

# Sequence correction for Gaussian probability

Patrick Ding

2021-03-30

Expectation propagation for Gaussian probability estimation works well for rectangular integration regions, but poorly for polyhedral regions. We provide a way to rewrite the polyhedral region Gaussian probability problem as a rectangular problem.

## Cast polyhedral problem as rectangular one

We write the multivariate Gaussian density with covariance  $\Sigma$  truncated to a polytope region  $l \leq Ax \leq u$ , where  $x \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ , and  $l, u \in \mathbb{R}^m$ , as

$$\begin{aligned} p(x) &= N(x; 0, \Sigma) 1\{l \leq Ax \leq u\} \\ &= N(x; 0, \Sigma) \prod_{i=1}^m t_i(x) \end{aligned}$$

where  $t_i(x) = 1\{l_i \leq c_i^T x \leq u_i\}$ . For random variable  $X$  with this density, the normalizing constant of this density is  $P(l \leq AX \leq u)$ . Let's assume  $\Sigma = I$ , and call the random variable  $Z$  in this case, so we want  $P(l \leq AZ \leq u)$ .

Let  $A_{\cdot,k}$  be the  $k$ th column of  $A$ . Let  $\epsilon_i \sim_{iid} N(0, 1)$  for  $i \in \{1, \dots, n\}$ . We construct a sequence of random vectors  $\{X_1, X_2, \dots, X_n\}$

$$\begin{aligned} X_1 &= A_{\cdot,1}\epsilon_1 \\ X_2 &= X_1 + A_{\cdot,2}\epsilon_2 \\ &\vdots \\ X_{n-1} &= X_{n-2} + A_{\cdot,n-1}\epsilon_{n-1} \\ X_n &= X_{n-1} + A_{\cdot,n}\epsilon_n \end{aligned}$$

substituting  $X_k$  for  $k \in \{1, \dots, n-1\}$  into the last expression, observe that

$$\begin{aligned} X_n &= A_{\cdot,1}\epsilon_1 + \dots + A_{\cdot,n}\epsilon_n \\ &= A \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}. \end{aligned}$$

From properties of multivariate normal  $X_n$  has the same distribution as  $AZ$ ,  $N(0, AA^T)$ , so

$$P(l \leq X_n \leq u) = P(l \leq AZ \leq u).$$

The joint density of the  $X_i$  factorizes as

$$p(x_1, \dots, x_n) = p(x_n|x_{n-1})p(x_{n-1}|x_{n-2}) \cdots p(x_2|x_1)p(x_1)$$

where

$$\begin{aligned} p(x_i|x_{i-1}) &= N(x_i; x_{i-1}, A_{\cdot,i} A_{\cdot,i}^T) \\ p(x_1) &= N(x_1; 0, A_{\cdot,1} A_{\cdot,1}^T) \end{aligned} \quad i \in \{2, \dots, n\}$$

The covariance  $\Xi_i = A_{\cdot,i} A_{\cdot,i}^T$  is not invertible in general, it is rank 1. But from Khatri (1968) we can express this singular normal distribution in terms of the Moore-Penrose pseudoinverse of  $\Xi_i$ . Let  $\gamma_i$  be the sole eigenvector of  $\Xi_i$ , and let  $\lambda_i$  be the sole eigenvalue of  $\Xi_i$ . Then

$$\gamma_i = \frac{A_{\cdot,i}}{\|A_{\cdot,i}\|} \quad (1)$$

$$\lambda_i = \|A_{\cdot,i}\|^2 \quad (2)$$

$$\Xi^+ = \frac{1}{\lambda_i} \gamma_i \gamma_i^T \quad (3)$$

and the density is

$$\begin{aligned} p(x_i|x_{i-1}) &= (2\pi)^{-r/2} |\lambda_i|^{-1/2} \exp \left[ -\frac{\lambda_i}{2} (x_i - x_{i-1})^T \gamma_i \gamma_i^T (x_i - x_{i-1}) \right] \\ p(x_1) &= (2\pi)^{-r/2} |\lambda_1|^{-1/2} \exp \left[ -\frac{\lambda_1}{2} x_1^T \gamma_1 \gamma_1^T x_1 \right] \end{aligned}$$

Finally we add the indicator for the constraints on  $X_n$  and get an augmented truncated multivariate normal density proportional to:

$$\begin{aligned} p(x) &= \frac{1}{S} p(x_1, \dots, x_n) 1\{\ell \leq x_n \leq u\} \\ &= \frac{1}{S} p(x_1) \left[ \prod_{i=2}^n p(x_i|x_{i-1}) \right] 1\{\ell \leq x_n \leq u\} \end{aligned} \quad (4)$$

where  $x = (x_1, \dots, x_n)^T \in \mathbb{R}^{n^2}$  and  $S$  is a normalizing constant. The ratio of this density's normalizing constant and the normalizing constant of  $p(x_1, \dots, x_n)$  will give us  $P(\ell \leq X_n \leq u)$ .

