Gaussian GMs

So far, worked with $P[X_1^n]$ $X_i \in \mathcal{X} \leftarrow$ finite

- Any factor $f_a(x_1, \dots, x_\ell)$ can be written as a $|\mathcal{X}|^\ell$-table
- Algs used only $+, *,$ look-ups

Let's start talking about <u>continuous</u> R.V.'s / dist in $\mathbb{R}^n$.

- A <u>parametric</u> <u>family</u> allows us to $\left[\begin{array}{l}\text{store factors} \\ \text{compute messages}\end{array}\right.$

<u>Def</u> $X = (X_1, \dots, X_n)$ is Gaussian $\mathcal{N}(\mu, \Sigma)$ if

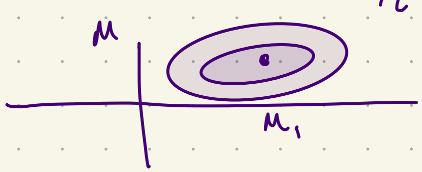
$$P(X) = \frac{1}{(2\pi)^{n/2} \, |\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2} \cdot (x-\mu)^T \Sigma^{-1} (X - \mu)\right)$$

PDF → $P(X)$; determinant → $|\Sigma|$; Σ^{-1} → PD Σ

where $\mu = \mathbb{E}[X]$, $\Sigma = \mathrm{Cov}(X) = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]$

- Σ is symmetric & positive-definite
  - ↑ $x^T \Sigma x > 0 \quad \forall x \neq 0$
  - $\sigma_{\min}(\Sigma) > 0$ (invertible)

ex: $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\right)$



Def The Covariance form of a MV gaussian is $X \sim N(\mu, \Sigma)$ (Σ PD)

$$\longleftrightarrow P(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\tfrac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right]$$

$$\hookrightarrow -\tfrac{1}{2} x^T \Sigma^{-1} x + \tfrac{1}{2}\left(x^T \Sigma^{-1}\mu + \mu^T \Sigma^{-1} x\right) + C$$

$$= -\tfrac{1}{2} x^T \Sigma^{-1} x + (\Sigma^{-1}\mu)^T x + C$$

Def the information form of a MV Gaussian

$$X \sim N^{-1}(h, J) \longleftrightarrow P(x) = \frac{1}{Z_{hJ}} \exp\left(-\tfrac{1}{2} x^T J x + h^T x\right)$$

(J PD)

Claim $N(\mu, \Sigma) = N^{-1}(h, J)$ iff $\begin{cases} \Sigma^{-1} = J \\ \Sigma^{-1}\mu = h \end{cases}$

Remark: Why have both?

(1) Marginalization is easy w. covariance form

(2) Conditioning is eay w. info form

(1) $\begin{bmatrix} X_1 \in \mathbb{R}^n \\ X_2 \in \mathbb{R}^m \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\right)$

Marginal form of X_1? $X_1 \sim N(\mu_1, \Sigma_1)$

Pf A $P(x_1) = \int P(x_1, x_2)\, dx_2$, B $\mathbb{E}[x_1] = \mu_1$

Now just NTS X_1 Gaussian. C $\mathbb{E}[(x_1-\mu_1)^T(x_1-\mu_1)] = \Sigma_{11}$

$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N^{-1}\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \left[\begin{array}{c|c} J_{11} & J_{12} \\ \hline J_{21} & J_{22} \end{array}\right]\right)$

Marginal form wrt h_1, h_2, J's?

$$X_1 \sim N^{-1}\left(h_1 - J_{12}J_{22}^{-1}h_2,\; J_{11} - J_{12} J_{22}^{-1} J_{21}\right)$$

Might be costly

② Conditioning $P(x_1|x_2)$?

Covariance form

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right] \right)$$

$$P(x_1) \sim N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

Σ_{22}^{-1} Might be costly

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N^{-1}\left(\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \left[\begin{array}{c|c} J_{11} & J_{12} \\ \hline J_{21} & J_{22} \end{array}\right] \right)$$

$$P(x_1|x_2) \sim N^{-1}(h_1 - J_{12}x_2, J_{11})$$

Pf

$$P(x_1|x_2) \propto \exp\left(-\tfrac{1}{2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^T \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} h_1^T & h_2^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right)$$

$$\propto \exp\left(-\tfrac{1}{2} x_1^T J_{11} x_1 \right.$$
$$\left. \left(-\tfrac{1}{2} J_{12} x_2 - \tfrac{1}{2} J_{12} x_2 + h_1 \right)^T x_1 \right.$$

Covariance makes $\perp$ easy, Info form makes $\perp\!\!\!\perp$ easy.

