Accelerated Min-Sum for Consensus

Patrick Rebeschini

(joint work with Sekhar Tatikonda)

Yale Institute For
Network Science

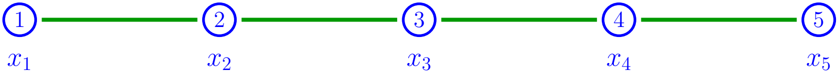
Large-Scale and Distributed Optimization: Workshop Program
LCCC, Lund University

June 14, 2017

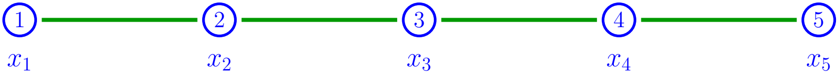
(NSF Grant: *Locality in Network Optimization*, Award no. 1609484, ECCS)

Min-Sum

Min-sum: path graph

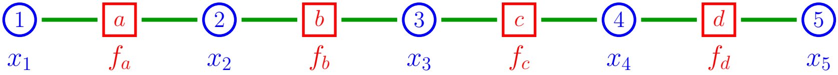


Min-sum: path graph



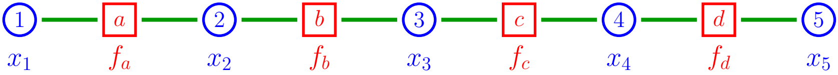
$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \min_{x_5} \underbrace{\left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + f_d(x_4, x_5) \right)}_{f(x)}$$

Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \min_{x_5} \underbrace{\left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + f_d(x_4, x_5) \right)}_{f(x)}$$

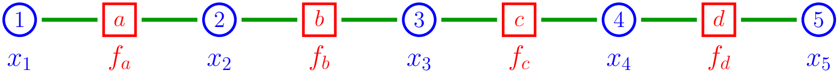
Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \min_{x_5} \underbrace{\left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + f_d(x_4, x_5) \right)}_{f(x)}$$

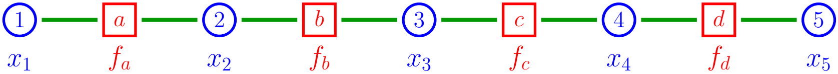
$x_i \in \{0, 1\}$ **naive algorithm** $O(2^n)$

Min-sum: path graph



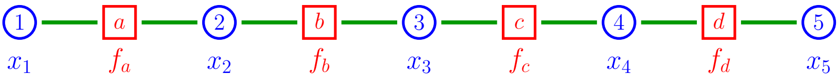
$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \min_{x_5} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + f_d(x_4, x_5) \right)$$

Min-sum: path graph



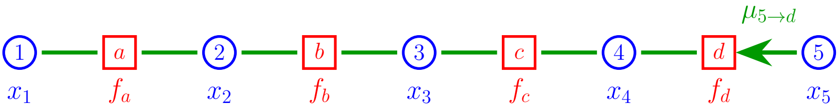
$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + \boxed{\min_{x_5}} f_d(x_4, x_5) \right)$$

Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + \boxed{\min_{x_5} f_d(x_4, x_5)} \right)$$

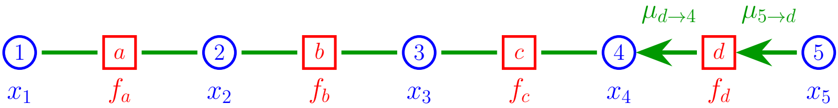
Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + \boxed{\min_{x_5} f_d(x_4, x_5)} \right)$$

$\mu_{5\to d} = 0$

Min-sum: path graph

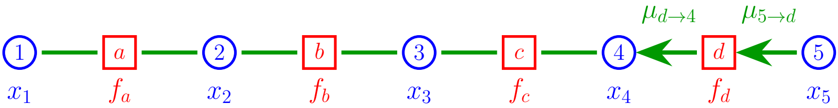


$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + \boxed{\min_{x_5} f_d(x_4, x_5)} \right)$$

$\mu_{5\to d} = 0$

$$\mu_{d\to 4}(\ \star\ ) = \min_{x_5} \big(f_d(\ \star\ , x_5) + \mu_{5\to d}(x_5) \big)$$

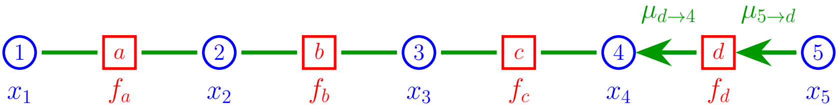
Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + \mu_{d\to 4}(x_4) \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

Min-sum: path graph

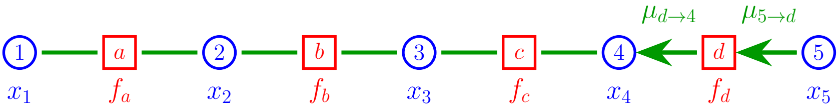


$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \Big(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + \mu_{d\to 4}(x_4) \Big)$$

$\mu_{5\to d} = 0$

$$\mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

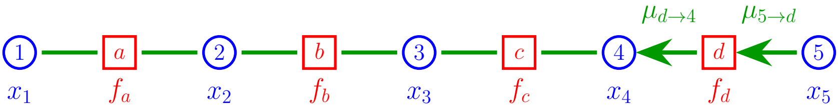
Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + \boxed{\min_{x_4}} \left(f_c(x_3, x_4) + \mu_{d\to4}(x_4) \right) \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to4}(\ \star\ ) = \min_{x_5} \left(f_d(\ \star\ , x_5) + \mu_{5\to d}(x_5) \right)$$

Min-sum: path graph

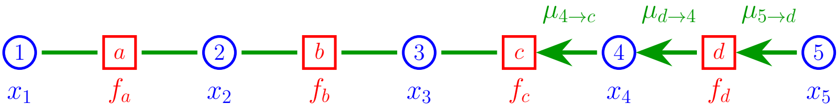


$$\min_{x_1} \min_{x_2} \min_{x_3} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + \boxed{\min_{x_4} \Big(f_c(x_3, x_4) + \mu_{d\to 4}(x_4) \Big)} \right)$$

$\mu_{5\to d} = 0$

$$\mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

Min-sum: path graph

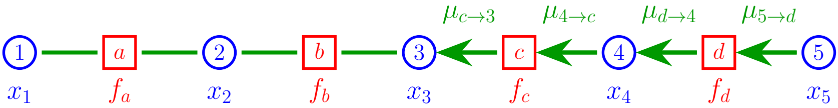


$$\min_{x_1} \min_{x_2} \min_{x_3} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + \boxed{\min_{x_4} \left(f_c(x_3, x_4) + \mu_{d\to 4}(x_4) \right)} \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

$$\mu_{4\to c} = \mu_{d\to 4}$$

Min-sum: path graph



$$\min_{x_1} \min_{x_2} \min_{x_3} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + \boxed{\min_{x_4} \Big(f_c(x_3, x_4) + \mu_{d\to 4}(x_4) \Big)} \right)$$

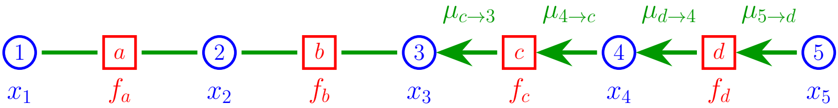
$\mu_{5\to d} = 0$

$$\mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

$\mu_{4\to c} = \mu_{d\to 4}$

$$\mu_{c\to 3}(\,\star\,) = \min_{x_4} \big(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \big)$$

Min-sum: path graph

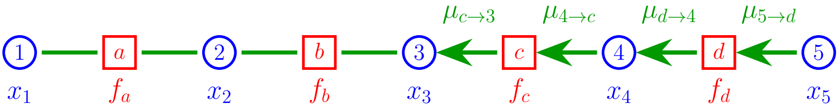


$$\min_{x_1} \min_{x_2} \min_{x_3} \left(f_a(x_1, x_2) + f_b(x_2, x_3) + \mu_{c\to 3}(x_3) \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\,\star\,) = \min_{x_4} \big(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \big)$$

Min-sum: path graph



$$\min_{x_1} \min_{x_2} \boxed{\min_{x_3}} \Big(f_a(x_1, x_2) + f_b(x_2, x_3) + \mu_{c\to 3}(x_3) \Big)$$

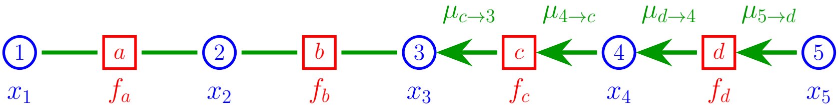
$\mu_{5\to d} = 0$

$$\mu_{d\to 4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

$\mu_{4\to c} = \mu_{d\to 4}$

$$\mu_{c\to 3}(\,\star\,) = \min_{x_4} \big(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \big)$$

Min-sum: path graph

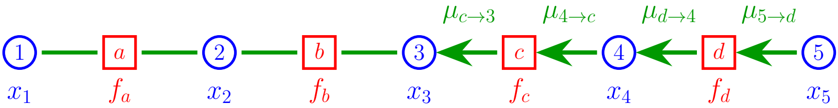


$$\min_{x_1} \min_{x_2} \left(f_a(x_1, x_2) + \min_{x_3} \left(f_b(x_2, x_3) + \mu_{c\to3}(x_3) \right) \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to4}(\star) = \min_{x_5} \left(f_d(\star, x_5) + \mu_{5\to d}(x_5) \right)$$

$$\mu_{4\to c} = \mu_{d\to4} \qquad \mu_{c\to3}(\star) = \min_{x_4} \left(f_c(\star, x_4) + \mu_{4\to c}(x_4) \right)$$

Min-sum: path graph

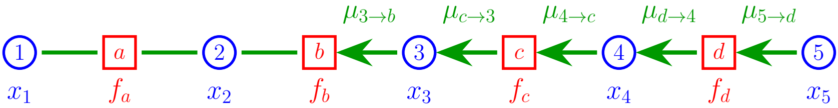


$$\min_{x_1} \min_{x_2} \left(f_a(x_1, x_2) + \boxed{\min_{x_3} \left(f_b(x_2, x_3) + \mu_{c\to 3}(x_3) \right)} \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\,\star\,) = \min_{x_5} \left(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \right)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\,\star\,) = \min_{x_4} \left(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \right)$$

Min-sum: path graph



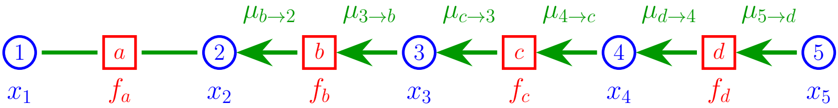
$$\min_{x_1} \min_{x_2} \left(f_a(x_1, x_2) + \boxed{\min_{x_3} \left(f_b(x_2, x_3) + \mu_{c\to 3}(x_3) \right)} \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\,\star\,) = \min_{x_5} \left(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \right)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\,\star\,) = \min_{x_4} \left(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \right)$$

$$\mu_{3\to b} = \mu_{c\to 3}$$

Min-sum: path graph



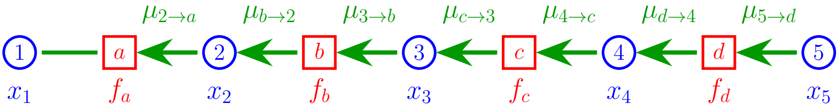
$$\min_{x_1} \min_{x_2} \left(f_a(x_1, x_2) + \boxed{\min_{x_3} \left(f_b(x_2, x_3) + \mu_{c\to 3}(x_3) \right)} \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\,\star\,) = \min_{x_5} \left(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \right)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\,\star\,) = \min_{x_4} \left(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \right)$$

$$\mu_{3\to b} = \mu_{c\to 3} \qquad \mu_{b\to 2}(\,\star\,) = \min_{x_3} \left(f_b(\,\star\,, x_3) + \mu_{c\to 3}(x_3) \right)$$

Min-sum: path graph



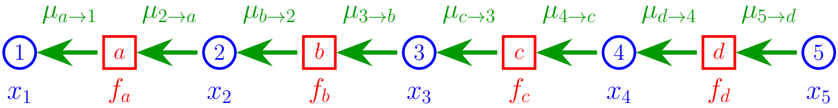
$$\min_{x_1} \min_{x_2} \left(f_a(x_1, x_2) + \mu_{b\to 2}(x_2) \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\star) = \min_{x_5} \left(f_d(\star, x_5) + \mu_{5\to d}(x_5) \right)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\star) = \min_{x_4} \left(f_c(\star, x_4) + \mu_{4\to c}(x_4) \right)$$

$$\mu_{3\to b} = \mu_{c\to 3} \qquad \mu_{b\to 2}(\star) = \min_{x_3} \left(f_b(\star, x_3) + \mu_{c\to 3}(x_3) \right)$$

Min-sum: path graph



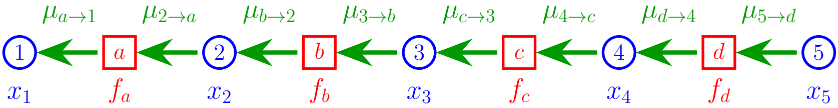
$$\min_{x_1} \boxed{\min_{x_2}} \left(f_a(x_1, x_2) + \mu_{b\to 2}(x_2) \right)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\star) = \min_{x_5} \big(f_d(\star, x_5) + \mu_{5\to d}(x_5) \big)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\star) = \min_{x_4} \big(f_c(\star, x_4) + \mu_{4\to c}(x_4) \big)$$

$$\mu_{3\to b} = \mu_{c\to 3} \qquad \mu_{b\to 2}(\star) = \min_{x_3} \big(f_b(\star, x_3) + \mu_{c\to 3}(x_3) \big)$$