## EP form

This density is intractable because of the term  $1\{\ell \leq x_n \leq u\}$ , but we can approximate the density with a Gaussian using the EP method of Cunningham et al (2011). This will be efficient and accurate since the truncation region is axis aligned, and their method performs well in that case. Equation 4 is exactly of the form that the Cunningham et al method handles:

$$p(x) = p_0(x) \prod_{k=1}^n t_k(x) \quad (5)$$

where

$$p_0(x) = p(x_1) \left[ \prod_{i=2}^n p(x_i|x_{i-1}) \right] \quad \text{prior} \quad (6)$$

$$t_k(x) = 1\{\ell_k \leq c_k^T x \leq b_k\} = 1\{\ell_k \leq x_{n_k} \leq b_k\} \quad \text{likelihood factor} \quad (7)$$

$$c_k = (0_{1 \times n(n-1)} \quad 0_{1 \times k-1} \quad 1 \quad 0_{1 \times n-k})^T \quad (8)$$

The density of the prior is singular Gaussian,

$$\begin{aligned}
p_0(x) &= N(x; m, K) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^n |\lambda_i|^{-1/2} \right) \exp(x^T K^{-1} x) \\
m &= 0 \\
K^{-1} &= \begin{pmatrix}
\Xi_1^+ + \Xi_2^+ & \Xi_2^+ & & & & \\
\Xi_2^+ & \Xi_2^+ + \Xi_3^+ & \Xi_3^+ & & & \\
& \Xi_3^+ & \Xi_3^+ + \Xi_4^+ & & & \\
& & & \ddots & & \\
& & & & \Xi_{n-1}^+ + \Xi_n^+ & \Xi_n^+ \\
& & & & \Xi_n^+ & \Xi_n^+
\end{pmatrix}
\end{aligned}$$

where unspecified blocks in  $K^{-1}$  are 0.

Only the bottom right  $n \times n$  corner of  $K^{-1}$ ,  $\Xi_n^+$ , will be involved in the EP updates. In general for a  $n^2 \times n^2$  matrix  $B$ , denote that sub matrix as  $B_{-n}$ .

We do not have easy access to  $K$ , which would require inverting the singular and  $d^2 \times d^2$  dimensional precision  $K^{-1}$ . In EP updates we would like to work only with  $K^{-1}$ .

## EP factor updates

We approximate the  $t_k(x)$  with unnormalized Gaussian densities with  $\tilde{S}_k$ ,  $\tilde{\mu}_k$ , and  $\tilde{\sigma}_k^2$  the approximating normalizing constant, mean, and variance for the  $k$ th factor:

$$\tilde{t}_k(x) = \tilde{S}_k N(c_k^T x; \tilde{\mu}_k, \tilde{\sigma}_k)$$

leading to the overall approximating unnormalized Gaussian with normalizing constant  $S$ , mean  $\mu$ , and covariance  $\Sigma$ :

$$q(x) = p_0(x) \prod_{k=1}^n \tilde{t}_k(x) = N(x; \mu, \Sigma)$$

Initialize  $\mu = m$ ,  $\Sigma^{-1} = \Lambda = K^{-1}$ . The EP cavity distribution is singular Gaussian

$$q^{\setminus k}(x) = Z^{\setminus k} N(x; u^{\setminus k}, V^{\setminus k}) \quad (9)$$

$$V^{\setminus k} = \left( \Sigma^{-1} - \frac{1}{\tilde{\sigma}_k} c_k c_k^T \right)^{-1} \quad (10)$$

$$u^{\setminus k} = V^{\setminus k} \left( \Sigma^{-1} \mu - \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2} c_k \right) \quad (11)$$

