Correlation Decay in Random Decision Networks

T. Weber (joint work with D. Gamarnik)

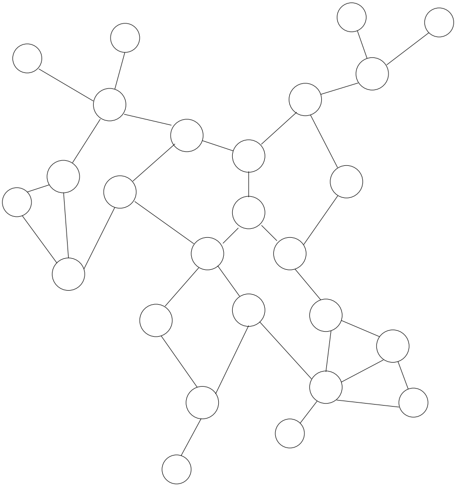
MIT

Stochastic Networks 08

Outline

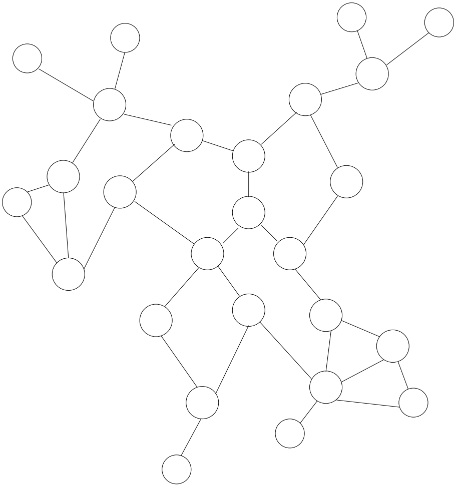
Outline

Optimization on Graphical Models



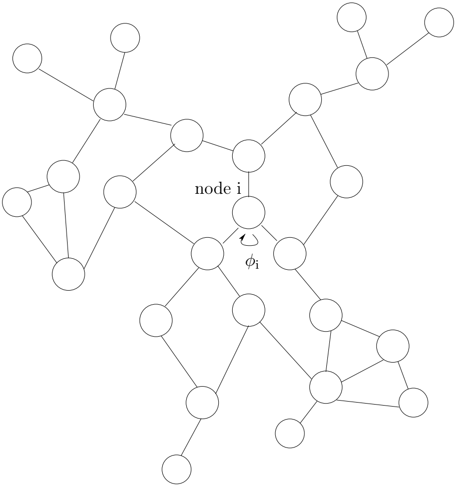
- Graph (V, E)
- In each node i, decision x_i is made.
- Cost functions on nodes $\phi_i(x_i)$ and edges $\phi_{i,j}(x_i, x_j)$
- Objective: find x^* s.t. $f(x^*) = \max_x \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{i,j}(x_i, x_j)$
- Our setting: (Φ_e), (ϕ_i) are random i.i.d functions.

Optimization on Graphical Models



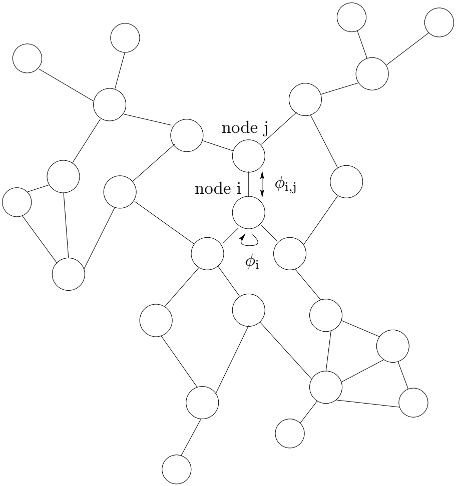
- Graph (V, E)
- In each node i, decision x_i is made.
- Cost functions on nodes $\phi_i(x_i)$ and edges $\phi_{i,j}(x_i, x_j)$
- Objective: find x^* s.t. $f(x^*) = \max_x \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{i,j}(x_i, x_j)$
- Our setting: (Φ_e), (ϕ_i) are random i.i.d functions.

Optimization on Graphical Models



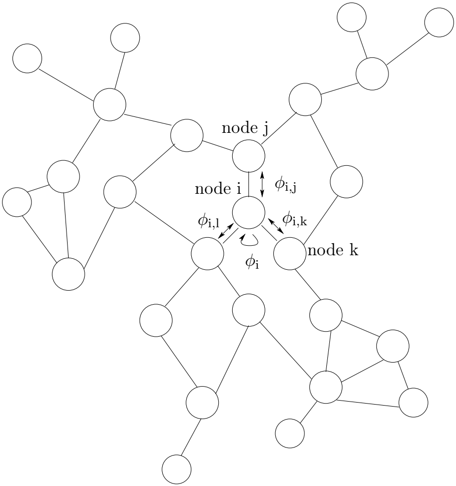
- Graph (V, E)
- In each node i, decision x_i is made.
- Cost functions on nodes $\phi_i(x_i)$ and edges $\phi_{i,j}(x_i, x_j)$
- Objective: find x^* s.t. $f(x^*) = \max_x \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{i,j}(x_i, x_j)$
- Our setting: (Φ_e), (ϕ_i) are random i.i.d functions.

Optimization on Graphical Models



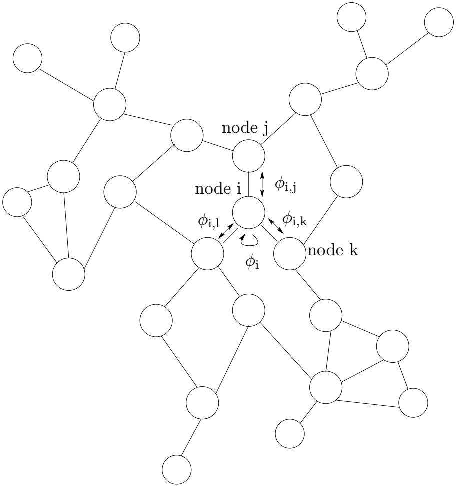
- Graph (V, E)
- In each node i, decision x_i is made.
- Cost functions on nodes $\phi_i(x_i)$ and edges $\phi_{i,j}(x_i, x_j)$
- Objective: find x^* s.t. $f(x^*) = \max_x \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{i,j}(x_i, x_j)$
- Our setting: (Φ_e), (ϕ_i) are random i.i.d functions.

Optimization on Graphical Models



- Graph (V, E)
- In each node i, decision x_i is made.
- Cost functions on nodes $\phi_i(x_i)$ and edges $\phi_{i,j}(x_i, x_j)$
- Objective: find x^* s.t. $f(x^*) = \max_x \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{i,j}(x_i, x_j)$
- Our setting: (Φ_e), (ϕ_i) are random i.i.d functions.

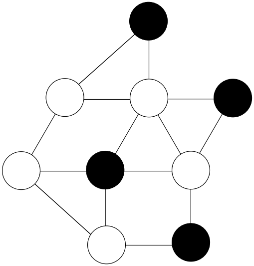
Optimization on Graphical Models



- Graph (V, E)
- In each node i, decision x_i is made.
- Cost functions on nodes $\phi_i(x_i)$ and edges $\phi_{i,j}(x_i, x_j)$
- Objective: find x^* s.t. $f(x^*) = \max_x \sum_{i \in V} \phi_i(x_i) + \sum_{(i,j) \in E} \phi_{i,j}(x_i, x_j)$
- Our setting: (Φ_e), (ϕ_i) are random i.i.d functions.

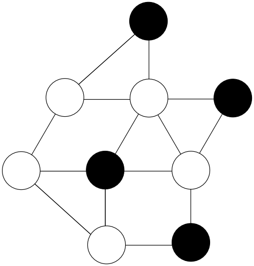
Outline

Combinatorial optimization



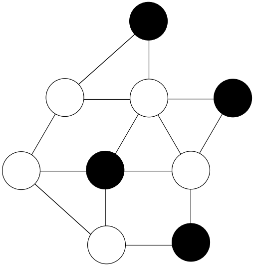
- Graph (V, E), weights $W \in \mathbb{R}^{|V|}$
- Independent set U: $\forall (u, v) \in E, (u \notin U) \cup (v \notin V)$
- Maximum Weigthed Independent Set: find an independent set U which maximizes $\sum_{u \in U} W_u$
- Decision network formulation: $\Phi_e(1, 1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$

Combinatorial optimization



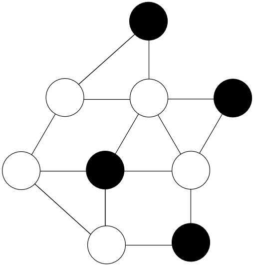
- Graph (V, E), weights $W \in \mathbb{R}^{|V|}$
- Independent set U: $\forall (u, v) \in E, (u \notin U) \cup (v \notin V)$
- Maximum Weigthed Independent Set: find an independent set U which maximizes $\sum_{u \in U} W_u$
- Decision network formulation: $\Phi_e(1, 1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$

Combinatorial optimization



- Graph (V, E), weights $W \in \mathbb{R}^{|V|}$
- Independent set U: $\forall (u, v) \in E, (u \notin U) \cup (v \notin V)$
- Maximum Weigthed Independent Set: find an independent set U which maximizes $\sum_{u \in U} W_u$
- Decision network formulation: $\Phi_e(1, 1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$

