# * Gaussian Graphical Models.

**Covariance Form**

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

**Information Form**

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

**Marginal**

$$X_1 \sim N(\mu_1, \Sigma_{11})$$

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

Independence

$$X_1 \perp\!\!\!\perp X_2 \iff \Sigma_{12} = 0$$

uncorrelated $\Leftrightarrow$ independence

**Conditioning**

$$X_1 | X_2 \sim N\left( \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2),\; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \right)$$

$$X_1 | X_2 \sim N^{-1}(h_1 - J_{12}X_2,\; J_{11})$$

conditional Independence

$$X_i \perp\!\!\!\perp X_j \mid X_{rest} \iff J_{ij} = 0.$$

Schur Complement $\triangleq S = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

why?

$$\exp\left( -\frac{1}{2} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} (\Sigma_{11} - \Sigma_{12}\Sigma_{22}\Sigma_{21})^{-1}, & -S^{-1}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}S^{-1}, & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}S^{-1}\Sigma_{12}\Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \right)$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \Big| X_{rest} \sim N^{-1}\left( \;,\; \underbrace{\begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}}_{A} \right)$$

$$\sim N(\;,\; A^{-1})$$

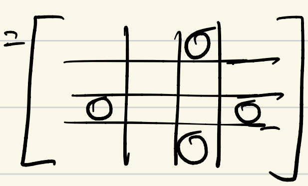
$$X_1 \perp\!\!\!\perp X_2 \mid X_{rest} \longleftrightarrow (A^{-1})_{12} = 0$$

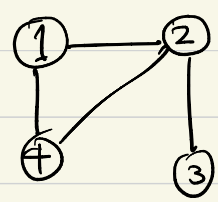
$$J_{12} = 0$$

## Def. Undirected Gaussian Graphical Model

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

$$P(x) = \frac{1}{Z} \cdot \prod_{i=1}^{n} \underbrace{e^{-\frac{1}{2}x_i^T J_{ii} x_i + h_i^T x_i}}_{f_i(x_i)} \prod_{(i,j)\in E} \underbrace{e^{-x_i^T J_{ij} x_j}}_{f_{ij}(x_i, x_j)}$$

$J =$ 





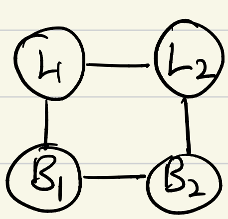
$X_1 \perp\!\!\!\perp X_3 \mid X_2, X_4$

$X_3 \perp\!\!\!\perp X_4 \mid X_1, X_2$

ex> [Whittaker 1990]

$(L_1, W_1, L_2, W_2)$

$L_1$: length of head of 1st child; $W_1$: width; $L_2$: length of head of 2nd child; $W_2$: width

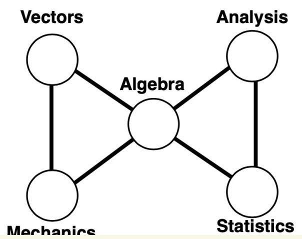




$\leftarrow$ Graphical Model learned from data.

- Examination scores of 88 students in 5 subjects
- empirical information matrix (diagonal and above) covariance (below diagonal)

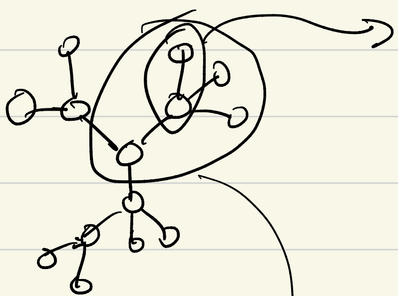
| | Mechanics | Vectors | Algebra | Analysis | Statistics |
|---|---|---|---|---|---|
| Mechanics | 5.24 | -2.44 | -2.74 | 0.01 | -0.14 |
| Vectors | 0.33 | 10.43 | -4.71 | -0.79 | -0.17 |
| Algebra | 0.23 | 0.28 | 26.95 | -7.05 | -4.70 |
| Analysis | 0.00 | 0.08 | 0.43 | 9.88 | -2.02 |
| Statistics | 0.02 | 0.02 | 0.36 | 0.25 | 6.45 |

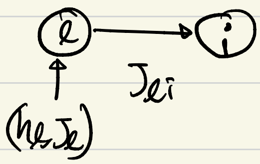



* Gaussian Belief Propagation

in the same way as we derived BP for discrete R.V.s

we design an elimination algorithm for a tree $\Longrightarrow$ parallel BP.





