Algorithme somme produit

Francis Bach

15 novembre 2006

Soit $T = (V, E)$ un arbre non orienté à n sommets et $n - 1$ arêtes. Soit $u(x)$ un produit de potentiels positifs se factorisant dans T, i.e.,

$$u(x) = \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j) \prod_{i\in V} \psi_i(x_i).$$

Les marginales sur les singletons et sur les paires d'éléments voisins dans T sont définies par

$$\forall i \in V,\ u_i(x_i) = \sum_{x_j, j\neq i} u(x)$$

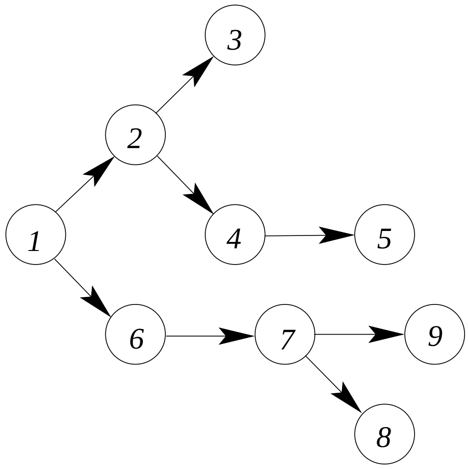
$$\forall (i,j) \in E,\ u_{ij}(x_i, x_j) = \sum_{x_k, k\neq i,j} u(x)$$

On considère d'abord que chaque variables x_i prend un nombre fini de valeurs. Le message passé de i vers j est défini par :

$$\boxed{m_{ij}(x_j) = \sum_{x_i} \psi_i(x_i)\psi_{ij}(x_i, x_j) \prod_{k\in\mathcal{N}(i)\backslash\{j\}} m_{ki}(x_i)} \tag{1}$$

où $\mathcal{N}(i)$ est l'ensemble des voisins de i dans T.

On considère la procédure suivante (dite "en série") pour l'ordre de passage des messages : on définit une racine en choisissant un des sommet (au hasard) ; cette racine permet de définir un unique arbre orienté. On considère un ordre topologique quelconque (i.e., tel qu'un noeud apparaiî toujours après ses parents), et on suppose que les noeuds sont etiquettés en suivant l'ordre topologique, i.e., $V = \{1, \ldots, n\}$. Voir figure.



Les $2(n-1)$ messages sont passés dans l'ordre suivant :

$$n \to \pi_n, n-1 \to \pi_{n-1}, \dots, 2 \to \pi_2, \pi_2 \to 2, \dots, \pi_n \to n$$

Cette procédure respecte bien le protocole de passage des messages, i.e., un sommet n'envoie un message que quand il a recu un message de tous ses autres voisins.

Nous allons démontrer qu'une fois les $2(n-1)$ messages passés, les marginales sont égales à

$$\boxed{\forall i \in V,\ u(x_i) = \psi_i(x_i) \prod_{k \in \mathcal{N}(i)} m_{k,i}(x_i),} \tag{2}$$

$$\boxed{\forall (i,j) \in E,\ u(x_i, x_j) = \psi_i(x_i)\psi_j(x_j)\psi_{ij}(x_i, x_j) \prod_{k \in \mathcal{N}(i)\setminus\{j\}} m_{k,i}(x_i) \prod_{k \in \mathcal{N}(j)\setminus\{i\}} m_{k,j}(x_j),} \tag{3}$$

1 Preuve

Le résultat se démontre par récurrence sur le nombre de sommets. L'hypothèse de récurrence est que pour tout arbre de taille n, toute famille de potentiels, toute racine, et tout ordre topologique correspondant, les équations (2) et (3) sont vraies.

$n = 2$

Si T a deux sommets 1 et 2, alors deux messages sont passés, de 1 vers 2 : $m_{12}(x_2) = \sum_{x_1} \psi_1(x_1)\psi_{12}(x_1, x_2)$, puis de 2 vers 1 : $m_{21}(x_1) = \sum_{x_2} \psi_2(x_2)\psi_{12}(x_1, x_2)$. D'autre part, on a par définition

$$\begin{aligned} u(x_1) &= \sum_{x_2} \psi_1(x_1)\psi_2(x_2)\psi_{12}(x_1, x_2) \\ &= \psi_1(x_1) \sum_{x_2} \psi_2(x_2)\psi_{12}(x_1, x_2) \\ u(x_1) &= \psi_1(x_1) m_{21}(x_1), \end{aligned}$$

et de façon analogue, $u(x_2) = \psi(x_2)m_{12}(x_2)$. Ainsi, (2) est vérifiée.