Min-sum: path graph



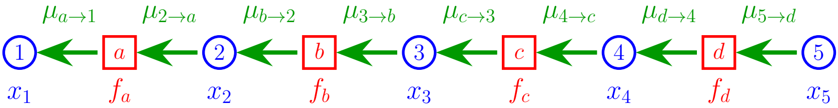
$$\min_{x_1} \boxed{\min_{x_2} \left(f_a(x_1, x_2) + \mu_{b\to2}(x_2) \right)}$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to4}(\,\star\,) = \min_{x_5} \big(f_d(\,\star\,, x_5) + \mu_{5\to d}(x_5) \big)$$

$$\mu_{4\to c} = \mu_{d\to4} \qquad \mu_{c\to3}(\,\star\,) = \min_{x_4} \big(f_c(\,\star\,, x_4) + \mu_{4\to c}(x_4) \big)$$

$$\mu_{3\to b} = \mu_{c\to3} \qquad \mu_{b\to2}(\,\star\,) = \min_{x_3} \big(f_b(\,\star\,, x_3) + \mu_{c\to3}(x_3) \big)$$

Min-sum: path graph



$$\min_{x_1} \boxed{\min_{x_2} \left(f_a(x_1, x_2) + \mu_{b\to 2}(x_2) \right)}$$

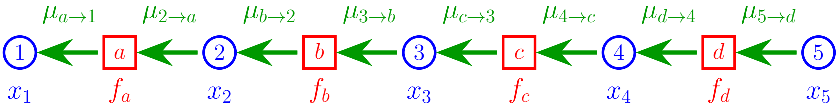
$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\star) = \min_{x_5} \left(f_d(\star, x_5) + \mu_{5\to d}(x_5) \right)$$

$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\star) = \min_{x_4} \left(f_c(\star, x_4) + \mu_{4\to c}(x_4) \right)$$

$$\mu_{3\to b} = \mu_{c\to 3} \qquad \mu_{b\to 2}(\star) = \min_{x_3} \left(f_b(\star, x_3) + \mu_{c\to 3}(x_3) \right)$$

$$\mu_{2\to a} = \mu_{b\to 2}$$

Min-sum: path graph



$$\min_{x_1} \boxed{\min_{x_2} \left(f_a(x_1, x_2) + \mu_{b\to 2}(x_2) \right)}$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to 4}(\star) = \min_{x_5} \left(f_d(\star, x_5) + \mu_{5\to d}(x_5) \right)$$

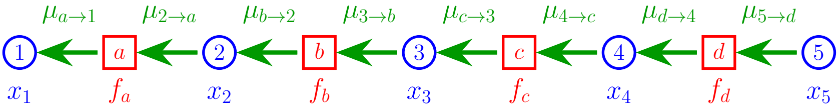
$$\mu_{4\to c} = \mu_{d\to 4} \qquad \mu_{c\to 3}(\star) = \min_{x_4} \left(f_c(\star, x_4) + \mu_{4\to c}(x_4) \right)$$

$$\mu_{3\to b} = \mu_{c\to 3} \qquad \mu_{b\to 2}(\star) = \min_{x_3} \left(f_b(\star, x_3) + \mu_{c\to 3}(x_3) \right)$$

$$\mu_{2\to a} = \mu_{b\to 2} \qquad \mu_{a\to 1}(\star) = \min_{x_2} \left(f_a(\star, x_2) + \mu_{b\to 2}(x_2) \right)$$

Min-sum: path graph

$O(n)$ **Dynamic Programming**



$$\min_{x_1} \mu_{a\to1}(x_1)$$

$$\mu_{5\to d} = 0 \qquad \mu_{d\to4}(\star) = \min_{x_5}\left(f_d(\star, x_5) + \mu_{5\to d}(x_5)\right)$$

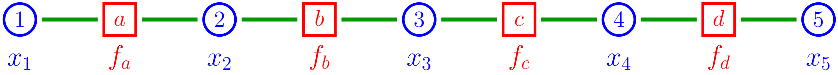
$$\mu_{4\to c} = \mu_{d\to4} \qquad \mu_{c\to3}(\star) = \min_{x_4}\left(f_c(\star, x_4) + \mu_{4\to c}(x_4)\right)$$

$$\mu_{3\to b} = \mu_{c\to3} \qquad \mu_{b\to2}(\star) = \min_{x_3}\left(f_b(\star, x_3) + \mu_{c\to3}(x_3)\right)$$

$$\mu_{2\to a} = \mu_{b\to2} \qquad \mu_{a\to1}(\star) = \min_{x_2}\left(f_a(\star, x_2) + \mu_{b\to2}(x_2)\right)$$

Min-sum: path graph

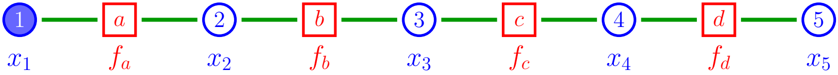
$O(n)$ **Dynamic Programming**



$$\min_{x_1} \min_{x_2} \min_{x_3} \min_{x_4} \min_{x_5} \underbrace{\left(f_a(x_1, x_2) + f_b(x_2, x_3) + f_c(x_3, x_4) + f_d(x_4, x_5) \right)}_{f(x)}$$

Min-sum: path graph

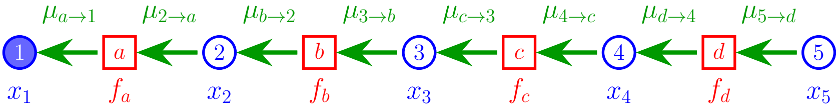
$O(n)$ **Dynamic Programming**



$$\min_{x_1} \min_{x_2} \left(f_a(x_1, x_2) + \min_{x_3} \left(f_b(x_2, x_3) + \min_{x_4} \left(f_c(x_3, x_4) + \min_{x_5} f_d(x_4, x_5) \right) \right) \right)$$

Min-sum: path graph

$O(n)$ **Dynamic Programming**

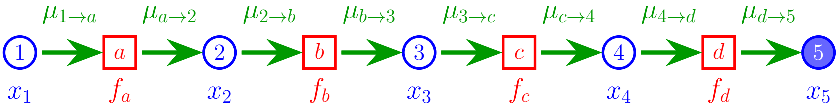


$$\min_{x_1} \underbrace{\min_{x_2}\left(f_a(x_1,x_2) + \underbrace{\min_{x_3}\left(f_b(x_2,x_3)+\underbrace{\min_{x_4}\left(f_c(x_3,x_4)+\underbrace{\min_{x_5} f_d(x_4,x_5)}_{\mu_{d\to4}(x_4)}\right)}_{\mu_{c\to3}(x_3)}\right)}_{\mu_{b\to2}(x_2)}\right)}_{\mu_{a\to1}(x_1)}$$

$$\boxed{= \min_{x_1} \mu_{a\to1}(x_1)}$$

Min-sum: path graph

$O(n)$ **Dynamic Programming**

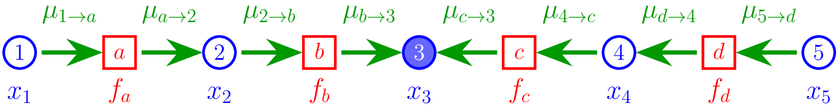


$$\min_{x_5}\underbrace{\min_{x_4}\left(\underbrace{\min_{x_3}\left(\underbrace{\min_{x_2}\left(\underbrace{\min_{x_1} f_a(x_1,x_2)}_{\mu_{a\to 2}(x_2)}+f_b(x_2,x_3)\right)}_{\mu_{b\to 3}(x_3)}+f_c(x_3,x_4)\right)}_{\mu_{c\to 4}(x_4)}+f_d(x_4,x_5)\right)}_{\mu_{d\to 5}(x_5)}$$

$$= \min_{x_5} \mu_{d\to 5}(x_5)$$

Min-sum: path graph

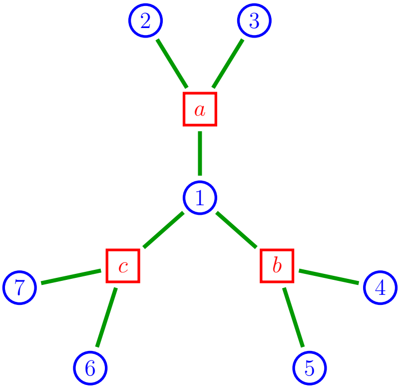
$O(n)$ **Dynamic Programming**



$$\min_{x_3}\Bigg(\underbrace{\min_{x_2}\Big(\underbrace{\min_{x_1} f_a(x_1,x_2)}_{\mu_{a\to 2}(x_2)} + f_b(x_2,x_3)\Big)}_{\mu_{b\to 3}(x_3)} + \underbrace{\min_{x_4}\Big(f_c(x_3,x_4) + \underbrace{\min_{x_5} f_d(x_4,x_5)}_{\mu_{d\to 4}(x_4)}\Big)}_{\mu_{c\to 3}(x_3)}\Bigg)$$

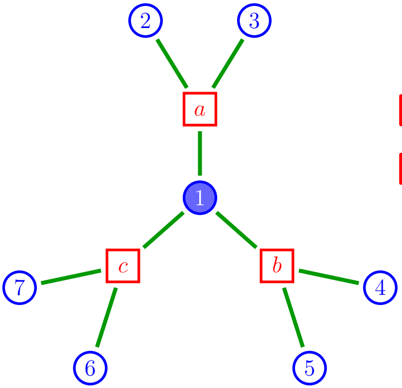
$$\boxed{= \min_{x_3}\Big(\mu_{b\to 3}(x_3) + \mu_{c\to 3}(x_3)\Big)}$$

Min-sum: trees



$$\min_{x}\left(f_a(x_1, x_2, x_3) + f_b(x_1, x_4, x_5) + f_c(x_1, x_6, x_7) \right)$$

Min-sum: trees



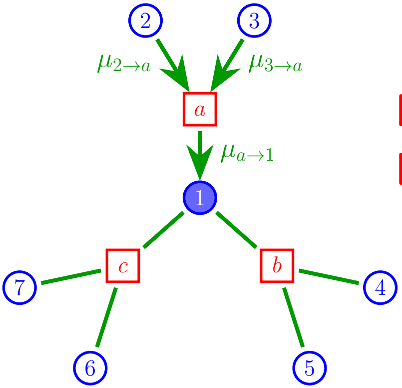
Dynamic Programming

$O(n)$

$$\min_{x}\left(f_a(x_1,x_2,x_3)+f_b(x_1,x_4,x_5)+f_c(x_1,x_6,x_7)\right)$$

$$=\min_{x_1}\left(\min_{x_2,x_3} f_a(x_1,x_2,x_3)+\min_{x_4,x_5} f_b(x_1,x_4,x_5)+\min_{x_6,x_7} f_c(x_1,x_6,x_7)\right)$$

Min-sum: trees



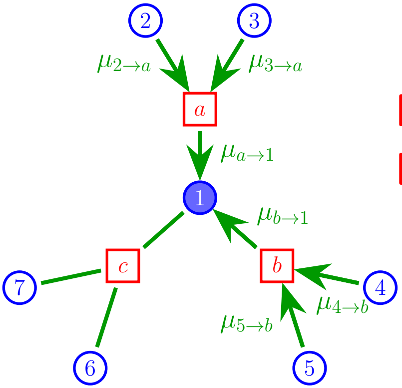
Dynamic Programming

$O(n)$

$$\min_x \left(f_a(x_1, x_2, x_3) + f_b(x_1, x_4, x_5) + f_c(x_1, x_6, x_7) \right)$$

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

Min-sum: trees



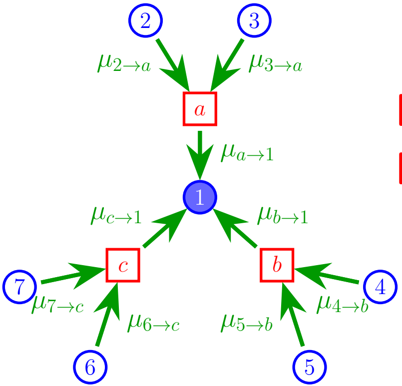
Dynamic Programming

$O(n)$

$$\min_x \left(f_a(x_1, x_2, x_3) + f_b(x_1, x_4, x_5) + f_c(x_1, x_6, x_7) \right)$$

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

Min-sum: trees



Dynamic Programming

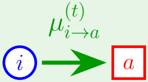
$O(n)$

$$\min_x \left(f_a(x_1,x_2,x_3) + f_b(x_1,x_4,x_5) + f_c(x_1,x_6,x_7) \right)$$

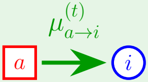
$$= \min_{x_1} \left(\underbrace{\min_{x_2,x_3} f_a(x_1,x_2,x_3)}_{\mu_{a\to 1}(x_1)} + \underbrace{\min_{x_4,x_5} f_b(x_1,x_4,x_5)}_{\mu_{b\to 1}(x_1)} + \underbrace{\min_{x_6,x_7} f_c(x_1,x_6,x_7)}_{\mu_{c\to 1}(x_1)} \right)$$

