INFERENCE IN PGM

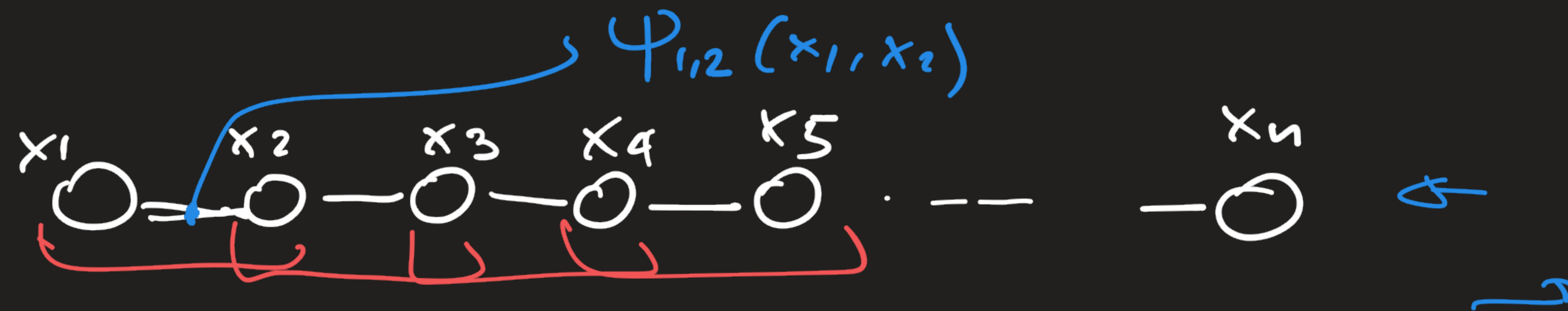
$P(x_1, \dots, x_n) = P(X)$ $\quad \underbrace{Y \subseteq X}_{\text{observed}}$ $\quad Z \subseteq X$, $\quad Y \cap Z = \emptyset$ $\quad Y \cup Z \subsetneq X$ $\quad X' : X' \cup Y \cup Z = X$

$Y = \bar{y}$

$$P(Z \mid Y = \bar{y}) = \sum_{X'} P(Z, X' \mid Y = \bar{y}) \quad \text{on} \quad |X'| = m \quad O(K^m)$$

COMPUTATIONALLY CHALLENGING

$$= \int P(Z, X' \mid Y = \bar{y})\, dx'$$



$$p(x) = \frac{1}{Z}\Psi_{1,2}(x_1,x_2)\,\Psi_{2,3}(x_2,x_3) \;\cdots\; \Psi_{n-1,n}(x_{n-1},x_n)$$

$$p(x) = \underbrace{p(x_1)\,p(x_2|x_1)}_{\Psi_{1,2}(x_1,x_2)}\;\underbrace{p(x_3|x_2)}_{\Psi_{2,3}(x_2,x_3)} \;\cdots\; \underbrace{p(x_n|x_{n-1})}_{\Psi_{n-1,n}(x_{n-1},x_n)}$$

$1 < k < n$

$$p(x_k) = \sum_{x_1,\dots,x_{k-1},\; x_{k+1},\dots,x_n} p(x_1,\dots,x_n) = \sum_{x\setminus k} \frac{1}{Z}\Psi_{1,2}(x_1,x_2)\cdots\Psi_{n-1,n}(x_{n-1},x_n)$$

$O(K^{n-1})$

$$= \frac{1}{Z}\cdot\sum_{x_{k-1}}\Psi_{k-1,k}(x_{k-1},x_k)\cdots\left[\sum_{x_2}\Psi_{2,3}(x_2,x_3)\left[\sum_{x_1}\Psi_{1,2}(x_1,x_2)\right]\right]$$

$O(nK)$

$$\sum_{x_{k+1}}\Psi_{k,k+1}(x_k,x_{k+1})\cdots\sum_{x_{n-1}}\Psi_{n-2,n-1}(x_{n-2},x_{n-1})\left[\sum_{x_n}\Psi_{n-1,n}(x_{n-1},x_n)\right]$$

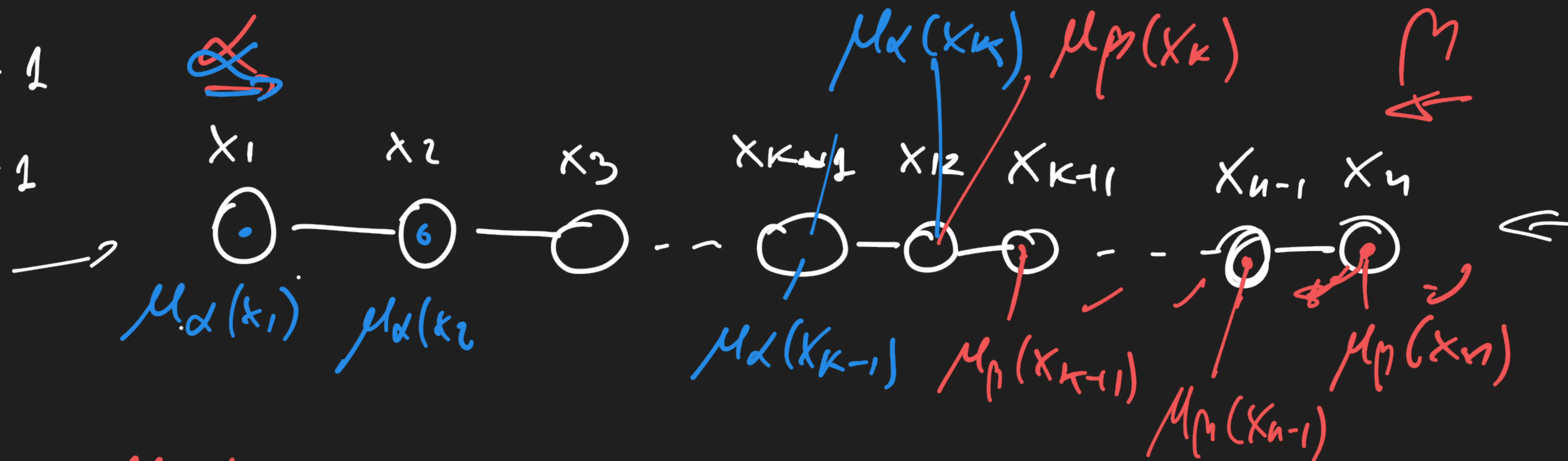
MESSAGE-PASSING ALGORITHMS

$$\mu_\alpha(x_k) := \sum_{x_{k-1}} \psi_{k-1,k}(x_{k-1}, x_k)\, \mu_\alpha(x_{k-1})$$

$$\mu_\beta(x_k) = \sum_{x_{k+1}} \psi_{k,k+1}(x_k, x_{k+1})\, \mu_\beta(x_{k+1})$$