Combinatorial optimization



- Graph (V, E), weights $\mathrm{W} \in \mathbb{R}^{|\mathrm{V}|}$
- Independent set U: $\forall(\mathrm{u}, \mathrm{v}) \in \mathrm{E}, (\mathrm{u} \notin \mathrm{U}) \cup (\mathrm{v} \notin \mathrm{V})$
- Maximum Weigthed Independent Set: find an independent set U which maximizes $\sum_{\mathrm{u} \in \mathrm{U}} \mathrm{W}_{\mathrm{u}}$
- Decision network formulation: $\Phi_{\mathrm{e}}(1, 1) = -\infty$ for all $\mathrm{e} \in \mathrm{E}$, and 0 otherwise. $\Phi_{\mathrm{u}}(1) = \mathrm{W}_{\mathrm{u}}$ for all $\mathrm{u} \in \mathrm{U}$

Some hardness facts

- NP-hard, even for $\Delta = 3$, where $\Delta = \max_{v \in V} |\mathcal{N}_v|$
- Hastad[1996] NP-hard to approximate within

$$n^{\alpha}, \forall \alpha < 1, |V| = n$$

- Trevisan[2001] NP-hard to approximate to within

$$\frac{\Delta}{2^{O(\sqrt{\log \Delta})}}$$

- Berman & Karpinski[1999]. NP-hard to approximate within

$$1.0071 \quad \text{for } \Delta = 3$$

$$1.0136 \quad \text{for } \Delta = 4$$

$$1.0148 \quad \text{for } \Delta = 5$$

Some hardness facts

- NP-hard, even for $\Delta = 3$, where $\Delta = \max_{v \in V} |\mathcal{N}_v|$
- Hastad[1996] NP-hard to approximate within

$$n^{\alpha}, \forall \alpha < 1, |V| = n$$

- Trevisan[2001] NP-hard to approximate to within

$$\frac{\Delta}{2^{O(\sqrt{\log \Delta})}}$$

- Berman & Karpinski[1999]. NP-hard to approximate within

$$1.0071 \quad \text{for } \Delta = 3$$

$$1.0136 \quad \text{for } \Delta = 4$$

$$1.0148 \quad \text{for } \Delta = 5$$

Some hardness facts

- NP-hard, even for $\Delta = 3$, where $\Delta = \max_{v \in V} |\mathcal{N}_v|$
- Hastad[1996] NP-hard to approximate within

$$n^{\alpha}, \forall \alpha < 1, |V| = n$$

- Trevisan[2001] NP-hard to approximate to within

$$\frac{\Delta}{2^{O(\sqrt{\log \Delta})}}$$

- Berman & Karpinski[1999]. NP-hard to approximate within

$$1.0071 \quad \text{for } \Delta = 3$$

$$1.0136 \quad \text{for } \Delta = 4$$

$$1.0148 \quad \text{for } \Delta = 5$$

Some hardness facts

- NP-hard, even for $\Delta = 3$, where $\Delta = \max_{v \in V} |\mathcal{N}_v|$
- Hastad[1996] NP-hard to approximate within

$$n^{\alpha}, \forall \alpha < 1, |V| = n$$

- Trevisan[2001] NP-hard to approximate to within

$$\frac{\Delta}{2^{O(\sqrt{\log \Delta})}}$$

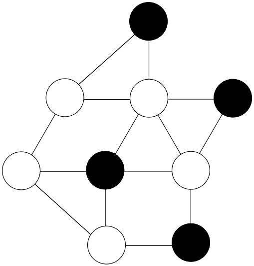
- Berman & Karpinski[1999]. NP-hard to approximate within

$$1.0071 \quad \text{for } \Delta = 3$$

$$1.0136 \quad \text{for } \Delta = 4$$

$$1.0148 \quad \text{for } \Delta = 5$$

Contribution



Theorem

Suppose W_u are independent exponential random variables, and $\Delta \leq 3$. Then there exists an EPRAS for solving the MWIS problem: In $O(|V|2^{\epsilon^{-1.1}})$ time, one can find U such that $1 - \epsilon \leq \frac{W^* - W_U}{W^*} \leq 1$

In fact, for any Δ, one can find a distribution such that there exists a FPTAS.

Statistical Inference

- Family of random variables $(X_1, \ldots, X_n)$ with factored probability distribution $P(X) = \prod_{(i,j) \in E} \Psi_{i,j}(x_i, x_j) \prod_{i \in V} \Psi_i(x_i)$
- Maximum likelihood problem: Find X which maximizes P(X). Equivalent to optimization problem on a graph, with $\Phi_{i,j} = \log(\Psi_{i,j})$ and $\Phi_i = \log(\Psi_i)$.
- If $(X_i)_{i \in S}$, where $S \subset \{1, \ldots, n\}$ are known :
  Maximum likelihood $\rightarrow$ Maximum a posteriori problem
- Edge cost functions are modified by the values of $(X_i)_{i \in S}$

Statistical Inference

- Family of random variables $(X_1, \ldots, X_n)$ with factored probability distribution $P(X) = \prod_{(i,j)\in E} \Psi_{i,j}(x_i, x_j) \prod_{i\in V} \Psi_i(x_i)$
- Maximum likelihood problem: Find X which maximizes P(X). Equivalent to optimization problem on a graph, with $\Phi_{i,j} = \log(\Psi_{i,j})$ and $\Phi_i = \log(\Psi_i)$.
- If $(X_i)_{i\in S}$, where $S \subset \{1, \ldots, n\}$ are known :
  Maximum likelihood → Maximum a posteriori problem
- Edge cost functions are modified by the values of $(X_i)_{i\in S}$

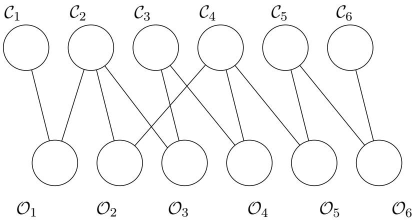
Statistical Inference

- Family of random variables $(X_1, \dots, X_n)$ with factored probability distribution $P(X) = \prod_{(i,j)\in E} \Psi_{i,j}(x_i, x_j) \prod_{i \in V} \Psi_i(x_i)$
- Maximum likelihood problem: Find X which maximizes P(X). Equivalent to optimization problem on a graph, with $\Phi_{i,j} = \log(\Psi_{i,j})$ and $\Phi_i = \log(\Psi_i)$.
- If $(X_i)_{i \in S}$, where $S \subset \{1, \dots, n\}$ are known : Maximum likelihood $\rightarrow$ Maximum a posteriori problem
- Edge cost functions are modified by the values of $(X_i)_{i \in S}$

Statistical Inference

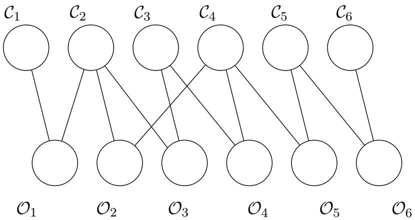
- Family of random variables $(X_1, \ldots, X_n)$ with factored probability distribution $P(X) = \prod_{(i,j) \in E} \Psi_{i,j}(x_i, x_j) \prod_{i \in V} \Psi_i(x_i)$
- Maximum likelihood problem: Find X which maximizes P(X). Equivalent to optimization problem on a graph, with $\Phi_{i,j} = \log(\Psi_{i,j})$ and $\Phi_i = \log(\Psi_i)$.
- If $(X_i)_{i \in S}$, where $S \subset \{1, \ldots, n\}$ are known :
  Maximum likelihood → Maximum a posteriori problem
- Edge cost functions are modified by the values of $(X_i)_{i \in S}$

Example



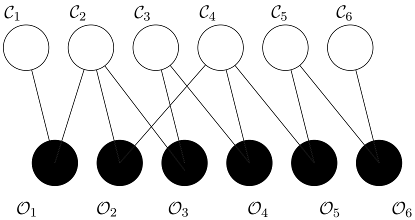
- Probability model: random variable $(\mathcal{C}_\mathrm{i})_{\mathrm{i}=1,\ldots,\mathrm{n}}$ and $(\mathcal{O})_{\mathrm{i}=1,\ldots,\mathrm{m}}$. Each $\mathcal{O}_\mathrm{i}$ is connected to a set of parents $\mathcal{P}(\mathrm{i})$ (with $|\mathcal{P}(\mathrm{i})| \leq 2$)
- Probability distribution $\mathrm{P}(\mathcal{C}, \mathcal{O}) = \prod_{\mathrm{i}=1,\ldots,\mathrm{n}} \mathrm{P}(\mathcal{C}_\mathrm{i}) \prod_{\mathrm{k}=1,\ldots,\mathrm{m}} \mathrm{P}(\mathcal{O}_\mathrm{k} | \mathcal{C}_\mathrm{j}, \mathrm{j} \in \mathcal{P}(\mathrm{k}))$
- MAP is equivalent to a maximization within our framework, with $\mathrm{V} = \{\mathcal{C}_\mathrm{i}\}$, $\mathrm{E} = \{\mathcal{P}(\mathrm{i})\}$, $\Psi_\mathrm{i} = \mathrm{P}(\mathcal{C}_\mathrm{i})$ and for any two i, j parents of obersvation k, $\Psi_{\mathrm{i},\mathrm{j}} = \log(\mathrm{P}(\mathcal{O}_\mathrm{k} | \mathcal{C}_\mathrm{j}, \mathrm{j} \in \mathcal{P}(\mathrm{k})))$.
- Since $\mathcal{O}$ are random, $\Psi_{\mathrm{i},\mathrm{j}}$ is naturally random as well; but the structure of the network is not.