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left[\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right]$$

$$x_1 \sim N(\mu_1, \Sigma_{11})$$

$$x_1 \perp x_2 \iff \Sigma_{12} = 0$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} | x_{rest} \sim N^{-1}\left(h, \left[\begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \right] = \tilde{J} \right)$$

$$\sim N(\tilde{\mu}, \tilde{\Sigma} = \tilde{J}^{-1})$$

$$\tilde{\Sigma}_{12} = 0 \longleftrightarrow \tilde{J}_{12} = 0$$

Def Undirected gaussian GMs

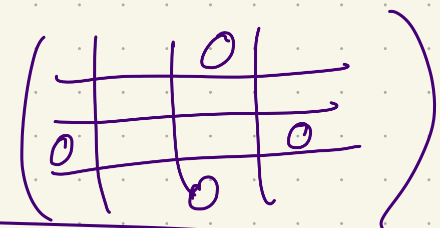
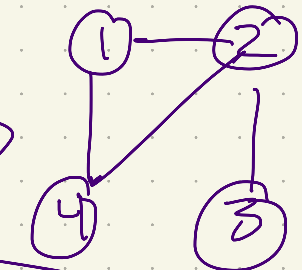
$$X \sim N^{-1}(h, J)$$

$$P(X) = \frac{1}{Z} \prod_{(i,j)\in E} e^{-x_i^T J_{ij} x_j} \times \prod_{i \in V} e^{(-\frac{1}{2} x_i^T J_{ii} x_i + h_i^T x_i)}$$

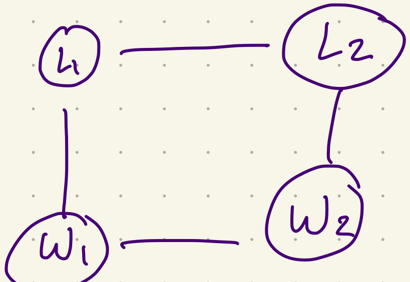
J is called the precision matrix

h the information vector

$$E = \{(i,j) \mid J_{ij} \neq 0\}$$

$J =$  $\rightarrow$ 

Examples: ① Suppose there are 2 children (siblings)



