

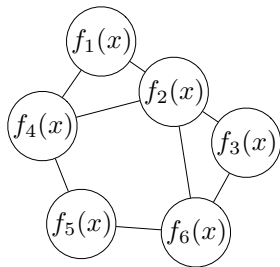
Distributed Dual Averaging in Networks

John C. Duchi

Alekh Agarwal

Martin J. Wainwright

University of California, Berkeley



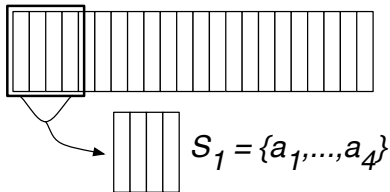
Setup: Function f_i at node i

Goal: Solve

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{n} \sum_{i=1}^n f_i(x)$$

Example: Data $\{a_j\}_{j=1}^N$,
shard onto processors

$$f_i(x) = \sum_{j \in \mathcal{S}_i} \ell(\langle x, a_j \rangle)$$



Distributed Dual Averaging Algorithm

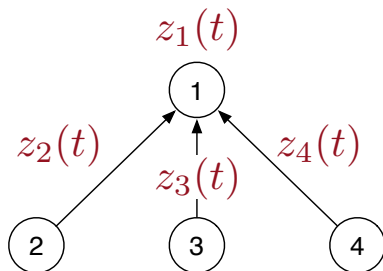
At node i : Primal $x_i(t) \in \mathcal{X}$, dual $z_i(t) \in \mathbb{R}^d$.

Dual Update:

$$z_i(t+1) = \sum_j P_{ji} z_j(t) + \nabla f_i(x_i(t))$$

Primal Update:

$$x_i(t+1) = \arg \min_{x \in \mathcal{X}} \left\{ \alpha \langle z_i(t+1), x \rangle + \frac{1}{2} \|x\|_2^2 \right\}.$$



Main Results

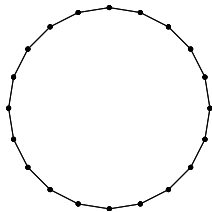
Theorem: At each node i , optimization error at most

$$f(x_i(T)) - f(x^*) = \mathcal{O} \left(\frac{1}{\sqrt{T} \sqrt{1 - \sigma_2(P)}} \right).$$

Dependence on convergence time of **random walk** on graph.

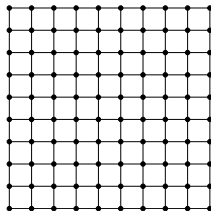
Examples:

Cycle



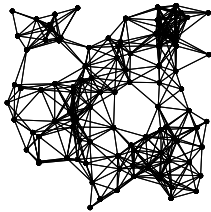
$$\frac{n}{\sqrt{T}}$$

Grid



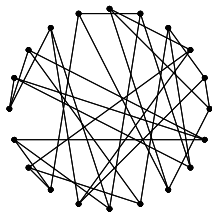
$$\frac{\sqrt{n}}{\sqrt{T}}$$

Geometric



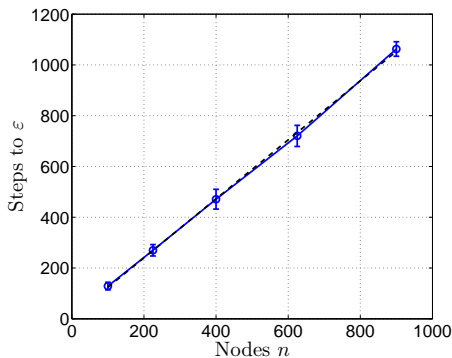
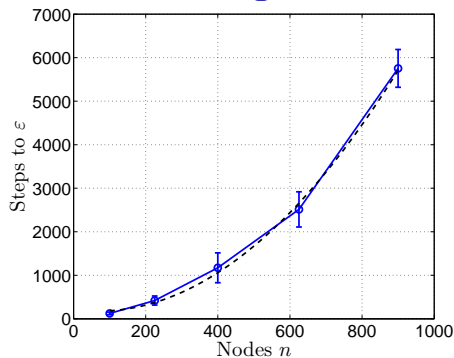
$$\sqrt{\frac{n}{\log n}} \frac{1}{\sqrt{T}}$$

Expander



$$\frac{\sqrt{\log n}}{\sqrt{T}}$$

Network scaling and extensions



- ▶ **Random Communication:** robustness to edge failures, high probability convergence even with randomness
- ▶ **Stochastic optimization:** streaming and online applications, see stochastic versions of each f_i
- ▶ **Simulations and lower bound:** match theoretical predictions