Example



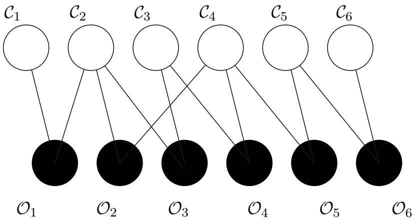
- Probability model: random variable $(\mathcal{C}_{\mathrm{i}})_{\mathrm{i}=1,\ldots,\mathrm{n}}$ and $(\mathcal{O})_{\mathrm{i}=1,\ldots,\mathrm{m}}$. Each $\mathcal{O}_{\mathrm{i}}$ is connected to a set of parents $\mathcal{P}(\mathrm{i})$ (with $|\mathcal{P}(\mathrm{i})| \leq 2$)
- Probability distribution $\mathrm{P}(\mathcal{C}, \mathcal{O}) = \prod_{\mathrm{i}=1,\ldots,\mathrm{n}} \mathrm{P}(\mathcal{C}_{\mathrm{i}}) \prod_{\mathrm{k}=1,\ldots,\mathrm{m}} \mathrm{P}(\mathcal{O}_{\mathrm{k}} | \mathcal{C}_{\mathrm{j}}, \mathrm{j} \in \mathcal{P}(\mathrm{k}))$
- MAP is equivalent to a maximization within our framework, with $\mathrm{V} = \{\mathcal{C}_{\mathrm{i}}\}$, $\mathrm{E} = \{\mathcal{P}(\mathrm{i})\}$, $\Psi_{\mathrm{i}} = \mathrm{P}(\mathcal{C}_{\mathrm{i}})$ and for any two i, j parents of obersvation k, $\Psi_{\mathrm{i,j}} = \log(\mathrm{P}(\mathcal{O}_{\mathrm{k}} | \mathcal{C}_{\mathrm{j}}, \mathrm{j} \in \mathcal{P}(\mathrm{k})))$.
- Since $\mathcal{O}$ are random, $\Psi_{\mathrm{i,j}}$ is naturally random as well; but the structure of the network is not.

Example



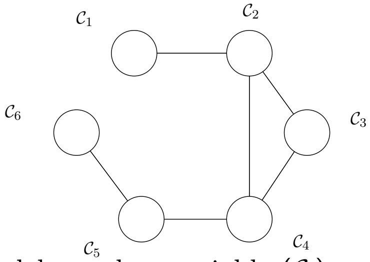
- Probability model: random variable $(\mathcal{C}_i)_{i=1,\dots,n}$ and $(\mathcal{O})_{i=1,\dots,m}$. Each $\mathcal{O}_i$ is connected to a set of parents $\mathcal{P}(i)$ (with $|\mathcal{P}(i)| \leq 2$)
- Probability distribution $P(\mathcal{C},\mathcal{O}) = \prod_{i=1,\dots,n} P(\mathcal{C}_i) \prod_{k=1,\dots,m} P(\mathcal{O}_k|\mathcal{C}_j, j \in \mathcal{P}(k))$
- MAP is equivalent to a maximization within our framework, with $V = \{\mathcal{C}_i\}$, $E = \{\mathcal{P}(i)\}$, $\Psi_i = P(\mathcal{C}_i)$ and for any two i, j parents of obersvation k, $\Psi_{i,j} = \log(P(\mathcal{O}_k|\mathcal{C}_j, j \in \mathcal{P}(k)))$.
- Since $\mathcal{O}$ are random, $\Psi_{i,j}$ is naturally random as well; but the structure of the network is not.

Example



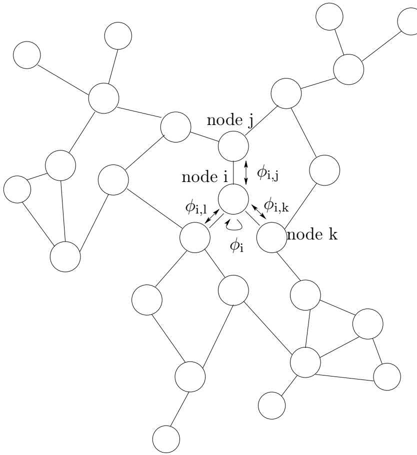
- Probability model: random variable $(\mathcal{C}_\mathrm{i})_{\mathrm{i}=1,\ldots,\mathrm{n}}$ and $(\mathcal{O})_{\mathrm{i}=1,\ldots,\mathrm{m}}$. Each $\mathcal{O}_\mathrm{i}$ is connected to a set of parents $\mathcal{P}(\mathrm{i})$ (with $|\mathcal{P}(\mathrm{i})| \leq 2$)
- Probability distribution $\mathrm{P}(\mathcal{C}, \mathcal{O}) = \prod_{\mathrm{i}=1,\ldots,\mathrm{n}} \mathrm{P}(\mathcal{C}_\mathrm{i}) \prod_{\mathrm{k}=1,\ldots,\mathrm{m}} \mathrm{P}(\mathcal{O}_\mathrm{k} | \mathcal{C}_\mathrm{j}, \mathrm{j} \in \mathcal{P}(\mathrm{k}))$
- MAP is equivalent to a maximization within our framework, with $\mathrm{V} = \{\mathcal{C}_\mathrm{i}\}$, $\mathrm{E} = \{\mathcal{P}(\mathrm{i})\}$, $\Psi_\mathrm{i} = \mathrm{P}(\mathcal{C}_\mathrm{i})$ and for any two i, j parents of obersvation k, $\Psi_{\mathrm{i},\mathrm{j}} = \log(\mathrm{P}(\mathcal{O}_\mathrm{k} | \mathcal{C}_\mathrm{j}, \mathrm{j} \in \mathcal{P}(\mathrm{k})))$.
- Since $\mathcal{O}$ are random, $\Psi_{\mathrm{i},\mathrm{j}}$ is naturally random as well; but the structure of the network is not.

Example



- Probability model: random variable $(\mathcal{C}_i)_{i=1,\dots,n}$ and $(\mathcal{O})_{i=1,\dots,m}$. Each $\mathcal{O}_i$ is connected to a set of parents $\mathcal{P}(i)$ (with $|\mathcal{P}(i)| \leq 2$)
- Probability distribution $P(\mathcal{C},\mathcal{O}) = \prod_{i=1,\dots,n} P(\mathcal{C}_i) \prod_{k=1,\dots,m} P(\mathcal{O}_k|\mathcal{C}_j, j \in \mathcal{P}(k))$
- MAP is equivalent to a maximization within our framework, with $V = \{\mathcal{C}_i\}$, $E = \{\mathcal{P}(i)\}$, $\Psi_i = P(\mathcal{C}_i)$ and for any two i, j parents of obersvation k, $\Psi_{i,j} = \log(P(\mathcal{O}_k|\mathcal{C}_j, j \in \mathcal{P}(k)))$.
- Since $\mathcal{O}$ are random, $\Psi_{i,j}$ is naturally random as well; but the structure of the network is not.

Decentralized Optimization and Objectives



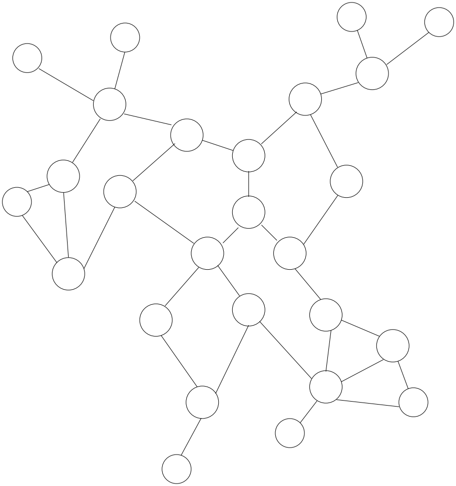
- Each node i represents an agent which makes a decision x_i
- Agents try to collaborate towards a common objective

Decentralized optimization:

How to find x_i^* for each i, using only <u>local information</u>?