Since we are dealing with rank-one factors and the approximating factors are Gaussians and therefore closed under marginalization, we can moment match the tilted and approximating densities in the one dimension of interest, where the cavity distribution is univariate Gaussian:

$$q_{\setminus k}(c_k^T x) = q_{\setminus k}(x_{k'}) = S^{\setminus k} N(x_{k_k}; \mu_{\setminus k}, \sigma_{\setminus k}) \quad (12)$$

$$\sigma_{\setminus k} = \left( \frac{1}{\Sigma_{k'k'}} - \frac{1}{\tilde{\sigma}_k^2} \right)^{-1} \quad (13)$$

$$\mu_{\setminus k} = \sigma_{\setminus k} \left( \frac{\mu_{k'}}{\Sigma_{k',k'}} - \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2} \right) \quad (14)$$

$$k' = n(n+1) + k \quad (15)$$

However this depends on the diagonal elements of  $\Sigma$ , which we do not have. One might suppose working in the natural parameterization would help us here, but that is not the case. To see this, we have the canonical parameters of the full cavity distribution  $q(x) = N(x; e^{\setminus k}, R^{\setminus k})$ :

$$\begin{aligned} R^{\setminus k} &= (V^{\setminus k})^{-1} = \Sigma^{-1} - \frac{1}{\tilde{\sigma}_k} c_k c_k^T \\ e^{\setminus k} &= (V^{\setminus k})^{-1} u^{\setminus k} = \Sigma^{-1} \mu - \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2} c_k \end{aligned}$$

Then the marginal distribution is normal with precision and mean precision:

$$\tau_{\setminus k} = R_{k'k'}^{\setminus k} - R_{k'k'}^{\setminus k} (R_{-k', -k'}^{\setminus k})^{-1} R_{\cdot k'}^{\setminus k}$$

where  $R_{-k', -k'}^{\setminus k}$  is the  $n^2 - 1 \times n^2 - 1$  matrix formed by removing the  $k'$ th row and column of  $R^{\setminus k}$ . We cannot compute the inverse of this matrix.

Setting aside the issues with the cavity distribution, the moments of the tilted distribution  $t_k(x)q_{\setminus k}(x)$  are

$$\hat{S}_k = \frac{1}{2}(\Phi(\beta) - \Phi(\alpha)) \quad (16)$$

$$\hat{\mu}_k = \mu_{\setminus k} + \frac{1}{\hat{S}_k} \frac{\sigma_{\setminus k}}{\sqrt{2\pi}} (\exp(-\alpha^2) - \exp(-\beta^2)) \quad (17)$$

$$\hat{\sigma}_k^2 = \mu_{\setminus k}^2 + \sigma_{\setminus k} + \frac{1}{\hat{S}_k} \frac{\sigma_{\setminus k}}{\sqrt{2\pi}} (-(b_k + \mu_{\setminus k}) \exp(-\beta^2)) - \hat{\mu}_k^2 \quad (18)$$

$$\alpha = \frac{\ell_k - \mu_{\setminus k}}{\sqrt{2\sigma_{\setminus k}^2}} \quad (19)$$

$$\beta = \frac{u_k - \mu_{\setminus k}}{\sqrt{2\sigma_{\setminus k}^2}} \quad (20)$$

We match the moments of  $\tilde{t}_k(x)q_{\setminus k}(x)$  to the moments of  $t_k(x)q_{\setminus k}(x)$ , leading to factor updates:

$$\tilde{\sigma}_k^2 = (\hat{\sigma}_k^{-2} - \sigma_{\setminus k}^{-2})^{-1} \quad (21)$$

$$\tilde{\mu}_k = \tilde{\sigma}_k^2 (\hat{\sigma}_k^{-2} \hat{\mu}_k - \sigma_{\setminus k}^{-2} \mu_{\setminus k}) \quad (22)$$

$$\tilde{S}_k = \hat{S}_k \sqrt{2\pi} \sqrt{\sigma_{\setminus k}^2 + \tilde{\sigma}_k^2} \exp\left(\frac{1}{2} \frac{(\mu_{\setminus k} - \tilde{\mu}_k)^2}{(\sigma_{\setminus k}^2 + \tilde{\sigma}_k^2)}\right) \quad (23)$$

## Overall approximation update

Finally we can update the the overall approximation. To avoid inverting  $d^2 \times d^2$  matrices, we work with the natural parameterization of the approximating Gaussian:

$$\begin{aligned} q(x) &= \exp\left(a + \eta^T x - \frac{1}{2} x^T \Lambda x\right) \\ \Lambda &= \Sigma^{-1} \\ \eta &= \Lambda \mu \\ a &= -\frac{1}{2} n \log(2\pi) - \log |\Lambda| + \eta^T \Lambda \eta \end{aligned}$$