$\mu_\alpha(x_1) = 1$

$\mu_\beta(x_n) = 1$



$$p(x_k) \propto \underbrace{\mu_\beta(x_k)\, \mu_\alpha(x_k)}$$

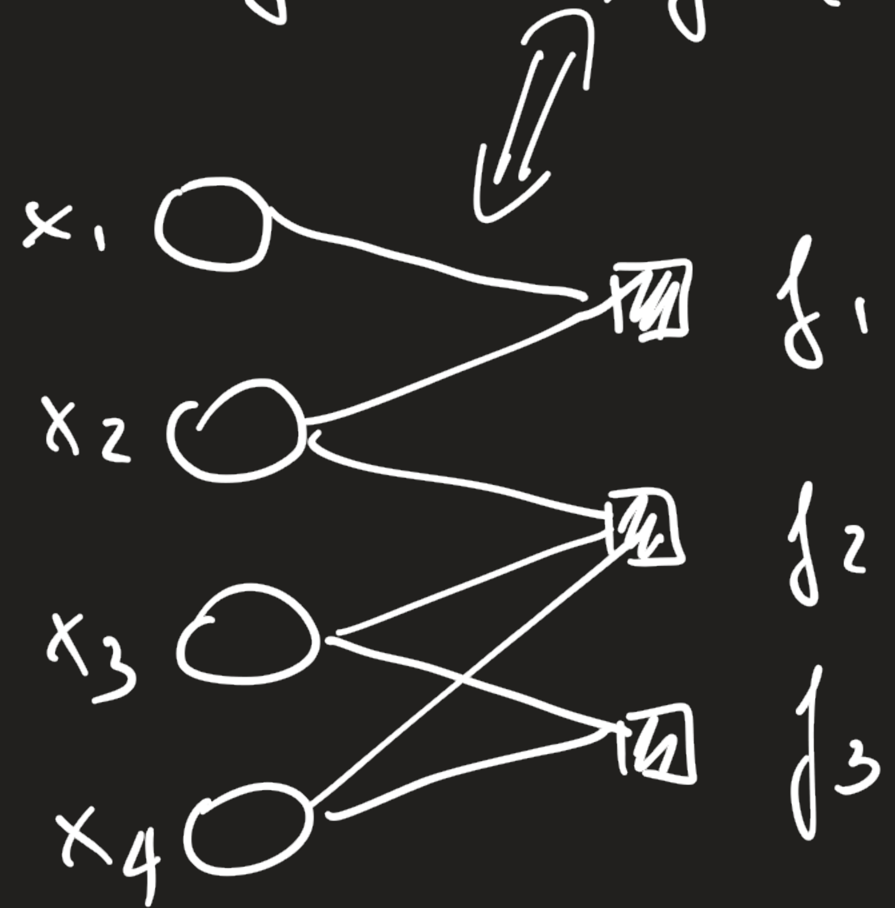
$$= \frac{1}{Z_k} \mu_\alpha(x_k)\, \mu_\beta(x_k) \qquad Z_k = \sum_{x_k} \mu_\alpha(x_k)\, \mu_\beta(x_k)$$

