Recap:

Inference Algorithms:

	Marginalization	Maximization			
✓	Elimination	Elim.	Exact	any G	$O(\lvert\mathcal{X}\rvert^{\text{Treewidth}})$ also NP-hard to find best ordering.
✓	sum-product	Max-product	Approx	Pairwise MRF	$O(\lvert\mathcal{X}\rvert^{2})$
	sum-Product on Factor Graphs	Max-product on Factor Graphs	Approx	any G	$O(\lvert\mathcal{X}\rvert^{\text{max-degree}})$
	Elimination on Junction Tree	Elimination on Junction Tree	Exact.	Junction Tree	$O(\lvert\mathcal{X}\rvert^{\text{Treewidth}})$ np-hard to find best ordering.

---

* Belief Propagation on factor graphs.

Def. factor graph $G = (V \cup F, E)$ ← bipartite graph

V: variables , F: factors

a probability factorizes w.r.t G

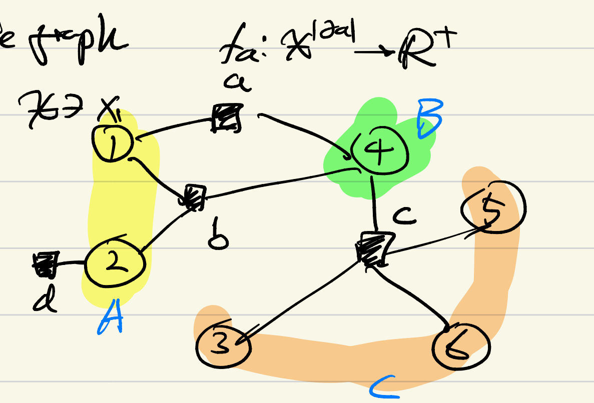
$$P(x) = \frac{1}{Z} \prod_{a \in F} f_a(x_a), \quad x_a = x_{\partial a}$$

$\mathcal{X} \ni x_1 \quad x_4$

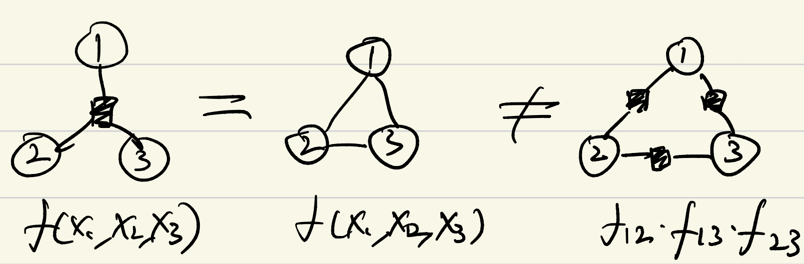
$V = \{1, \ldots n\}$

$F = \{a, b, c \ldots\}$

$f_a: \mathcal{X}^{|\partial a|} \to \mathbb{R}_+$



A − B − C separated iff any $a \in A$, $c \in C$ path $a \to c$ goes through B.



- FGs are strict generalization of MRFs. encode more fine grained factorizations.
- FGs have the same Markov Property as MRF.

$$A - B - C \text{ separation} \Longrightarrow X_A \perp\!\!\!\perp X_C \mid X_B$$

* Sum-Product on FGs

$E = \{(\underset{\text{variable}}{i}, \underset{\text{factor}}{a})\}$

Input: $G = (V \cup F, E)$, T

Output: $\{\hat{P}(X_i)\}_{i \in V}$

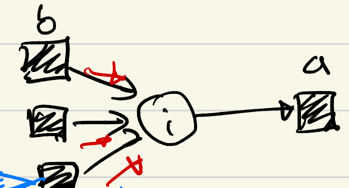
Initialize: $\{(m_{i\to a}, \tilde{m}_{a\to i})\}_{(i,a)\in E} = \mathbb{1}\frac{1}{|\mathcal{X}|}$ or RANDOM

$[\ ]\}|\mathcal{X}|$

Repeat $t = 1, \cdots, T$

① Update $m_{i\to a}(X_i)$: for all $(i,a) \in E$

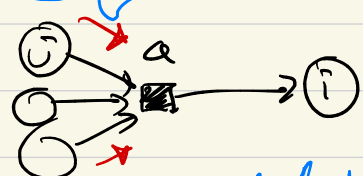
$$m_{i\to a}(X_i) = \prod_{b \in \partial i \setminus \{a\}} \tilde{m}_{b\to i}(X_i)$$



Intuition: independent belief about $\hat{P}(X_i)$

② Update $\tilde{m}_{a\to i}(X_i)$ for all $(i,a) \in E$.

