

PROBABILISTIC ROBOTICS: THE SPARSE EXTENDED INFORMATION FILTER

Pierre-Paul TACHER

2

We consider a probability density $p(x_1, x_2, x_3)$ in $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p$. With the same abuse of notation of the book, we consider also its approximation defined by

$$\begin{aligned}\tilde{p}(x_1, x_2, x_3) &= \frac{p(x_1, x_3)p(x_2, x_3)}{p(x_3)} \\ &= p(x_1, x_3)p(x_2 | x_3) \\ &= p(x_1 | x_3)p(x_2, x_3)\end{aligned}$$

Each p refer to different density of marginals. Let $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ and $\tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \tilde{X}_3 \end{bmatrix}$ two r.v. in $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p$ with density p and \tilde{p} respectively.

$$\begin{aligned}E[\tilde{X}_1] &= \int_{\Omega} \tilde{X}_1 dP \\ &= \int_{\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p} x_1 \tilde{p}(x_1, x_2, x_3) dx_1 dx_2 dx_3 \\ &= \int_{\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p} x_1 p(x_1, x_3) p(x_2 | x_3) dx_1 dx_2 dx_3 \\ &= \int_{\mathbb{R}^m \times \mathbb{R}^p} x_1 p(x_1, x_3) \underbrace{\left(\int_{\mathbb{R}^n} p(x_2 | x_3) dx_2 \right)}_{=1} dx_1 dx_3 \\ &= \int_{\mathbb{R}^m} x_1 \underbrace{\left(\int_{\mathbb{R}^p} p(x_1, x_3) dx_3 \right)}_{=p(x_1)} dx_1 \\ &= \int_{\mathbb{R}^m} x_1 p(x_1) dx_1 \\ &= \int_{\mathbb{R}^m} x_1 p(x_1) dx_1 \\ E[\tilde{X}_1] &= E[X_1]\end{aligned}$$

Similarly,

$$E[\tilde{X}_2] = E[X_2]$$

Also,

$$\begin{aligned}
E[\tilde{X}_3] &= \int_{\Omega} \tilde{X}_3 dP \\
&= \int_{\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p} x_3 \tilde{p}(x_1, x_2, x_3) dx_1 dx_2 dx_3 \\
&= \int_{\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p} x_3 \frac{p(x_1, x_3)p(x_2, x_3)}{p(x_3)} dx_1 dx_2 dx_3 \\
&= \int_{\mathbb{R}^p} x_3 \underbrace{\left(\int_{\mathbb{R}^m} p(x_1, x_3) dx_1 \right)}_{=p(x_3)} \underbrace{\left(\int_{\mathbb{R}^n} p(x_2, x_3) dx_2 \right)}_{=p(x_3)} \frac{1}{p(x_3)} dx_3 \\
&= \int_{\mathbb{R}^p} x_3 p(x_3) dx_3 \\
E[\tilde{X}_3] &= E[X_3]
\end{aligned}$$

This shows both distributions have the same mean. We now determine the information matrix for \tilde{p} . Let the canonical representation of gaussian distribution p be

$$\begin{aligned}
\xi &= \left[\begin{array}{c} \xi_1 \\ \xi_2 \\ \xi_3 \end{array} \right] \left\{ \begin{array}{l} m \\ n \\ p \end{array} \right. \\
\Omega &= \left[\begin{array}{ccc} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{12}^T & \Omega_{22} & \Omega_{23} \\ \Omega_{13}^T & \Omega_{23}^T & \Omega_{33} \end{array} \right] \left\{ \begin{array}{l} m \\ n \\ p \end{array} \right. \\
&\quad \underbrace{\hspace{1.5cm}}_m \quad \underbrace{\hspace{1.5cm}}_n \quad \underbrace{\hspace{1.5cm}}_p
\end{aligned}$$