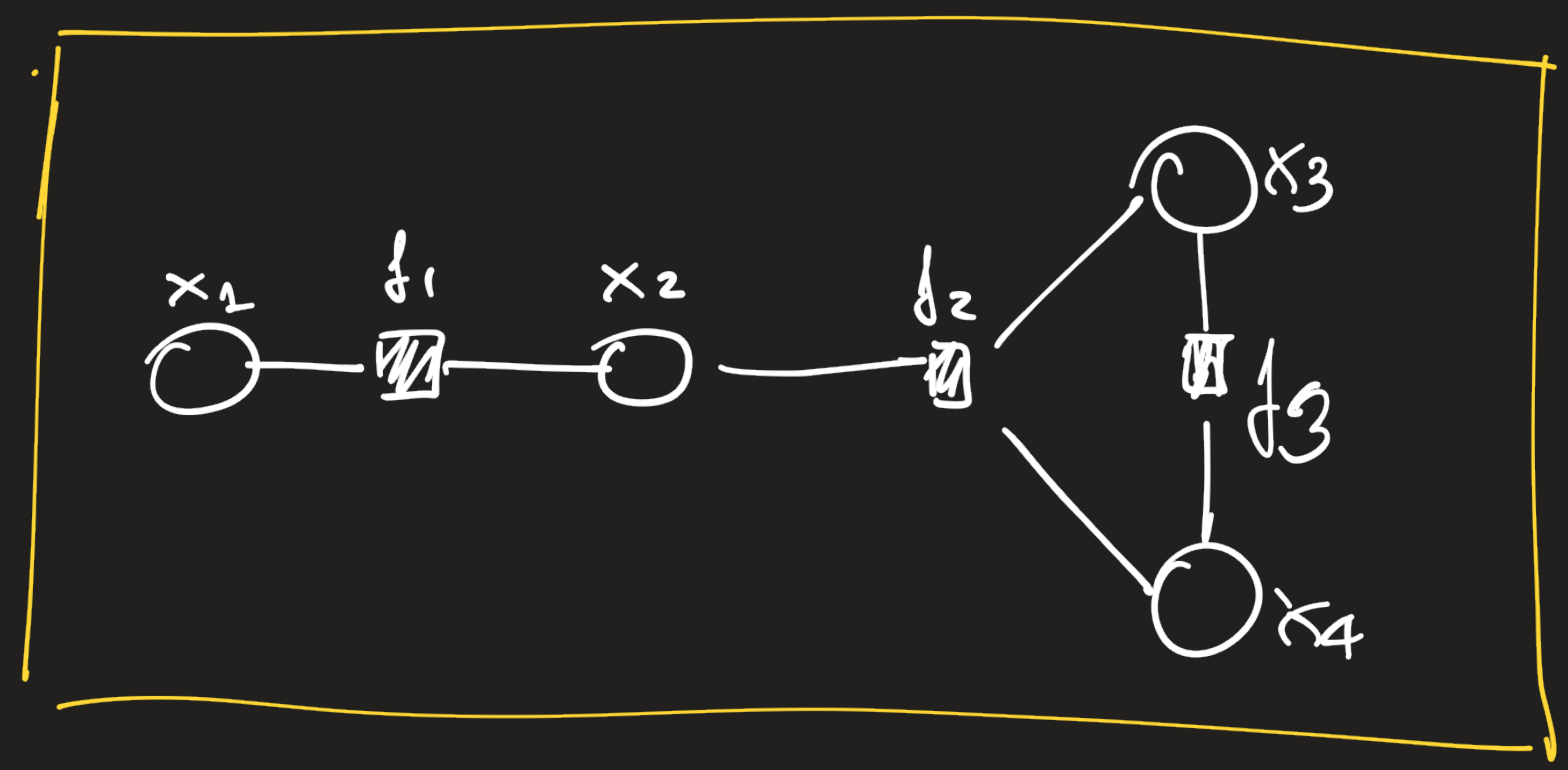
FACTOR GRAPHS

BIPARTITE GRAPHS :

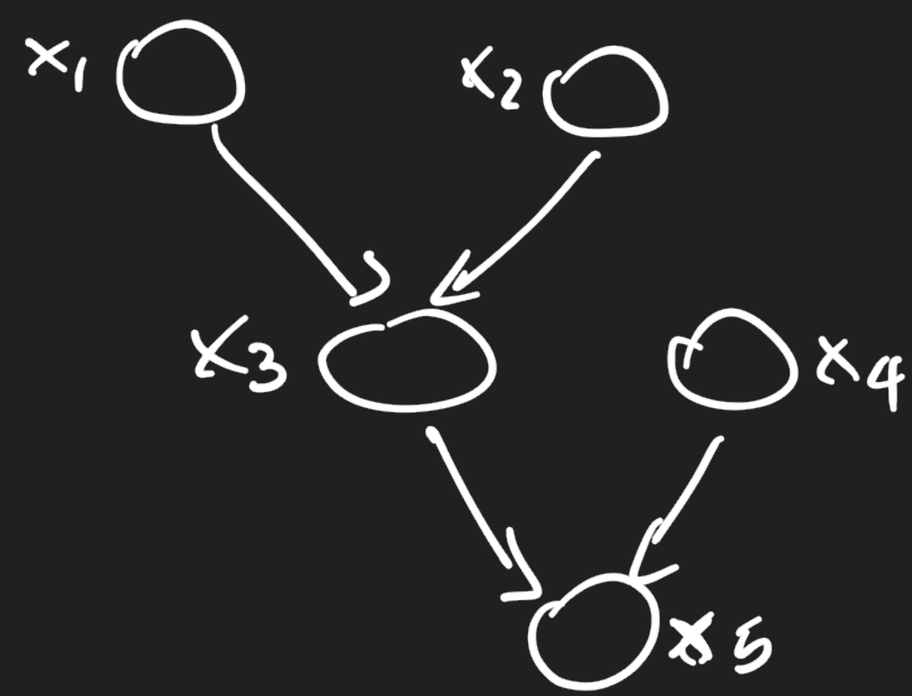
- ○ VARIABLE NODES
- ▨ FACTOR NODES

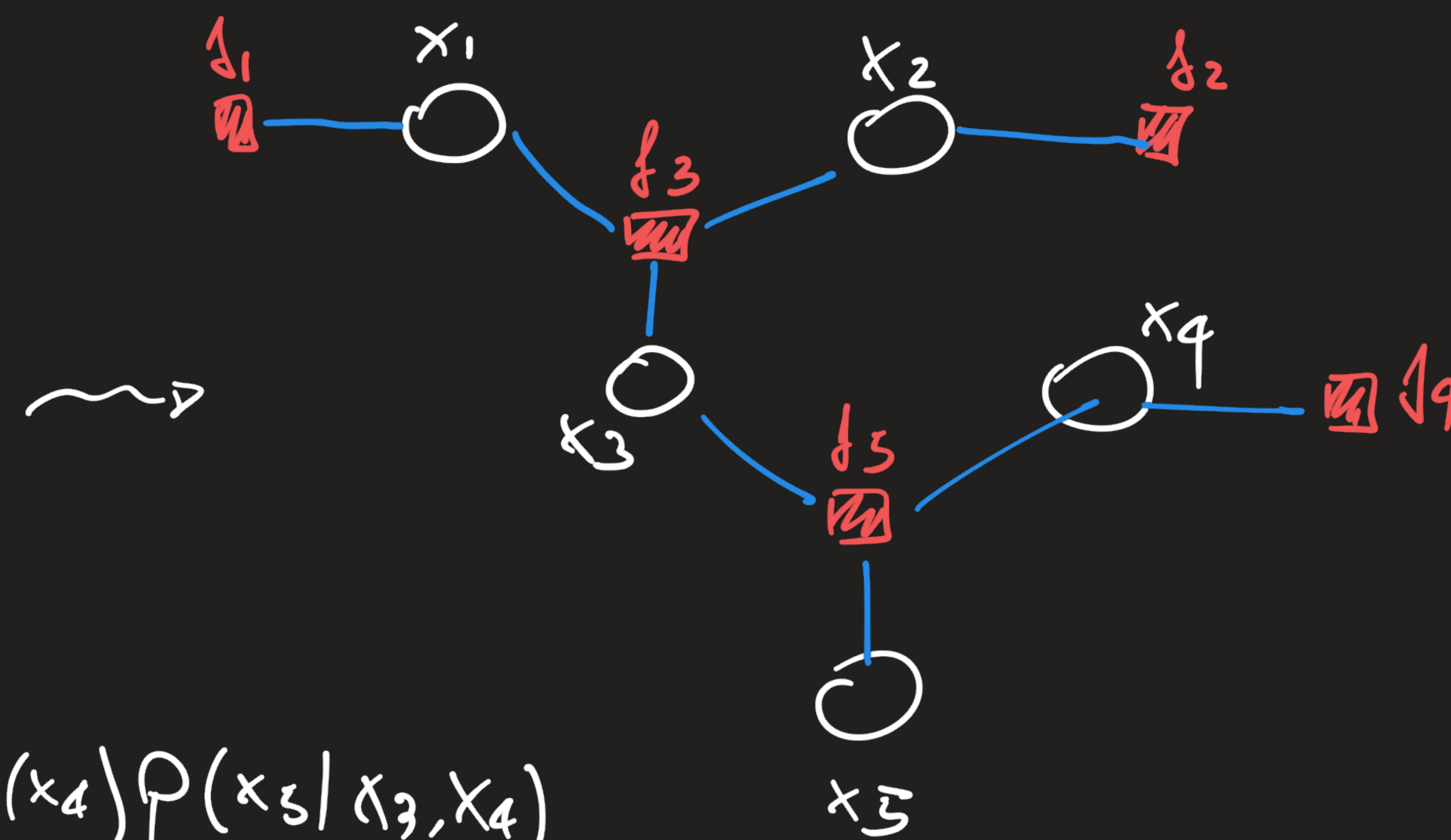
$$p(x) = f_1(x_1, x_2)\, f_2(x_2, x_3, x_4)\, f_3(x_3, x_4)$$