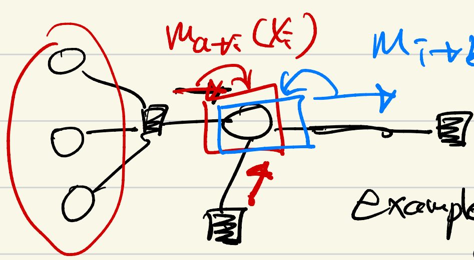
$$\underbrace{\tilde{m}_{a\to i}(X_i)}_{\hat{P}(X_i)} = \underbrace{\sum_{X_{j \in \partial a \setminus \{i\}}} f_a(X_a)}_{\hat{P}(X_i \mid X_{j_1}, X_{j_2}, \cdots)} \cdot \prod_{j \in \partial a \setminus \{i\}} m_{j\to a}(X_j)$$



Intuition: indept belief $\hat{P}(X_j)$

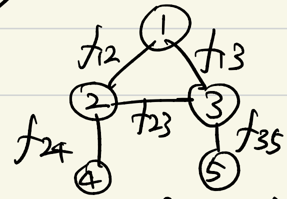
Compute marginals: $M_i(X_i) = \prod_{a \in \partial i} \tilde{m}_{a\to i}(X_i)$

$$\tilde{m}_{a\to 1}(X_1) = \sum_{X_2, X_3, X_4} f_a(X_1, X_2, X_3, X_4) \cdot \underbrace{m_{2\to a}(X_2)\, m_{3\to a}(X_3)\, m_{4\to a}(X_4)}_{\prod_{j \in \partial a \setminus \{i\}}}$$



$m_{i\to b}(X_i) = F(\tilde{m}_{a\to i}(X_i), \tilde{m}_{c\to i}(X_i), \cdots)$

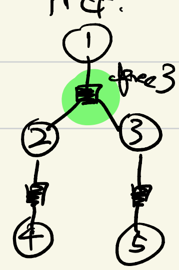
example > pairwise MRF



not a tree
approx

$O(|\mathcal{X}|^2)$ — Pairwise BP algorithm

FG.



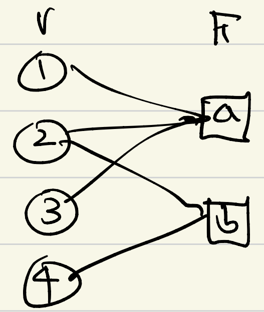
FG BP
- Exact
- Tree
- $O(|\mathcal{X}|^3)$

Example> Decoding LDPC codes.

Def. LDPC (Low Density Parity Check) codes.

a family of codes defined as follows.

$G = (V \cup F, E)$, $\mathcal{X} = \{0, 1\}$



factors are parity checking.

$$f_a(X_a) = \mathbb{I}(x_1 \underset{\text{XOR}}{\underset{\uparrow}{\oplus}} x_2 \oplus x_3 = 0)$$

$$f_b(X_b) = \mathbb{I}(x_2 \oplus x_4 = 0)$$

Def. $\text{Codebook} = \{ \underset{\text{Codeword}}{\underset{\uparrow}{X}} \in \mathcal{X}^n \mid \text{satisfy all parities} \} = \left\{ \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{array} \right\}$

$$|\text{Codebook}| = 2^{\underbrace{|V| - |F|}_{K}} = 2^2 = 4$$

At transmission, one of the codeword from the codebook is sent over a noisy channel, defined by $\mathbb{P}(Y_i | X_i)$

Goal: to recover X from Y. e.g. $P(Y_i | X_i) = \begin{array}{cc|c} \multicolumn{2}{c}{X_i} & \\ 0 & 1 & \\ \end{array}$ $\begin{bmatrix} .7 & .3 \\ .3 & .7 \end{bmatrix} \begin{array}{c} 0 \\ 1 \end{array} Y_i$

Binary Symmetric channel.

