

# IID Draws from the DLM

Matt Simpson

April 12, 2014

## 1 Introduction

In this note we investigate the implications of a theorem of ? which suggest the possibility of obtaining IID draws from the posterior distribution of a DLM.

## 2 The Theorem

The following theorem is established by ? (Theorem 1):

**Theorem 2.0.1.** *Given a posterior distribution of interest  $p(\phi|y)$ ,  $\theta \in \Theta$ , suppose we have two augmentation schemes  $\theta$  and  $\tilde{\theta}$  such that their joint distribution, conditioning on both  $\theta$  and  $y$ , is well defined for  $\theta \in \Theta$  (a.s. w.r.t.  $p(\theta|y)$ ). Denote the geometric rate of convergence of the DA algorithm under  $\theta$  by  $r_1$  and under  $\tilde{\theta}$  by  $r_2$ . Then*

$$r_{1\&2} \leq \mathcal{R}_{1,2} \sqrt{r_1 r_2}$$

where  $r_{1\&2}$  is the geometric rate of the GIS sampler interweaving between  $\theta$  and  $\tilde{\theta}$  and  $\mathcal{R}_{1,2}$  is the maximal correlation between  $\theta$  and  $\tilde{\theta}$  in their joint posterior distribution  $p(\theta, \tilde{\theta}|y)$ .

One implication of this theorem, as ? note, is that if  $\theta$  and  $\tilde{\theta}$  are independent in the posterior, then the GIS sampler will obtain iid draws from  $p(\phi|y)$ .

## 3 Characterizing the posterior distribution of interest

The dynamic linear model for a time series of  $k$ -dimensional vectors  $y_t$  is defined by

$$\begin{aligned} y_t &= F_t \theta_t + v_t \\ \theta_t &= G_t \theta_{t-1} + w_t \end{aligned}$$

for  $t = 1, 2, \dots, T$  where

$$\begin{aligned} v_t &\sim N_k(0, V) \\ w_t &\sim N_p(0, W). \end{aligned}$$

Let  $\phi$  denote the unknown parameter. Here we allow that  $\theta_0$ ,  $F_{1:T}$ ,  $G_{1:T}$ ,  $V$ , and  $W$  all may depend on  $\phi$ . Typically  $\theta_0$  is treated as part of the data augmentation and not as part of the set of parameters to be estimated, but we will see that for computational purposes thinking of  $\theta_0$  as a parameter is natural.

The above defines the usual augmented data model, but we ultimately need to characterize the posterior  $p(\phi|y)$  not  $p(\phi, \theta|y)$ , so we need find the marginal distribution of  $y$  given  $\phi$ . We can rewrite the model by recursively substituting as

$$y_t = v_t + F_t (w_t + G_t w_{t-1} + G_t G_{t-1} w_{t-2} + \dots + G_t G_{t-1} \dots G_2 w_1 + G_t G_{t-1} \dots G_1 \theta_0)$$

Now each  $y_t$  is a linear combination of normal distributions, so  $y_{1:T}$  has a marginal normal distribution such that

$$\begin{aligned} \mathbb{E}[y_t|\phi] &= F_t \prod_{s=t}^1 G_s \theta_0 \\ \text{Var}[y_t|\phi] &= V + F_t H_t W \\ \text{Cov}[y_s, y_t|\phi] &= F_t H_t W \end{aligned}$$

where

$$\prod_{s=t}^1 G_s = G_t G_{t-1} \cdots G_1$$

and

$$H_t = I_p + G_t + G_t G_{t-1} + \cdots + G_t G_{t-1} \cdots G_2.$$

Now define

$$\mu = \begin{bmatrix} F_1 G_1 \theta_0 \\ F_2 G_2 G_1 \theta_0 \\ \vdots \\ F_T G_T G_{T-1} \cdots G_1 \theta_0 \end{bmatrix}, \quad \tilde{V}_{k \times k} = \begin{bmatrix} V & 0 & 0 & \ddots & 0 \\ 0 & V & 0 & \ddots & 0 \\ 0 & 0 & V & \ddots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & \ddots & V \end{bmatrix}, \quad \tilde{W}_{k \times k} = \begin{bmatrix} F_1 H_1 \\ F_2 H_2 \\ \vdots \\ F_T H_T \end{bmatrix} W \begin{bmatrix} H_1' F_1' & H_2' F_2' & \cdots & H_T' F_T' \end{bmatrix}.$$

Then we have

$$y_{1:T} \sim N_{p \times k}(\mu, \tilde{V} + \tilde{W}).$$

Now given a prior  $p(\phi)$ , this defines the posterior distribution of interest  $p(\phi|y_{1:T})$ .