BN $\to$ FG

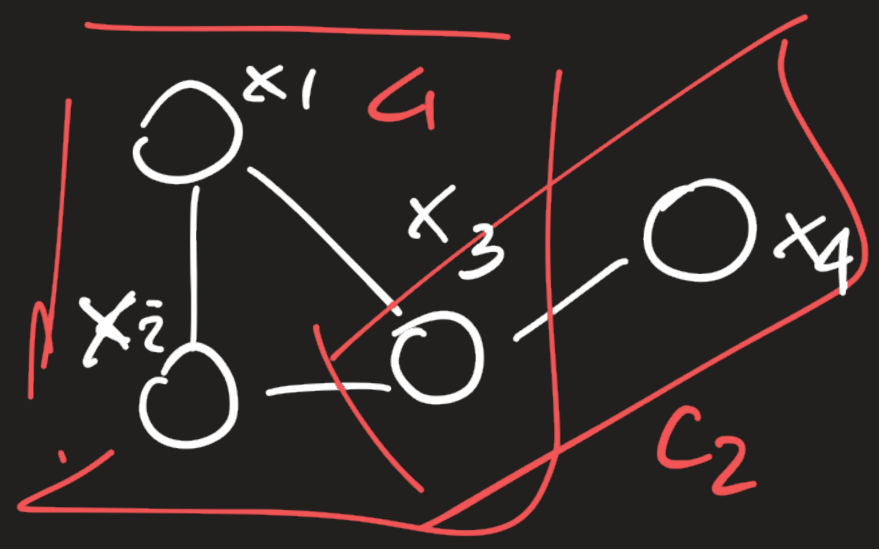


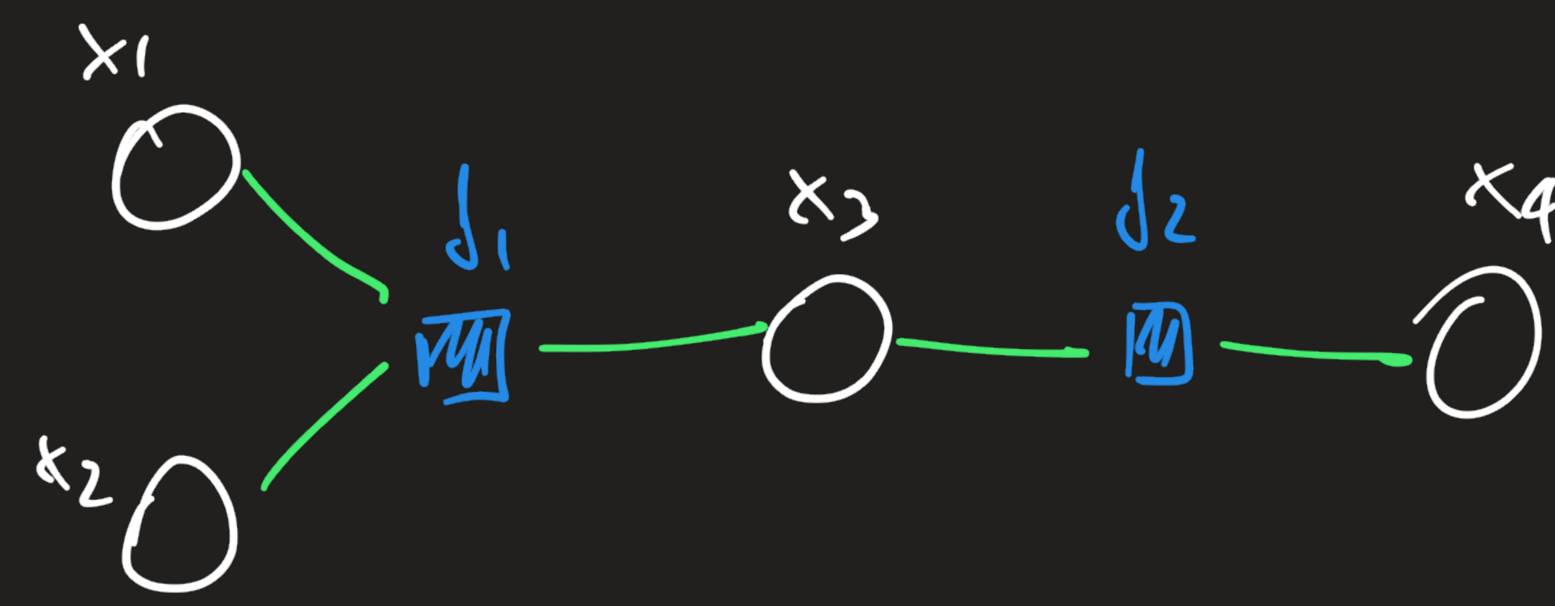


$$p(x) = \underbrace{p(x_1)}_{f_1}\,\underbrace{p(x_2)}_{f_2}\,\underbrace{p(x_3 | x_1, x_2)}_{f_3}\,\underbrace{p(x_4)}_{f_4}\,\underbrace{p(x_5 | x_3, x_4)}_{f_5}$$