· strategy: Use belief propagation to estimate $\hat{P}(X_i | Y_1 \cdots Y_n)$, $\forall i \in V$

output $\hat{x}_i = \begin{cases} 1 & \text{if } \log \dfrac{\hat{P}(X_i = 1 | Y)}{\hat{P}(X_i = 0 | Y)} > 0 \\ 0 & \text{otherwise} \end{cases}$

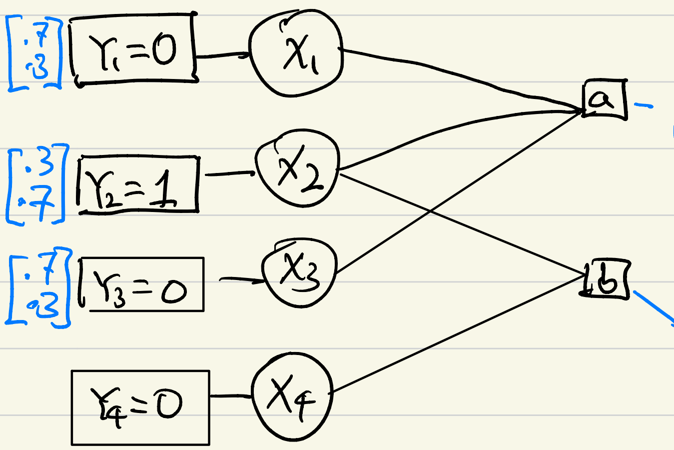
* How to turn BN → FG.

* How to include observation Y in inference

Factor Graph representation of LDPC

$$\boxed{\text{maximize } P(X|Y)} \propto P(X,Y) = \frac{1}{Z} \underbrace{\prod_{a\in F} \mathbb{I}(\oplus x_a = 0)}_{\text{codeword } X \text{ satisfy parity.}} \underbrace{\prod_{i\in V} P(Y_i|X_i)}_{\text{independent noisy channel}}$$

$$P(X) \times P(Y|X)$$



$$f_a = \begin{array}{c|cccc} x_1 \backslash (x_2,x_3) & 00 & 01 & 11 & 10 \\ \hline 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{array}$$

$$f_b = \begin{array}{c|cc} x_4 \backslash x_2 & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & 0 & 1 \end{array} = \mathbb{I}(X_2 \oplus X_4 = 0)$$

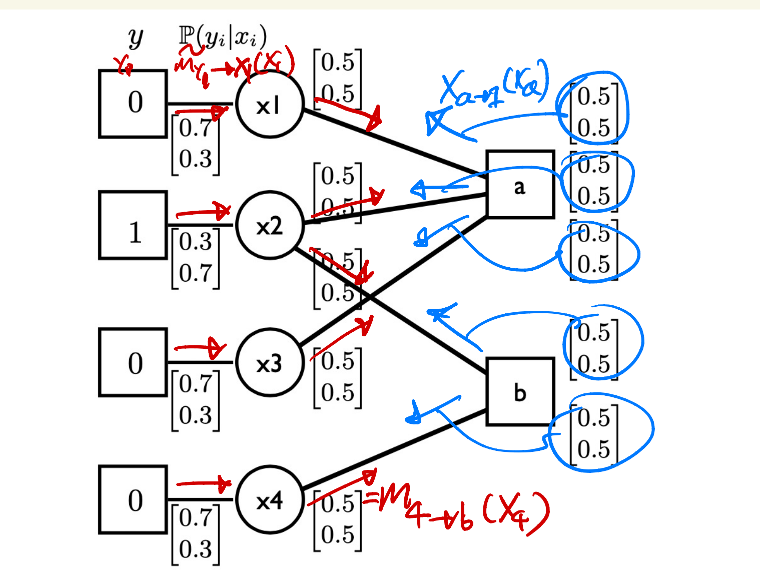
$P(Y_4|X_4) = \begin{array}{c} \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix} \end{array} \begin{array}{c} 0 \\ 1 \end{array}$ (X_4)

Codebook $= \{0000, 0111, 1010, 1101\} = \dfrac{2 = k = |V| - |F|}{4 = n = |V|}$

0000 is the MAP estimation.