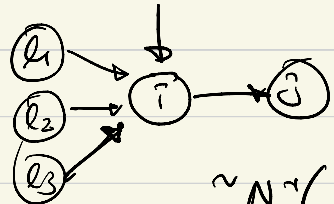

message

$$N^{-1}(h_{l\to i}, J_{l\to i}) \sim m_{l\to i}(x_i)$$

Given $\begin{pmatrix} x_l \\ x_i \end{pmatrix} \sim N^{-1}\left(\begin{bmatrix} h_l \\ 0 \end{bmatrix}, \begin{bmatrix} J_l & J_{li} \\ J_{il} & 0 \end{bmatrix}\right)$

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

$-J_{il}J_l^{-1}h_l$ $\qquad$ $-J_{il}J_l^{-1}J_{li}$




$$m_{i\to j}(x_j) \sim N^{-1}(h_{i\to j}, J_{i\to j})$$

$$\sim N^{-1}\Big( -J_{ji}\big(J_i + \sum_{l\in \partial i\setminus\{j\}} J_{l\to i}\big)^{-1}\big(h_i + \sum_{l\in\partial i\setminus\{j\}} h_{l\to i}\big),\ -J_{ji}\big(J_i + \sum_{l\in\partial i\setminus\{j\}} J_{l\to i}\big)^{-1} J_{ij} \Big)$$

---

*GBP

parallel update: $h_{i\to j} = -J_{ji}\big(J_i + \sum_{l\in\partial i\setminus\{j\}} J_{l\to i}\big)^{-1} \cdot \big(h_i + \sum_{l\in\partial i\setminus\{j\}} h_{l\to i}\big)$

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

Decision/Marginal : $\hat{h}_i = h_i + \sum_{l\in\partial i} h_{l\to i}$

$$\hat{J}_i = J_i + \sum_{l\in\partial i} J_{l\to i}$$

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

* Analogous to discrete, there are 2 versions of BP on Gaussian Graphical model

① $m_{i\to j}(X_j)$ we just derived above.

② $m_{i\to j}(X_i)$ can be derived analogously.

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

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

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

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

Consider $n$ variables $X_1, \dots, X_n$ each in $\mathbb{R}^d$, then

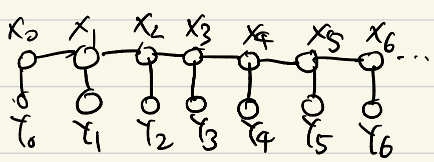
GBP takes $(|E|\cdot O(d^3))$-time per iteration

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

* at the end of GBP, we get $X_i \sim N^{-1}(\hat{h}_i, \hat{J}_i) \xrightarrow{\text{Compute}} N(\underbrace{\hat{J}_i^{-1}\hat{h}_i}_{\text{marginal}}, \hat{J}_i^{-1})$

* MAP or Maximization is $\hat{\mu}_i = \hat{J}_i^{-1}\hat{h}_i$ (the same algorithm GBP)

Gaussian Hidden Markov Models.




State $X_i \in\mathbb{R}^d$

State transition matrix: $A\in\mathbb{R}^{d\times d}$

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

$$X_0 \sim N(0, \Sigma_0)$$

$$X_{t+1} = AX_t + BV_t$$

observation: $Y_t \in\mathbb{R}^{d'}$, noise $W_t \sim N(0, W)$.

$$Y_t = CX_t + W_t.$$

$$x_0 \sim N(0, \Sigma_0)$$
$$x_{t+1} | x_t \sim N(Ax_t, H = BVB^T)$$
$$y_t | x_t \sim N(Cx_t, W)$$

· factorization

$$p(x) = \frac{1}{Z} \cdot \exp\left(-\frac{1}{2} x_0^T \Sigma_0^{-1} x_0\right) \cdot \exp\left(-\frac{1}{2}(x_1 - Ax_0)^T H^{-1}(x_1 - Ax_0)\right) \cdot \exp(\cdots)$$
$$\cdot \exp\left(-\frac{1}{2}(y_0 - Cx_0)^T W^{-1}(y_0 - Cx_0)\right) \cdot \exp\left(-\frac{1}{2}(y_1 - Cx_1)^T W^{-1}(y_1 - Cx_1)\right) \cdots$$

information form → $= \frac{1}{Z} \prod_{i=0}^{n} \exp\left(-\frac{1}{2} x_i^T J_i x_i\right) \cdot \prod_{i=0}^{n-1} \exp\left(-x_i^T J_{i,i+1} x_{i+1}\right) \cdot \prod_{i=0}^{n} \exp(h_i^T x_i)$

$$J_i = \begin{cases} \Sigma_0^{-1} + C^T W^{-1} C + A^T H^{-1} A, & i = 0 \\ H^{-1} + C^T W^{-1} C + A^T H^{-1} A, & 0 < i < n \\ H^{-1} + C^T W^{-1} C, & i = n \end{cases}$$

$$J_{i,i+1} = -A^T H^{-1}, \qquad h_i = y_i^T W^{-1} C$$

* Gaussian BP.