Messages

variable → factor

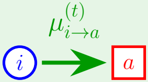


factor → variable



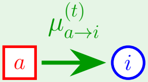
Messages

variable $\rightarrow$ factor



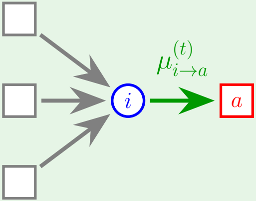
$$\mu_{i\to a}^{(t)}(x_i) =$$

factor $\rightarrow$ variable



Messages

variable $\rightarrow$ factor



$\mu_{i\to a}^{(t)}$

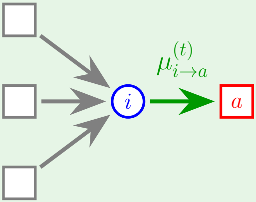
$\mu_{i\to a}^{(t)}(x_i) =$

factor $\rightarrow$ variable

$\mu_{a\to i}^{(t)}$

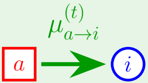
Messages

variable → factor



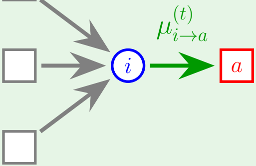
$$\mu^{(t)}_{i\to a}(x_i) = \sum_{b\in\partial i\setminus a} \mu^{(t-1)}_{b\to i}(x_i)$$

factor → variable



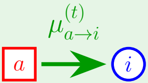
Messages

variable → factor



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

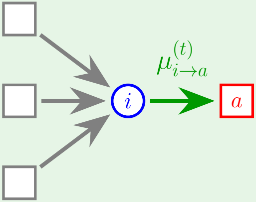
factor → variable



$$\mu^{(t)}_{a\to i}(x_i) =$$

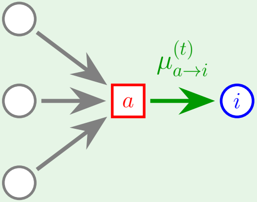
Messages

variable → factor



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

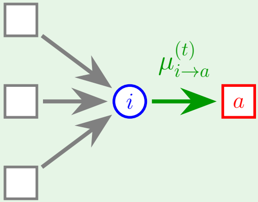
factor → variable



$$\mu^{(t)}_{a\to i}(x_i) =$$

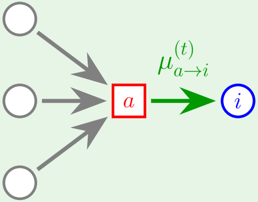
Messages

variable $\rightarrow$ factor



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

factor $\rightarrow$ variable

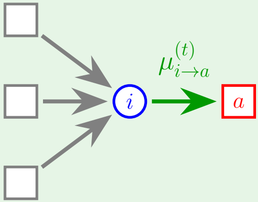


$$\mu^{(t)}_{a\rightarrow i}(x_i) =$$

$$\sum_{j\in\partial a\setminus i} \mu^{(t-1)}_{j\rightarrow a}(x_j)$$

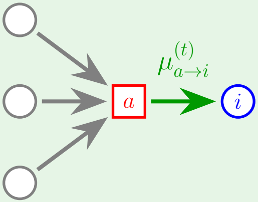
Messages

variable → factor



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

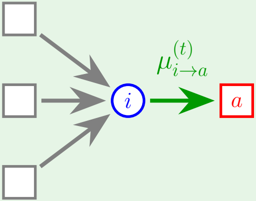
factor → variable



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

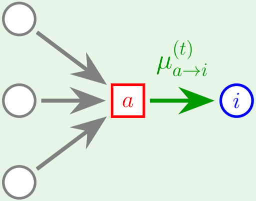
Messages

variable → factor



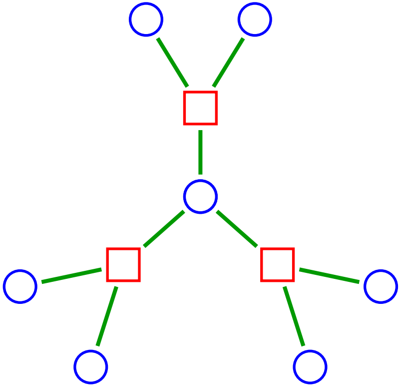
$$\mu^{(t)}_{i\to a}(x_i) = \sum_{b\in\partial i\setminus a} \mu^{(t-1)}_{b\to i}(x_i)$$

factor → variable



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

Min-sum: loopy graphs

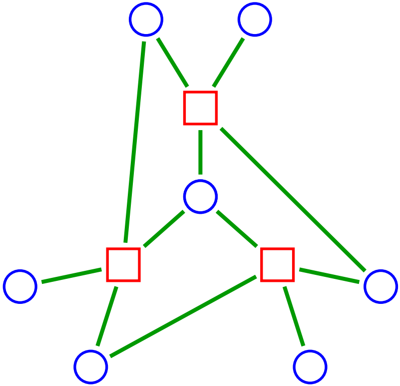


Optimal solution:

$$x^{\star} := \arg\min_{x} \sum_{a \in \text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

?

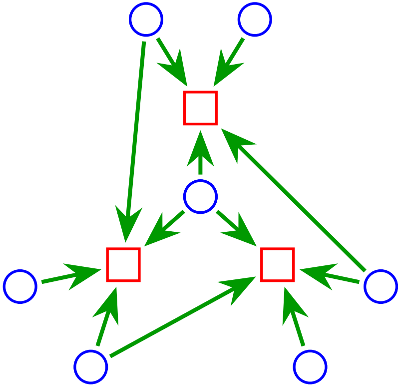


Optimal solution:

$$x^\star := \arg\min_x \sum_{a \in \text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

time 1: $\{\mu_{i\to a}^{(1)} = 0\}$



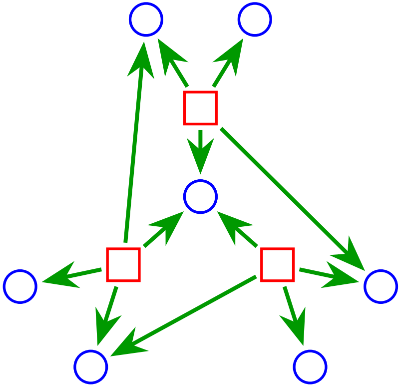
Optimal solution:

$$x^\star := \arg\min_x \sum_{a \in \text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

time 1: $\{\mu_{i\to a}^{(1)} = 0\}$

time 2: $\{\mu_{a\to i}^{(2)}\}$



Optimal solution:

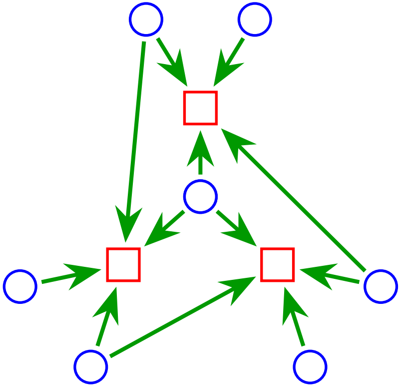
$$x^\star := \arg\min_x \sum_{a\in\text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

time 1: $\{\mu^{(1)}_{i\to a} = 0\}$

time 2: $\{\mu^{(2)}_{a\to i}\}$

time 3: $\{\mu^{(3)}_{i\to a}\}$



Optimal solution:

$$x^\star := \arg\min_x \sum_{a \in \text{factors}} f_a(x_{\partial a})$$

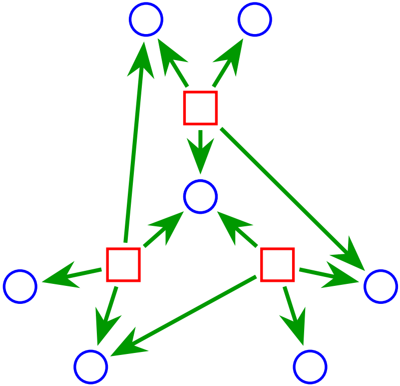
Min-sum: loopy graphs

time 1: $\{\mu^{(1)}_{i\to a} = 0\}$

time 2: $\{\mu^{(2)}_{a\to i}\}$

time 3: $\{\mu^{(3)}_{i\to a}\}$

time 4: $\{\mu^{(4)}_{a\to i}\}$



Optimal solution:

$$x^\star := \arg\min_x \sum_{a\in\text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

time 1: $\{\mu^{(1)}_{i\to a} = 0\}$

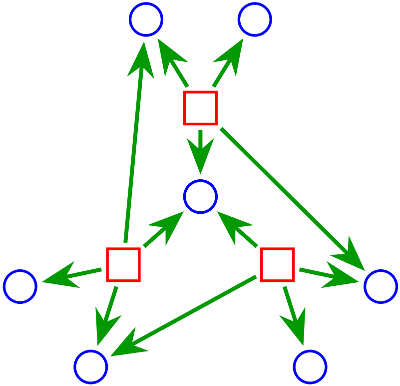
time 2: $\{\mu^{(2)}_{a\to i}\}$

time 3: $\{\mu^{(3)}_{i\to a}\}$

time 4: $\{\mu^{(4)}_{a\to i}\}$

$\vdots$

time t: $\{\mu^{(t)}_{a\to i}\}$



Optimal solution:

$$x^\star := \arg\min_x \sum_{a\in \text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

time 1: $\{\mu^{(1)}_{i\to a} = 0\}$

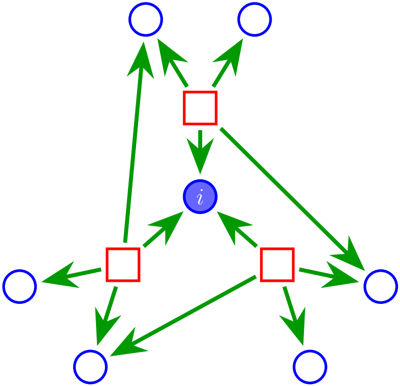
time 2: $\{\mu^{(2)}_{a\to i}\}$

time 3: $\{\mu^{(3)}_{i\to a}\}$

time 4: $\{\mu^{(4)}_{a\to i}\}$

⋮

time t: $\{\mu^{(t)}_{a\to i}\}$



Estimate time t:

$$\hat{x}^{(t)}_i := \arg\min_{x_i} \sum_{a\in\partial i} \mu^{(t)}_{a\to i}(x_i)$$

Optimal solution:

$$x^\star := \arg\min_{x} \sum_{a\in \text{factors}} f_a(x_{\partial a})$$

Min-sum: loopy graphs

time 1: $\{\mu^{(1)}_{i\to a} = 0\}$

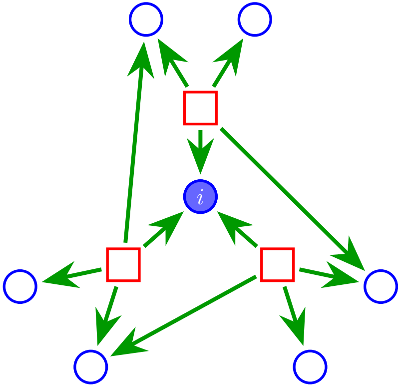
time 2: $\{\mu^{(2)}_{a\to i}\}$

time 3: $\{\mu^{(3)}_{i\to a}\}$

time 4: $\{\mu^{(4)}_{a\to i}\}$

⋮

time t: $\{\mu^{(t)}_{a\to i}\}$



Estimate time t:

$$\hat{x}_i^{(t)} := \arg\min_{x_i} \sum_{a\in\partial i} \mu^{(t)}_{a\to i}(x_i)$$

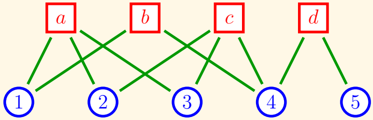
? →

Optimal solution:

$$x^\star := \arg\min_{x} \sum_{a\in\text{factors}} f_a(x_{\partial a})$$

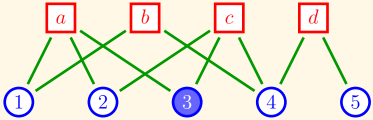
Computation tree

Graph



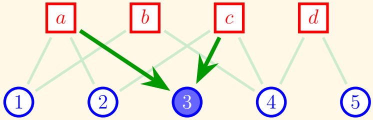
Computation tree

Graph



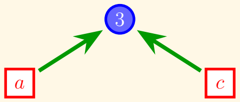
Computation tree

Graph

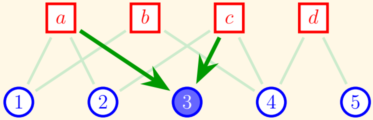


Computation tree

Computation Tree

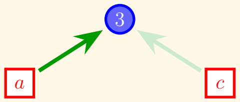


Graph

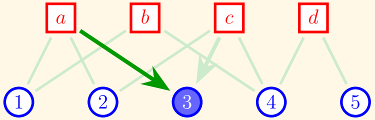


Computation tree

Computation Tree

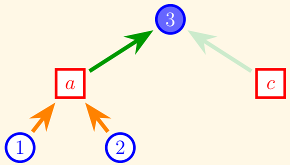


Graph

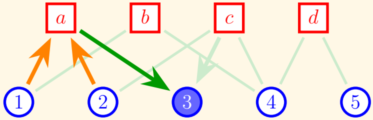


Computation tree

Computation Tree

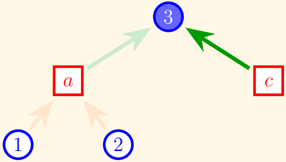


Graph

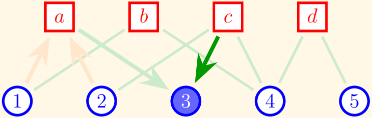


Computation tree

Computation Tree

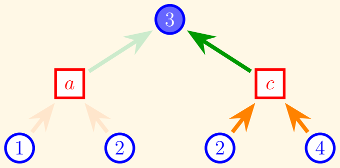


Graph

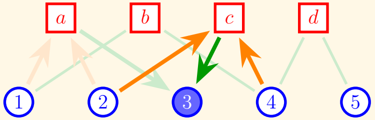


Computation tree

Computation Tree

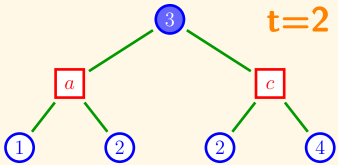


Graph

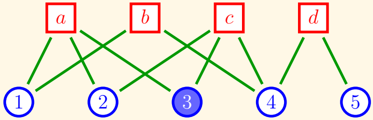


Computation tree

Computation Tree

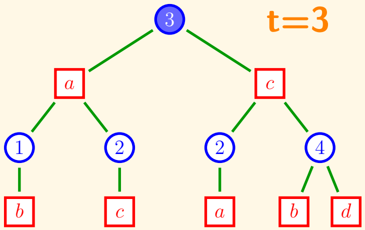


Graph

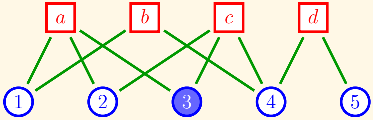


Computation tree

Computation Tree

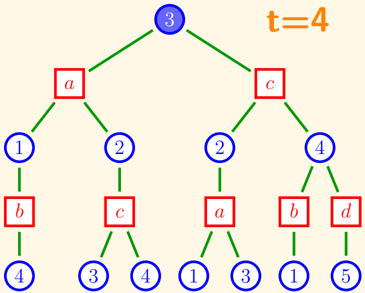


Graph

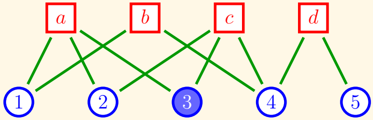


Computation tree

Computation Tree

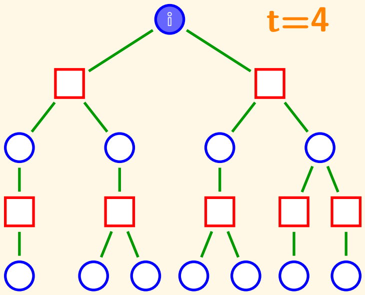


Graph

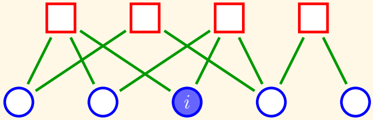


Computation tree

Computation Tree

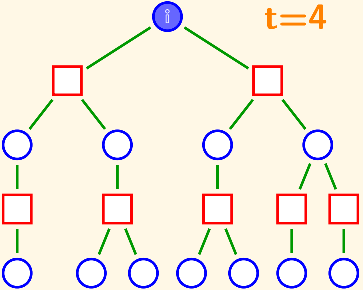


Graph

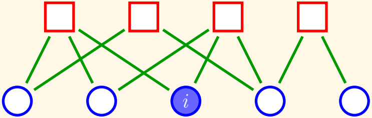


Computation tree

Computation Tree



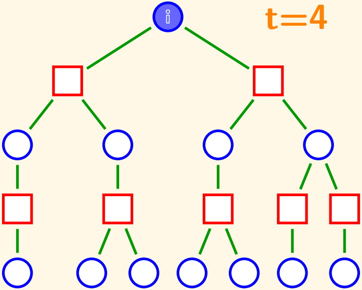
Graph



$$\overbrace{\sum_{\substack{a\in\text{factors}\\ \text{graph}}} f_a(x_{\partial a})}^{f(x)}$$

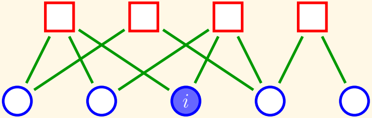
Computation tree

Computation Tree



$$\overbrace{\sum_{\mathbb{a}\in\text{factors tree}} f_{\mathbb{a}}(\mathbb{x}_{\partial\mathbb{a}})}^{\mathbb{f}(\mathbb{x})}$$

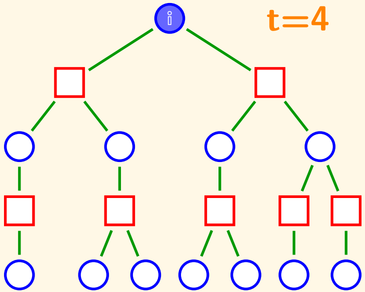
Graph



$$\overbrace{\sum_{a\in\text{factors graph}} f_a(x_{\partial a})}^{f(x)}$$

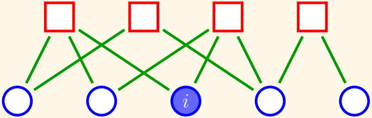
Computation tree

Computation Tree



$$\mathbb{x}^\star := \arg\min_{\mathbb{x}} \overbrace{\sum_{\substack{\mathbb{a}\in\text{factors}\\ \text{tree}}} f_{\mathbb{a}}(\mathbb{x}_{\partial\mathbb{a}})}^{\mathbb{f}(\mathbb{x})}$$

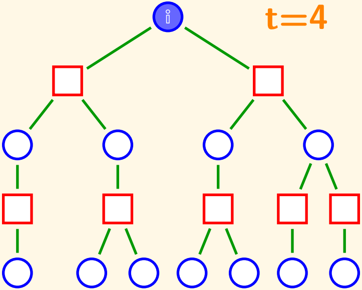
Graph



$$x^\star := \arg\min_{x} \overbrace{\sum_{\substack{a\in\text{factors}\\ \text{graph}}} f_a(x_{\partial a})}^{f(x)}$$

Computation tree

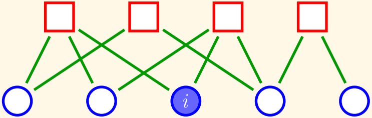
Computation Tree



$$\mathbb{x}^\star := \arg\min_{\mathbb{x}} \overbrace{\sum_{\substack{\mathbb{a}\in\text{factors}\\ \text{tree}}} f_{\mathbb{a}}(\mathbb{x}_{\partial\mathbb{a}})}^{\mathbb{f}(\mathbb{x})}$$

Lemma: $\boxed{\hat{x}_i^{(t)} = \mathbb{x}_{\mathbb{i}}^\star}$

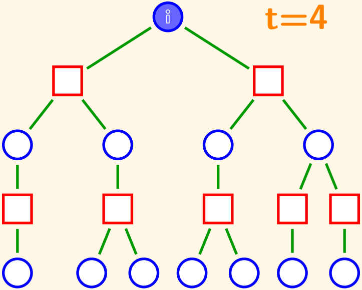
Graph



$$x^\star := \arg\min_{x} \overbrace{\sum_{\substack{a\in\text{factors}\\ \text{graph}}} f_a(x_{\partial a})}^{f(x)}$$

Computation tree

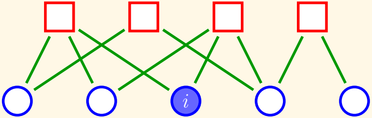
Computation Tree



$$\mathbb{x}^\star := \arg\min_{\mathbb{x}} \overbrace{\sum_{\substack{\mathbb{a}\in\text{factors}\\ \text{tree}}} f_{\mathbb{a}}(\mathbb{x}_{\partial\mathbb{a}})}^{\mathbb{f}(\mathbb{x})}$$

Lemma: $\boxed{\hat{x}_i^{(t)} = \mathbb{x}_{\mathbb{i}}^\star}$

Graph

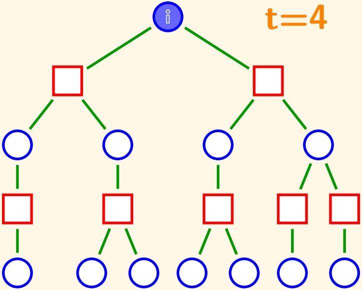


$$x^\star := \arg\min_{x} \overbrace{\sum_{\substack{a\in\text{factors}\\ \text{graph}}} f_a(x_{\partial a})}^{f(x)}$$

$x_i^\star$

Computation tree

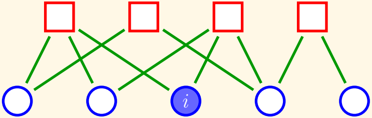
Computation Tree



$$\mathbb{x}^\star := \arg\min_{\mathbb{x}} \overbrace{\sum_{\substack{\mathbb{a}\in\text{factors}\\ \text{tree}}} f_{\mathbb{a}}(\mathbb{x}_{\partial\mathbb{a}})}^{\mathbb{f}(\mathbb{x})}$$

Lemma: $\boxed{\hat{x}_i^{(t)} = \mathbb{x}_{\mathbb{i}}^\star}$ → ? $x_i^\star$

Graph

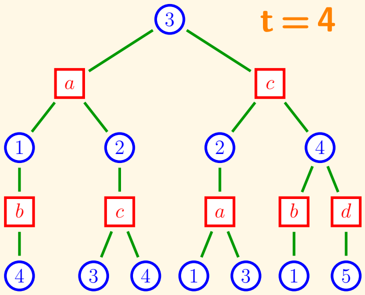


$$x^\star := \arg\min_{x} \overbrace{\sum_{\substack{a\in\text{factors}\\ \text{graph}}} f_a(x_{\partial a})}^{f(x)}$$

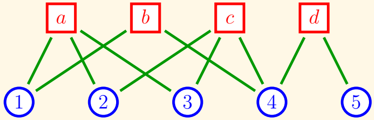
$x_i^\star$

Correctness

Computation Tree

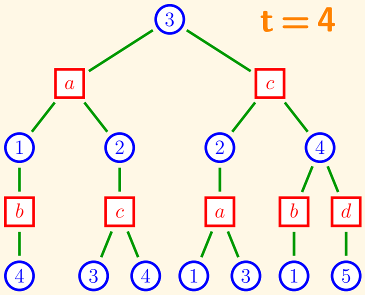


Graph

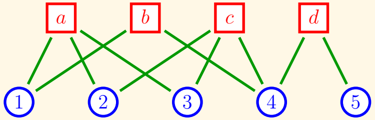


Correctness

Computation Tree



Graph

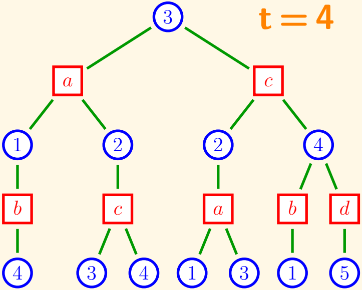


Optimality condition:

$$\overbrace{\sum_{a \in \partial j} \frac{d}{dx_j} f_a(x^{\star}_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

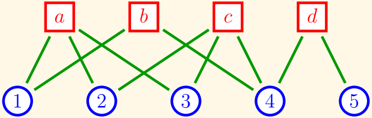
Computation Tree



Optimality condition:

$$\overbrace{\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{d\mathbb{x}_{\mathbb{j}}} f_{\mathbb{a}}(\mathbb{x}^{\star}_{\partial\mathbb{a}})}^{\frac{d}{d\mathbb{x}_{\mathbb{j}}}\mathbb{f}(\mathbb{x})} = 0 \qquad \forall \mathbb{j}$$

Graph

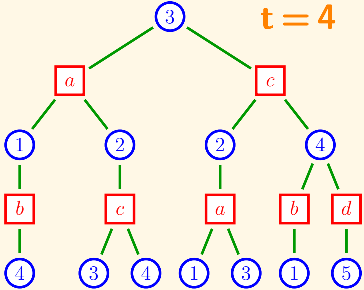


Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^{\star}_{\partial a})}^{\frac{d}{dx_j}f(x)} = 0 \qquad \forall j$$

Correctness

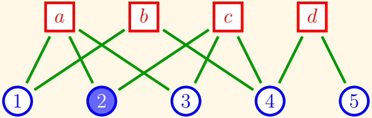
Computation Tree



Optimality condition:

$$\overbrace{\sum_{a \in \partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Graph



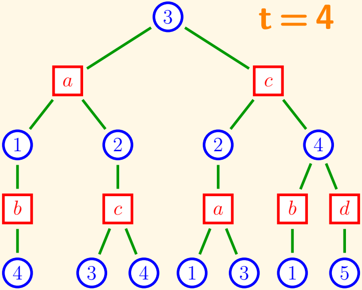
$$\frac{d f_a(x_1^\star, x_2^\star, x_3^\star)}{dx_2} + \frac{d f_c(x_2^\star, x_3^\star, x_4^\star)}{dx_2} = 0 \quad j = 2$$

Optimality condition:

$$\overbrace{\sum_{a \in \partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

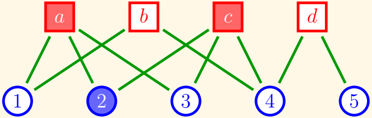
Computation Tree



Optimality condition:

$$\overbrace{\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{d\mathbb{x}_{\mathbb{j}}} f_{\mathbb{a}}(\mathbb{x}^{\star}_{\partial\mathbb{a}})}^{\frac{d}{d\mathbb{x}_{\mathbb{j}}}\mathbb{f}(\mathbb{x})} = 0 \qquad \forall \mathbb{j}$$

Graph



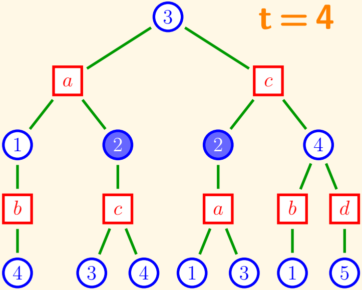
$$\frac{d f_a(x_1^{\star}, x_2^{\star}, x_3^{\star})}{dx_2} + \frac{d f_c(x_2^{\star}, x_3^{\star}, x_4^{\star})}{dx_2} = 0 \quad j = 2$$

Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^{\star}_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

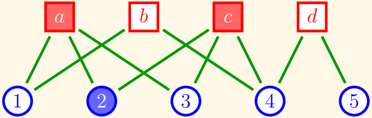
Computation Tree



Optimality condition:

$$\overbrace{\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{d\mathbb{x}_{\mathbb{j}}} f_{\mathbb{a}}(\mathbb{x}^{\star}_{\partial\mathbb{a}})}^{\frac{d}{d\mathbb{x}_{\mathbb{j}}}\mathbb{f}(\mathbb{x})} = 0 \qquad \forall \mathbb{j}$$

