. 4. N. Ruozzi and S. Tatikonda. Message-passing algorithms: Reparameterizations and splittings. *IEEE Transactions on Information Theory*, 59(9):5860–5881, 2013. 5. Devavrat Shah. Gossip algorithms. *Foundations and Trends® in Networking*, 3(1):1–125, 2009. 6. Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65 – 78, 2004.