Preuve du calcul de la postérieure pour Thompson Sampling linéaire

On souhaite définir la distribution postérieure des paramètres θ d'un modèle linéaire tel que : $\mathbb{E}(r_{i,t}|x_{i,t},\theta) =< \theta, x_{i,t} >$, pour tout t et tout i, avec $x_{i,t}$ le contexte observé pour i à l'instant t et $r_{i,t}$ la récompense associée.

On suppose au temps t un ensemble d'observations passées $\mathcal{D}_t = \{(s, i_s, x_{i,s}, r_s)_{s=1}^{t-1}\}$, dont chaque reward r_s suit une loi normale en fonction de son contexte correspondant x_s et des paramètres θ du modèle linéaire : $P(r_s|\theta, x_s) = \mathcal{N}(\theta^T x_s, v^2)$ pour tout $s \in [1; t[$. On suppose également un prior gaussien sur les paramètres θ : $P(\theta) = \mathcal{N}(0, \sigma^2)$.

$$
\begin{aligned}
\log P(\theta|\mathcal{D}_t) &\propto \log P(\mathcal{D}_t|\theta) + \log P(\theta) && \text{(car } P(\mathcal{D}_t) \text{ constant par rapport à } \theta) \\
&\propto \sum_t \log P(r_t|\theta, x_t) + \log P(\theta) && \text{(car } P(x_t) \text{ indépendant de } \theta) \\
&\propto -\frac{1}{2}\sum_t \frac{|\theta^T x_t - r_t|^2}{v^2} - \frac{1}{2}\frac{||\theta||^2}{\sigma^2} && \text{(constante gaussienne ne dépend pas de } \theta) \\
&\propto -\frac{1}{2v^2}\sum_t (\theta^T x_t x_t^T \theta - 2\theta^T x_t r_t) - \frac{1}{2\sigma^2}\theta^T\theta && \text{(car } r_t^2 \text{ constant par rapport à } \theta) \\
&= -\frac{1}{2}\theta^T \left(\frac{1}{v^2}\sum_t (x_t x_t^T) + \frac{1}{\sigma^2} I\right)\theta - \theta^T\left(\frac{1}{v^2}\sum_t (x_t r_t)\right) \\
&= -\frac{1}{2}\theta^T A_t \theta - \theta^T b_t \\
&\propto -\frac{1}{2}\left(\theta^T - b_t A_t^{-1}\right)^T A_t \left(\theta^T - b_t A_t^{-1}\right) && \text{(car } b_t^T b_t A_t^{-1} \text{ indépendant de } \theta)
\end{aligned}
$$

avec $A_t = \frac{1}{v^2}\sum_t (x_t x_t^T) + \frac{1}{\sigma^2} I$, I la matrice identité et $b_t = \frac{1}{v^2}\sum_t (x_t r_t)$.

On reconnait alors une normale multivariée : $\theta|\mathcal{D}_t \sim \mathcal{N}(b_t A_t^{-1}; A_t^{-1})$.