Graph



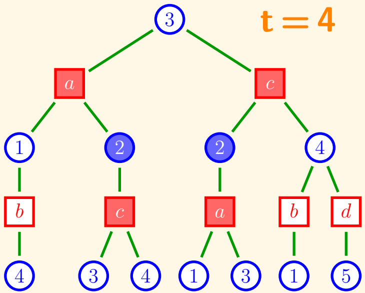
$$\frac{d f_a(x_1^\star, x_2^\star, x_3^\star)}{dx_2} + \frac{d f_c(x_2^\star, x_3^\star, x_4^\star)}{dx_2} = 0 \quad j = 2$$

Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^{\star}_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

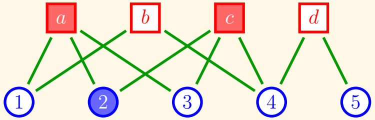
Computation Tree



Optimality condition:

$$\overbrace{\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{d\mathbb{x}_{\mathbb{j}}} f_{\mathbb{a}}(\mathbb{x}^{\star}_{\partial\mathbb{a}})}^{\frac{d}{d\mathbb{x}_{\mathbb{j}}}\mathbb{f}(\mathbb{x})} = 0 \qquad \forall\mathbb{j}$$

Graph



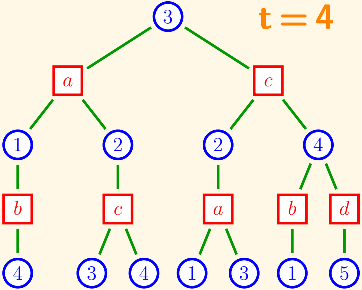
$$\frac{d f_a(x_1^\star, x_2^\star, x_3^\star)}{dx_2} + \frac{d f_c(x_2^\star, x_3^\star, x_4^\star)}{dx_2} = 0 \quad j = 2$$

Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^{\star}_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

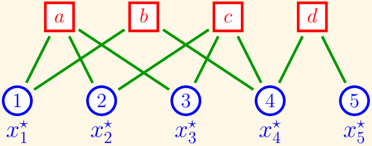
Computation Tree



Optimality condition:

$$\overbrace{\sum_{a\in\partial j}\frac{d}{dx_j}f_a(x^\star_{\partial a})}^{\frac{d}{dx_j}f(x)} = 0 \qquad \forall j$$

Graph

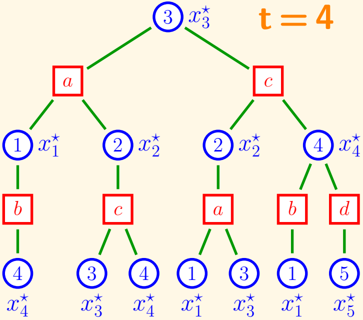


Optimality condition:

$$\overbrace{\sum_{a\in\partial j}\frac{d}{dx_j}f_a(x^\star_{\partial a})}^{\frac{d}{dx_j}f(x)} = 0 \qquad \forall j$$

Correctness

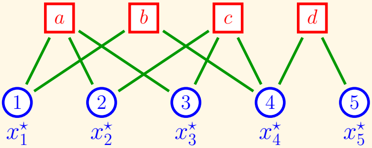
Computation Tree



Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Graph

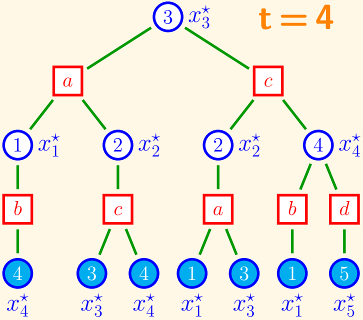


Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

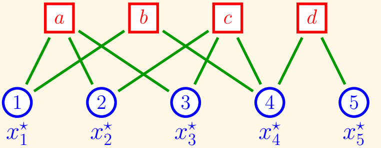
Computation Tree



Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Graph

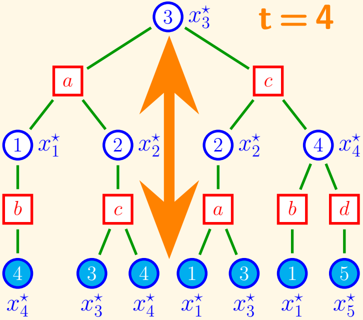


Optimality condition:

$$\overbrace{\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Correctness

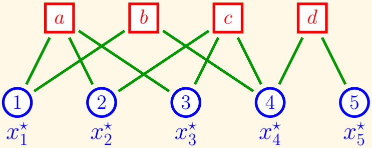
Computation Tree



Optimality condition:

$$\overbrace{\sum_{a \in \partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Graph

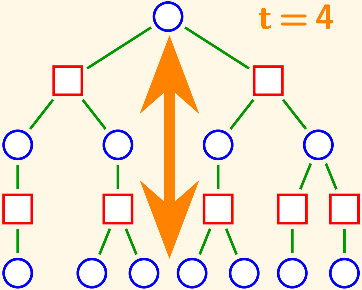


Optimality condition:

$$\overbrace{\sum_{a \in \partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a})}^{\frac{d}{dx_j} f(x)} = 0 \qquad \forall j$$

Convergence

Computation Tree

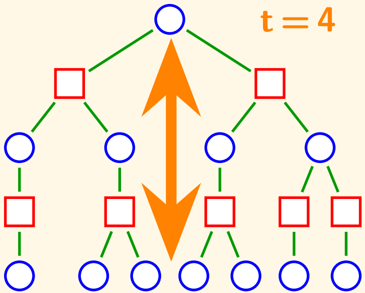


Convergence condition:

$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

Convergence

Computation Tree



Convergence condition:

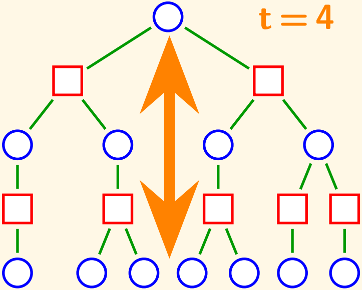
$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left|\frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x})\right| \le \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \qquad \forall \mathbb{i}, \mathbb{x}$$

$$\boxed{\left|\frac{\partial}{\partial p_{j\to a}} \mathbb{x}^\star(p)_{\mathbb{i}}\right| \le \frac{\max_{\mathbb{i}\in\mathbb{V}} \omega_{\mathbb{i}}}{\min_{\mathbb{i}\in\mathbb{V}} \omega_{\mathbb{i}}} \frac{\lambda^t}{1-\lambda}}$$

(Algorithmic) Locality

Convergence

Computation Tree

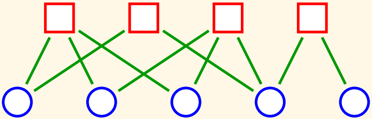


Convergence condition:

$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

$$\boxed{\left| \frac{\partial}{\partial p_{j\to a}} \mathbb{x}^\star(p)_{\mathbb{i}} \right| \leq \frac{\max_{\mathbb{i}\in\mathbb{V}} \omega_{\mathbb{i}}}{\min_{\mathbb{i}\in\mathbb{V}} \omega_{\mathbb{i}}} \frac{\lambda^t}{1-\lambda}}$$

Graph



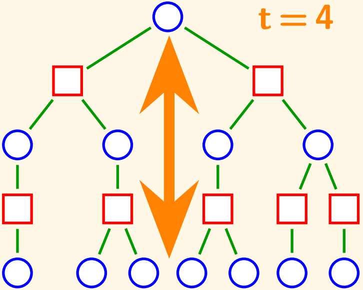
Convergence condition:

$$\sum_{j\in V\setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda\, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

(λ, ω)-scaled diagonal dominance

Convergence

Computation Tree

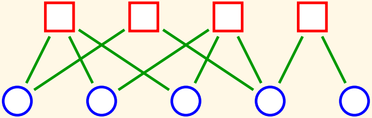


Convergence condition:

$$\sum_{\mathbb{j} \in \mathbb{V} \setminus \mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda \, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

$$\boxed{\left| \frac{\partial}{\partial p_{j \to a}} \mathbb{x}^\star(p)_{\mathbb{i}} \right| \leq \frac{\max_{\mathbb{i} \in \mathbb{V}} \omega_{\mathbb{i}}}{\min_{\mathbb{i} \in \mathbb{V}} \omega_{\mathbb{i}}} \, \frac{\lambda^t}{1 - \lambda}}$$

Graph



Convergence condition:

$$\sum_{j \in V \setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda \, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

(λ, ω)-scaled diagonal dominance

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j}\in\mathbb{V}\backslash\mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

Graph

$$\sum_{j\in V\backslash i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda\, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

←

Graph

$$\sum_{j\in V\setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda\, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

⇕

$$\rho(|R(x)|) < 1$$

with $R(x) := I - D^{-1/2} \nabla^2 f(x) D^{-1/2}$

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j} \in \mathbb{V} \setminus \mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda \, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

$\Updownarrow$

$$\rho(|\mathbb{R}(\mathbb{x})|) < 1$$

with $\mathbb{R}(\mathbb{x}) := I - \mathbb{D}^{-1/2} \nabla^2 \mathbb{f}(\mathbb{x}) \mathbb{D}^{-1/2}$

$\leftarrow$

Graph

$$\sum_{j \in V \setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda \, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

$\Updownarrow$

$$\rho(|R(x)|) < 1$$

with $R(x) := I - D^{-1/2} \nabla^2 f(x) D^{-1/2}$

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left|\frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}\partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x})\right| \le \lambda\,\omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2}\mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i},\mathbb{x}$$

$$\Updownarrow$$

$$\rho(|\mathbb{R}(\mathbb{x})|) < 1$$

with $\mathbb{R}(\mathbb{x}) := I - \mathbb{D}^{-1/2}\nabla^2\mathbb{f}(\mathbb{x})\mathbb{D}^{-1/2}$

Graph

$$\sum_{j\in V\setminus i} \omega_{j} \left|\frac{\partial^2}{\partial x_{i}\partial x_{j}} f(x)\right| \le \lambda\,\omega_{i} \frac{\partial^2}{\partial x_{i}^2}f(x) \quad \forall i,x$$

$$\Updownarrow$$

$$\rho(|R(x)|) < 1$$

with $R(x) := I - D^{-1/2}\nabla^2 f(x)D^{-1/2}$

(Weiss, Freeman, 2001) (Malioutov, Johnson, Willsky, 2006) (Moallemi, Van Roy, 2010)

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

$\Updownarrow$

$$\rho(|\mathbb{R}(\mathbb{x})|) < 1$$

with $\mathbb{R}(\mathbb{x}) := I - \mathbb{D}^{-1/2} \nabla^2 \mathbb{f}(\mathbb{x}) \mathbb{D}^{-1/2}$

$\Leftarrow$

Graph

$$\sum_{j\in V\setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda\, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

$\Updownarrow$

$$\rho(|R(x)|) < 1$$

with $R(x) := I - D^{-1/2} \nabla^2 f(x) D^{-1/2}$

(Weiss, Freeman, 2001) (Malioutov, Johnson, Willsky, 2006) (Moallemi, Van Roy, 2010)

Limitations:

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j}\in\mathbb{V}\setminus\mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda\, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

$\Updownarrow$

$$\rho(|\mathbb{R}(\mathbb{x})|) < 1$$

with $\mathbb{R}(\mathbb{x}) := I - \mathbb{D}^{-1/2} \nabla^2 \mathbb{f}(\mathbb{x}) \mathbb{D}^{-1/2}$

$\Longleftarrow$

Graph

$$\sum_{j\in V\setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda\, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

$\Updownarrow$

$$\rho(|R(x)|) < 1$$

with $R(x) := I - D^{-1/2} \nabla^2 f(x) D^{-1/2}$

(Weiss, Freeman, 2001) (Malioutov, Johnson, Willsky, 2006) (Moallemi, Van Roy, 2010)

Limitations:

- Inheritance does not capture convergence behavior on the **tree**.

Scaled diagonal dominance or walk summability

Computation Tree

$$\sum_{\mathbb{j} \in \mathbb{V} \setminus \mathbb{i}} \omega_{\mathbb{j}} \left| \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}} \partial \mathbb{x}_{\mathbb{j}}} \mathbb{f}(\mathbb{x}) \right| \leq \lambda \, \omega_{\mathbb{i}} \frac{\partial^2}{\partial \mathbb{x}_{\mathbb{i}}^2} \mathbb{f}(\mathbb{x}) \quad \forall \mathbb{i}, \mathbb{x}$$

$$\rho(|\mathbb{R}(\mathbb{x})|) < 1$$

with $\mathbb{R}(\mathbb{x}) := I - \mathbb{D}^{-1/2} \nabla^2 \mathbb{f}(\mathbb{x}) \mathbb{D}^{-1/2}$

Graph

$$\sum_{j \in V \setminus i} \omega_j \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| \leq \lambda \, \omega_i \frac{\partial^2}{\partial x_i^2} f(x) \quad \forall i, x$$

$$\rho(|R(x)|) < 1$$

with $R(x) := I - D^{-1/2} \nabla^2 f(x) D^{-1/2}$

(Weiss, Freeman, 2001) (Malioutov, Johnson, Willsky, 2006) (Moallemi, Van Roy, 2010)

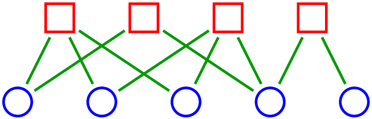
Limitations:

- Inheritance does not capture convergence behavior on the **tree**.
- Condition can **not** be applied to **constrained problems**:

$$\begin{aligned} \text{minimize} \quad & \sum_a f_a(x_{\partial a}) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

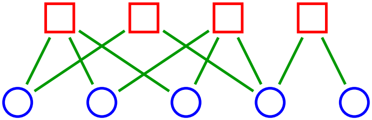
Problems with constraints

minimize $\sum_{a} f_a(x_{\partial a})$



Problems with constraints

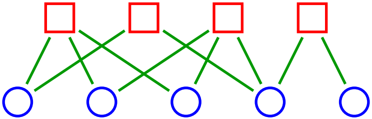
minimize $\sum_{a} f_a(x_{\partial a})$



Condition for convergence and correctness: scaled-diagonal dominance.