$\forall \; p(x_j | pa_j)$ s.t. $f_j \to \{x_j\} \cup pa_j$

MRF $\rightsquigarrow$ FG

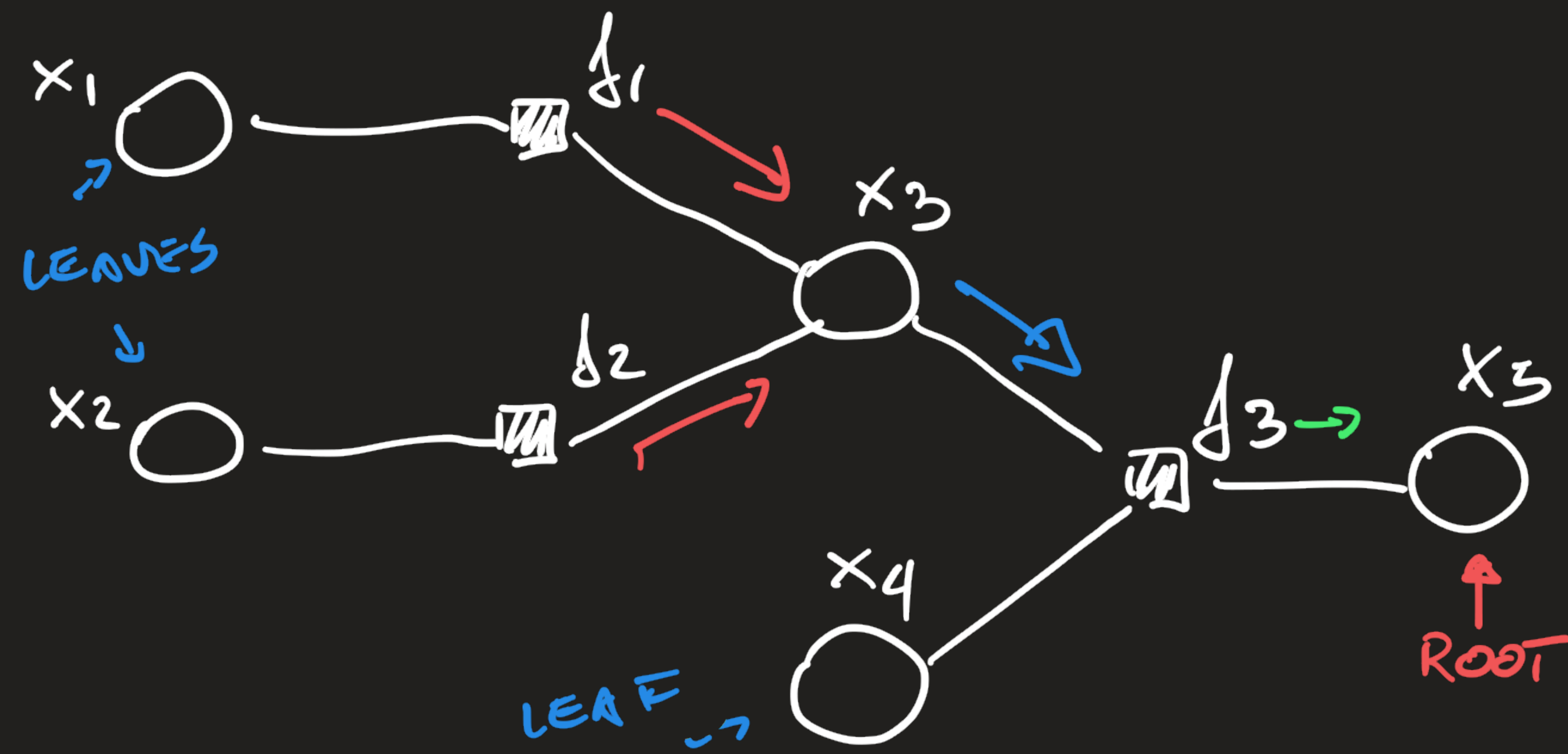




$$p(x) = \frac{1}{Z} \underbrace{\psi_1(x_1, x_2, x_3)}_{f_1} \underbrace{\psi_2(x_3, x_4)}_{f_2}$$

$$\forall c \in \mathcal{C} \rightsquigarrow f_c(x_c) = \psi_c(x_c)$$

MESSAGE PASSING IN FACTOR GRAPHS



BELIEF PROPAGATION ON FG. WHICH ARE TREES

(NO LOOPS IN F.G.)

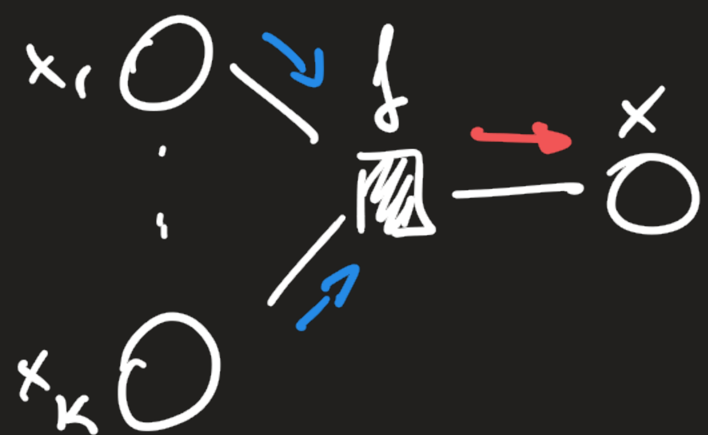
$$p(x) = f_1(x_1, x_3)\, f_2(x_2, x_3)\, f_3(x_3, x_4, x_5)$$