If, instead, we assume  $\theta_0 \sim N(m_0, C_0)$  in the prior, independent of any other parameters, and also integrate out  $\theta_0$ , we get

$$y_{1:T} \sim N_{p \times k}(\mu^*, \tilde{V} + \tilde{W} + \tilde{C}_0)$$

where  $\mu^*$  is defined analogously to  $\mu$  with  $m_0$  replacing  $\theta_0$  and

$$\tilde{C}_0 = \begin{bmatrix} G_1 \\ G_2 G_1 \\ \vdots \\ G_T G_{T-1} \cdots G_1 \end{bmatrix} C_0 \begin{bmatrix} G_1 & G_1' G_2' & \cdots & G_1' G_2' \cdots G_T' \end{bmatrix}$$

This will ultimately show why thinking of  $\theta_0$  as a “parameter” instead of part of the data augmentation is natural — mainly because it results in a simpler covariance matrix for  $y_{1:T}$  which we will see has advantages when trying to find independent DAs.

## 4 Finding DAs which are independent in the posterior

Now in order to obtain iid draws from this posterior, we need to find data augmentations  $\theta$  and  $\tilde{\theta}$  which are independent in the posterior. Here  $\theta$  need not be the same as the  $\theta_{1:T}$  used to define the model. Another

requirement is that the joint posterior of  $(\theta, \tilde{\theta})$  be well define. So to start we shall assume that  $(y, \theta, \tilde{\theta})$  has a joint normal distribution conditional on the parameter, i.e.

$$\begin{bmatrix} \theta \\ \tilde{\theta} \\ y \end{bmatrix} \Big| \phi \sim N \left( \begin{bmatrix} \alpha_\theta \\ \alpha_{\tilde{\theta}} \\ \alpha_y \end{bmatrix}, \begin{bmatrix} \Omega_\theta & \Omega'_{\theta, \tilde{\theta}} & \Omega'_{y, \theta} \\ \Omega_{\theta, \tilde{\theta}} & \Omega_{\tilde{\theta}} & \Omega'_{y, \tilde{\theta}} \\ \Omega_{y, \theta} & \Omega_{y, \tilde{\theta}} & \Omega_y \end{bmatrix} \right).$$

This implies that

$$\begin{bmatrix} \theta \\ \tilde{\theta} \end{bmatrix} \Big| \phi, y \sim N \left( \begin{bmatrix} \alpha_\theta + \Omega'_{y, \theta} \Omega_y^{-1} (y - \alpha_y) \\ \alpha_{\tilde{\theta}} + \Omega'_{y, \tilde{\theta}} \Omega_y^{-1} (y - \alpha_y) \end{bmatrix}, \begin{bmatrix} \Omega_\theta - \Omega'_{y, \theta} \Omega_y^{-1} \Omega_{y, \theta} & \Omega'_{\theta, \tilde{\theta}} - \Omega'_{y, \tilde{\theta}} \Omega_y^{-1} \Omega_{y, \theta} \\ \Omega_{\theta, \tilde{\theta}} - \Omega'_{y, \theta} \Omega_y^{-1} \Omega_{y, \tilde{\theta}} & \Omega_{\tilde{\theta}} - \Omega'_{y, \tilde{\theta}} \Omega_y^{-1} \Omega_{y, \tilde{\theta}} \end{bmatrix} \right).$$

Now we want to define the various components of this distribution such that  $\theta$  and  $\tilde{\theta}$  are independent in the posterior, but we also are constrained to be consistent with the marginal distribution of  $y$  implied by the data model. This constraint gives

$$\begin{aligned} \alpha_y &= \mu \\ \Omega_y &= \tilde{V} + \tilde{W}. \end{aligned}$$

Our strategy will be to make  $\theta$  and  $\tilde{\theta}$  conditionally independent given  $y$  in the data model, then ensure that they are still independent in the posterior. The first part of this strategy forces

$$\Omega_{\theta, \tilde{\theta}} = \Omega'_{y, \theta} \Omega_y^{-1} \Omega_{y, \tilde{\theta}}.$$

We also need to remove  $\alpha_y = \mu$  from the means of  $\theta$  and  $\tilde{\theta}$  since  $\mu$  depends on  $\phi$  through the  $G_t$ 's. Setting

$$\begin{aligned} \alpha_\theta &= \Omega'_{y, \theta} \Omega_y^{-1} \alpha_y \\ \alpha_{\tilde{\theta}} &= \Omega'_{y, \tilde{\theta}} \Omega_y^{-1} \alpha_y \end{aligned}$$

will suffice to do this. So now we have

$$\begin{bmatrix} \theta \\ \tilde{\theta} \end{bmatrix} \Big| \phi, y \sim N \left( \begin{bmatrix} \Omega'_{y, \theta} \Omega_y^{-1} y \\ \Omega'_{y, \tilde{\theta}} \Omega_y^{-1} y \end{bmatrix}, \begin{bmatrix} \Omega_\theta - \Omega'_{y, \theta} \Omega_y^{-1} \Omega_{y, \theta} & 0 \\ 0 & \Omega_{\tilde{\theta}} - \Omega'_{y, \tilde{\theta}} \Omega_y^{-1} \Omega_{y, \tilde{\theta}} \end{bmatrix} \right).$$