Enfin, on a directement par définition $u(x_1, x_2) = \psi(x_1)\psi(x_2)\psi(x_1, x_2)$. D'où, (3) est également vérifiée.

On a donc montré que le résultat est vrai pour $n = 2$.

$n - 1 \to n$

On suppose que le résultat est vrai pour les arbres de taille $n-1$, et on considère un arbre de taille n, avec racine et ordre correspondant.

Comme le sommet n (dernier dans l'ordre topologique) n'a qu'un seul voisin π_n, le premier message passé, de n vers π_n est

$$m_{n\pi_n}(x_{\pi_n}) = \sum_{x_n} \psi_n(x_n)\psi_{n\pi_n}(x_n, x_{\pi_n})$$

Le dernier message passé, de π_n vers n est égal à :

$$m_{\pi_n n}(x_n) = \sum_{x_{\pi_n}} \psi_{\pi_n}(x_{\pi_n})\psi_{n\pi_n}(x_n, x_{\pi_n}) \prod_{k \in \mathcal{N}(\pi_n)\backslash\{n\}} m_{k\pi_n}(x_{\pi_n})$$

Nous allons construire un arbre $\tilde{T}$ de taille $n-1$, ainsi qu'une famille de potentiels, de telle sorte que les $2(n-2)$ messages passés dans T (i.e., tous les messages exceptés le premier et le dernier) soient égaux aux $2(n-2)$ messages passés dans $\tilde{T}$. On définit l'arbre et les potentiels comme suit :

- $\tilde{T} = (\tilde{V}, \tilde{E})$ avec $\tilde{V} = \{1, \ldots, n-1\}$ et $\tilde{E} = E\backslash\{n, \pi_n\}$ (i.e., c'est le sous-arbre correspondant aux $n-1$ premiers sommets).
- Les potentiels sont tous les mêmes que ceux de T, excepté le potentiel $\tilde{\psi}_{\pi_n}(x_{\pi_n}) = \psi_{\pi_n}(x_{\pi_n})m_{n\pi_n}(x_{\pi_n})$.
- La racine est inchangée et l'ordre topologique est aussi conservé.

Le produit des potentiels de l'arbre de taille $n-1$, est égale à :

$$\begin{aligned}
\tilde{u}(x_1, \ldots, x_{n-1}) &= \prod_{i=1}^{n-1} \psi_i(x_i) \left(\prod_{(i,j)\in E\backslash\{n,\pi_n\}} \psi_{ij}(x_i, x_j) \right) m_{n\pi_n}(x_{\pi_n}) \\
&= \prod_{i=1}^{n-1} \psi_i(x_i) \left(\prod_{(i,j)\in E\backslash\{n,\pi_n\}} \psi_{ij}(x_i, x_j) \right) \sum_{x_n} \psi_n(x_n)\psi_{n\pi_n}(x_n, x_{\pi_n}) \\
&= \sum_{x_n} \prod_{i=1}^{n} \psi_i(x_i) \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j) \\
&= \sum_{x_n} u(x)
\end{aligned}$$

et est donc égal à la marginalisation de $u(x)$ aux $n-1$ premiers sommets.

Par construction, tous les messages passés dans $\tilde{T}$ correspondent aux messages passés dans T (hormis le premier et le dernier). Nous montrons maintenant que les équations (2) et (3) sont vraies.

Cas 1 : $i \neq n, i \neq \pi_n$

Les messages correspondant à ces noeuds sont les mêmes dans les deux arbres T et $\tilde{T}$, et comme les marginales sont les mêmes (car $\tilde{u}$ est elle-même la marginalisation de u sur les $n-1$ premiers sommets), les équations (2) et (3) sont vraies par hypothèse de récurrence. En particulier, (2) est vraie pour tout $i \notin \{n, \pi_n\}$ et (3) est vraie pour tout $i \notin \{n, \pi_n\}$ et j voisin de i.