Problems with constraints

minimize $\sum_{a} f_a(x_{\partial a})$



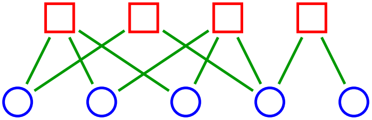
Condition for convergence and correctness: scaled-diagonal dominance.

minimize $\sum_{a} f_a(x_{\partial a})$

subject to $h_b(x_{\partial b}) = 0, \quad \forall b$

Problems with constraints

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



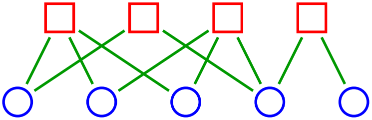
Condition for convergence and correctness: scaled-diagonal dominance.

minimize $\sum_a f_a(x_{\partial a}) + \sum_b \chi(h_b(x_{\partial b}))$

$$\chi(z) := \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{if } z \neq 0 \end{cases}$$

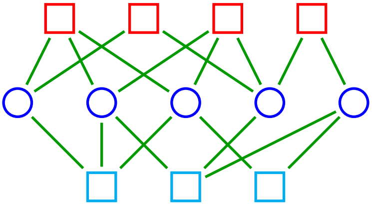
Problems with constraints

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



Condition for convergence and correctness: scaled-diagonal dominance.

minimize $\sum_a f_a(x_{\partial a}) + \sum_b \chi(h_b(x_{\partial b}))$

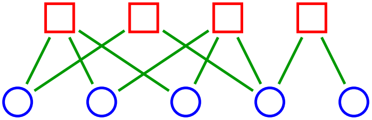


$$\chi(z) := \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{if } z \neq 0 \end{cases}$$

We can run Message Passing! Need new tools for analysis.

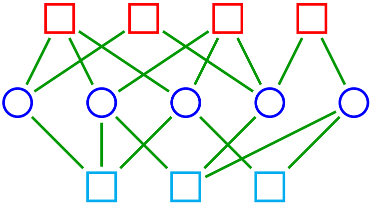
Problems with constraints

minimize $\sum_{a} f_a(x_{\partial a})$



Condition for convergence and correctness: scaled-diagonal dominance.

minimize $\overbrace{\sum_{a} f_a(x_{\partial a}) + \sum_{b} \chi(h_b(x_{\partial b}))}^{g(x)}$

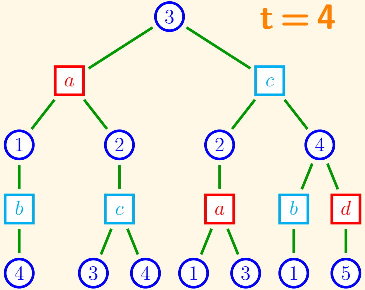


$$\chi(z) := \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{if } z \neq 0 \end{cases}$$

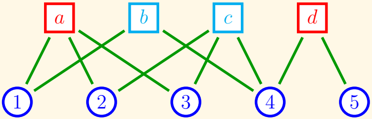
We can run Message Passing! Need new tools for analysis.

Correctness (with constraints!)

Computation Tree

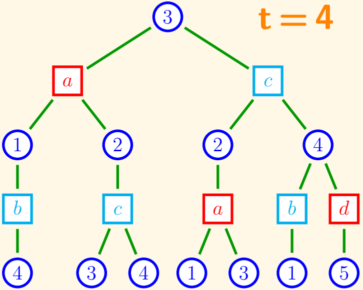


Graph

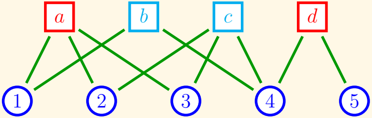


Correctness (with constraints!)

Computation Tree



Graph



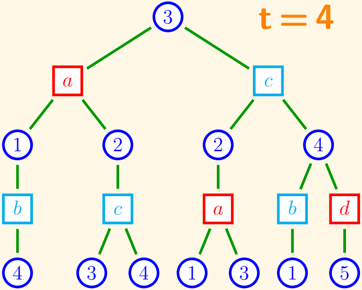
KKT optimality conditions:

$$\sum_{a \in \partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a}) + \sum_{b \in \partial j} \nu^\star_b \frac{d}{dx_j} h_b(x^\star_{\partial b}) = 0 \quad \forall j$$

$$h_b(x^\star_{\partial b}) = 0 \quad \forall b$$

Correctness (with constraints!)

Computation Tree

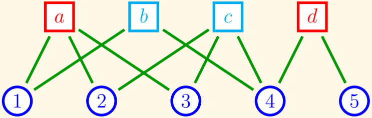


KKT optimality conditions:

$$\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{dx_{\mathbb{j}}} f_{\mathbb{a}}(x^{\star}_{\partial\mathbb{a}}) + \sum_{\mathbb{b}\in\partial\mathbb{j}} \nu^{\star}_{\mathbb{b}} \frac{d}{dx_{\mathbb{j}}} h_{\mathbb{b}}(x^{\star}_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{j}$$

$$h_{\mathbb{b}}(x^{\star}_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{b}$$

Graph



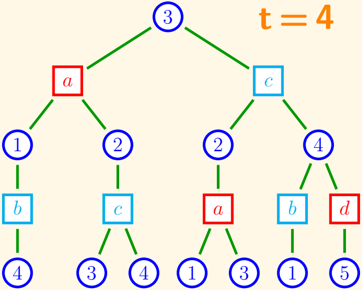
KKT optimality conditions:

$$\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^{\star}_{\partial a}) + \sum_{b\in\partial j} \nu^{\star}_{b} \frac{d}{dx_j} h_b(x^{\star}_{\partial b}) = 0 \quad \forall j$$

$$h_b(x^{\star}_{\partial b}) = 0 \quad \forall b$$

Correctness (with constraints!)

Computation Tree

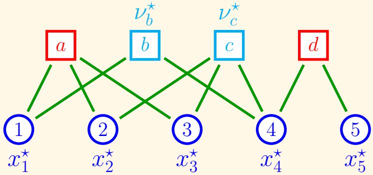


KKT optimality conditions:

$$\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{dx_{\mathbb{j}}} f_{\mathbb{a}}(x^\star_{\partial\mathbb{a}}) + \sum_{\mathbb{b}\in\partial\mathbb{j}} \nu^\star_{\mathbb{b}} \frac{d}{dx_{\mathbb{j}}} h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{j}$$

$$h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{b}$$

Graph



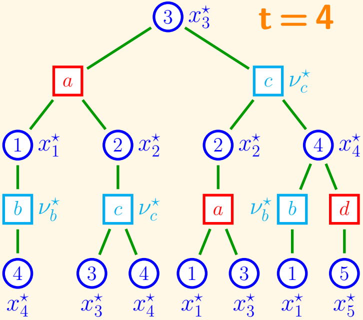
KKT optimality conditions:

$$\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a}) + \sum_{b\in\partial j} \nu^\star_b \frac{d}{dx_j} h_b(x^\star_{\partial b}) = 0 \quad \forall j$$

$$h_b(x^\star_{\partial b}) = 0 \quad \forall b$$

Correctness (with constraints!)

Computation Tree

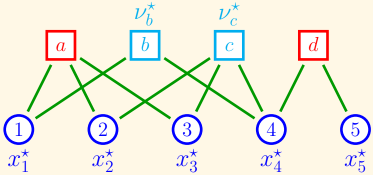


KKT optimality conditions:

$$\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{dx_{\mathbb{j}}} f_{\mathbb{a}}(x^\star_{\partial\mathbb{a}}) + \sum_{\mathbb{b}\in\partial\mathbb{j}} \nu^\star_{\mathbb{b}} \frac{d}{dx_{\mathbb{j}}} h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{j}$$

$$h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{b}$$

Graph



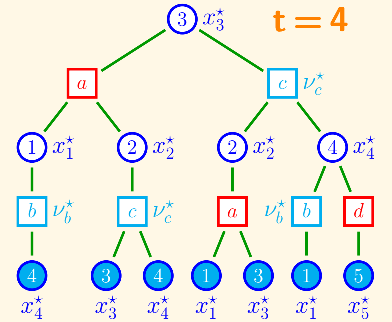
KKT optimality conditions:

$$\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a}) + \sum_{b\in\partial j} \nu^\star_b \frac{d}{dx_j} h_b(x^\star_{\partial b}) = 0 \quad \forall j$$

$$h_b(x^\star_{\partial b}) = 0 \quad \forall b$$

Correctness (with constraints!)

Computation Tree

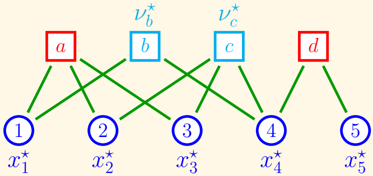


KKT optimality conditions:

$$\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{dx_{\mathbb{j}}} f_{\mathbb{a}}(x^\star_{\partial\mathbb{a}}) + \sum_{\mathbb{b}\in\partial\mathbb{j}} \nu^\star_{\mathbb{b}} \frac{d}{dx_{\mathbb{j}}} h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{j}$$

$$h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{b}$$

Graph



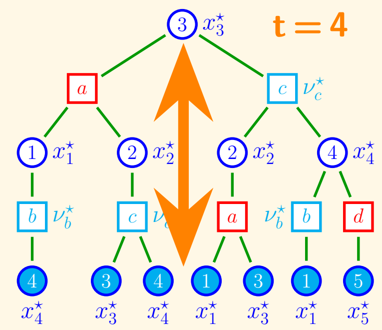
KKT optimality conditions:

$$\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a}) + \sum_{b\in\partial j} \nu^\star_b \frac{d}{dx_j} h_b(x^\star_{\partial b}) = 0 \quad \forall j$$

$$h_b(x^\star_{\partial b}) = 0 \quad \forall b$$

Correctness (with constraints!)

Computation Tree

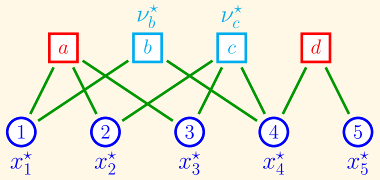


KKT optimality conditions:

$$\sum_{\mathbb{a}\in\partial\mathbb{j}} \frac{d}{dx_{\mathbb{j}}} f_{\mathbb{a}}(x^\star_{\partial\mathbb{a}}) + \sum_{\mathbb{b}\in\partial\mathbb{j}} \nu^\star_{\mathbb{b}} \frac{d}{dx_{\mathbb{j}}} h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{j}$$

$$h_{\mathbb{b}}(x^\star_{\partial\mathbb{b}}) = 0 \quad \forall \mathbb{b}$$

Graph



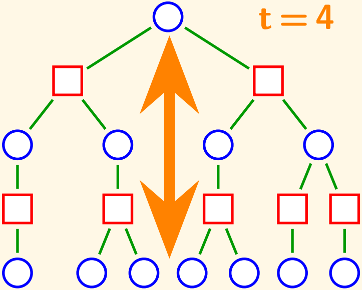
KKT optimality conditions:

$$\sum_{a\in\partial j} \frac{d}{dx_j} f_a(x^\star_{\partial a}) + \sum_{b\in\partial j} \nu^\star_b \frac{d}{dx_j} h_b(x^\star_{\partial b}) = 0 \quad \forall j$$

$$h_b(x^\star_{\partial b}) = 0 \quad \forall b$$

Convergence (with constraints!)

Computation Tree

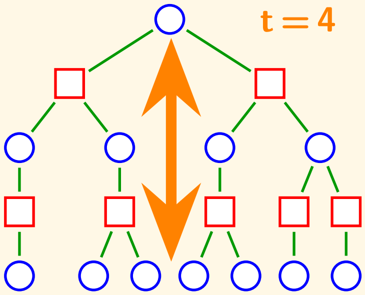


Object of interest:

$$\boxed{\frac{\partial}{\partial p_{j \to a}} \mathbb{x}^{\star}(p)_{\mathbb{i}}}$$

Convergence (with constraints!)