The marginals density are given by

$$\begin{aligned}
p(x_1, x_3) &= \alpha e^{-\frac{1}{2} \begin{bmatrix} x_1^T & x_3^T \end{bmatrix} \Lambda \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} + \xi^T \begin{bmatrix} x_1 \\ x_3 \end{bmatrix}} \\
p(x_2, x_3) &= \alpha' e^{-\frac{1}{2} \begin{bmatrix} x_2^T & x_3^T \end{bmatrix} \Lambda' \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} + \xi'^T \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}} \\
p(x_3) &= \alpha'' e^{-\frac{1}{2} \begin{bmatrix} x_3^T \end{bmatrix} \Lambda'' \begin{bmatrix} x_3 \end{bmatrix} + \xi''^T \begin{bmatrix} x_3 \end{bmatrix}}
\end{aligned}$$

where

$$\begin{aligned}
\Lambda &= \left[\begin{array}{cc} \Omega_{11} & \Omega_{13} \\ \Omega_{13}^T & \Omega_{33} \end{array} \right] - \left[\begin{array}{c} \Omega_{12} \\ \Omega_{23}^T \end{array} \right] \Omega_{22}^{-1} \left[\begin{array}{cc} \Omega_{12}^T & \Omega_{23} \end{array} \right] \\
&= \left[\begin{array}{cc} \Omega_{11} & \Omega_{13} \\ \Omega_{13}^T & \Omega_{33} \end{array} \right] - \left[\begin{array}{cc} \Omega_{12} \Omega_{22}^{-1} \Omega_{12}^T & \Omega_{12} \Omega_{22}^{-1} \Omega_{23} \\ \Omega_{23}^T \Omega_{22}^{-1} \Omega_{12}^T & \Omega_{23}^T \Omega_{22}^{-1} \Omega_{23} \end{array} \right]
\end{aligned}$$

$$\begin{aligned}\Lambda' &= \begin{bmatrix} \Omega_{22} & \Omega_{23} \\ \Omega_{23}^T & \Omega_{33} \end{bmatrix} - \begin{bmatrix} \Omega_{12}^T \\ \Omega_{13}^T \end{bmatrix} \Omega_{11}^{-1} \begin{bmatrix} \Omega_{12} & \Omega_{13} \end{bmatrix} \\ &= \begin{bmatrix} \Omega_{22} & \Omega_{23} \\ \Omega_{23}^T & \Omega_{33} \end{bmatrix} - \begin{bmatrix} \Omega_{12}^T \Omega_{11}^{-1} \Omega_{12} & \Omega_{12}^T \Omega_{11}^{-1} \Omega_{23} \\ \Omega_{13}^T \Omega_{11}^{-1} \Omega_{12} & \Omega_{13}^T \Omega_{11}^{-1} \Omega_{13} \end{bmatrix}\end{aligned}$$

and

$$\Lambda'' = \Omega_{33} - \begin{bmatrix} \Omega_{13}^T & \Omega_{23}^T \end{bmatrix} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21}^T & \Omega_{22} \end{bmatrix}^{-1} \begin{bmatrix} \Omega_{13} \\ \Omega_{23} \end{bmatrix}$$

The distribution \tilde{p} can be written , grouping linear and quadratic terms :

$$\tilde{p}(x_1, x_2, x_3) = \alpha e^{-\frac{1}{2} \begin{bmatrix} x_1^T & x_2^T & x_3^T \end{bmatrix} \tilde{\Omega} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \tilde{\xi}^T \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}$$

with

$$\tilde{\Omega} = \begin{bmatrix} \Omega_{11} & 0 & \Omega_{13} \\ 0 & \Omega_{22} & \Omega_{23} \\ \Omega_{13}^T & \Omega_{23}^T & \Omega_{33} \end{bmatrix} - \begin{bmatrix} \Omega_{12} \Omega_{22}^{-1} \Omega_{12}^T & 0 & \Omega_{12} \Omega_{22}^{-1} \Omega_{23} \\ 0 & \Omega_{12}^T \Omega_{11}^{-1} \Omega_{12} & \Omega_{12}^T \Omega_{11}^{-1} \Omega_{23} \\ \Omega_{23}^T \Omega_{22}^{-1} \Omega_{12} & \Omega_{23}^T \Omega_{11}^{-1} \Omega_{12} & A \end{bmatrix}$$

(note the 0 blocks between 1 and 2), the matrix A being

$$A = \begin{bmatrix} \Omega_{13}^T & \Omega_{23}^T \end{bmatrix} \left(\begin{bmatrix} \Omega_{11}^{-1} & 0 \\ 0 & \Omega_{22}^{-1} \end{bmatrix} - \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21}^T & \Omega_{22} \end{bmatrix}^{-1} \right) \begin{bmatrix} \Omega_{13} \\ \Omega_{23} \end{bmatrix}$$