Cas 2 : $i = \pi_n$

Dans ce cas, par hypothèse de récurrence,

$$\begin{aligned}
\tilde{u}(x_{\pi_n}) &= \tilde{\psi}_{\pi_n}(x_{\pi_n}) \prod_{k \in \tilde{\mathcal{N}}(\pi_n)} m_{k\pi_n}(x_{\pi_n}) \quad \text{(produit sur les voisins de } \pi_n \text{ dans } \tilde{T}) \\
&= \tilde{\psi}_{\pi_n}(x_{\pi_n}) \prod_{k \in \mathcal{N}(\pi_n) \setminus \{n\}} m_{k\pi_n}(x_{\pi_n}) \\
&= \psi(x_{\pi_n}) m_{n\pi_n}(x_{\pi_n}) \prod_{k \in \mathcal{N}(\pi_n) \setminus \{n\}} m_{k\pi_n}(x_{\pi_n}) \text{ par définition de } \tilde{\psi}_{\pi_n} \\
&= \psi(x_{\pi_n}) \prod_{k \in \mathcal{N}(\pi_n)} m_{k\pi_n}(x_{\pi_n})
\end{aligned}$$

Comme $\tilde{u}(x_{\pi_n}) = u(x_{\pi_n})$, l'équation (2) est vérifiée.

Cas 3 : $i = n$

Il reste à montrer que les marginales sur n et (n, π_n) sont effectivement correctes. On a :

$$\begin{aligned}
u(x_n, x_{\pi_n}) &= \sum_{\substack{x_i \\ i \neq n, i \neq \pi_n}} u(x) \\
&= \psi_n(x_n) \psi_{\pi_n}(x_{\pi_n}) \psi_{n\pi_n}(x_n, x_{\pi_m}) \underbrace{\sum_{\substack{x_i \\ i \neq n, i \neq \pi_n}} \frac{u(x)}{\psi_n(x_n) \psi_{\pi_n}(x_{\pi_n}) \psi_{n\pi_n}(x_n, x_{\pi_n})}}_{\alpha(x_{\pi_n})},
\end{aligned}$$

et par conséquent

$$\begin{aligned}
u(x_{\pi_n}) &= \sum_{x_n} u(x_n, x_{\pi_n}) \\
&= \left(\sum_{x_n} \psi_n(x_n)\psi_{n\pi_n}(x_n, x_{\pi_n})\right)\psi_{\pi_n}(x_{\pi_n})\alpha(x_{\pi_n}) \\
&= m_{n\pi_n}(x_{\pi_n})\psi_{\pi_n}(x_{\pi_n})\alpha(x_{\pi_n}).
\end{aligned}$$

On en déduit alors

$$\alpha(x_{\pi_n}) = \frac{1}{\psi_{\pi_n}(x_{\pi_n})}\frac{u(x_{\pi_n})}{m_{n\pi_n}(x_{\pi_n})}.$$

et

$$u(x_n, x_{\pi_n}) = \psi_n(x_n)\psi_{n\pi_n}(x_n, x_{\pi_n})\frac{u(x_{\pi_n})}{m_{n\pi_n}(x_{\pi_n})}.$$

En utilisant le résultat montré pour le cas 2, on obtient :

$$\begin{aligned}
u(x_n, x_{\pi_n}) &= \psi_n(x_n)\psi_{n\pi_n}(x_n, x_{\pi_n})\frac{\psi_{\pi_n}(x_{\pi_n})\prod_{k\in\mathcal{N}(\pi_n)} m_{k\pi_n}(x_{\pi_n})}{m_{n\pi_n}(x_{\pi_n})} \\
&= \psi_n(x_n)\psi_{\pi_n}(x_{\pi_n})\psi_{n\pi_n}(x_n, x_{\pi_n}) \prod_{k\in\mathcal{N}(\pi_n)\setminus\{n\}} m_{k\pi_n}(x_{\pi_n})
\end{aligned}$$

ce qui montre le résultat pour la probabilité jointe sur x_n, x_{π_n}. En sommant par rapport à x_{π_n}, on obtient immédiatement le résultat pour $u(x_n)$:

$$\begin{aligned}
u(x_n) &= \sum_{x_{\pi_n}} u(x_n, x_{\pi_n}) \\
&= \sum_{x_{\pi_n}} \psi_n(x_n)\psi_{\pi_n}(x_{\pi_n})\psi_{n\pi_n}(x_n, x_{\pi_n}) \prod_{k\in\mathcal{N}(\pi_n)\setminus\{n\}} m_{k\pi_n}(x_{\pi_n}) \\
&= \psi_n(x_n)\sum_{x_{\pi_n}} \psi_{\pi_n}(x_{\pi_n})\psi_{n\pi_n}(x_n, x_{\pi_n}) \prod_{k\in\mathcal{N}(\pi_n)\setminus\{n\}} m_{k\pi_n}(x_{\pi_n}) \\
&= \psi_n(x_n)m_{\pi_n n}(x_n) \text{ par définition de } m_{\pi_n n}
\end{aligned}$$