Computation Tree



Object of interest:

$$\frac{\partial}{\partial p_{j\to a}} \mathbb{x}^\star(p)_{\mathbb{i}}$$

Application:
Network Flows and Laplacian Solvers

Directed graph $G = (V, E)$

$$\text{minimize} \quad \sum_{e\in E} \overbrace{f_e(x_e)}^{(x_e)^2}$$

$$\text{subject to} \quad Ax = b$$

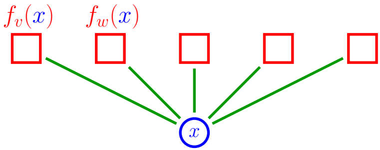
$$A_{ve} := \begin{cases} 1 & \text{if } e \text{ leaves } v, \\ -1 & \text{if } e \text{ enters } v, \\ 0 & \text{otherwise.} \end{cases}$$

"A new approach to Laplacian solvers and flow problems," **arXiv:1611.07138** (2016)

Consensus & Acceleration Mechanisms

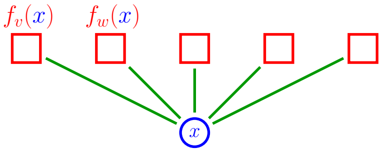
Consensus

minimize $\sum_{v \in V} f_v(x)$

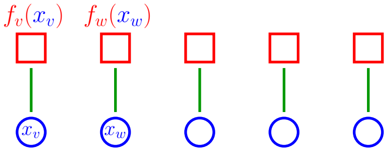


Consensus

minimize $\sum_{v \in V} f_v(x)$

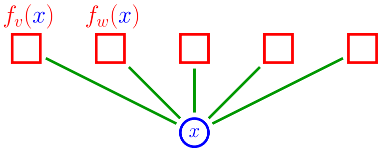


minimize $\sum_{v \in V} f_v(x_v)$

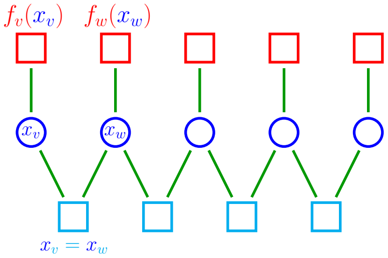


Consensus

$$\text{minimize} \quad \sum_{v \in V} f_v(x)$$

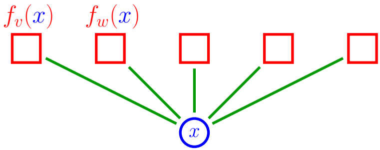


$$\begin{array}{ll} \text{minimize} & \sum_{v \in V} f_v(x_v) \\ \text{subject to} & x_v = x_w, \{v, w\} \in E \end{array}$$



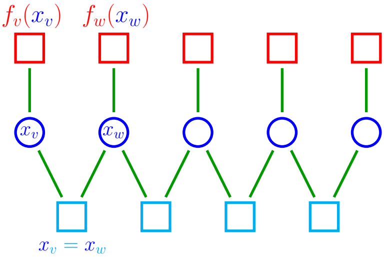
Consensus

minimize $\sum_{v \in V} f_v(x)$



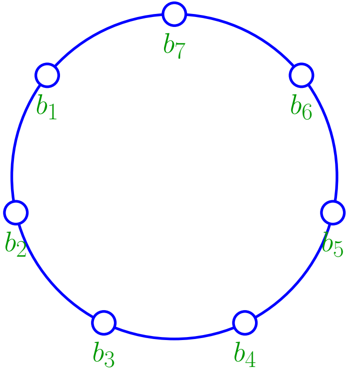
minimize $\sum_{v \in V} f_v(x_v)$

subject to $x_v = x_w, \{v, w\} \in E$

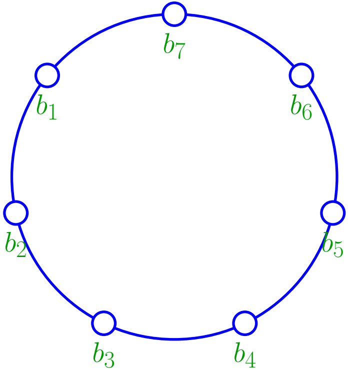


Classical setting: $f_v(x_v) := (x_v - b_v)^2$ (**network averaging problem**)

Classical algorithms for Consensus

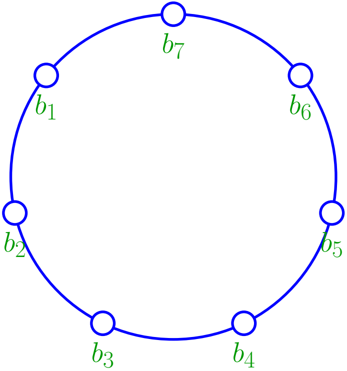


Classical algorithms for Consensus



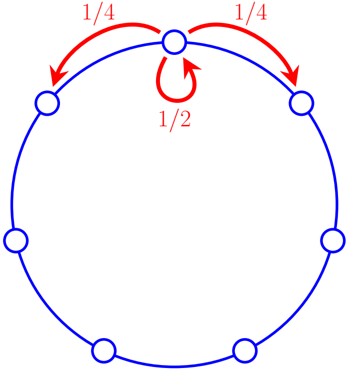
- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$

Classical algorithms for Consensus



- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$
- $\lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$ (Boyd, Xiao, 2004)

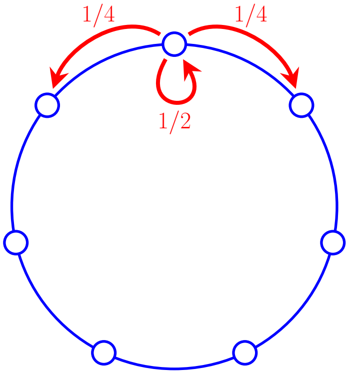
Classical algorithms for Consensus



- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$
- $\lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$ (Boyd, Xiao, 2004)
- Common choice:

$$W_{ij} = \begin{cases} 1/(2d_{\text{max}}) & \text{if } \{i,j\} \in E \\ 1 - d_i/(2d_{\text{max}}) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Classical algorithms for Consensus

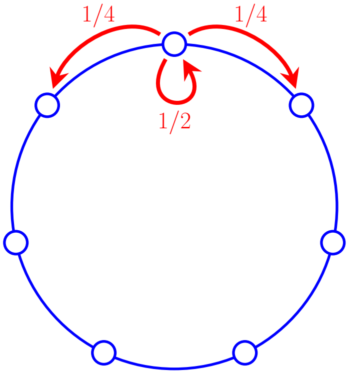


- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$
- $\lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$ (Boyd, Xiao, 2004)
- Common choice:

$$W_{ij} = \begin{cases} 1/(2d_{\mathsf{max}}) & \text{if } \{i,j\}\in E \\ 1-d_i/(2d_{\mathsf{max}}) & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$$

- $\min\{\rho(W-\mathbf{1}\mathbf{1}^T/n) : W \text{ symm}\}$ is SDP

Classical algorithms for Consensus

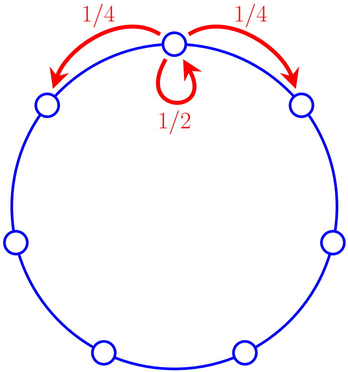


- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$
- $\lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$ (Boyd, Xiao, 2004)
- Common choice:

$$W_{ij} = \begin{cases} 1/(2d_{\mathsf{max}}) & \text{if } \{i,j\}\in E \\ 1-d_i/(2d_{\mathsf{max}}) & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$$

- $\min\{\rho(W-\mathbf{1}\mathbf{1}^T/n) : W \text{ symm}\}$ is SDP
- **Diffusive convergence time** $O(D^2)$ achieved by W (D is graph diameter)

Classical algorithms for Consensus

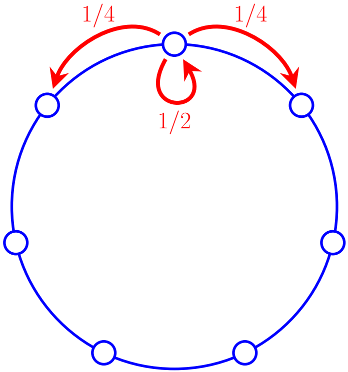


- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$
- $\lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$ (Boyd, Xiao, 2004)
- Common choice:

$$W_{ij} = \begin{cases} 1/(2d_{\mathsf{max}}) & \text{if } \{i,j\} \in E \\ 1 - d_i/(2d_{\mathsf{max}}) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

- $\min\{\rho(W - \mathbf{1}\mathbf{1}^T/n) : W \text{ symm}\}$ is SDP
- **Diffusive convergence time** $O(D^2)$ achieved by W (D is graph diameter)
- Lower-bound: $\Omega(D)$

Classical algorithms for Consensus



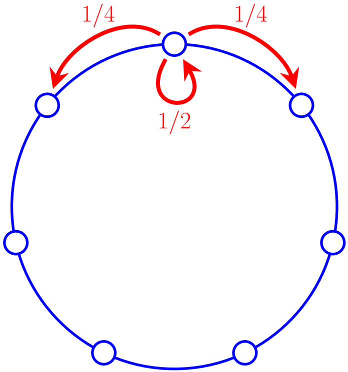
- $x^{(0)} = b$ and $x^{(t)} = Wx^{(t-1)}$
- $\lim_{t\to\infty} W^t \to \mathbf{1}\mathbf{1}^T/n$ (Boyd, Xiao, 2004)
- Common choice:

$$W_{ij} = \begin{cases} 1/(2d_{\max}) & \text{if } \{i,j\} \in E \\ 1 - d_i/(2d_{\max}) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

- $\min\{\rho(W - \mathbf{1}\mathbf{1}^T/n) : W \text{ symm}\}$ is SDP
- **Diffusive convergence time** $O(D^2)$ achieved by W (D is graph diameter)
- Lower-bound: $\Omega(D)$

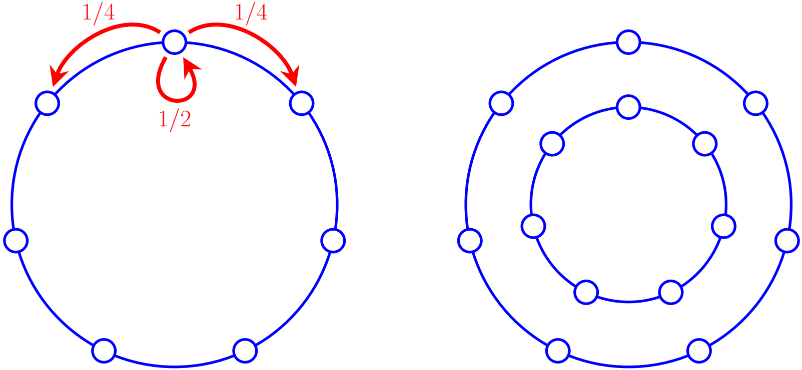
Q. Subdiffusive rates?

Lifted Markov Chains



Original communication graph

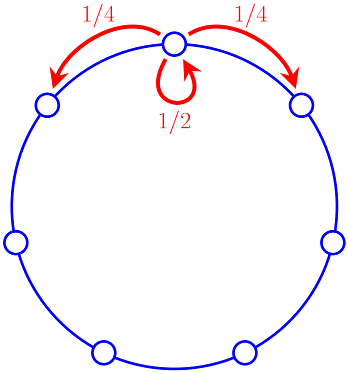
Lifted Markov Chains



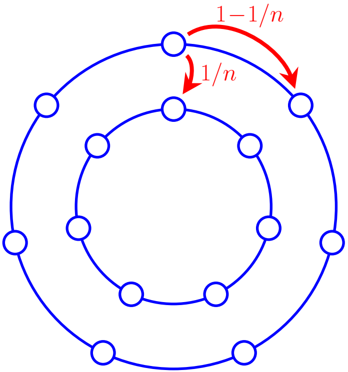
Original communication graph

Lifted communication graph
(Diaconis et al., 2000) (Chen et al., 1999)

Lifted Markov Chains



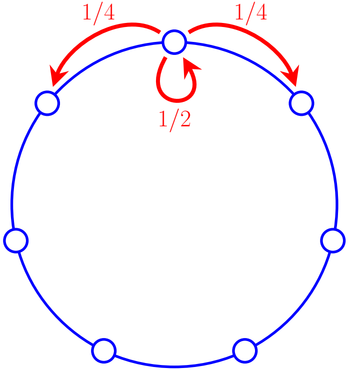
Original communication graph



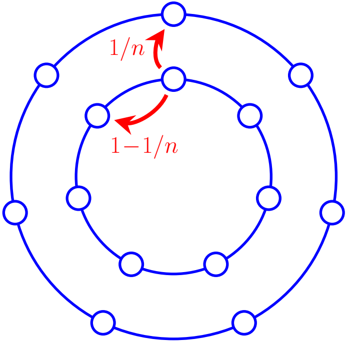
Lifted communication graph

(Diaconis et al., 2000) (Chen et al., 1999)

Lifted Markov Chains



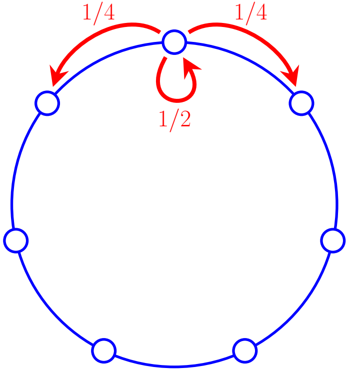
Original communication graph



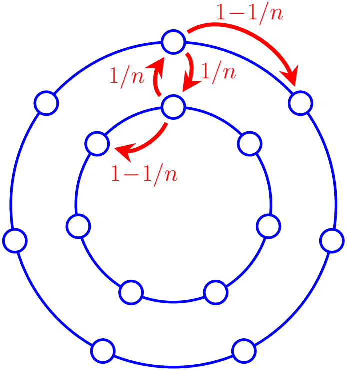
Lifted communication graph

(Diaconis et al., 2000) (Chen et al., 1999)

Lifted Markov Chains



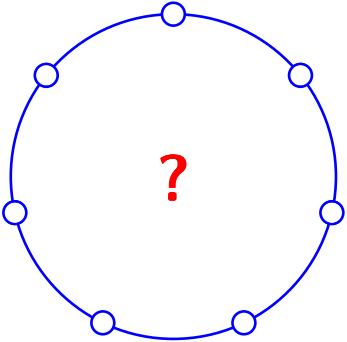
Original communication graph



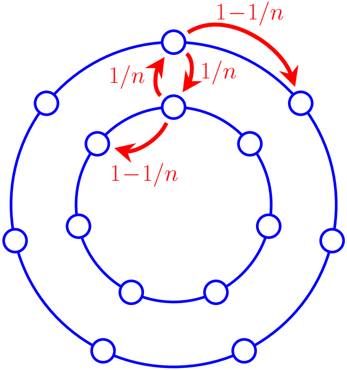
Lifted communication graph

(Diaconis et al., 2000) (Chen et al., 1999)