Lengths of heads

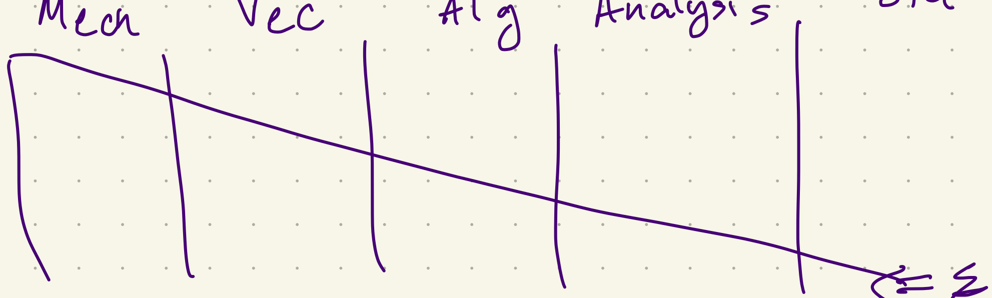
width of heads

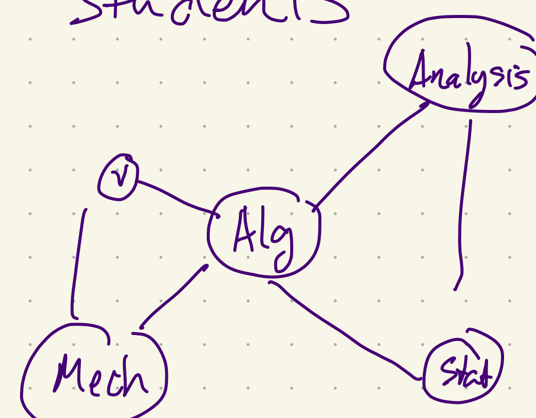
Child1 Child2

② Exam scores in 5 subjects of 88 students

Mech Vec Alg Analysis Stat

Mech
Vectors
Alg
Analysis
stat

 $= \Sigma \Rightarrow J$

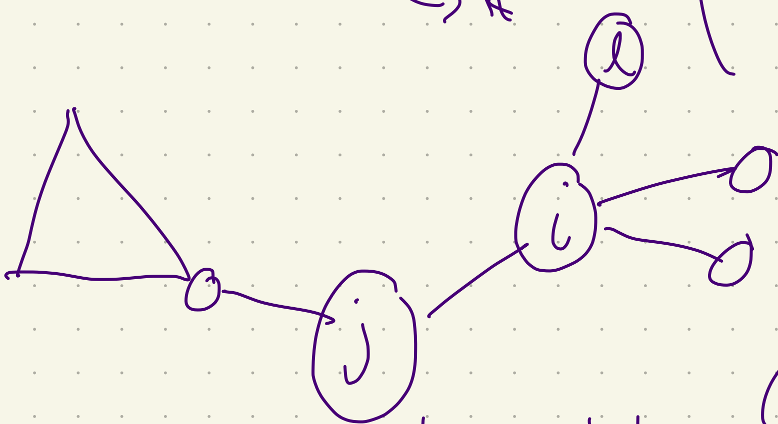


Gaussian BP

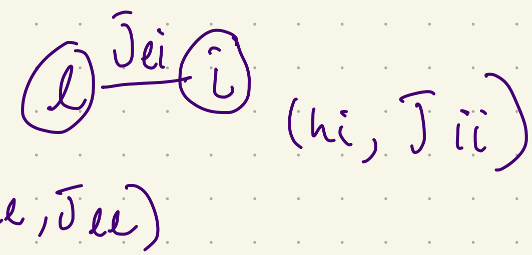
Given $\begin{pmatrix} x_\ell \\ x_i \end{pmatrix} \sim N^{-1}\left(\begin{bmatrix} h_\ell \\ 0 \end{bmatrix}, \begin{bmatrix} J_{\ell\ell} & J_{\ell i} \\ J_{i\ell} & 0 \end{bmatrix}\right)$

$\in \mathbb{R}^{d\times d}$

$\to \mathbb{R}^d$

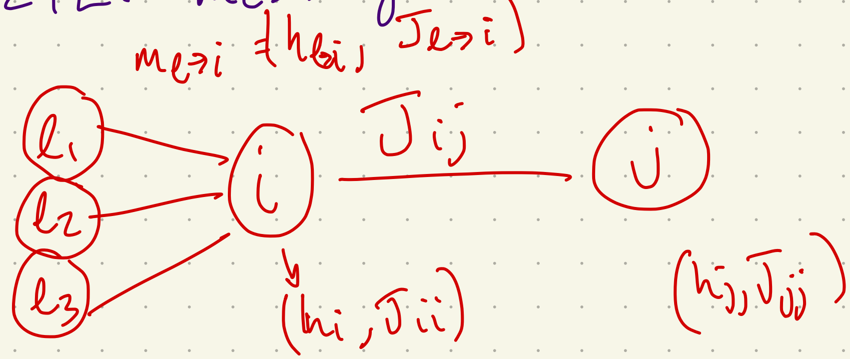


$$\boxed{m_{\ell\to i}(x_i) \propto \exp\left(-\tfrac{1}{2} x_i^T J_{\ell\to i} x_i + h_{\ell\to i}^T x_i\right) \triangleq \left(h_{\ell\to i}, J_{\ell\to i}\right)}$$



message : $\mathbb{R}^d \times \mathbb{R}^{d\times d}$

$2|E|$ messages



$$x_i \sim N^{-1}\left(h_{\ell\to i}, J_{\ell\to i}\right)$$

$$h_{\ell\to i} = -J_{i\ell} J_{\ell\ell}^{-1} h_\ell \qquad J_{\ell\to i} = -J_{i\ell} J_{\ell\ell}^{-1} J_{\ell i}$$

$$h_{i\to j} = -J_{ji}\Big(J_{ii} + \sum_{\ell\in\partial i\setminus\{j\}} J_{\ell\to i}\Big)^{-1}\Big(h_i + \sum_{\ell\in\partial i\setminus\{j\}} h_{\ell\to i}\Big)$$

$$J_{i\to j} = -J_{ji}\Big(J_i + \sum_{\ell\in\partial j\setminus\{i\}} J_{\ell\to i}\Big)^{-1} J_{ij}$$

Decision / Marginal

$$\hat{h}_i = h_i + \sum_{l \in \partial i} \hat{h}_{l \to i}$$

$$x_i \sim N^{-1}(\hat{h}_i, \hat{J}_{ii})$$

$$N(\hat{J}_i^{-1} \hat{h}_i, \hat{J}_{ii}^{-1}) \hookrightarrow O(d^3)$$

$$\hat{J}_{ii} = J_{ii} + \sum_{l \in \partial i} J_{l \to i}$$

T steps of BP takes $O(d^3 \underbrace{|E|}_{\nearrow n} \cdot \underbrace{T}_{\nearrow \log n})$

