MFSS Estimation Methods

We are interested in estimating state space models with time-varying parameters of the form

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{N}(0, H_t)$$
$$\alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t \qquad \eta_t \sim \mathcal{N}(0, Q_t) \qquad \alpha_1 \sim \mathcal{N}(a_1, P_1)$$

The elements of the parameter matrices Z_t, d_t, H_t, T_t, c_t, R_t, and Q_t are divided into those that are known and those that are to be estimated as a function of the parameter vector θ. We are interested in the maximum likelihood estimate of those parameters given data $Y_n = \{y_1, \ldots, y_n\}$.

Instructions for how to set up the state space and estimate the maximum likelihood of the unknown parameters can be found in the MFSS user guide. This document attempts to detail how the likelihood and its gradient are calculated for a given system, and how the parameters in θ will be estimated.

1 Kalman Filtering and Smoothing Calculations

Multivariate Filter

As shown in [1] (§4.3), the filtered estimates of the state, $a_t = \mathbb{E}(\alpha_t|Y_t)$ and $P_t = \text{Var}(\alpha_t|Y_t)$, are given by

$$a_{t+1} = T_{t+1}a_t + c_{t+1} + K_t v_t \qquad P_{t+1} = T_{t+1}P_t L_t' + R_{t+1}Q_{t+1}R_{t+1}'$$
$$v_t = y_t - Z_t a_t - d_t \qquad K_t = T_{t+1}P_t Z_t' F_t^{-1}$$
$$F_t = Z_t P_t Z_t' + H_t \qquad L_t = T_{t+1} - K_t Z_t$$

and the smoothed estimates, $\hat{\alpha}_t = \mathbb{E}(\alpha_t|Y_n)$ and $V_t = \text{Var}(\alpha_t|Y_n)$ are given by

$$\hat{\alpha}_t = a_t + P_t r_t \qquad V_t = P_t - P_t N_t P_t$$
$$r_t = Z_t' F_t^{-1} v_t + L_t' r_{t+1} \qquad N_t = Z_t' F_t^{-1} Z_t + L_t' N_{t+1} L_t$$

where $r_{n+1} = 0$ and $N_{n+1} = 0$.

Univariate Filter

Substantial computational gains are available using the univariate Kalman filter by avoiding the inversion of the F_t matrix above (as well as enabling the use of the exact initial filter, see below). For more details, see [1] (§6.4).

When any H_t is non-diagonal, the observation equation is transformed by taking the LDL factorization of the H_t matrices. The transformed parameters are marked with $*$ to denote that they have been transformed.

$$y_t^* = Z_t^* \alpha_t + d_t^* + \varepsilon_t^* \qquad \varepsilon_t^* \sim N(0, H_t^*)$$
$$y_t^* = C_t^{-1} y_t \qquad Z_t^* = C_t^{-1} Z_t \qquad d_t^* = C_t^{-1} d_t \qquad \varepsilon_t^* = C_t^{-1} \varepsilon_t \qquad H_t = C_t H_t^* C_t'$$

The multivariate filter and smoother is then modified by computing several quantities via the univariate filter on the transformed system:

$$a_{t,i+1} = a_{t,i} + K_{t,i} v_{t,i} \qquad P_{t,i+1} = P_{t,i} - K_{t,i} F_{t,i} K_{t,i}'$$
$$v_{t,i} = y_{t,i} - Z_{t,i} a_{t,i} - d_{t,i} \qquad F_{t,i} = Z_{t,i} P_{t,i} Z_{t,i}' + H_{t,i} \qquad K_{t,i} = P_{t,i} Z_{t,i}' F_{t,i}^{-1}$$
$$a_{t+1,1} = T_{t+1} a_{t,p+1} + c_{t+1} \qquad P_{t+1,1} = T_{t+1} P_{t,p+1} T_{t+1}' + R_{t+1} Q_{t+1} R_{t+1}'$$