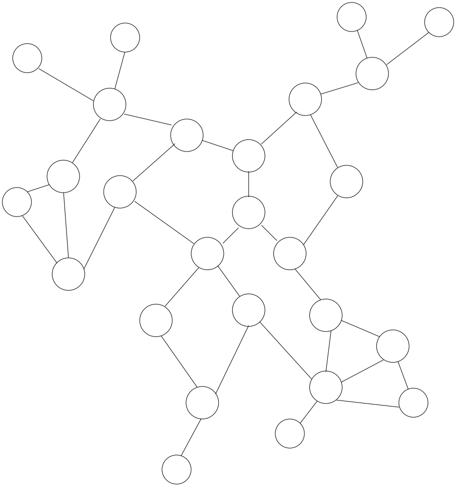
Outline

Correlation decay



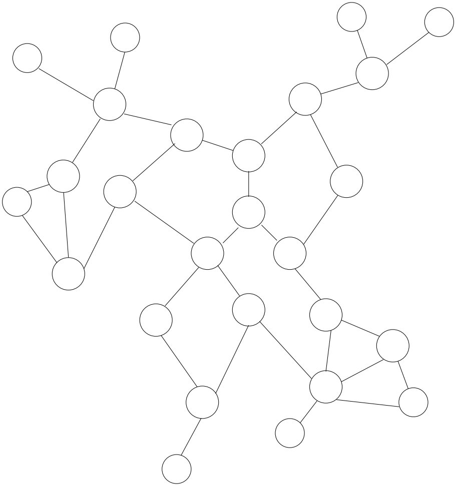
- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Correlation decay



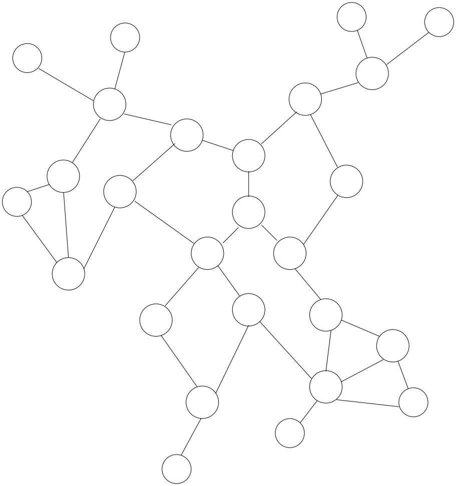
- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Correlation decay



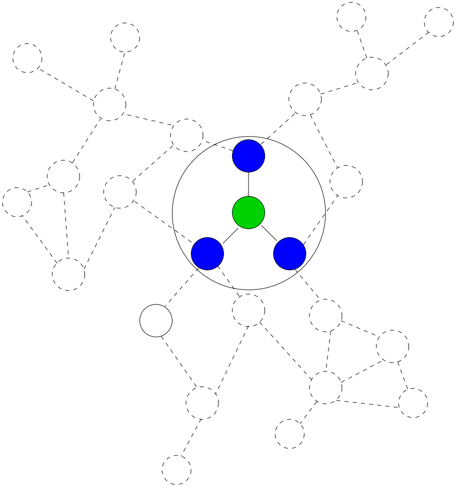
- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Correlation decay



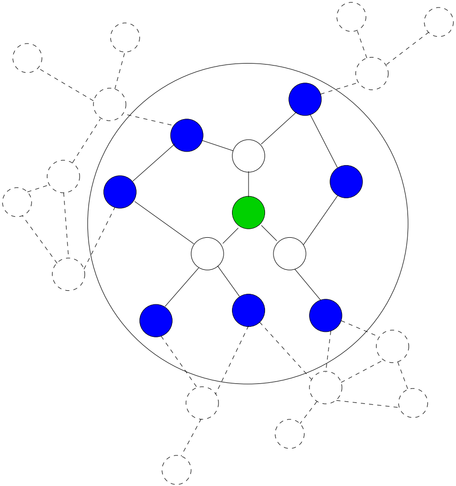
- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Correlation decay



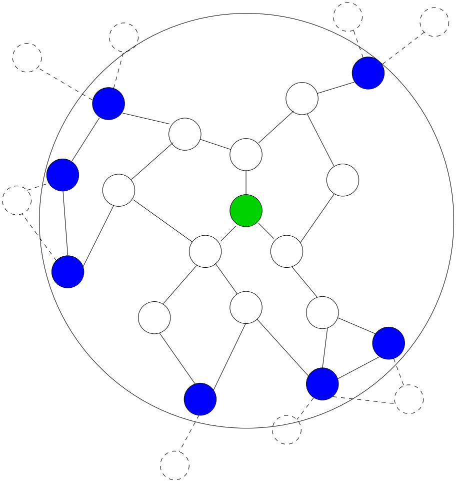
- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Correlation decay



- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Correlation decay



- $Z_i(x_i)$ quantity of interest to be computed. Example: marginal probability, count, value function.
- Boundary condition at distance k: $B_k = \{x_j, j : d(i,j) \geq k\}$.
- Correlation decay: $Z_i(x_i|B_k) \approx Z_i(x_i)$ as $k \to \infty$.

Contributions

- Cavity Expansion Algorithm (CEA) for optimization on general decision networks
- Correlation Decay implies near-optimality in poly-time for CEA
- Sufficient conditions for correlation decay
- PTAS for Independent set with Exp(1) weights

Outline

Related work

Decentralized decision making:

- Rusmevichientong & Van Roy (01,03)

Optimization in graphical models:

- Moallemi & Van Roy (06,07)
- Bayati, Sharma, Shah (06)
- Sanghavi & Shah (08)
- Jung & Shah (07)

Bonus-type recursion (“Objective method”)

- Aldous&Steel (02)

Related work

Computation trees, correlation decay

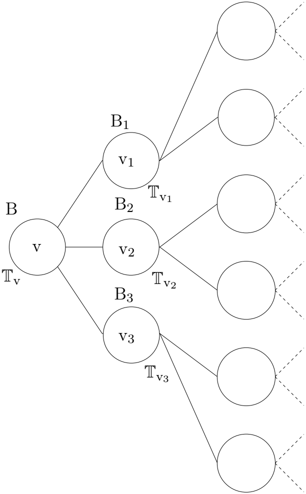
- Bayati, Gamarnik, Katz, Nair, Tetali (07)
- Gamarnik&Katz (07)
- Gamarnik, Nowicki, Swirscsz (05)
- Weitz (06)
- Nair&Tetali (07)

Statistical Physics and cavity algorithm

- Montanari
- Mezard and Montanari (book)

Outline

Bonus and cavity recursion on trees



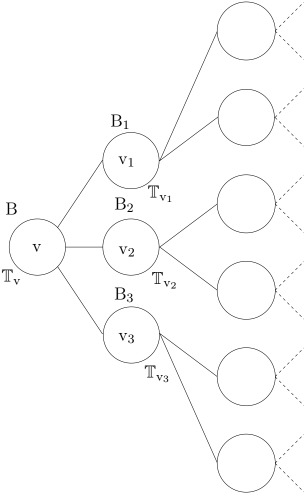
- $\mathbb{T}_v$ subtree rooted at v. $J_v(x)$ the optimal value when decision in v is forced to be x.
- Recursion: $J_v(1) = \phi_v(1) + \sum_1^d pJ(v \leftarrow v_i)$

$$pJ(v \leftarrow v_i) = \max(\quad \Phi_{v,v_i}(1,1) + J_{v_i}(1), \quad \Phi_{v,v_i}(1,0) + J_{v_i}(0))$$

- Bonus: $B_v \triangleq J_v(1) - J_v(0)$.
- $B_v = \phi_v(1) - \phi_v(0) + \sum_1^d \mu(v \leftarrow v_i)$ where μ (partial bonus or message) is:

$$\mu(v \leftarrow v_i) = \max(\Phi_{v,v_i}(1,1) + B_{v_i}, \Phi_{v,v_i}(1,0)) - \max(\Phi_{v,v_i}(0,1) + B_{v_i}, \Phi_{v,v_i}(0,0))$$

Bonus and cavity recursion on trees



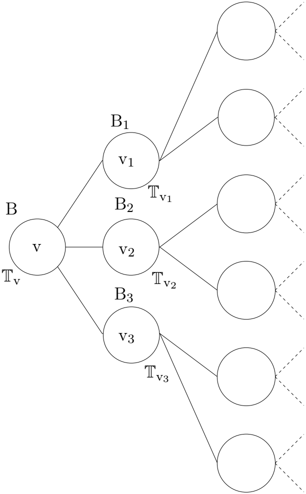
- $\mathbb{T}_v$ subtree rooted at v. $J_v(x)$ the optimal value when decision in v is forced to be x.
- Recursion: $J_v(1) = \phi_v(1) + \sum_1^d pJ(v \leftarrow v_i)$

$$pJ(v \leftarrow v_i) = \max\big(\Phi_{v,v_i}(1,1) + J_{v_i}(1), \; \Phi_{v,v_i}(1,0) + J_{v_i}(0)\big)$$

- Bonus: $B_v \triangleq J_v(1) - J_v(0)$.
- $B_v = \phi_v(1) - \phi_v(0) + \sum_1^d \mu(v \leftarrow v_i)$ where μ (partial bonus or message) is:

$$\mu(v \leftarrow v_i) = \max(\Phi_{v,v_i}(1,1) + B_{v_i}, \Phi_{v,v_i}(1,0)) - \max(\Phi_{v,v_i}(0,1) + B_{v_i}, \Phi_{v,v_i}(0,0))$$

Bonus and cavity recursion on trees



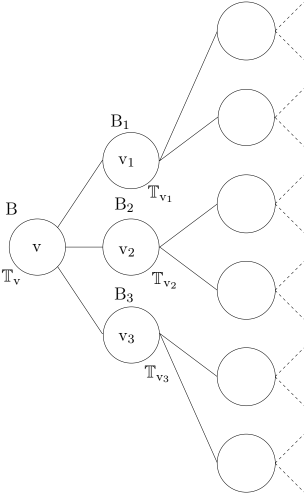
- $\mathbb{T}_v$ subtree rooted at v. $J_v(x)$ the optimal value when decision in v is forced to be x.
- Recursion: $J_v(1) = \phi_v(1) + \sum_1^d pJ(v \leftarrow v_i)$

$$pJ(v \leftarrow v_i) = \max(\quad \Phi_{v,v_i}(1,1) + J_{v_i}(1), \quad \Phi_{v,v_i}(1,0) + J_{v_i}(0))$$