inverting $J \in \mathbb{R}^{dn \times dn}$ takes $O(d^3 n^3)$

Alt version of GBP

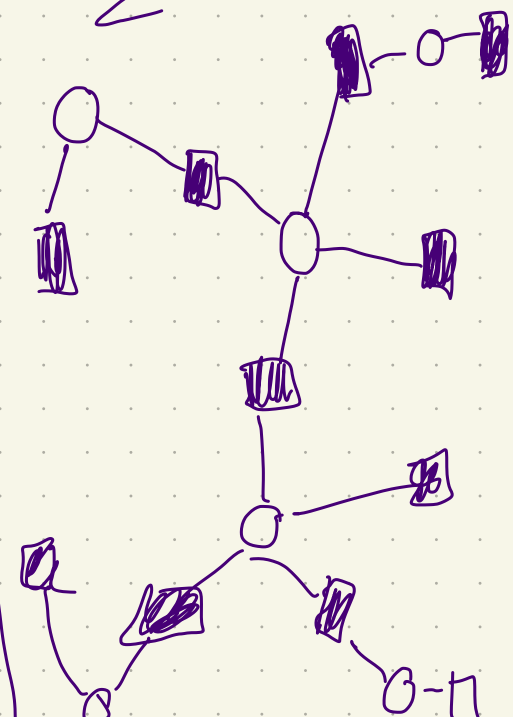
$$h_{i \to j} = h_i - \sum_{l \in \partial i \setminus j} J_{il} J_{l \to i}^{-1} h_{l \to i}$$

$$J_{i \to j} = J_{ii} - \sum_{l \in \partial i \setminus j} J_{il} J_{l \to i}^{-1} J_{li}$$

$$\hat{h}_i = h_i - \sum_{l \in \partial i} J_{il} J_{l \to i}^{-1} h_{l \to i}$$

$$\hat{J}_i = J_i - \sum_{l \in \partial i} J_{il} J_{l \to i}^{-1} J_{li}$$

$\Sigma \Rightarrow \int$



Maximization is equivalent $(\vec{\mu}_1, \ldots \vec{\mu}_n) = \vec{\mu}$

Q: $J = \begin{bmatrix} & \\ & \end{bmatrix}$

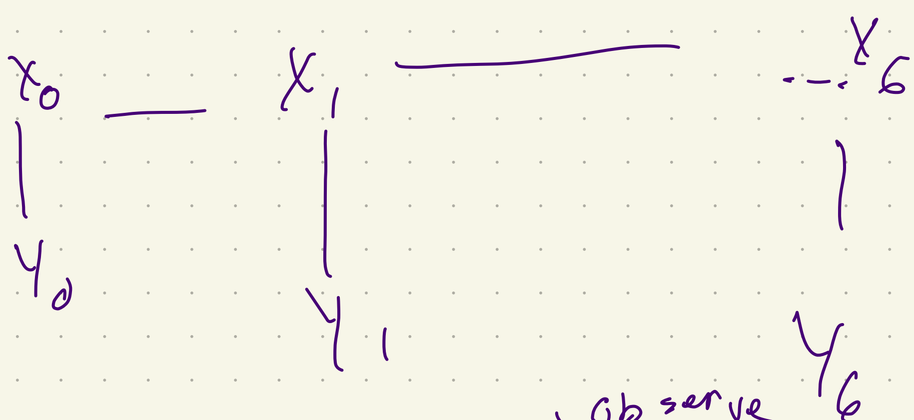
① how can we check if $J \succ 0$?

– $O(n^3)$:(

② $\exists$ suff conditions to $\Rightarrow J \succ 0$?

↳ Linear Dynamical Systems (Kalman Filtering)

G Hidden Markov Models



x_t : state $\in \mathbb{R}^d$

$A \in \mathbb{R}^{d \times d}$ state trans. matrix

$x_0 \sim N(0, \Sigma_0)$

$x_{t+1} = A x_t + B v_t$

observe $y_t \in \mathbb{R}^{J}$

noise $w_t \sim N(0, W)$

$y_t = C x_t + w_t$

Process noise $\begin{cases} v_t \in \mathbb{R}^p \sim N(0, V) \\ B \in \mathbb{R}^{d \times p} \end{cases}$

GGM $\quad x_0 \sim N(0, \Sigma_0)$

$x_{t+1} \mid x_t \sim N(A x_t, H = BVB^T)$

$y_t \mid x_t \sim N(C x_t, W)$

Factorization:

$$P_y(x) = \frac{1}{Z} \underbrace{\exp\left(-\frac{1}{2} x_0^T \Sigma_0^{-1} x_0\right)}_{P(x_0)} \underbrace{\exp\left(-\frac{1}{2}(x_1 - Ax_0)^T H^{-1} (x_1 - Ax_0)\right)}_{P(x_2 \mid x_0)}$$

$$\underbrace{\exp\left(-\frac{1}{2}\, y_0 - (x_0)^T W^{-1} (y_0 - Cx_0)\right)}_{P_{y_0 \mid x_0}} \cdots$$

<u>Information form</u>:

$$= \frac{1}{Z} \prod_{i=0}^{n} \exp\left(-\frac{1}{2} x_i J_i x_i + h_i^T\right.$$