where $y_{i,t}$ is the ith element of y_t^*, $Z_{t,i}$ is the ith row of Z_t^*, $d_{t,i}$ is the ith element of d_t^*, and $H_{t,i}$ is the ith diagonal element of H_t^*. The filtered estimates of the state $a_t = a_{t,1}$ and $P_t = P_{t,1}$ are equivalent to those computed above. Note that $F_{t,i}$ is a scalar, so that $F_{t,i}^{-1}$ is simply scalar division instead of more computationally expensive matrix inversion. However, this comes at the cost of there being no convenient transformation of the quantities $v_{t,i}$, $F_{t,i}$ or $K_{t,i}$ that recovers the multivariate versions computed above.

Similarly, the univariate smoother provides $r_t = r_{t,0}$ and $N_t = N_{t,0}$:

$$r_{t,i-1} = Z_{t,i}' F_{t,i}^{-1} v_{t,i} + L_{t,i}' r_{t,i} \qquad N_{t,i-1} = Z_{t,i}' F_{t,i}^{-1} Z_{t,i} + L_{t,i}' N_{t,i} L_{t,i} \qquad L_{t,i} = I_m - K_{t,i} Z_{t,i} F_{t,i}^{-1}$$
$$r_{t-1,p} = T_t' r_{t,0} \qquad N_{t-1,p} = T_{t-1}' N_{t,0} T_{t-1}$$

where $r_{t+1,p} = 0$ and $N_{t+1,p} = 0$.

Exact Initial Kalman Filter

When the state α_t is stationary, the initial values a_1 and P_1 can be computed as the unconditional mean and variance of the state given the system parameters by inverting the transition equation.

To handle cases where some states are nonstationary, the state is separated into states with known variance (those that are stationary) and those that are initialized as diffuse (the nonstationary states). Let $\tilde{R}$ be a selection matrix with columns from the identity such that the initial shock η_1 is applied to the states with known variance. The selection matrix A is composed of the columns of the identity matrix associated with the diffuse states such that

taking the limit as $\kappa \to \infty$ allows the diffuse states to have infinite initial variance where the initial values are given by

$$a_1 = a + \tilde{R}\eta_1 + A\delta \qquad \eta_1 \sim N(0, \tilde{Q}) \qquad \delta \sim N(0, \kappa I)$$
$$P_1 = P_{*,1} + \kappa P_{\infty,1} \qquad P_{*,1} = \tilde{R}\tilde{Q}\tilde{R}' \qquad P_{\infty,1} = AA'$$

The unconditional mean, a, and variance, $P_{*,1}$, of the state are computed by inverting the stationary portion of the system. Any elements of a associated with nonstationary or diffuse states are set to 0. Letting $\tilde{T} = \tilde{R}'T_1\tilde{R}$ and $\tilde{c} = \tilde{R}'c_1$, this is accomplished by

$$\tilde{R}'a = (I_m - \tilde{T})^{-1}\tilde{c}$$
$$\text{vec}(\tilde{R}\tilde{Q}\tilde{R}') = (I_{m^2} - \tilde{T} \otimes \tilde{T})^{-1}\text{vec}(\tilde{R}'R_1Q_1R_1'\tilde{R})$$

When all states are stationary, this initialization collapses down to the simple case where $\alpha_1 \sim N(a_1, P_1)$, where a_1 and P_1 are the unconditional mean and variance of the state, determined by inverting the full system.

Given this initialization, the univariate filter recursions must be altered to separate the states with finite and infinite variances (see [1] (§5.2) and [2]):

$$F_{*,t,i} = Z_{t,i}P_{*,t,i}Z_{t,i}' + H_{t,i} \qquad F_{\infty,t,i} = Z_{t,i}P_{\infty,t,i}Z_{t,i}'$$
$$K_{*,t,i} = P_{*,t,i}Z_{t,i}'F_{*,t,i}^{-1} \qquad K_{\infty,t,i} = P_{\infty,t,i}Z_{t,i}'F_{\infty,t,i}^{-1}$$