It follows that

$$\Omega - \tilde{\Omega} = \begin{bmatrix} \Omega_{12} \Omega_{22}^{-1} \Omega_{12}^T & \Omega_{12} & \Omega_{12} \Omega_{22}^{-1} \Omega_{23} \\ \Omega_{12}^T & \Omega_{12}^T \Omega_{11}^{-1} \Omega_{12} & \Omega_{12}^T \Omega_{11}^{-1} \Omega_{23} \\ \Omega_{23}^T \Omega_{22}^{-1} \Omega_{12} & \Omega_{23}^T \Omega_{11}^{-1} \Omega_{12} & A \end{bmatrix}$$

To ensure the *no overconfidence* property is preserved, we need to have

$$\begin{aligned} \forall x \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^p, \quad x^T \tilde{\Omega} x &\leq x^T \Omega x \\ \Leftrightarrow \quad x^T (\Omega - \tilde{\Omega}) x &\geq 0 \end{aligned}$$

i.e. the matrix $\Omega - \tilde{\Omega}$ is positive semidefinite. A can be made to be negative definite for some choice of Ω , this prevents $\Omega - \tilde{\Omega}$ to have the required property (cf. A.1). In conclusion, the enforced conditional independance approximation does not retain consistency.

Appendix A

A.1.

Proposition. The matrix

$$A = \begin{bmatrix} \Omega_{13}^T & \Omega_{23}^T \end{bmatrix} \left(\begin{bmatrix} \Omega_{11}^{-1} & 0 \\ 0 & \Omega_{22}^{-1} \end{bmatrix} - \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21}^T & \Omega_{22} \end{bmatrix}^{-1} \right) \begin{bmatrix} \Omega_{13} \\ \Omega_{23} \end{bmatrix}$$

can be made to be negative definite for some choice of symmetric positive definite matrix Ω .

Proof. We consider the case $m = n = p$ and the matrix Ω defined by

$$\Omega = \begin{bmatrix} I & -\lambda I & I \\ -\lambda I & I & I \\ I & I & \mu I \end{bmatrix}$$

with $\lambda \in]0, 1[$, $\mu = \frac{3}{1-\lambda}$. The eigenvalues of this matrix are

$$\begin{aligned} \lambda_1 &= 1 + \lambda \\ \lambda_2 &= \frac{\mu + 1 - \lambda - \sqrt{(\mu - 1 + \lambda)^2 + 8}}{2} \\ &= \frac{\mu + 1 - \lambda + \sqrt{(\mu - 1 + \lambda)^2 + 8}}{2} \\ \lambda_3 &= \frac{\mu + 1 - \lambda + \sqrt{(\mu - 1 + \lambda)^2 + 8}}{2} \end{aligned}$$

showing the matrix is symmetric positive definite, which is a necessary and sufficient condition to be a valid information matrix. Furthermore,

$$\begin{aligned} \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21}^T & \Omega_{22} \end{bmatrix}^{-1} &= \begin{bmatrix} I & -\lambda I \\ -\lambda I & I \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \frac{1}{1-\lambda^2} I & \frac{\lambda}{1-\lambda^2} I \\ \frac{\lambda}{1-\lambda^2} I & \frac{1}{1-\lambda^2} I \end{bmatrix} \end{aligned}$$

so,

$$\begin{bmatrix} \Omega_{11}^{-1} & 0 \\ 0 & \Omega_{22}^{-1} \end{bmatrix} - \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21}^T & \Omega_{22} \end{bmatrix}^{-1} = \begin{bmatrix} -\frac{\lambda^2}{1-\lambda^2} I & -\frac{\lambda}{1-\lambda^2} I \\ -\frac{\lambda}{1-\lambda^2} I & -\frac{\lambda^2}{1-\lambda^2} I \end{bmatrix}$$

and finally,

$$\begin{aligned} A &= -2 \frac{\lambda^2}{1-\lambda^2} I - 2 \frac{\lambda}{1-\lambda^2} I \\ &= -2 \frac{\lambda^2 + \lambda}{1-\lambda^2} I \end{aligned}$$

which is negative definite. □

Let now $x \in \mathbb{R}^n$, $x \neq 0$. If $x' = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ x \end{bmatrix} \in \mathbb{R}^{3n}$,

$$x'^T(\Omega - \tilde{\Omega})x' = x^T A x < 0$$

proves that $\Omega - \tilde{\Omega}$ is not positive semidefinite.