We will also parameterize the approximate factors in the natural parameterization:

$$\begin{aligned} \tilde{\tau}_k &= \frac{1}{\tilde{\sigma}_k^2} \\ \tilde{\nu}_k &= \frac{\tilde{\mu}_k}{\tilde{\sigma}_k^2} \end{aligned}$$

Initialize  $\Lambda = K^{-1}$ ,  $\eta = 0$ . We update only the parts of the parameters that are affected by the factor updates. This is the bottom right  $n \times n$  corner of  $\Lambda$ ,  $\Lambda_{nn}$ , and the last  $n$  elements of  $\eta$ ,  $\eta_n$ .

$$\begin{aligned} \Lambda_{nn}^{new} &= \Lambda_{nn}^{old} + (\tilde{\tau}_k^{new} - \tilde{\tau}_k^{old})c_k c_k^T \\ \implies \Lambda_{nn_{kk}}^{new} &= \Lambda_{nn_{kk}}^{old} + \tilde{\tau}_k^{new} - \tilde{\tau}_k^{old} \end{aligned} \quad (24)$$

$$\begin{aligned} \eta^{new} &= \eta^{old} + (\tilde{\nu}_k^{new} - \tilde{\nu}_k^{old})c_k \\ \implies \eta_{n_k}^{new} &= \eta_{n_k}^{old} + \tilde{\nu}_k^{new} - \tilde{\nu}_k^{old} \end{aligned} \quad (25)$$

Since only  $\eta_n$  is updated, the rest of  $\eta$  is 0. We will need to update the determinant  $|\Lambda|$  for the normalization constant. For each dimension  $k \in \{1, \dots, n\}$  the update is

$$|\Lambda^{new}| = |\Lambda^{old}| \left( 1 + (\tilde{\tau}_k^{new} - \tilde{\tau}_k^{old})(\Lambda^{old})_{k',k'}^{-1} \right) \quad (26)$$

For this update to work we need the initial determinant and the inverse precision at each step:

$$|\Lambda^0| = |K^{-1}| \quad (27)$$

$$(\Lambda^{old})^{-1} \quad (28)$$

Next we update the normalizing constant of the approximation:

$$\begin{aligned} \log S &= -\frac{1}{2} \sum_{k=1}^n \log |\lambda_k| \\ &\quad + \sum_{k=1}^n \log \tilde{S}_k - \frac{1}{2} \left( \frac{\tilde{\mu}_k^2}{\tilde{\sigma}_k^2} + \log \tilde{\sigma}_k^2 + \log(2\pi) \right) \\ &\quad + \frac{1}{2} (\eta_n^T \Lambda_{nn} \eta_n - \log |\Lambda|) \\ &= -\frac{1}{2} \sum_{k=1}^n \log |\lambda_k| \\ &\quad + \sum_{k=1}^n \log \hat{S}_k + \frac{1}{2} \log(\tilde{\tau}_k \sigma_{\setminus k}^2 + 1) + \frac{1}{2} \frac{\mu_{\setminus k}^2 \tilde{\tau}_k - 2\mu_{\setminus k} \tilde{\nu}_k - \tilde{\nu}_k^2 \sigma_{\setminus k}^2}{1 + \tilde{\tau}_k \sigma_{\setminus k}^2} \\ &\quad + \frac{1}{2} (\eta_n^T \Lambda_{nn} \eta_n - \log |\Lambda|) \end{aligned} \quad (29)$$

where we rewrite the middle term as Cunningham does.  $S$  is the desired probability  $P(AZ \leq b)$ .

## Algorithm

---

**Algorithm 1** Sequence corrected EP for Gaussian probability

---

```

Set  $\tilde{\sigma}_k = 0$ 
while not converged do
  for  $k = 1 : n$  do
    Form cavity distribution  $q^{\setminus k}(x) = Z^{\setminus k} N(x; u^{\setminus k}, V^{\setminus k})$ 
    Compute moments of tilted distribution  $t_k(x) q_{\setminus k}(x)$ 
    Update parameters of  $\tilde{t}_k(x)$ 
  end for
  Update overall approximation  $\mu, \Sigma$ 
end while
Calculate  $S$ 
return  $S$ 

```

---

# Appendix