$$a_{t,i+1} = \begin{cases} a_{t,i} + K_{*,t,i}v_{t,i} & F_{\infty,t,i} = 0 \\ a_{t,i} + K_{\infty,t,i}v_{t,i} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$P_{*,t,i+1} = \begin{cases} P_{*,t,i} - K_{*,t,i}K_{*,t,i}'F_{*,t,i} & F_{\infty,t,i} = 0 \\ P_{*,t,i} - (K_{*,t,i}K_{\infty,t,i}' + K_{\infty,t,i}K_{*,t,i}' - K_{\infty,t,i}K_{\infty,t,i}')F_{*,t,i} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$P_{\infty,t,i+1} = \begin{cases} P_{\infty,t,i} & F_{\infty,t,i} = 0 \\ P_{\infty,t,i} - K_{\infty,t,i}K_{\infty,t,i}'F_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$a_{t+1,1} = T_{t+1}a_{t,p+1} + c_{t+1}$$
$$P_{\infty,t+1,1} = T_{t+1}P_{\infty,t,p+1}T_{t+1}' \qquad P_{*,t+1,1} = T_{t+1}P_{*,t,p+1}T_{t+1}' + R_{t+1}Q_{t+1}R_{t+1}'$$

For any set of system parameters where the state can be identified, there exists some time d such that $F_{\infty,d,i} = 0$ for all i. For time $t \geq d$, the simpler Kalman filter recursion above can be employed with $P_{t,i} = P_{*,t,i}$.

The smoother must be similarly altered so that beginning at $t = d$, the computation of $r_{t,i}$ is expanded to account for the initialization:

$$L_{\infty,t,i} = I_m - K_{\infty,t,i}Z_{t,i} \qquad L_{*,t,i} = I_m - K_{*,t,i}Z_{t,i}$$
$$L_{t,i}^{(0)} = (K_{\infty,t,i} - K_{*,t,i})Z_{t,i}F_{*,t,i}F_{\infty,t,i}^{-1}$$

$$r_{t,i-1}^{(0)} = \begin{cases} Z'_{t,i}F_{*,t,i}^{-1}v_{t,i} + L'_{*,t,i}r_{t,i}^{(0)} & F_{\infty,t,i} = 0 \\ L'_{\infty,t,i}r_{t,i}^{(0)} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$r_{t,i-1}^{(1)} = \begin{cases} r_{t,i}^{(1)} & F_{\infty,t,i} = 0 \\ Z'_{t,i}F_{\infty,t,i}^{-1}v_{t,i} + L_{t,i}^{(0)\prime}r_{t,i}^{(0)} + L'_{\infty,t,i}r_{t,i}^{(1)} & F_{\infty,t,i} \neq 0 \end{cases}$$

$$r_{t-1,p}^{(0)} = T'_t r_{t,0}^{(0)} \qquad r_{t-1,p}^{(1)} = T'_t r_{t,0}^{(1)}$$

$$\hat{\alpha}_t = a_t + P_{*,t,1}r_{t,0}^{(0)} + P_{\infty,t,1}r_{t,0}^{(1)}$$

where $r_{d,p}^{(0)} = r_{d,p}$ and $r_{d,p}^{(1)} = 0$. Note that $L_{\infty,t,i}$ and $L_{t,i}^{(0)}$ only need to be computed when $F_{\infty,t,i} \neq 0$ and $L_{*,t,i}$ only needs to be computed when $F_{\infty,t,i} = 0$.