2 Potentiels Gaussiens

On fait maintenant l'hypothèse que chaque x_i est une variable **multivariée** réelle, et que les potentiels $\psi_i(x_i)$ et $\psi_{ij}(x_i, x_j)$ sont des exponentielles de fonctions quadratiques, i.e.,

$$\psi_i(x_i) \propto \exp\left(-\frac{1}{2}x_i^\top \Lambda_i x_i + \eta_i^\top x_i\right)$$

$$\psi_{ij}(x_i, x_j) \propto \exp\left(-\frac{1}{2}\begin{pmatrix} x_i \\ x_j \end{pmatrix}^\top \begin{pmatrix} A_{ij} & B_{ij} \\ B_{ij}^\top & C_{ij} \end{pmatrix}\begin{pmatrix} x_i \\ x_j \end{pmatrix} + \begin{pmatrix} a_{ij} \\ c_{ij} \end{pmatrix}^\top \begin{pmatrix} x_i \\ x_j \end{pmatrix}\right)$$

où la notation $q(x) \propto r(x)$ signifie $q(x) = ar(x)$ où a est une constante (indépendante de x), et les matrices sont de tailles appropriées (dépendant des dimensions de x_i et x_j).

Tous les messages passés dans l'algorithme somme-produit seront aussi des exponentielles de fonctions quadratiques, donc de la forme :

$$m_{ji}(x_i) \propto \exp\left(-\frac{1}{2}x_i^\top Q_{ji}x_i + \xi_{ji}^\top x_i\right)$$

Le passage des messages est équivalent à recalculer Q_{ji} et ξ_{ji} en fonction des messages recus en j. Plus précisément, on a :

$$\begin{aligned}
m_{ij}(x_j) &\propto \int_{x_i} \psi_i(x_i)\psi_{ij}(x_i,x_j) \prod_{k\in\mathcal{N}(i)\backslash\{j\}} m_{ki}(x_i) \\
&\propto \int \exp\Bigg[-\frac{1}{2}\begin{pmatrix} x_i \\ x_j \end{pmatrix}^\top \begin{pmatrix} A_{ij}+\Lambda_i+\sum_{k\in\mathcal{N}(i)\backslash\{j\}} Q_{ki} & B_{ij} \\ B_{ij}^\top & C_{ij} \end{pmatrix}\begin{pmatrix} x_i \\ x_j \end{pmatrix} \\
&= \qquad\qquad + \begin{pmatrix} a_{ij}+\eta_i+\sum_{k\in\mathcal{N}(i)\backslash\{j\}} \xi_{kj} \\ c_{ij} \end{pmatrix}^\top \begin{pmatrix} x_i \\ x_j \end{pmatrix}\Bigg] dx_i
\end{aligned}$$

En utilisant le Lemme 1, on obtient les passages de message suivants :

$$\boxed{Q_{ij} = C_{ij} - B_{ij}^\top \left(A_{ij} + \Lambda_i + \sum_{k\in\mathcal{N}(i)\backslash\{j\}} Q_{ki}\right)^{-1} B_{ij}}$$

$$\boxed{\xi_{ij} = c_{ij} - B_{ij}^\top \left(d_{ij} + \Lambda_i + \sum_{k\in\mathcal{N}(i)\backslash\{j\}} Q_{ki}\right)^{-1} \left(a_{ij} + \eta_i + \sum_{k\in\mathcal{N}(i)\backslash\{j\}} \xi_{kj}\right)}$$

Lemme 1 *Si* $u(x,y) \propto \exp\left(-\frac{1}{2}\begin{pmatrix} x \\ y \end{pmatrix}^\top \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ c \end{pmatrix}^\top \begin{pmatrix} x \\ y \end{pmatrix}\right)$, *alors*

$$\int u(x,y)dx \propto \exp\left(-\frac{1}{2}x^\top(C - B^\top A^{-1}B)x + x^\top(c - B^\top A^{-1}a)\right)$$