Now our goal is to ensure that the mean and variance of  $\theta$  conditional on  $\phi$  and  $y$  does not depend on  $\tilde{W}$  but possibly on  $\tilde{V}$  and similarly that the mean and variance of  $\tilde{\theta}$  conditional on  $\phi$  and  $y$  does not depend on  $\tilde{V}$  but possibly on  $\tilde{W}$ . This will guarantee that integrating  $\phi$  out of the conditional posterior of  $(\theta, \tilde{\theta})$  will preserve their independence so long as  $\tilde{V}$  and  $\tilde{W}$  are independent in the prior. In practice the latter is not hard to enforce —  $\tilde{V}$  only depends on  $V$ , while  $\tilde{W}$  depends on  $W$  and any unknown parameters in the  $F_t$ 's and  $G_t$ 's. So for example if the unknown parameters are  $\theta_0$ ,  $V$ ,  $W$ , and  $\psi$  where only the  $F_t$ 's and  $G_t$ 's depend on  $\psi$ , and the collection of these four parameters are independent in the prior, the the constraint is satisfied.

We also could have done the same thing with  $\theta_0$  treated as part of the augmentation instead of  $\phi$ , in which case the conditional distribution of  $\theta$  and  $\tilde{\theta}$  would be the same with the definitions of  $\alpha_\theta$  and  $\alpha_{\tilde{\theta}}$  changed accordingly and, more importantly,  $\Omega_y = \tilde{V} + \tilde{W} + \tilde{C}_0$  where we treat  $\tilde{C}_0$  as part of  $\tilde{W}$  for the purpose of ensuring that  $\theta$  and  $\tilde{\theta}$  are independent in the posterior.

In order to further simplify we will suppose  $\Omega_y = \Omega_{y, \theta} = \Omega_{y, \tilde{\theta}}$  and that  $\Omega_\theta = \Omega_y + \tilde{\Omega}_\theta$  and  $\Omega_{\tilde{\theta}} = \Omega_y + \tilde{\Omega}_{\tilde{\theta}}$ . This yields

$$\begin{bmatrix} \theta \\ \tilde{\theta} \end{bmatrix} \Big| \phi, y \sim N \left( \begin{bmatrix} y \\ y \end{bmatrix}, \begin{bmatrix} \tilde{\Omega}_\theta & 0 \\ 0 & \tilde{\Omega}_{\tilde{\theta}} \end{bmatrix} \right).$$

So long as  $\Omega_\theta^*$  depends on possibly  $\tilde{V}$  but not  $\tilde{W}$  and  $\Omega_{\tilde{\theta}}^*$  depends on possibly  $\tilde{W}$  but not  $\tilde{V}$ , then  $\theta$  and  $\tilde{\theta}$  will be independent in the posterior. The full joint distribution of  $(y, \theta, \tilde{\theta})$  is then given by

$$\begin{bmatrix} \theta \\ \tilde{\theta} \\ y \end{bmatrix} \Big| \phi \sim N \left( \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \tilde{\Omega}_\theta + \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} \\ \tilde{V} + \tilde{W} & \tilde{\Omega}_{\tilde{\theta}} + \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} \\ \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} \end{bmatrix} \right).$$

It is easy to show that this covariance matrix is nonsingular so long as  $\tilde{V}$ ,  $\tilde{W}$ ,  $\tilde{\Omega}_\theta$ , and  $\tilde{\Omega}_{\tilde{\theta}}$  are nonsingular.

## 5 Constructing the GIS algorithm

At this point, conditional on a choice of  $\tilde{\Omega}_\theta$  and  $\tilde{\Omega}_{\tilde{\theta}}$ , the only things that remain to be done in order to construct the algorithm is to determine the conditional posterior distributions  $p(\phi|\theta, y)$  and  $p(\phi|\tilde{\theta}, y)$ . The distributions  $p(\theta|\phi, y)$  and  $p(\tilde{\theta}|\phi, y)$  are already determined above. The “tilde”  $\Omega$ ’s should be chosen to primarily make determining the two different conditional posteriors of  $\phi$  easy, both in terms of analytically writing them down and in terms of computation. It will turn out that this constraint is very hard to meet.

In order to see this, we will impose a bit more structure on the model. Specifically, let  $F_t$  and  $G_t$  be constant for each  $t = 1, 2, \dots, T$ . We will suppose that  $p(V, W, \theta_0) = p(V)p(W)p(\theta_0)$  but will ignore any additional structure on the prior.