For the smoothed variance of the state,

$$V_t = P_{*,t,1} - P_{*,t,1}N_{t,0}^{(0)}P_{*,t,1} - \left(P_{\infty,t,1}N_{t,0}^{(1)}P_{*,t,1}\right)' - P_{\infty,t,1}N_{t,0}^{(1)}P_{*,t,1} - P_{\infty,t,1}N_{t,0}^{(2)}P_{\infty,t,1}$$

$$N_{t-1,p}^{(0)} = T'_t N_{t,0}^{(0)}T_t \qquad N_{t-1,p}^{(1)} = T'_t N_{t,0}^{(1)}T_t \qquad N_{t-1,p}^{(2)} = T'_t N_{t,0}^{(1)}T_t$$

where when $F_{\infty,t,i} = 0$,

$$N_{t,i-1}^{(0)} = Z'_{t,i}F_{*,t,i}^{-1}Z_{t,i} + L'_{*,t,i}N_{t,i}^{(0)}L_{*,t,i} \qquad N_{t,i-1}^{(1)} = N_{t,i}^{(1)} \qquad N_{t,i-1}^{(2)} = N_{t,i}^{(2)}$$

and when $F_{\infty,t,i} \neq 0$,

$$N_{t,i-1}^{(0)} = L'_{\infty,t,i}N_{t,i}^{(0)}L_{\infty,t,i}$$

$$N_{t,i-1}^{(1)} = Z_{t,i}^{*\prime}F_{\infty,t,i}^{-1}Z_{t,i} + L'_{\infty,t,i}N_{t,i}^{(0)}L_{t,i}^{(0)} + L'_{\infty,t,i}N_{t,i}^{(1)}L_{\infty,t,i}$$

$$N_{t,i-1}^{(2)} = Z'_{t,i}F_{\infty,t,i}^{-2}Z_{t,i}F_{*,t,i} + L_{t,i}^{(0)\prime}N_{t,i}^{(1)}L_{t,i}^{(0)} + L'_{\infty,t,i}N_{t,i}^{(1)}L_{t,i}^{(0)} + L_{t,i}^{(0)\prime}N_{t,i}^{(1)}L_{\infty,t,i} + L'_{\infty,t,i}N_{t,i}^{(2)}L_{\infty,t,i}$$

where $N_{d,p}^{(0)} = N_{d,p}$ and $N_{d,p}^{(1)} = N_{d,p}^{(2)} = 0$.

2   Likelihood and Gradient Calculation

The likelihood of of data $y_1, \ldots, y_n$ in the standard filter as shown in [1] (§7.2) is given by the prediction error decomposition:

$$\log L(Y_n) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \left(\log |F_t| + v_t' F_{t,i}^{-1} v_t \right)$$

For the univariate filter, the same decomposition works in the univariate context:

$$\log L(Y_n) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \sum_{i=1}^{p} \log F_{t,i} + v_{t,i}^2 / F_{t,i} \tag{1}$$

The likelihood for the exact initial filter allows for some simplifications in $F_{\infty,t,i}$ as shown in [1] (§7.2):

$$\log L_d(Y_n) = -\frac{1}{2} \sum_{t=1}^{n} \sum_{i=1}^{p} \iota_{t,i} \log 2\pi - \frac{1}{2} \sum_{t=1}^{d} \sum_{i=1}^{p} w_{t,i} - \frac{1}{2} \sum_{t=d}^{n} \sum_{i=1}^{p} \iota_{t,i} \left(\log F_{t,i} + v_{t,i}^2 / F_{t,i} \right) \tag{2}$$

where $\iota_{t,i} = 1$ if $F_{*,t,i} \neq 0$ or $t > d$, and

$$w_{t,i} = \begin{cases} \iota_{t,i}(\log(F_{*,t,i}) + v_{t,i}^{(0)2}/F_{*,t,i}) & F_{\infty,t,i} = 0 \\ \log F_{\infty,t,i} & F_{\infty,t,i} \neq 0 \end{cases}$$

Since these quantities are naturally produced by the Kalman filter, this is the preferred method to calculate the likelihood.

Univariate Gradient

Following the notation of [3], we denote $\partial \left[\mathrm{vec}(A)' \right] / \partial \theta$ by $G_\theta(A)$, where for an $m \times n$ matrix A and a θ of length n_θ, $G_\theta(A)$ will be a $n_\theta \times mn$ matrix of partial derivatives. See Appendix A for details and properties of $G_\theta(\cdot)$. For the calculation of the gradients of the state space parameters, see section 3.