- Bonus: $B_v \triangleq J_v(1) - J_v(0)$.
- $B_v = \phi_v(1) - \phi_v(0) + \sum_1^d \mu(v \leftarrow v_i)$ where μ (partial bonus or message) is:

$$\mu(v \leftarrow v_i) = \max(\Phi_{v,v_i}(1,1) + B_{v_i}, \Phi_{v,v_i}(1,0)) - \max(\Phi_{v,v_i}(0,1) + B_{v_i}, \Phi_{v,v_i}(0,0))$$

Bonus and cavity recursion on trees



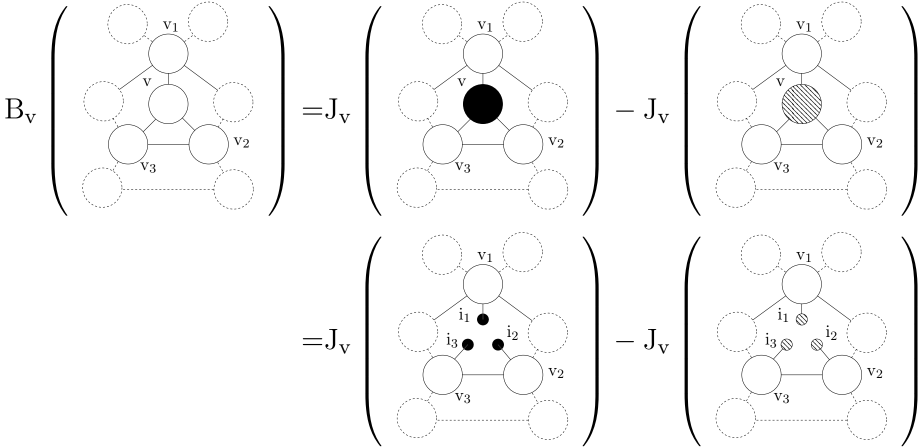
- $\mathbb{T}_v$ subtree rooted at v. $J_v(x)$ the optimal value when decision in v is forced to be x.
- Recursion: $J_v(1) = \phi_v(1) + \sum_1^d pJ(v \leftarrow v_i)$

$$pJ(v \leftarrow v_i) = \max(\quad \Phi_{v,v_i}(1,1) + J_{v_i}(1), \quad \Phi_{v,v_i}(1,0) + J_{v_i}(0))$$

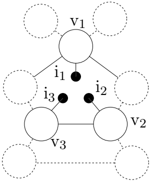
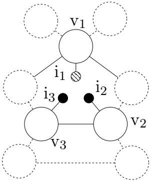
- Bonus: $B_v \triangleq J_v(1) - J_v(0)$.
- $B_v = \phi_v(1) - \phi_v(0) + \sum_1^d \mu(v \leftarrow v_i)$ where μ (partial bonus or message) is:

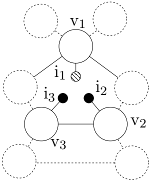
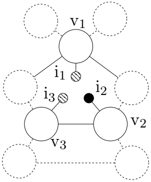
$$\mu(v \leftarrow v_i) = \max(\Phi_{v,v_i}(1,1) + B_{v_i}, \Phi_{v,v_i}(1,0)) - \max(\Phi_{v,v_i}(0,1) + B_{v_i}, \Phi_{v,v_i}(0,0))$$

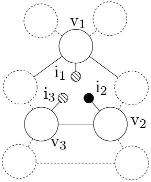
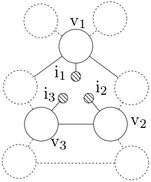
The cavity recursion on general graphs



The cavity recursion(continued)

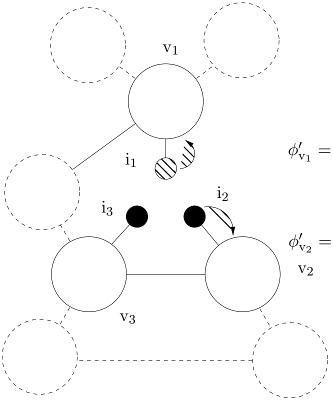
$B_v = J_v$ () $- J_v$ ()

$+J_v$ () $- J_v$ ()

$+J_v$ () $- J_v$ ()

The cavity recursion(continued)

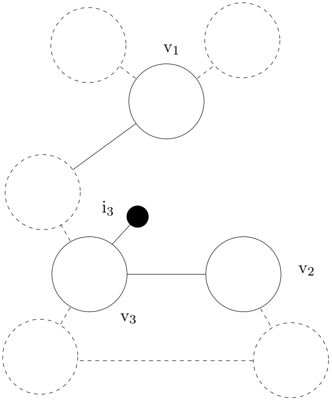
Note:



$$\phi'_{v_1} = \phi_{v_1} + \phi_{v,v_1}(0, x_{v_1})$$

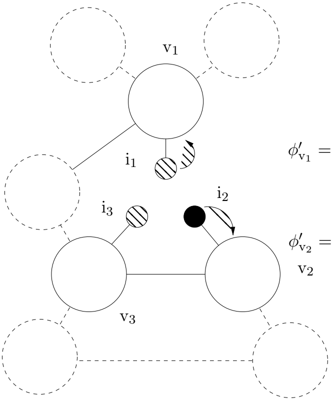
$$\Longleftrightarrow$$

$$\phi'_{v_2} = \phi_{v_2} + \phi_{v,v_2}(1, x_{v_2})$$



The cavity recursion(continued)

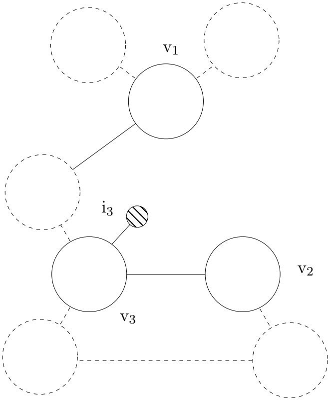
And:



$$\phi'_{v_1} = \phi_{v_1} + \phi_{v,v_1}(0, x_{v_1})$$

$$\Longleftrightarrow$$

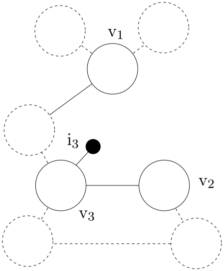
$$\phi'_{v_2} = \phi_{v_2} + \phi_{v,v_2}(1, x_{v_2})$$



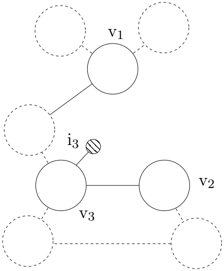
The cavity recursion(continued)

Defining $\mu(\mathrm{v} \leftarrow \mathrm{v}_3)$ as

$\mathrm{J_v}\Bigg($



$\Bigg) - \mathrm{J_v}\Bigg($



$\Bigg)$

We obtain a recursion:

$$\mathrm{B_v} = \phi_\mathrm{v}(1) - \phi_\mathrm{v}(0) + \sum_{1}^{\mathrm{d}} \mu(\mathrm{v} \leftarrow \mathrm{v_i})$$

If $\mathrm{B}_{\mathrm{v_i}}$ is the bonus of $\mathrm{v_i}$ in the modified graph, then we have the partial bonus equation:

$$\begin{aligned}\mu(\mathrm{v} \leftarrow \mathrm{v_i}) &= \max(\Phi_{\mathrm{v},\mathrm{v_i}}(1,1) + \mathrm{B}_{\mathrm{v_i}}, \Phi_{\mathrm{v},\mathrm{v_i}}(1,0)) \\ &\quad - \max(\Phi_{\mathrm{v},\mathrm{v_i}}(0,1) + \mathrm{B}_{\mathrm{v_i}}, \Phi_{\mathrm{v},\mathrm{v_i}}(0,0))\end{aligned}$$

Outline

Correlation decay

Definition (Correlation decay)

Let B be the bonus in the node, and B^r the approximation of the bonus resulting from the depth r cavity recursion. We say that correlation decay occurs if there exists $K_c \geq 0, \alpha_c < 1$ such that

$$E|B^r - B| \leq K_c \, \alpha_c^r$$

Theorem (Correlation decay implies decentralized optimum)

Suppose the system exhibits correlation decay with parameters K_c, α_c and has bounded degree Δ. Consider $\epsilon > 0$. If $r > P(|V|, \ln(\epsilon), \ln(\frac{1}{1-\alpha_c}))$ and $x_i = 1_{B^r > 0}$ (computable in poly time), then

$$P(x \text{ is optimal}) > 1 - \epsilon$$

Correlation decay

Definition (Correlation decay)

Let B be the bonus in the node, and B^r the approximation of the bonus resulting from the depth r cavity recursion. We say that correlation decay occurs if there exists $K_c \geq 0, \alpha_c < 1$ such that

$$E|B^r - B| \leq K_c \, \alpha_c^r$$

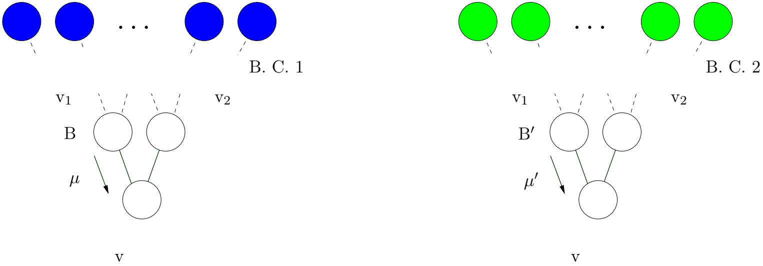
Theorem (Correlation decay implies decentralized optimum)

Suppose the system exhibits correlation decay with parameters K_c, α_c and has bounded degree Δ. Consider $\epsilon > 0$. If $r > P(|V|, \ln(\epsilon), \ln(\frac{1}{1-\alpha_c}))$ and $x_i = 1_{B^r > 0}$ (computable in poly time), then

$$P(x \text{ is optimal}) > 1 - \epsilon$$

Outline

Coupling technique



Definition

The network experiences coupling with parameters (a, b) if for any edge of the computation tree, we have

$$P(\mu = \mu' \mid B, B') \geq (1 - a) - b|B - B'|$$

Theorem

Suppose that the network exhibits coupling with parameters (a, b) and has degree bounded by Δ. Then, if

$$a(\Delta - 1) + \sqrt{4bK}(\Delta - 1)^{3/2} < 1 \tag{1}$$

the network experiences exponential correlation decay.
Supose additionally that the network is triangle free and that for cost functions are symmetric. Then, if

$$(\Delta - 1)(a + \sqrt{4bK}) < 1 \tag{2}$$

the network experiences exponential correlation decay.