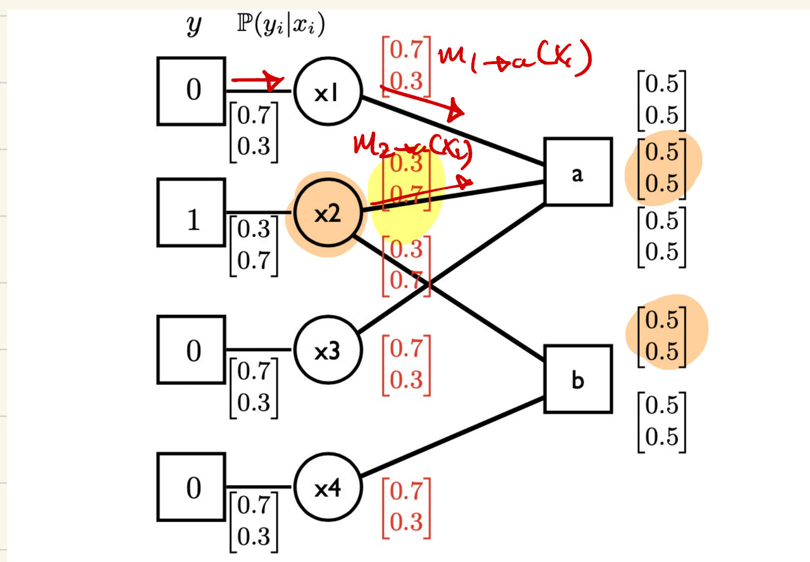
we received 0100

$\boxed{\text{rate}} = \frac{1}{2}$



$$\nu^{(t+1)}_{i\to a}(x_i) = \mathbb{P}(y_i|x_i) \textstyle\prod_{b\in\partial i\backslash\{a\}} \tilde{\nu}^{(t)}_{b\to i}(x_i)$$

$$\tilde{\nu}^{(t+1)}_{a\to i}(x_i) = \textstyle\sum_{x_{\partial a\backslash\{i\}}} \prod_{j\in\partial a\backslash\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a} = 0)$$



$$\nu^{(t+1)}_{i\to a}(x_i) = \mathbb{P}(y_i|x_i) \textstyle\prod_{b\in\partial i\backslash\{a\}} \tilde{\nu}^{(t)}_{b\to i}(x_i)$$

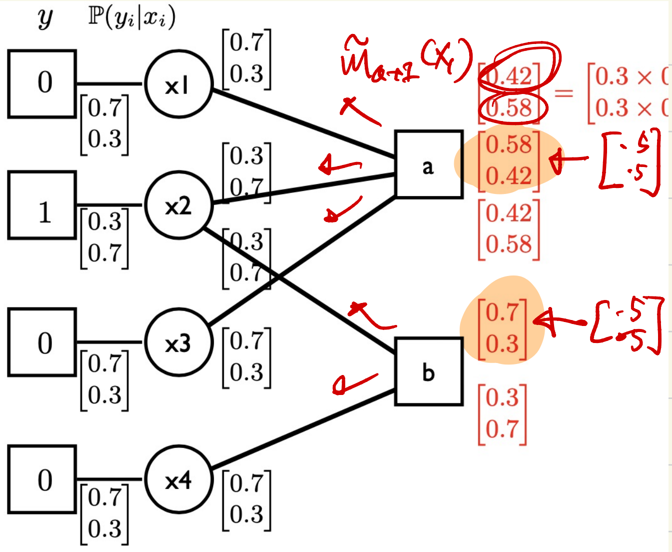
$$\tilde{\nu}^{(t+1)}_{a\to i}(x_i) = \textstyle\sum_{x_{\partial a\backslash\{i\}}} \prod_{j\in\partial a\backslash\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a} = 0)$$

Sum-product algorithm

$$m_{1\to a}(X_1) = \tilde{m}_{Y_1\to 1}(X_1) = \begin{bmatrix} .7 \\ .3 \end{bmatrix}$$

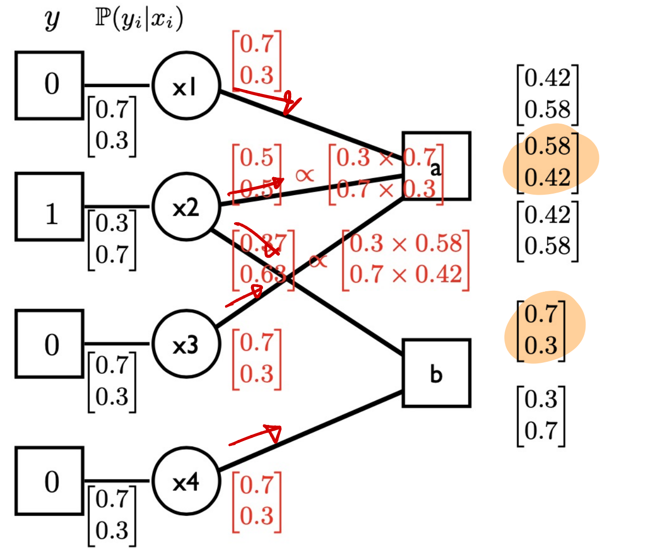
$$m_{2\to a}(X_2) = \underset{\begin{bmatrix} .7 \\ .3 \end{bmatrix}}{\tilde{m}_{Y_2\to 2}(X_2)} \cdot \underset{\begin{bmatrix} .5 \\ .5 \end{bmatrix}}{\tilde{m}_{b\to 2}(X_2)}$$

$$= \begin{bmatrix} .7\times.5 \\ .3\times.5 \end{bmatrix} = \begin{bmatrix} .7 \\ .3 \end{bmatrix}$$