Now consider  $(\theta, y)$ . The joint distribution is

$$\begin{bmatrix} \theta \\ y \end{bmatrix} \Big| \phi \sim N \left( \begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \tilde{\Omega}_\theta + \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} \\ \tilde{V} + \tilde{W} & \tilde{V} + \tilde{W} \end{bmatrix} \right).$$

and we have

$$\begin{aligned} \theta|y, \phi &\sim N(y, \Omega_\theta) \\ y|\phi &\sim N(\mu, \tilde{V} + \tilde{W}) \end{aligned}$$

with joint density

$$p(y, \theta|\phi) \propto |\tilde{\Omega}_\theta|^{-1/2} \exp \left[ -\frac{1}{2}(\theta - y)' \tilde{\Omega}_\theta^{-1} (\theta - y) \right] |\tilde{V} + \tilde{W}|^{-1/2} \exp \left[ -\frac{1}{2}(y - \mu)' (\tilde{V} + \tilde{W})^{-1} (y - \mu) \right].$$

So in order to make the posterior for  $\phi$  as simple as possible, it seems like a good idea to make  $\tilde{\Omega}_\theta$  independent of any element of  $\phi$ , i.e.  $\tilde{\Omega}_\theta = I$ , the identity matrix. However, this gives

$$p(\phi|y, \theta) \propto |\tilde{V} + \tilde{W}|^{-1/2} \exp \left[ -\frac{1}{2}(y - \mu)' (\tilde{V} + \tilde{W})^{-1} (y - \mu) \right] p(\phi)$$

which is the original posterior distribution we wish to sample from.

This problem does not appear intrinsic to trying to find two data augmentations which are independent in the posterior, but any attempt to construct a DA by assuming a simple structure for the various  $\Omega$ ’s appears to run into this problem. Further, treating  $\theta_0$  as part of the DA instead of a parameter will not help the problem — the only thing changed in this reasoning is that the matrix  $\tilde{V} + \tilde{W} + \tilde{C}_0$  replaces  $\tilde{V} + \tilde{W}$ , in particular in  $p(\phi|y, \theta)$ . This, yet again, is the original posterior we wish to sample from.

## 6 No “Sufficient Augmentation” for $(V, W)$ exists

A sufficient augmentation  $\theta$  is a data augmentation such that  $p(y|\theta, \phi) = p(y|\theta)$ , i.e.  $y$  and  $\phi$  are conditionally independent given  $\phi$ . If we suppose that  $\theta$  and  $y$  have a joint normal distribution as above, then if  $\theta$  is a

sufficient augmentation then  $p(\phi|\theta, y) = p(\phi|y)$ . To see this, start with

$$\begin{bmatrix} \theta \\ y \end{bmatrix} \Big| \phi \sim N \left( \begin{bmatrix} \alpha_\theta \\ \mu \end{bmatrix}, \begin{bmatrix} \Omega_\theta & \Omega'_{y,\theta} \\ \Omega_{y,\theta} & \tilde{V} + \tilde{W} \end{bmatrix} \right)$$

which implies

$$\begin{aligned} y|\theta, \phi &\sim N(\mu + \Omega'_{y,\theta}\Omega_\theta^{-1}(\theta - \alpha_\theta), \tilde{V} + \tilde{W} - \Omega'_{y,\theta}\Omega_\theta^{-1}\Omega_{y,\theta}) \\ \theta|\phi &\sim N(\alpha_\theta, \Omega_\theta). \end{aligned}$$

Now for  $\theta$  to be a sufficient augmentation we need

$$\mu + \Omega'_{y,\theta}\Omega_\theta^{-1}(\theta - \alpha_\theta)$$

and

$$\tilde{V} + \tilde{W} - \Omega'_{y,\theta}\Omega_\theta^{-1}\Omega_{y,\theta}$$

to be independent of  $\phi$ . This requires that

$$\mu + \Omega'_{y,\theta}\Omega_\theta^{-1}(\theta - \alpha_\theta) = A\theta$$

where  $A$  a matrix which does not depend on  $\phi$ . Rearranging, this gives  $A = \Omega'_{y,\theta}\Omega_\theta^{-1}$  so that  $\mu = A\alpha_\theta$  and  $\Omega_{y,\theta} = \Omega_\theta A'$ . Then using the second equation, we now require  $\Sigma = \tilde{V} + \tilde{W} - A\Omega_\theta A'$  free of  $\phi$ . This gives  $A\Omega_\theta A' = \tilde{V} + \tilde{W} - \Sigma$ . Consider  $\tilde{\theta} = A\theta$ . Then we have

$$\tilde{\theta} \sim N(\mu, A\Omega_\theta A').$$