Given the univariate likelihood function (1), the identities for the gradient in Appendix A produce the expression for the univariate gradient:

$$G_\theta(\log L(Y_n)) = G_\theta \left(-\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{n} \sum_{i=1}^{p} \log F_{t,i} + v_{t,i}^2 F_{t,i}^{-1} \right)$$

$$= -\frac{1}{2} \sum_{t=1}^{n} \sum_{i=1}^{p} G_\theta(\log F_{t,i}) + G_\theta(v_{t,i}^2 F_{t,i}^{-1})$$

$$= -\sum_{t=1}^{n} \sum_{i=1}^{p} \frac{1}{2} G_\theta(F_{t,i}) \left[F_{t,i}^{-1} - F_{t,i}^{-2} v_{t,i}^2 \right] + G_\theta(v_{t,i}) F_{t,i}^{-1} v_{t,i}$$

The quantities used in the gradient recursion are computed similarly using the definitions from the univariate filter:

$$
\begin{aligned}
G_\theta(v_{t,i}) &= G_\theta(y_{t,i} - Z_{t,i}a_{t,i} - d_{t,i}) \\
&= G_\theta(y_{t,i}) - G_\theta(Z_{t,i})a_{t,i} - G_\theta(a_{t,i})Z'_{t,i} - G_\theta(d_{t,i}) \\
G_\theta(F_{t,i}) &= G_\theta(Z_{t,i}P_{t,i}Z'_{t,i} + H_{t,i}) \\
&= 2G_\theta(Z_{t,i})P_{t,i}Z'_{t,i} + G_\theta(P_{t,i})(Z'_{t,i} \otimes Z'_{t,i}) + G_\theta(H_{t,i}) \\
G_\theta(K_{t,i}) &= G_\theta(P_{t,i}Z'_{t,i}F^{-1}_{t,i}) \\
&= G_\theta(P_{t,i}Z'_{t,i})F^{-1}_{t,i} + G_\theta(F^{-1}_{t,i})Z_{t,i}P_{t,i} \\
&= G_\theta(P_{t,i})(Z'_{t,i}F^{-1}_{t,i} \otimes I_m) + G_\theta(Z_{t,i})P_{t,i}F^{-1}_{t,i} - G_\theta(F_{t,i})F^{-2}_{t,i}Z_{t,i}P_{t,i} \\
G_\theta(a_{t+1,1}) &= G_\theta(T_{t+1}a_{t,p+1} + c_{t+1}) \\
&= G_\theta(T_{t+1})(a_{t,p+1} \otimes I_m) + G_\theta(a_{t,p+1})T'_{t+1} + G_\theta(c_{t+1}) \\
G_\theta(a_{t,i+1}) &= G_\theta(a_{t,i} + K_{t,i}v_{t,i}) \\
&= G_\theta(a_{t,i}) + G_\theta(K_{t,i})v_{t,i} + G_\theta(v_{t,i})K'_{t,i} \\
G_\theta(P_{t+1,1}) &= G_\theta(T_{t+1}P_{t,p+1}T'_{t+1} + R_{t+1}Q_{t+1}R'_{t+1}) \\
&= G_\theta(T_{t+1})(P_{t,p+1}T'_{t+1} \otimes I_m)N_m + G_\theta(P_{t,p+1})(T'_{t+1} \otimes T'_{t+1}) \\
&\quad + G_\theta(R_{t+1})(Q_{t+1}R'_{t+1} \otimes I_m)N_m + G_\theta(Q_{t+1})(R'_{t+1} \otimes R'_{t+1}) \\
G_\theta(P_{t,i+1}) &= G_\theta(P_{t,i} - K_{t,i}F_{t,i}K'_{t,i}) \\
&= G_\theta(P_{t,i}) - G_\theta(K_{t,i})(F_{t,i}K'_{t,i} \otimes I_m)N_m - G_\theta(F_{t,i})(K'_{t,i} \otimes K'_{t,i})
\end{aligned}
$$