$$p(x_5) = \sum_{x_1, x_2, x_3, x_4} p(x_1, x_2, x_3, x_4, x_5) =$$

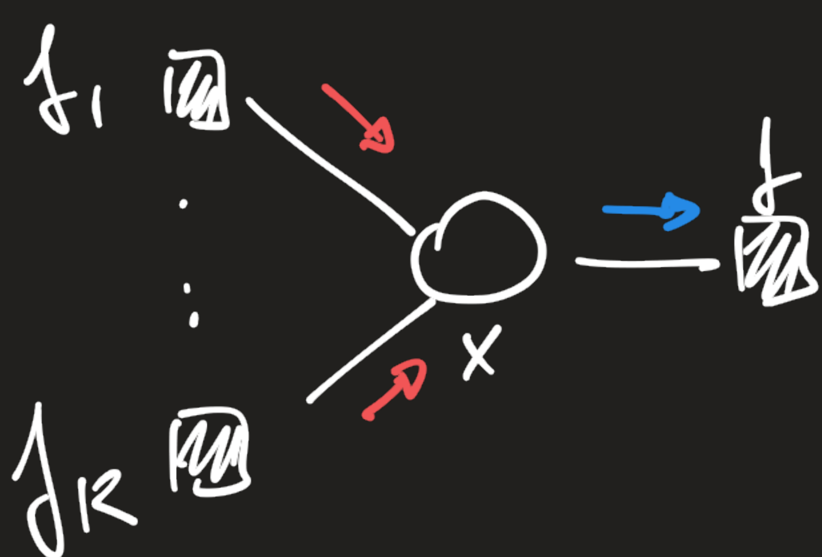
$$= \underbrace{\sum_{x_3, x_4} f_3(x_3, x_4, x_5) \cdot \overbrace{\underbrace{\left[\sum_{x_1} f_1(x_1, x_3)\right]}_{\mu_{f_1 \to x_3}} \underbrace{\left[\sum_{x_2} f_2(x_2, x_3)\right]}_{\mu_{f_2 \to x_3}}}^{\mu_{x_3 \to f_3}}}_{\mu_{f_3 \to x_5}}$$

$N(x) = \{f_1, \dots, f_k \mid f_i - x\}$

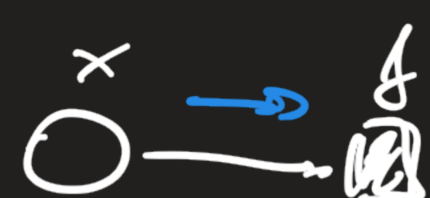
$N(f) : \{x_1, \dots, x_k \mid x_i - f\} \quad (N(f) \setminus x)$

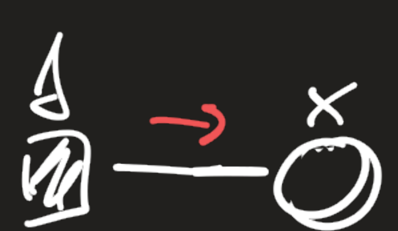


$$\mu_{f \to x}(x) = \sum_{x_1, \dots, x_k \in N(f) \setminus x} f(x, x_1, \dots, x_k) \cdot \prod_{x_i \in N(f) \setminus x} \mu_{x_i \to f}(x_i)$$



$$\mu_{x \to f}(x) = \prod_{f_i \in N(x) \setminus f} \mu_{f_i \to x}(x)$$





$$\left.\begin{array}{ll} \mu_{x \to f}(x) = 1 & x \text{ leaf} \\ \mu_{f \to x}(x) = f(x) & f \text{ leaf} \end{array}\right\} \text{base cases}$$

LEAVES → ROOT forward pass

ROOT → LEAVES backward pass

→ ALL MESSAGES $\mu_{f \to x}$ $\mu_{x \to f}$ are computed

SUM-PRODUCT ALGORITHM

- MARGINAL OF A VAR NODE x

$$p(x) = \prod_{f \in \mathcal{N}(x)} \mu_{f \to x}(x)$$

- MARGINAL AT A FACTOR NODE $f(\bar{x})$

$$p(\bar{x}) = f(\bar{x}) \cdot \prod_{x \in \mathcal{N}(f)} \mu_{x \to f}(x)$$

WE OBTAIN $p(x \mid y = \hat{y})$ by NORMALIZATION

CONDITION ON $y = \hat{y}$? or $p(x \mid y = \hat{y})$

WE CLAMP the value of y_j to $\hat{y}_j$ $p(x, y = \hat{y})$: THIS IS what we compute

$$\sum_{x_1 \dots x_K} p(\cancel{x} x_1 \dots, x_K, y = \hat{y}) = p(x, y = \hat{y})$$

IT IS UNNORMALISED

$$p(x \mid \hat{y} = y) = \frac{p(x, y = \hat{y})}{p(y = \hat{y})}$$

HOW TO CONDITION ON $y = \hat{y}$ $\quad y \subseteq \{x_1, \dots, x_n\}$

$$p(x \mid y = \hat{y})$$

$$p(x, x_1, \dots, x_k, y = \hat{y})$$

CLAMPING y to $\hat{y}$

$$p(x, y = \hat{y}) = \sum_{x_1, \dots, x_k} p(x, x_1, \dots, x_k, y = \hat{y})$$

Sum product with y_j FIXED TO $\hat{y}_j$ FOR ALL y_j in y.

$$p(x \mid y = \hat{y}) = \frac{p(x, y = \hat{y})}{p(\hat{y})} \qquad p(\hat{y}) = \sum_{x} p(x, y = \hat{y})$$

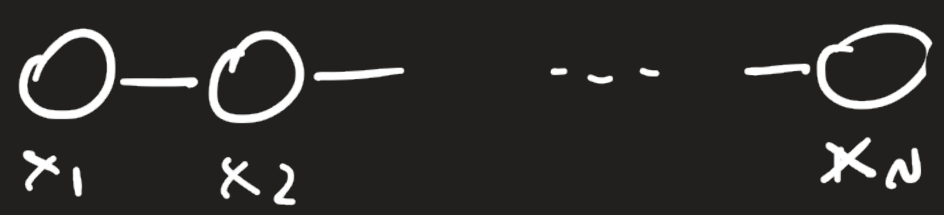
unnormalized conditional

normalization constant

MAX-PLUS ALGORITHM

$p(x)$

$x^M = \underset{x}{\operatorname{argmax}}\, p(x)$



$p(x) = \frac{1}{Z} \psi_{1,2}(x_1, x_2) \cdots \psi_{n-1,n}(x_{n-1}, x_n)$

$\max_x p(x) = \max_{x_1} \cdots \max_{x_n} p(x)$

$= \frac{1}{Z} \max_{x_1} \max_{x_2} \psi_{1,2}(x_1, x_2) \cdots \max_{x_{n-1}} \psi_{n-2,n-1}(x_{n-2}, x_{n-1}) \max_{x_n} \psi_{n-1,n}(x_{n-1}, x_n)$