$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i)\prod_{b\in\partial i\setminus\{a\}} \tilde\nu_{b\to i}^{(t)}(x_i)$

$\tilde\nu_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}}\prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a}=0)$



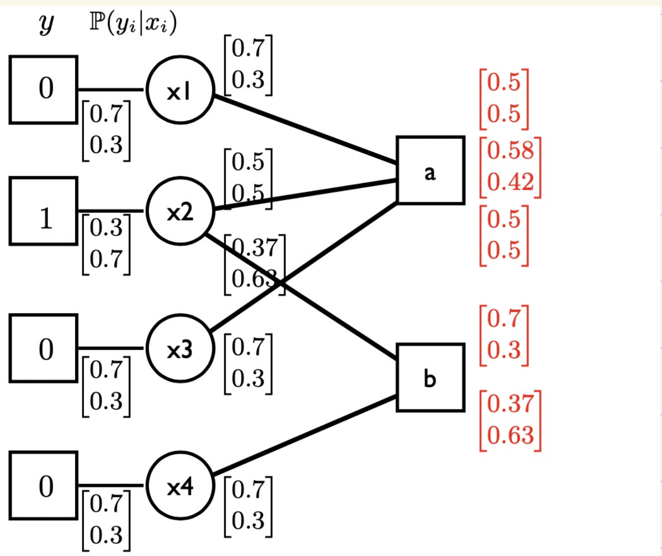
$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i)\prod_{b\in\partial i\setminus\{a\}} \tilde\nu_{b\to i}^{(t)}(x_i)$

$\tilde\nu_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}}\prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a}=0)$

$\tilde m_{a\to 1}(x_1=0) = m_{2\to a}(x_2=0)\cdot m_{3\to a}(x_3=0) + m_{2\to a}(x_2=1)\, m_{3\to a}(x_3=1)$

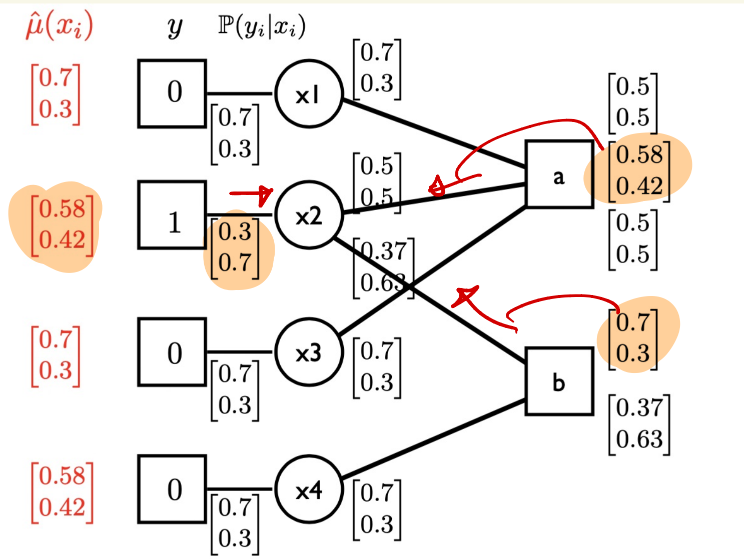
$= .3\times .7 + .7\times .3 = 0.42$

$(x_1=1) =$



$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i)\prod_{b\in\partial i\setminus\{a\}} \tilde\nu_{b\to i}^{(t)}(x_i)$

$\tilde\nu_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}}\prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a}=0)$



$\nu_{i\to a}^{(t+1)}(x_i) = \mathbb{P}(y_i|x_i)\prod_{b\in\partial i\setminus\{a\}} \tilde\nu_{b\to i}^{(t)}(x_i)$

$\tilde\nu_{a\to i}^{(t+1)}(x_i) = \sum_{x_{\partial a\setminus\{i\}}}\prod_{j\in\partial a\setminus\{i\}} \nu_{j\to a}(x_j)\mathbb{I}(\oplus x_{\partial a}=0)$

$\dfrac{\|m^{(t)} - m^{(t+1)}\|}{\|m^{(t)}\|} < \varepsilon.$

$\hat P(x_2) = \begin{bmatrix}.58\\ .42\end{bmatrix} \to \hat x_2 = 0.$

*In practice normalize all messages to sum one.