Lifted Markov Chains



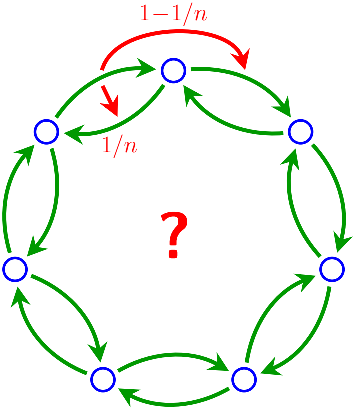
Original communication graph



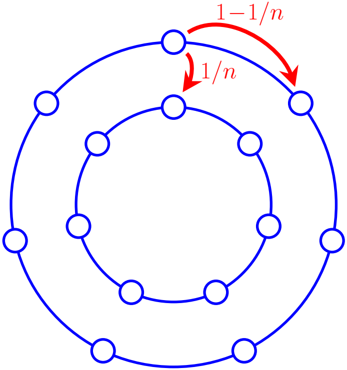
Lifted communication graph
(Diaconis et al., 2000) (Chen et al., 1999)

Q. Can define process on edges?

Lifted Markov Chains



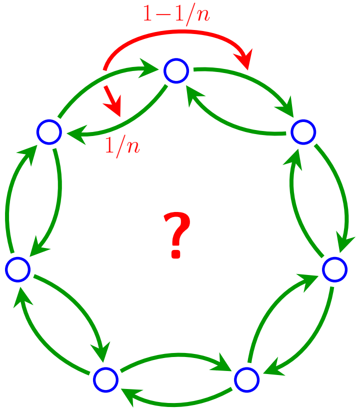
Original communication graph



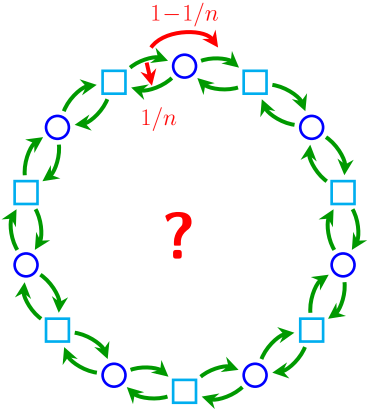
Lifted communication graph
(Diaconis et al., 2000) (Chen et al., 1999)

Q. Can define process on edges?

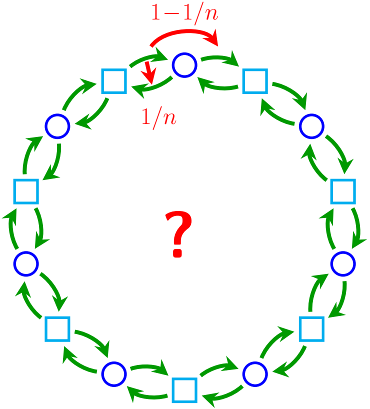
Min-Sum for Consensus

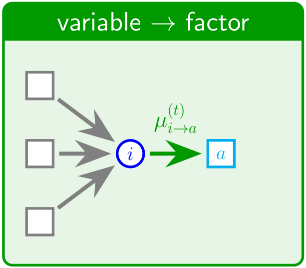


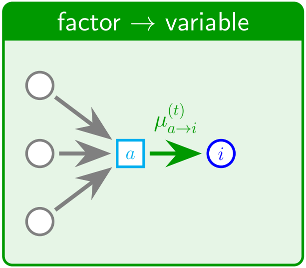
Min-Sum for Consensus



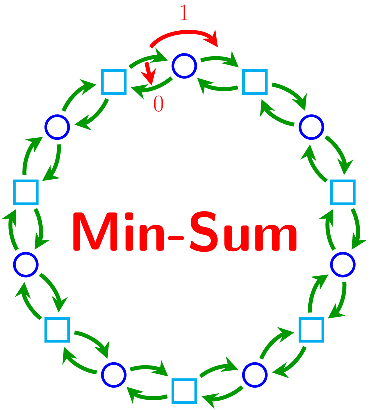
Min-Sum for Consensus



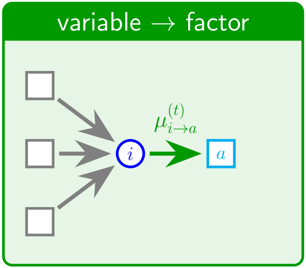


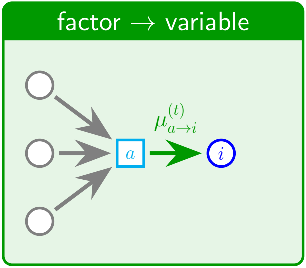


Min-Sum for Consensus

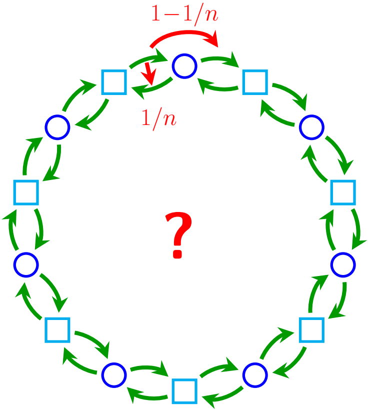


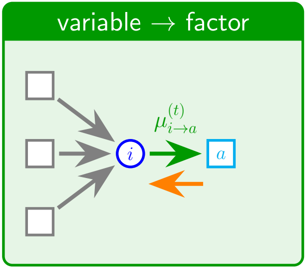
Min-Sum **does not converge**
in graphs with loops
(Moallemi, Van Roy, 2006)

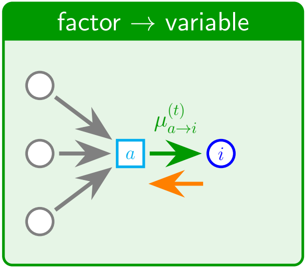




Min-Sum for Consensus





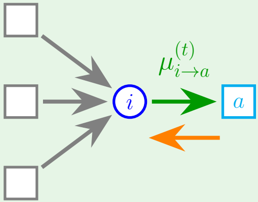


Q. Can modify Min-Sum?

Min-Sum Splitting

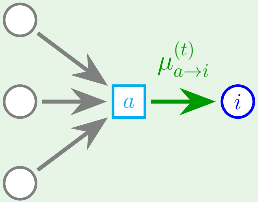
(Ruozzi and Tatikonda, 2013)

variable → factor



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

factor → variable



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

Min-Sum Splitting

Min-Sum Splitting $\equiv$ Min-Sum applied to new formulation of obj. function.

Original objective function

$$\text{minimize} \quad \sum_{v \in V} f_v(x_v) + \sum_{\{v,w\} \in E} f_{vw}(x_v, x_w)$$

Min-Sum Splitting

Min-Sum Splitting $\equiv$ Min-Sum applied to new formulation of obj. function.

Original objective function

$$\text{minimize} \quad \sum_{v \in V} f_v(x_v) + \sum_{\{v,w\} \in E} f_{vw}(x_v, x_w)$$

Recall Consensus problem:

$$\begin{aligned} \text{minimize} \quad & \sum_{v \in V} (x_v - b_v)^2 \\ \text{subject to} \quad & x_v = x_w, \{v, w\} \in E \end{aligned}$$

Min-Sum Splitting

Min-Sum Splitting $\equiv$ Min-Sum applied to new formulation of obj. function.

Original objective function

$$\text{minimize} \quad \sum_{v \in V} f_v(x_v) + \sum_{\{v,w\} \in E} f_{vw}(x_v, x_w)$$

Recall Consensus problem:

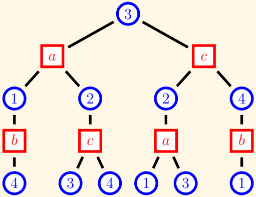
$$\begin{aligned} \text{minimize} \quad & \sum_{v \in V} (x_v - b_v)^2 \\ \text{subject to} \quad & x_v = x_w, \{v, w\} \in E \end{aligned}$$

New formulation

$$\text{minimize} \quad \sum_{v \in V} f_v(x_v) + \sum_{\{v,w\} \in E} \sum_{k=1}^{\Gamma_{vw}} \frac{f_{vw}(x_v, x_w)}{\Gamma_{vw}}$$

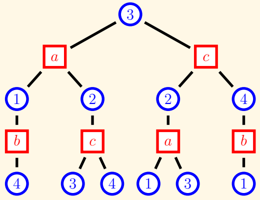
Min-Sum Splitting - Computation Tree

Computation Tree for Min-Sum

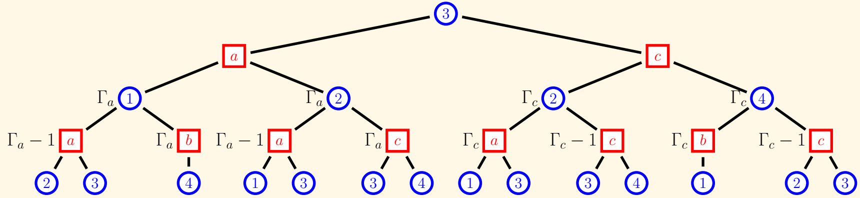


Min-Sum Splitting - Computation Tree

Computation Tree for Min-Sum

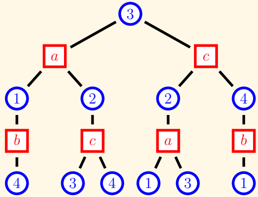


Computation Tree for Min-Sum Splitting

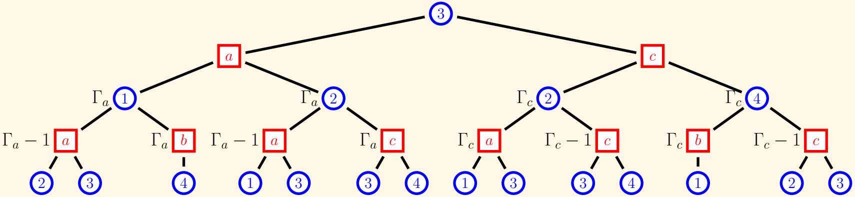


Min-Sum Splitting - Computation Tree

Computation Tree for Min-Sum



Computation Tree for Min-Sum Splitting



Optimality conditions are inherited from original graph also with Splitting.

Min-Sum Splitting for Consensus

Define:

$$\hat{h}_{(w,v)} := b_w$$

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

Algorithm 1: Min-Sum Splitting for Consensus

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

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

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

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

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

Results

- Let $W \in \mathbb{R}^{n\times n}$ be symmetric with $W\mathbf{1} = \mathbf{1}$ and $\rho_W := \rho(W - \mathbf{1}\mathbf{1}^T/n) < 1$.
- Let $\Gamma = \gamma W$, with $\gamma = 2/(1 + \sqrt{1-\rho_W^2})$.

Results

- Let $W \in \mathbb{R}^{n\times n}$ be symmetric with $W\mathbf{1} = \mathbf{1}$ and $\rho_W := \rho(W - \mathbf{1}\mathbf{1}^T/n) < 1$.
- Let $\Gamma = \gamma W$, with $\gamma = 2/(1+\sqrt{1-\rho_W^2})$.

Theorem

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, $\boxed{\|x^{(t)} - \bar{b}\mathbf{1}\| \le \frac{4\sqrt{2n}}{2-\gamma}\|(K-K^\infty)^t\|}$.

The asymptotic rate of convergence is given by

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

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

Results

- Let $W \in \mathbb{R}^{n\times n}$ be symmetric with $W\mathbf{1} = \mathbf{1}$ and $\rho_W := \rho(W - \mathbf{1}\mathbf{1}^T/n) < 1$.
- Let $\Gamma = \gamma W$, with $\gamma = 2/(1+\sqrt{1-\rho_W^2})$.

Theorem

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, $\boxed{\|x^{(t)} - \bar{b}\mathbf{1}\| \le \frac{4\sqrt{2n}}{2-\gamma}\|(K-K^\infty)^t\|}$.

The asymptotic rate of convergence is given by

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

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

Same rate as shift-register methods (Ghadimi, Shames, and Johansson, 2013).

Results

- Let $W \in \mathbb{R}^{n\times n}$ be symmetric with $W\mathbf{1} = \mathbf{1}$ and $\rho_W := \rho(W - \mathbf{1}\mathbf{1}^T/n) < 1$.
- Let $\Gamma = \gamma W$, with $\gamma = 2/(1 + \sqrt{1 - \rho_W^2})$.

Theorem

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, $\boxed{\|x^{(t)} - \bar{b}\mathbf{1}\| \leq \frac{4\sqrt{2n}}{2-\gamma}\|(K - K^\infty)^t\|}$.

The asymptotic rate of convergence is given by

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

which satisfies $\frac{1}{2}\sqrt{1/(1-\rho_W)} \leq 1/(1-\rho_K) \leq \sqrt{1/(1-\rho_W)}$.

Same rate as shift-register methods (Ghadimi, Shames, and Johansson, 2013).

(Asymptotic) convergence time $O(D \log D)$.

MAIN MESSAGE:

General toolbox for Min-Sum in convex optimization.

Consensus problem:

- First analysis for Min-Sum Splitting (rate of convergence).
- Quasi-optimal convergence time $O(D \log D)$, improving previous results in (Moallemi, Van Roy, 2006) ($\Theta(D^{2(d-1)/d})$ for $(d/2)$-dimensional grids).
- Connection Min-Sum Splitting and accelerated methods:

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

with $\widehat{K}(\delta, \Gamma) := \begin{pmatrix} (1-\delta)I - (1-\delta)\text{diag}(\Gamma \mathbf{1}) + \delta\Gamma & \delta I \\ \delta I - \delta\text{diag}(\Gamma \mathbf{1}) + (1-\delta)\Gamma & (1-\delta)I \end{pmatrix}$.

"Accelerated Consensus via Min-Sum Splitting," **arXiv:1706.03807** (2017)