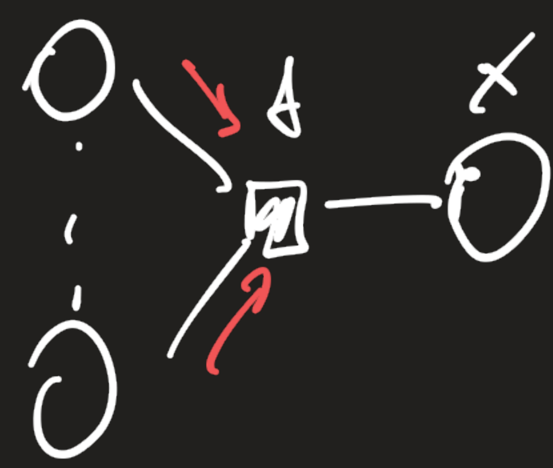
$a > 0 \quad \max(ab, ac) = a \max(b, c)$

$\max(a+b, a+c) = a + \max(b, c)$

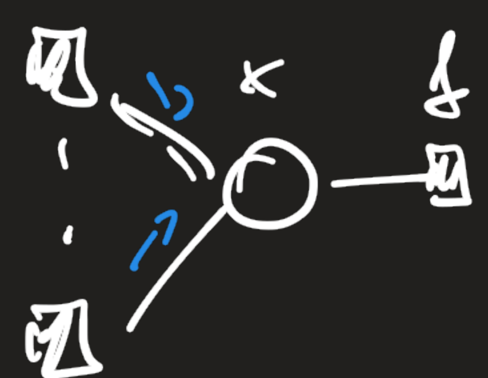
distributive prop of max

We maximize $\log p(x) = \sum_{i} \log \psi_{i,i+1}(x_i, x_{i+1}) + \log(Z)$

max-plus ~ sum-product $\begin{bmatrix} \max & - \text{sum} \\ \text{plus} & - \text{product} \end{bmatrix}$



$$\hat{\mu}_{f\to x}(x) = \max_{x_1,\dots,x_k \in N(f)\setminus x}\left[\log f(x, x_1, \dots, x_k) + \sum_{x_i \in N(f)\setminus x} \mu_{x_i \to f}(x_i)\right]$$



$$\hat{\mu}_{x\to f}(x) = \sum_{f_i \in N(x)\setminus f} \mu_{f_i \to x}(x)$$

$$\hat{\mu}_{x\to f}(x) = 0$$

$$\hat{\mu}_{f\to x}(x) = \log f(x)$$

FIX THE ROOT, do the forward pass

$$x_{root}^{max} = \arg\max_{x_{root}}$$

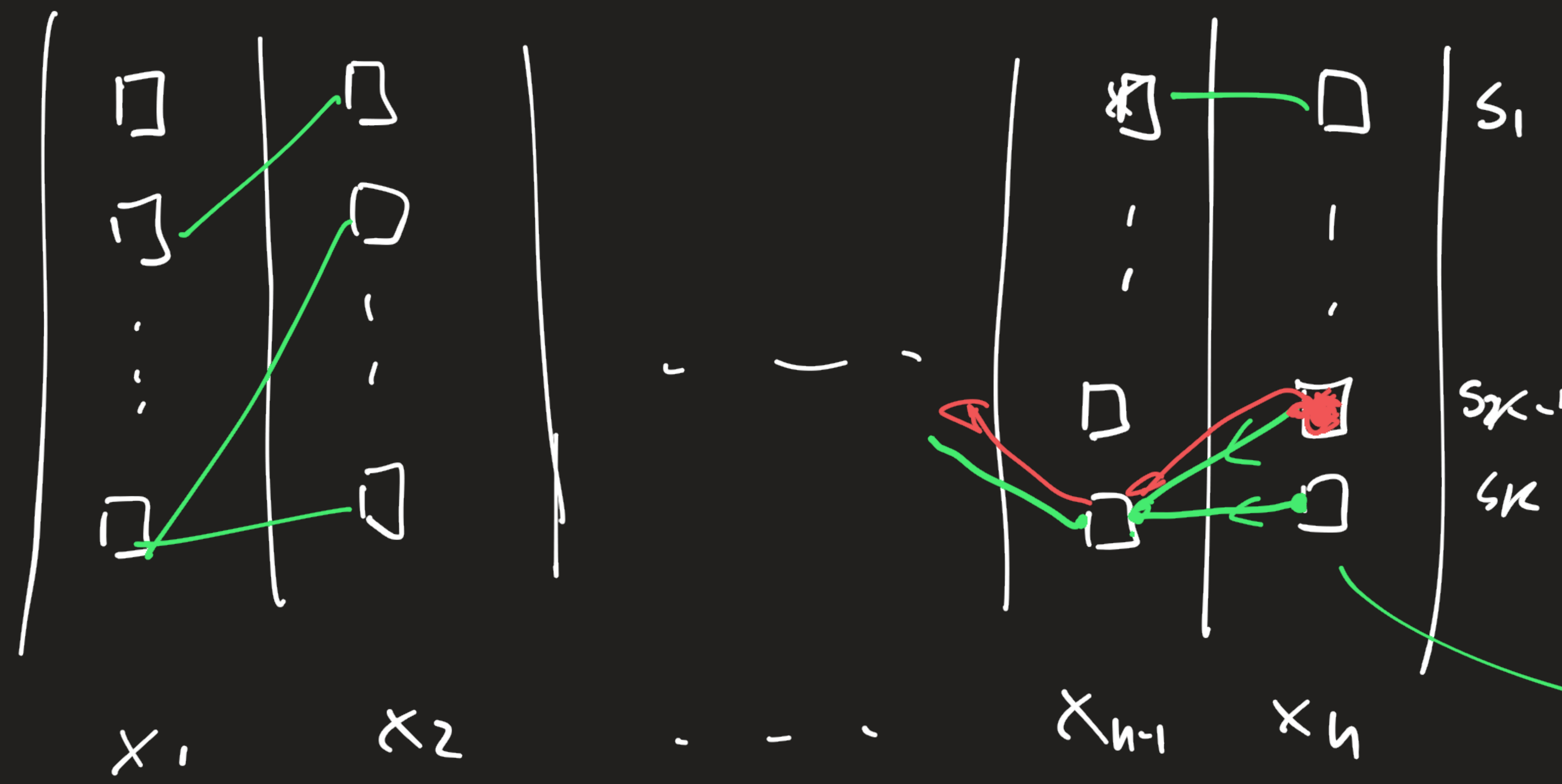
$$\max_{x_{root}} \boxed{\left[\sum_{f \in N(x_{root})} \mu_{f\to x_{root}}(x_{root})\right]} = p_{max}$$