Outline

Gaussian costs

- Example: zero cost on nodes, Gaussian cost on edges:

$$f \sim \mathcal{N}(\vec{\mu} = \begin{pmatrix} \mu_{00} = 0 \\ \mu_{01} = 0 \\ \mu_{10} = 0 \\ \mu_{11} = 0 \end{pmatrix}, C = \begin{pmatrix} 1 & \alpha & \alpha & -\alpha \\ \alpha & 1 & -\alpha & \alpha \\ \alpha & -\alpha & 1 & \alpha \\ -\alpha & \alpha & \alpha & 1 \end{pmatrix})$$

Theorem (Sufficient condition for correlation decay)

Suppose

$$\frac{1}{3} - \frac{2\pi}{9\Delta} < \alpha < 1/3$$

Then decay of correlation occurs.

Gaussian costs

More generally:

- Let $X = \Phi(1,1) - \Phi(0,0) - \Phi(0,1) + \Phi(1,0))$,
  $Y = (\Phi(1,1) + \Phi(0,0) - \Phi(0,1) - \Phi(1,0))$.
- Define $\beta = \beta_1 + \beta_2$, where:

$$\beta_1 = \frac{2}{\pi\sqrt{1-\rho^2}} \frac{\sigma_Y}{\sigma_X}$$

and

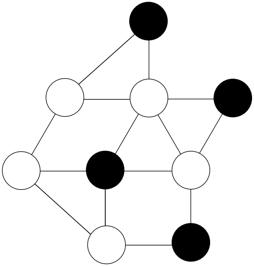
$$\beta_2 = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{(1-\rho^2)}} \frac{E|Y|}{\sigma_X}$$

Theorem (Condition for correlation decay)

Then correlation decay occurs provided

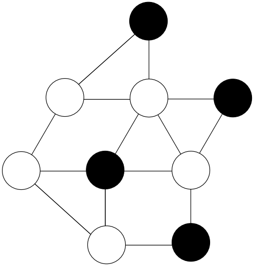
$$\Delta(\beta + \sqrt{\Delta\beta}) < 1$$

Cavity and bonus recursion



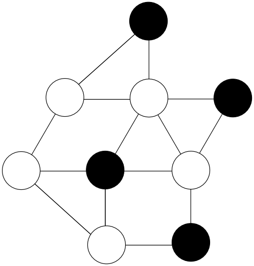
- Decision network formulation: $\Phi_e(1,1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$
- Message equation:
  $\mu(B_j) = \max(\Phi_j(1,1) + B_j, \Phi_j(1,0)) - \max(\Phi_j(0,1) + B_j, \Phi_j(0,0)$
- For IS: $\mu(B_j) = -\max(0, B_j)$ and so $B = W - \sum_j \max(0, B_j)$
- Let $\tilde{B} = \max(0, B)$. The equation becomes $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$

Cavity and bonus recursion



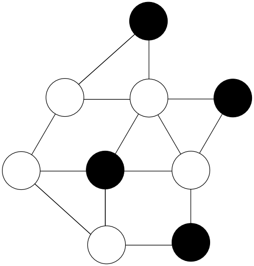
- Decision network formulation: $\Phi_e(1,1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$
- Message equation:
  $\mu(B_j) = \max(\Phi_j(1,1) + B_j, \Phi_j(1,0)) - \max(\Phi_j(0,1) + B_j, \Phi_j(0,0)$
- For IS: $\mu(B_j) = -\max(0, B_j)$ and so $B = W - \sum_j \max(0, B_j)$
- Let $\tilde{B} = \max(0, B)$. The equation becomes $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$

Cavity and bonus recursion



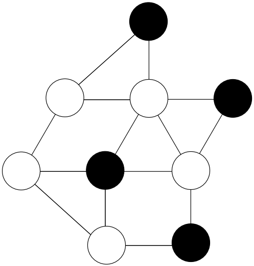
- Decision network formulation: $\Phi_e(1,1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$
- Message equation:
  $\mu(B_j) = \max(\Phi_j(1,1) + B_j, \Phi_j(1,0)) - \max(\Phi_j(0,1) + B_j, \Phi_j(0,0))$
- For IS: $\mu(B_j) = -\max(0, B_j)$ and so $B = W - \sum_j \max(0, B_j)$
- Let $\tilde{B} = \max(0, B)$. The equation becomes $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$

Cavity and bonus recursion



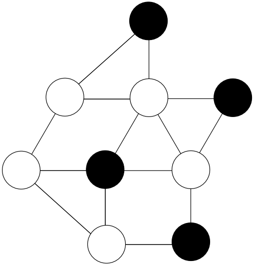
- Decision network formulation: $\Phi_e(1,1) = -\infty$ for all $e \in E$, and 0 otherwise. $\Phi_u(1) = W_u$ for all $u \in U$
- Message equation:
  $\mu(B_j) = \max(\Phi_j(1,1) + B_j, \Phi_j(1,0)) - \max(\Phi_j(0,1) + B_j, \Phi_j(0,0)$
- For IS: $\mu(B_j) = -\max(0, B_j)$ and so $B = W - \sum_j \max(0, B_j)$
- Let $\tilde{B} = \max(0, B)$. The equation becomes $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$

Density evolution



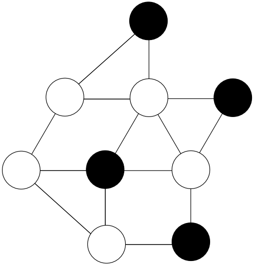
- $\tilde{B} = \max(0, W - \sum_j \tilde{B_j})$
- W and $\sum_j \tilde{B}_j$ are independent, even for general graphs (subtree independence property)
- Assume $W \sim \text{Exp}(1)$. Conditional on $\sum_j \tilde{B}_j = x$, if $W < x$, then $\tilde{B} = 0$, otherwise, $\tilde{B} = W - x$ is exponential
- Therefore, $\tilde{B}$ is always the mixture of a mass in 0 and an exponential random variable

Density evolution



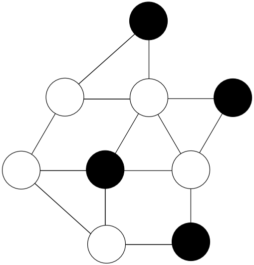
- $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$
- W and $\sum_j \tilde{B}_j$ are independent, even for general graphs (subtree independence property)
- Assume $W \sim \text{Exp}(1)$. Conditional on $\sum_j \tilde{B}_j = x$, if $W < x$, then $\tilde{B} = 0$, otherwise, $\tilde{B} = W - x$ is exponential
- Therefore, $\tilde{B}$ is always the mixture of a mass in 0 and an exponential random variable

Density evolution



- $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$
- W and $\sum_j \tilde{B}_j$ are independent, even for general graphs (subtree independence property)
- Assume $W \sim \mathrm{Exp}(1)$. Conditional on $\sum_j \tilde{B}_j = x$, if $W < x$, then $\tilde{B} = 0$, otherwise, $\tilde{B} = W - x$ is exponential
- Therefore, $\tilde{B}$ is always the mixture of a mass in 0 and an exponential random variable

Density evolution



- $\tilde{B} = \max(0, W - \sum_j \tilde{B}_j)$
- W and $\sum_j \tilde{B}_j$ are independent, even for general graphs (subtree independence property)
- Assume $W \sim \text{Exp}(1)$. Conditional on $\sum_j \tilde{B}_j = x$, if $W < x$, then $\tilde{B} = 0$, otherwise, $\tilde{B} = W - x$ is exponential
- Therefore, $\tilde{B}$ is always the mixture of a mass in 0 and an exponential random variable

Correlation decay in Independent Sets

- Let v be a node, and v_1, v_2 its neighbors in the graph.
- $\tilde{B}_v = \max(0, W - \tilde{B}_{v_1} - \tilde{B}_{v_2})$
- Let $M_v = \exp(-B_v)$ and $x = \tilde{B}_{v_1} + \tilde{B}_{v_2}$
- Then:

$$
\begin{aligned}
E[M_v|x] =& P(W_v \leq x)e^0 + P(W_v > x)E[e^{W_v - x}|W_v > x] \\
=& (1 - e^{-x}) + e^{-x}\frac{1}{2} \\
=& 1 - \frac{1}{2}e^{-x} \\
E[M_v] =& 1 - \frac{1}{2}E[M_{v_1}M_{v_2}]
\end{aligned}
$$