where $N_m = I_{m^2} + K_{mm}$ for a commutation matrix K_{mm}. The initial period gradients $G_\theta(a_{1,1}) = G_\theta(a_1)$ and $G_\theta(P_{1,1}) = G_\theta(P_1)$ will be given below. The values of $G_\theta(y_{t,i})$ are computed according to

$$
\begin{aligned}
G_\theta(y^*_t) &= G_\theta(C^{-1}_t y_t) \\
&= -G_\theta(C_t)(C^{-1}_t \otimes C'^{-1}_t)(y_t \otimes I_p)
\end{aligned}
$$

where $G_\theta(y_{t,i})$ is the ith column of $G_\theta(y^*_t)$.

Exact Initial Gradient

The process for computing the gradient for use with the exact initial filter is similar. Throughout, it is assumed that the boolean outcome of the tests $F_{*,t,i} = 0$ or $F_{\infty,t,i} = 0$ are unaffected by changes to θ. Beginning with the likelihood function,

$$
\begin{aligned}
G_\theta(\log L_d(Y_n)) &= G_\theta\left(-\frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{p}\iota_{t,i}\log 2\pi - \frac{1}{2}\sum_{t=1}^{d}\sum_{i=1}^{p}w_{t,i} - \frac{1}{2}\sum_{t=d}^{n}\sum_{i=1}^{p}\iota_{t,i}\left(\log F_{t,i} + v^2_{t,i}/F_{t,i}\right)\right) \\
&= -\frac{1}{2}\sum_{t=1}^{d}\sum_{i=1}^{p}G_\theta(w_{t,i}) - \frac{1}{2}\sum_{t=d}^{n}\sum_{i=1}^{p}G_\theta(F_{t,i})\left[F^{-1}_{t,i} - F^{-2}_{t,i}v^2_{t,i}\right] + 2G_\theta(v_{t,i})F^{-1}_{t,i}v_{t,i}
\end{aligned}
$$

where

$$G_\theta(w_{t,i}) = \begin{cases} G_\theta(\iota_{t,i}(\log(F_{*,t,i}) + v_{t,i}^{(0)2}/F_{*,t,i})) & F_{\infty,t,i} = 0 \\ G_\theta(\log F_{\infty,t,i}) & F_{\infty,t,i} \neq 0 \end{cases}$$

$$= \begin{cases} \iota_{t,i} G_\theta(F_{*,t,i}) \left[F_{*,t,i}^{-1} - F_{*,t,i}^{-2} v_{t,i}^{(0)2} \right] + \iota_{t,i} 2 G_\theta(v_{t,i}^{(0)}) F_{*,t,i}^{-1} v_{t,i}^{(0)} & F_{\infty,t,i} = 0 \\ G_\theta(F_{\infty,t,i}) F_{\infty,t,i}^{-1} & F_{\infty,t,i} \neq 0 \end{cases}$$

The required quantities in the recursion come from the definition of the filter and are similar to the basic univariate definitions:

$$\begin{aligned}
G_\theta(F_{*,t,i}) &= G_\theta(Z_{t,i} P_{*,t,i} Z_{t,i}' + H_{t,i}) \\
&= 2 G_\theta(Z_{t,i}) P_{*,t,i} Z_{t,i}' + G_\theta(P_{*,t,i})(Z_{t,i}' \otimes Z_{t,i}') + G_\theta(H_{t,i}) \\
G_\theta(F_{\infty,t,i}) &= G_\theta(Z_{t,i} P_{\infty,t,i} Z_{t,i}') \\
&= 2 G_\theta(Z_{t,i}) P_{\infty,t,i} Z_{t,i}' + G_\theta(P_{\infty,t,i})(Z_{t,i}' \otimes Z_{t,i}') \\
G_\theta(K_{*,t,i}) &= G_\theta(P_{*,t,i} Z_{t,i}' F_{*,t,i}^{-1}) \\
&= G_\theta(P_{*,t,i})(Z_{t,i}' F_{*,t,i}^{-1} \otimes I_m) + G_\theta(Z_{t,i}) P_{*,t,i} F_{*,t,i}^{-1} - G_\theta(F_{*,t,i}) F_{*,t,i}^{-2} Z_{t,i} P_{*,t,i} \\
G_\theta(K_{\infty,t,i}) &= G_\theta(P_{\infty,t,i} Z_{t,i}' F_{\infty,t,i}^{-1}) \\
&= G_\theta(P_{\infty,t,i})(Z_{t,i}' F_{\infty,t,i}^{-1} \otimes I_m) + G_\theta(Z_{t,i}) P_{\infty,t,i} F_{\infty,t,i}^{-1} - G_\theta(F_{\infty,t,i}) F_{\infty,t,i}^{-2} Z_{t,i} P_{\infty,t,i}
\end{aligned}$$

Between periods,

$$\begin{aligned}
G_\theta(a_{t+1,1}) &= G_\theta(T_{t+1} a_{t,p+1} + c_{t+1}) \\
&= G_\theta(T_{t+1})(a_{t,p+1} \otimes I_m) + G_\theta(a_{t,p+1}) T_{t+1}' + G_\theta(c_{t+1}) \\
G_\theta(P_{*,t+1,1}) &= G_\theta(T_{t+1} P_{*,t,p+1} T_{t+1}' + R_{t+1} Q_{t+1} R_{t+1}') \\
&= G_\theta(T_{t+1})(P_{*,t,p+1} T_{t+1}' \otimes I_m) N_m + G_\theta(P_{*,t,p+1})(T_{t+1}' \otimes T_{t+1}') \\
&\quad + G_\theta(R_{t+1})(Q_{t+1} R_{t+1}' \otimes I_m) N_m + G_\theta(Q_{t+1})(R_{t+1}' \otimes R_{t+1}') \\
G_\theta(P_{\infty,t+1,1}) &= G_\theta(T_{t+1} P_{\infty,t,p+1} T_{t+1}') \\
&= G_\theta(T_{t+1})(P_{\infty,t,p+1} T_{t+1}' \otimes I_m) N_m + G_\theta(P_{\infty,t,p+1})(T_{t+1}' \otimes T_{t+1}')
\end{aligned}$$

When $F_{\infty,t,i} = 0$,

$$\begin{aligned}
G_\theta(a_{t,i+1}) &= G_\theta(a_{t,i} + K_{*,t,i} v_{t,i}) \\
&= G_\theta(a_{t,i}) + G_\theta(K_{*,t,i}) v_{t,i} + G_\theta(v_{t,i}) K_{*,t,i}' \\
G_\theta(P_{*,t,i}) &= G_\theta(P_{*,t,i} - K_{*,t,i} K_{*,t,i}' F_{*,t,i}) \\
&= G_\theta(P_{*,t,i}) - G_\theta(K_{*,t,i})(F_{*,t,i} K_{*,t,i}' \otimes I_m) N_m - G_\theta(F_{*,t,i})(K_{*,t,i}' \otimes K_{*,t,i}') \\
G_\theta(P_{\infty,t,i+1}) &= G_\theta(P_{\infty,t,i})
\end{aligned}$$