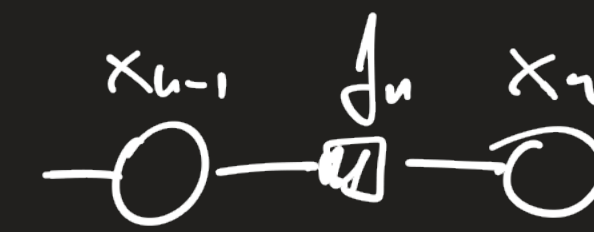
$$\Phi_{f \to x}(x) = \operatorname*{argmax}_{x_1, \dots, x_K \in N(f) \setminus x} \left[\log f(x, x_1, \dots, x_K) + \sum_{x_i \in N(f) \setminus x} \mu_{x_i \to f}(x_i) \right]$$

$$\Phi_{f \to x_{root}}\left(x_{root}^{max}\right) = \begin{cases} x_1 \to x_1^{max} \\ \vdots \\ x_K \to x_K^{max} \end{cases}$$

$$\Phi_{f \to x_i}\left(x_i^{max}\right) \; \dots \; \_\_ \;)$$

LATTICE / TRELLIS diagram



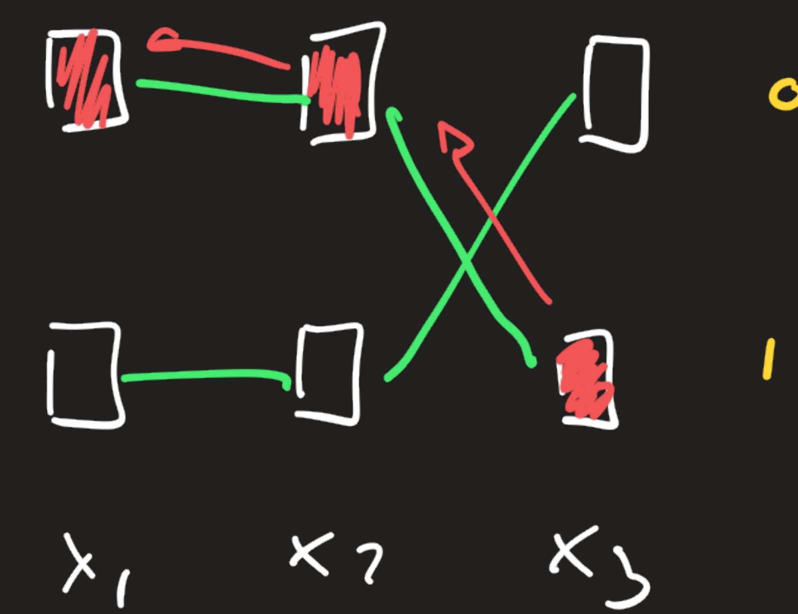


$\Phi_{f_n \to x_n}(s_k) \in \{s_1 \to s_k\}$

(HMM) - VITERBI ALGORITHM.

EXAMPLE $\quad x: x_1, x_2, x_3 \in \{0,1\}$

$$f_1(x_1,x_2) = \left[\begin{array}{c}0.3\\0.2\\0.1\\0.4\end{array}\right]\begin{array}{c}00\\01\\10\\11\end{array} \qquad f_2(x_2,x_3) = \left[\begin{array}{c}0.1\\0.5\\0.2\\0.2\end{array}\right]\begin{array}{c}00\\01\\10\\11\end{array}$$



$(0 \ 0 \ 1)$

$$\mu_{x_1 \to f_1}(x_1) = 0$$

$$\mu_{f_1 \to x_2}(x_2) = \max_{x_1}\left[\log f_1(x_1,x_2) + \mu_{x_1 \to f_1}(x_1)\right] = \left[\begin{array}{c}\log 0.3\\ \log 0.4\end{array}\right]\begin{array}{c}0\\1\end{array}$$

$$\phi_{f_1 \to x_2} = \left[\begin{array}{c}0\\1\end{array}\right]\begin{array}{c}0\\1\end{array}$$

$$\mu_{x_2 \to f_2}(x_2) = \mu_{f_1 \to x_2}(x_2)$$

$$\mu_{f_2 \to x_3}(x_3) = \max_{x_2}\left[\log f_2(x_2,x_3) + \mu_{x_2 \to f_2}(x_2)\right] = \left[\begin{array}{c}\log 0.2 + \log 0.4\\ \log 0.5 + \log 0.3\end{array}\right]\begin{array}{c}0\\1\end{array}$$

$$\phi_{f_2 \to x_3}(x_3) = \left[\begin{array}{c}1\\0\end{array}\right]\begin{array}{c}0\\1\end{array}$$

$$\max_{x_3} \mu_{f_2 \to x_3}(x_3) = \log 0.5 + \log 0.3$$

$$\arg\max \ x_3 = 1$$

INFERENCE IN GENERAL PGM

- JUNCTION TREE ALGORITHM : any F.G.
  - exponential in the size of longest clique.
- MONTE CARLO SAMPLING → next topic
- VARIATIONAL INFERENCE
- BELIEF - LOOPY PROPAGATION : iterative application of sum-product but may not converge.