initialize: $J_{i \to j} = J_i, \quad h_{i \to j} = h_i$

forward update: $J_{i \to i+1} = J_i - J_{i,i-1} J_{i-1 \to i}^{-1} J_{i,i-1}^T$

$h_{i \to i+1} = h_i - J_{i,i-1} J_{i-1 \to i}^{-1} \cdot h_{i-1 \to i}$

backward update: $J_{i \to i-1} = J_i - J_{i,i+1} J_{i+1 \to i}^{-1} J_{i,i+1}^T$

$h_{i \to i-1} = h_i - J_{i,i+1} J_{i+1 \to i}^{-1} h_{i+1 \to i}$

Decision: $\hat{J}_i = J_i + J_{i,i-1} J_{i-1 \to i}^{-1} J_{i,i-1}^T + J_{i,i+1} J_{i+1 \to i}^{-1} J_{i,i+1}^T$

$\hat{h}_i = h_i + J_{i,i-1} J_{i-1 \to i}^{-1} h_{i-1 \to i} + J_{i,i+1} J_{i+1 \to i}^{-1} h_{i+1 \to i}$

* Q. Given a $N^{-1}(h, J)$, how do we check if $J$ is positive definite?

P.D ⟷ · $x^T J x > 0, \ \forall x \neq 0$

· all eigenvalues of $J$ are positive

· has a Cholesky decomposition: there exist a (unique) lower triangular matrix $L$ with strictly positive diagonal entries s.t. $J = L^T L$.

*checking PD is computationally expensive $O(d^3)$

· Satisfies Sylvester's criterion: leading principal minors are all positive, where k-th leading principal minor of $J$ is the determinant of its upper left $k \times k$ submatrix.

*Claim: [Weiss, Freeman 2001, Rusmevichientong, Van Roy 2001]

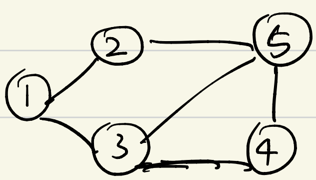
If Gaussian BP converges, then the expectations are computed correctly. Formally let

$$\hat{\mu}_i^{(\ell)} = (\hat{J}_i^{(\ell)})^{-1} \hat{h}_i^{(\ell)}$$

after $\ell$ steps of GBP.

If $\hat{\mu}_i^{(\infty)} \triangleq \lim_{\ell\to\infty} \hat{\mu}_i^{(\ell)}$ exists then $\hat{\mu}_i^{(\infty)} = \mu_i$.

Def. computation tree. $CT_G(i;\ell)$ is the tree of $\ell$-steps of non-reversing walks on $G$ starting at $i$



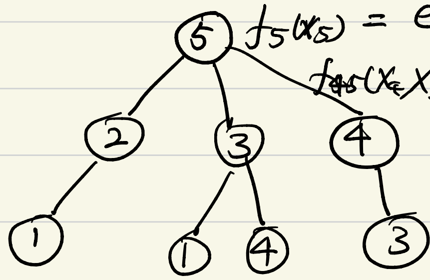
G.M $G$. $f_i(x_i)$'s

$f_{ij}(x_i, x_j)$'s.

$\phi$ = root = $i$ = 5

$f_5(x_5) = e^{(h_5^T x_5 - \frac{1}{2} x_5^T J_5 x_5)}$

$f_{45}(x_4, x_5) = e^{-x_4^T J_{45} x_5}$



CT

- CT consists of copies of variable nodes & copies of singleton & pairwise factors,

$$\hat{m}_i^{(\ell)}(G) = \hat{m}_\phi^{(\ell)}(CT_G(i;\ell))$$

proof: by induction we can show that BP update is exactly the same.

since $J \cdot \mu = h$ $\quad$ $J$ is invertible

$\mu = [\mu_1, \cdots, \mu_n]$ is the unique solution to

$$J_i \mu_i + \sum_{j \in \partial i} J_{ij} \mu_j = h_i$$

$CT_G(i;\ell)$

consider $t \geq 2\ell$, and BP marginals

$$\{\hat{\mu}_i^{(t)}\}$$

Because this is correct ($\leftarrow$ Tree) marginal of a (possibly bigger) graphical model, they also satisfy

$$\tilde{J} \cdot \tilde{\mu} = \tilde{h} \quad \leftarrow \text{on } CT_G(i;\ell)$$

$$\Rightarrow J_i \cdot \hat{\mu}_{\phi}^{(t)} + \sum_{j \in \partial i} J_{ij} \hat{\mu}_j^{(t)} = h_i$$

Now we grow this tree & # steps  let $t \uparrow \infty$

$$\Rightarrow J_i \hat{\mu}_{\phi}^{(\infty)} + \sum_{j \in \partial i} J_{ij} \hat{\mu}_j^{(\infty)} = h_i$$

convergence

Since $\hat{\mu}_i^{(\infty)}$'s satisfy the same set of linear equations as $\mu_i$'s they are the same.