- General graph: $M_{v_1}M_{v_2}$ are not independent

Correlation decay in Independent Sets

- Let v be a node, and v_1, v_2 its neighbors in the graph.
- $\tilde{B_v} = \max(0, W - \tilde{B_{v_1}} - \tilde{B_{v_2}})$
- Let $M_v = \exp(-B_v)$ and $x = \tilde{B_{v_1}} + \tilde{B_{v_2}}$
- Then:

$$\begin{aligned}
E[M_v|x] =& P(W_v \leq x)e^0 + P(W_v > x)E[e^{W_v - x}|W_v > x] \\
=& (1 - e^{-x}) + e^{-x}\frac{1}{2} \\
=& 1 - \frac{1}{2}e^{-x} \\
E[M_v] =& 1 - \frac{1}{2}E[M_{v_1}M_{v_2}]
\end{aligned}$$

- General graph: $M_{v_1}M_{v_2}$ are not independent

Correlation decay in Independent Sets

- Let v be a node, and v_1, v_2 its neighbors in the graph.
- $\tilde{B_v} = \max(0, W - \tilde{B_{v_1}} - \tilde{B_{v_2}})$
- Let $M_v = \exp(-B_v)$ and $x = \tilde{B_{v_1}} + \tilde{B_{v_2}}$
- Then:

$$\begin{aligned} E[M_v|x] =& P(W_v \leq x)e^0 + P(W_v > x)E[e^{W_v - x}|W_v > x] \\ =& (1 - e^{-x}) + e^{-x}\frac{1}{2} \\ =& 1 - \frac{1}{2}e^{-x} \\ E[M_v] =& 1 - \frac{1}{2}E[M_{v_1}M_{v_2}] \end{aligned}$$

- General graph: $M_{v_1}M_{v_2}$ are not independent

Correlation decay in Independent Sets

- Let v be a node, and v_1, v_2 its neighbors in the graph.
- $\tilde{B_v} = \max(0, W - \tilde{B_{v_1}} - \tilde{B_{v_2}})$
- Let $M_v = \exp(-B_v)$ and $x = \tilde{B_{v_1}} + \tilde{B_{v_2}}$
- Then:

$$\begin{aligned} E[M_v|x] =& P(W_v \leq x)e^0 + P(W_v > x)E[e^{W_v - x}|W_v > x] \\ =& (1 - e^{-x}) + e^{-x}\frac{1}{2} \\ =& 1 - \frac{1}{2}e^{-x} \\ E[M_v] =& 1 - \frac{1}{2}E[M_{v_1}M_{v_2}] \end{aligned}$$

- General graph: $M_{v_1}M_{v_2}$ are not independent

Correlation decay in Independent Sets

- Let v be a node, and v_1, v_2 its neighbors in the graph.
- $\tilde{B_v} = \max(0, W - \tilde{B_{v_1}} - \tilde{B_{v_2}})$
- Let $M_v = \exp(-B_v)$ and $x = \tilde{B_{v_1}} + \tilde{B_{v_2}}$
- Then:

$$
\begin{aligned}
E[M_v|x] =& P(W_v \leq x)e^0 + P(W_v > x)E[e^{W_v - x}|W_v > x] \\
=& (1 - e^{-x}) + e^{-x}\frac{1}{2} \\
=& 1 - \frac{1}{2}e^{-x} \\
E[M_v] =& 1 - \frac{1}{2}E[M_{v_1}M_{v_2}]
\end{aligned}
$$

- General graph: $M_{v_1}M_{v_2}$ are not independent

Correlation decay in Independent Sets

- Consider two extreme boundary conditions + ($B^+_{BC} = +\infty$) and − ($B^-_{BC} = 0$)
- $M^+_v = \exp(-B^+_v)$ and $M^-_v = \exp(-B^-_v)$
- We also have:

$$E[M^+_v] = 1 - \frac{1}{2}E[M^+_{v_1}M^+_{v_2}]$$
$$E[M^-_v] = 1 - \frac{1}{2}E[M^-_{v_1}M^-_{v_2}]$$

- Combining equations:

$$E[|M^+_v - M^-_v|] \le \max(E|M^+_{v_1} - M^-_{v_1}|, |M^+_{v_2} - M^-_{v_2}|)$$

- In general, if v had $d \le 2$ neighbors:

$$E[|M^+_v - M^-_v|] \le \frac{d}{2}\max_{1\le i\le d}(E|M^+_{v_i} - M^-_{v_i}|)$$

Correlation decay in Independent Sets

- Consider two extreme boundary conditions + ($B^{+}_{BC} = +\infty$) and − ($B^{-}_{BC} = 0$)
- $M^{+}_{v} = \exp(-B^{+}_{v})$ and $M^{-}_{v} = \exp(-B^{-}_{v})$
- We also have:

$$E[M^{+}_{v}] = 1 - \frac{1}{2} E[M^{+}_{v_1} M^{+}_{v_2}]$$
$$E[M^{-}_{v}] = 1 - \frac{1}{2} E[M^{-}_{v_1} M^{-}_{v_2}]$$

- Combining equations:

$$E[|M^{+}_{v} - M^{-}_{v}|] \le \max(E|M^{+}_{v_1} - M^{-}_{v_1}|, |M^{+}_{v_2} - M^{-}_{v_2}|)$$

- In general, if v had $d \le 2$ neighbors:

$$E[|M^{+}_{v} - M^{-}_{v}|] \le \frac{d}{2} \max_{1 \le i \le d} (E|M^{+}_{v_i} - M^{-}_{v_i}|)$$

Correlation decay in Independent Sets

- Consider two extreme boundary conditions $+$ ($B^+_{BC} = +\infty$) and $-$ ($B^-_{BC} = 0$)
- $M^+_v = \exp(-B^+_v)$ and $M^-_v = \exp(-B^-_v)$
- We also have:

$$E[M^+_v] = 1 - \frac{1}{2}E[M^+_{v_1}M^+_{v_2}]$$
$$E[M^-_v] = 1 - \frac{1}{2}E[M^-_{v_1}M^-_{v_2}]$$

- Combining equations:

$$E[|M^+_v - M^-_v|] \leq \max(E|M^+_{v_1} - M^-_{v_1}|, |M^+_{v_2} - M^-_{v_2}|)$$

- In general, if v had $d \leq 2$ neighbors:

$$E[|M^+_v - M^-_v|] \leq \frac{d}{2}\max_{1\leq i\leq d}(E|M^+_{v_i} - M^-_{v_i}|)$$

Correlation decay in Independent Sets

- Consider two extreme boundary conditions $+$ ($B^+_{BC} = +\infty$) and $-$ ($B^-_{BC} = 0$)
- $M^+_v = \exp(-B^+_v)$ and $M^-_v = \exp(-B^-_v)$
- We also have:

$$\begin{aligned} E[M^+_v] &= 1 - \frac{1}{2} E[M^+_{v_1} M^+_{v_2}] \\ E[M^-_v] &= 1 - \frac{1}{2} E[M^-_{v_1} M^-_{v_2}] \end{aligned}$$

- Combining equations:

$$E[|M^+_v - M^-_v|] \leq \max(E|M^+_{v_1} - M^-_{v_1}|, |M^+_{v_2} - M^-_{v_2}|)$$

- In general, if v had $d \leq 2$ neighbors:

$$E[|M^+_v - M^-_v|] \leq \frac{d}{2} \max_{1 \leq i \leq d} (E|M^+_{v_i} - M^-_{v_i}|)$$

Correlation decay in Independent Sets

- Consider two extreme boundary conditions + ($\mathrm{B}_{\mathrm{BC}}^{+} = +\infty$) and − ($\mathrm{B}_{\mathrm{BC}}^{-} = 0$)
- $\mathrm{M}_{\mathrm{v}}^{+} = \exp(-\mathrm{B}_{\mathrm{v}}^{+})$ and $\mathrm{M}_{\mathrm{v}}^{-} = \exp(-\mathrm{B}_{\mathrm{v}}^{-})$
- We also have:

$$\begin{aligned}\mathrm{E}[\mathrm{M}_{\mathrm{v}}^{+}] &= 1 - \frac{1}{2}\mathrm{E}[\mathrm{M}_{\mathrm{v}_1}^{+}\mathrm{M}_{\mathrm{v}_2}^{+}] \\ \mathrm{E}[\mathrm{M}_{\mathrm{v}}^{-}] &= 1 - \frac{1}{2}\mathrm{E}[\mathrm{M}_{\mathrm{v}_1}^{-}\mathrm{M}_{\mathrm{v}_2}^{-}]\end{aligned}$$

- Combining equations:

$$\mathrm{E}[|\mathrm{M}_{\mathrm{v}}^{+} - \mathrm{M}_{\mathrm{v}}^{-}|] \leq \max(\mathrm{E}|\mathrm{M}_{\mathrm{v}_1}^{+} - \mathrm{M}_{\mathrm{v}_1}^{-}|, |\mathrm{M}_{\mathrm{v}_2}^{+} - \mathrm{M}_{\mathrm{v}_2}^{-}|)$$

- In general, if v had $\mathrm{d} \leq 2$ neighbors:

$$\mathrm{E}[|\mathrm{M}_{\mathrm{v}}^{+} - \mathrm{M}_{\mathrm{v}}^{-}|] \leq \frac{\mathrm{d}}{2}\max_{1\leq \mathrm{i}\leq \mathrm{d}}(\mathrm{E}|\mathrm{M}_{\mathrm{v}_\mathrm{i}}^{+} - \mathrm{M}_{\mathrm{v}_\mathrm{i}}^{-}|)$$

Correlation decay in Independent Sets

- Remove a very small fraction δ of the neighbors.
- $\mathrm{E}[|\mathrm{M}_{\mathrm{v}}^{+} - \mathrm{M}_{\mathrm{v}}^{-}|] \leq \frac{\mathrm{d}(1-\delta)}{2} \max(\mathrm{E}|\mathrm{M}_{\mathrm{v}_1}^{+} - \mathrm{M}_{\mathrm{v}_1}^{-}|, |\mathrm{M}_{\mathrm{v}_2}^{+} - \mathrm{M}_{\mathrm{v}_2}^{-}|)$
- Correlation at depth r: $(1-\delta)^{\mathrm{r}}$
- Total error is proportional to $(1-\delta)^{\mathrm{r}} + \delta$
- In time 2^{r}, optimal δ gives suboptimality of $\mathrm{O}(\frac{\log(\mathrm{r})}{\mathrm{r}})$

Correlation decay in Independent Sets

- Remove a very small fraction δ of the neighbors.
- $E[|M_v^+ - M_v^-|] \leq \frac{d(1-\delta)}{2} \max(E|M_{v_1}^+ - M_{v_1}^-|, |M_{v_2}^+ - M_{v_2}^-|)$
- Correlation at depth r: $(1-\delta)^r$
- Total error is proportional to $(1-\delta)^r + \delta$
- In time 2^r, optimal δ gives suboptimality of $O(\frac{\log(r)}{r})$

Correlation decay in Independent Sets

- Remove a very small fraction δ of the neighbors.
- $\mathrm{E}[|\mathrm{M}_\mathrm{v}^+ - \mathrm{M}_\mathrm{v}^-|] \leq \frac{\mathrm{d}(1-\delta)}{2} \max(\mathrm{E}|\mathrm{M}_{\mathrm{v}_1}^+ - \mathrm{M}_{\mathrm{v}_1}^-|, |\mathrm{M}_{\mathrm{v}_2}^+ - \mathrm{M}_{\mathrm{v}_2}^-|)$
- Correlation at depth r: $(1-\delta)^\mathrm{r}$
- Total error is proportional to $(1-\delta)^\mathrm{r} + \delta$
- In time 2^r, optimal δ gives suboptimality of $\mathrm{O}(\frac{\log(\mathrm{r})}{\mathrm{r}})$

Correlation decay in Independent Sets

- Remove a very small fraction δ of the neighbors.
- $\mathrm{E}[|\mathrm{M}_{\mathrm{v}}^{+} - \mathrm{M}_{\mathrm{v}}^{-}|] \leq \frac{\mathrm{d}(1-\delta)}{2} \max(\mathrm{E}|\mathrm{M}_{\mathrm{v}_1}^{+} - \mathrm{M}_{\mathrm{v}_1}^{-}|, |\mathrm{M}_{\mathrm{v}_2}^{+} - \mathrm{M}_{\mathrm{v}_2}^{-}|)$
- Correlation at depth r: $(1-\delta)^{\mathrm{r}}$
- Total error is proportional to $(1-\delta)^{\mathrm{r}} + \delta$
- In time 2^{r}, optimal δ gives suboptimality of $\mathrm{O}(\frac{\log(\mathrm{r})}{\mathrm{r}})$

Correlation decay in Independent Sets

- Remove a very small fraction δ of the neighbors.
- $\mathrm{E}[|\mathrm{M}_\mathrm{v}^+ - \mathrm{M}_\mathrm{v}^-|] \leq \frac{\mathrm{d}(1-\delta)}{2} \max(\mathrm{E}|\mathrm{M}_{\mathrm{v}_1}^+ - \mathrm{M}_{\mathrm{v}_1}^-|, |\mathrm{M}_{\mathrm{v}_2}^+ - \mathrm{M}_{\mathrm{v}_2}^-|)$
- Correlation at depth r: $(1-\delta)^\mathrm{r}$
- Total error is proportional to $(1-\delta)^\mathrm{r} + \delta$
- In time 2^r, optimal δ gives suboptimality of $\mathrm{O}(\frac{\log(\mathrm{r})}{\mathrm{r}})$

Summary

- Correlation decay implies near-optimal decentralization
- A specific form of coupling can induce correlation decay
- Sufficient conditions can be computed for various problems
- Message: Stochasticity can make optimization easier

- Ongoing work
  - Game theoretical or dynamical versions of the problem
  - More general characterizations of conditions for correlation decay

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}} |B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a' e_{r-1} + b' e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then correlation decay occurs

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}}|B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a'e_{r-1} + b'e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then correlation decay occurs

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}} |B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a' e_{r-1} + b' e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then correlation decay occurs

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}}|B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a' e_{r-1} + b' e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then correlation decay occurs

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}} |B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a' e_{r-1} + b' e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then correlation decay occurs

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}} |B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a' e_{r-1} + b' e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then correlation decay occurs

Proof outline

- Bonus equations:

$$B = \Phi(1) - \Phi(0) + \sum_j \mu_j \qquad B' = \Phi(1) - \Phi(0) + \sum_j \mu'_j$$

- Continuity of messages: $|\mu_j - \mu'_j| \leq |B_j - B'_j|$
- Coupling event C_j : $|\mu_j - \mu'_j| \leq 1_{\overline{C_j}} |B'_j - B_j|$
- This leads to

$$E|B - B'| \leq \sum_j aE|B_j - B'_j| + bE|B_j - B_j|^2$$

- Let e_r be the maximum error for depth r
- Problem-dependent approaches (gaussian costs, bounded functions) allow to convert the inequality into:

$$e_r \leq a' e_{r-1} + b' e_{r-2}$$

- If $\sqrt{a'^2 + 4b'} < 1$ then <u>correlation decay occurs</u>

Proving coupling

- For any j, the partial bonus has the following form:

$$\mu \quad = \quad \mathrm{A}_{\mathrm{j}}^{3} + \max(\mathrm{B}_{\mathrm{j}}, \mathrm{A}_{\mathrm{j}}^{1}) - \max(\mathrm{B}_{\mathrm{j}}, \mathrm{A}_{\mathrm{j}}^{2})$$

  where

$$\begin{aligned} \mathrm{A}^{1} &= \Phi_{\mathrm{u},\mathrm{v}_{\mathrm{j}}}(1,0) - \Phi_{\mathrm{u},\mathrm{v}_{\mathrm{j}}}(1,1) \\ \mathrm{A}^{2} &= \Phi_{\mathrm{u},\mathrm{v}_{\mathrm{j}}}(0,0) - \Phi_{\mathrm{u},\mathrm{v}_{\mathrm{j}}}(0,1) \\ \mathrm{A}^{3} &= \Phi_{\mathrm{u},\mathrm{v}_{\mathrm{j}}}(1,1) - \Phi_{\mathrm{u},\mathrm{v}_{\mathrm{j}}}(0,1) \end{aligned}$$

- Consequence: If both B_{j}, B'_{j} are greater than $\mathrm{A}_{\mathrm{j}}^{1}$ and $\mathrm{A}_{\mathrm{j}}^{2}$, or smaller than $\mathrm{A}_{\mathrm{j}}^{1}$ and $\mathrm{A}_{\mathrm{j}}^{2}$, then $|\mu - \mu'| = 0$ ("coupling occurs")
- The $\mathrm{A}_{\mathrm{j}}^{\mathrm{i}}$ and B_{j} are independent so the probability can in principle be computed for any distribution

Proving coupling

- For any j, the partial bonus has the following form:

$$\mu \quad = \quad \mathrm{A}_\mathrm{j}^3 + \max(\mathrm{B}_\mathrm{j}, \mathrm{A}_\mathrm{j}^1) - \max(\mathrm{B}_\mathrm{j}, \mathrm{A}_\mathrm{j}^2)$$

  where

$$\begin{aligned} \mathrm{A}^1 &= \Phi_{\mathrm{u},\mathrm{v}_\mathrm{j}}(1,0) - \Phi_{\mathrm{u},\mathrm{v}_\mathrm{j}}(1,1) \\ \mathrm{A}^2 &= \Phi_{\mathrm{u},\mathrm{v}_\mathrm{j}}(0,0) - \Phi_{\mathrm{u},\mathrm{v}_\mathrm{j}}(0,1) \\ \mathrm{A}^3 &= \Phi_{\mathrm{u},\mathrm{v}_\mathrm{j}}(1,1) - \Phi_{\mathrm{u},\mathrm{v}_\mathrm{j}}(0,1) \end{aligned}$$

- Consequence: If both B_j, B'_j are greater than A_j^1 and A_j^2, or smaller than A_j^1 and A_j^2, then $|\mu - \mu'| = 0$ ("coupling occurs")
- The $\mathrm{A}_\mathrm{j}^\mathrm{i}$ and B_j are independent so the probability can in principle be computed for any distribution