and when $F_{\infty,t,i} \neq 0$,

$$
\begin{aligned}
G_\theta(a_{t,i+1}) &= G_\theta(a_{t,i} + K_{\infty,t,i}v_{t,i}) \\
&= G_\theta(a_{t,i}) + G_\theta(K_{\infty,t,i})v_{t,i} + G_\theta(v_{t,i})K'_{\infty,t,i} \\
G_\theta(P_{*,t,i+1}) &= G_\theta\left(P_{*,t,i} - (K_{*,t,i}K'_{\infty,t,i} + K_{\infty,t,i}K'_{*,t,i} - K_{\infty,t,i}K'_{\infty,t,i})F_{*,t,i}\right) \\
&= G_\theta(P_{*,t,i}) - G_\theta(F_{*,t,i})(K_{*,t,i}K'_{\infty,t,i} + K_{\infty,t,i}K'_{*,t,i} - K_{\infty,t,i}K'_{\infty,t,i}) \\
&\quad - G_\theta(K_{*,t,i})(K_{\infty,t,i} \otimes I_m)N_m F_{*,t,i} \\
&\quad - G_\theta(K_{\infty,t,i})[(K_{*,t,i} - K'_{\infty,t,i}) \otimes I_m]N_m F_{*,t,i} \\
G_\theta(P_{\infty,t,i+1}) &= G_\theta(P_{\infty,t,i} - K_{\infty,t,i}K'_{\infty,t,i}F_{\infty,t,i}) \\
&= G_\theta(P_{\infty,t,i}) - G_\theta(K_{\infty,t,i})(F_{\infty,t,i}K'_{\infty,t,i} \otimes I_m)N_m - G_\theta(F_{\infty,t,i})(K'_{\infty,t,i} \otimes K'_{\infty,t,i})
\end{aligned}
$$

Initial Conditions

The initial conditions for the recursion are given via the expressions for a_1 and P_1. We assume that the rank of A and R_0 are not affected by θ:

$$
\begin{aligned}
G_\theta(a_1) &= G_\theta(a) + G_\theta(R_0\eta_0) + G_\theta(A\delta) = G_\theta(a) \\
G_\theta(P_1) &= G_\theta(\tilde{R}\tilde{Q}\tilde{R}) + G_\theta(\kappa AA') = G_\theta(\tilde{R}\tilde{Q}\tilde{R})
\end{aligned}
$$

From the definitions of a and $\tilde{R}\tilde{Q}\tilde{R}$ above,

$$
\begin{aligned}
G_\theta(a) &= G_\theta([I_m - T_1]^{-1}c_1) \\
&= G_\theta([I_m - T_1]^{-1})(c_1 \otimes I_m) + G_\theta(c_1)(I_m - T_1)^{-1})' \\
&= -G_\theta(I_m - T_1)[(I_m - T_1)^{-1} \otimes (I_m - T_1)'^{-1}](c_1 \otimes I_m) + G_\theta(c_1)(I_m - T_1)'^{-1} \\
&= G_\theta(T_1)[(I_m - T_1)^{-1} \otimes (I_m - T_1)'^{-1}](c_1 \otimes I_m) + G_\theta(c_1)(I_m - T_1)'^{-1} \\
G_\theta(\tilde{R}\tilde{Q}\tilde{R}') &= G_\theta(S \operatorname{vec}(R_1 Q_1 R'_1)) \\
&= G_\theta(S)[\operatorname{vec}(R_1 Q_1 R'_1) \otimes I_{m^2}] + G_\theta(\operatorname{vec}(R_1 Q_1 R'_1))(I_1 \otimes S') \\
&= -G_\theta(I_{m^2} - T_1 \otimes T_1)(S \otimes S')[\operatorname{vec}(R_1 Q_1 R'_1) \otimes I_{m^2}] \\
&\quad + [G_\theta(R_1)(Q_1 R'_1 \otimes I_m)N_m + G_\theta(Q_1)(R'_1 \otimes R'_1)]S' \\
&= G_\theta(T_1 \otimes T_1)(S \otimes S')[\operatorname{vec}(R_1 Q_1 R'_1) \otimes I_{m^2}] \\
&\quad + [G_\theta(R_1)(Q_1 R'_1 \otimes I_m)N_m + G_\theta(Q_1)(R'_1 \otimes R'_1)]S'
\end{aligned}
$$

where $S = (I_{m^2} - T_1 \otimes T_1)^{-1}$. Note that $G_\theta(T_1 \otimes T_1)$ must be computed separately. For the diffuse filter, note that $G_\theta(P_{*,1}) = G_\theta(P_1)$ and $G_\theta(P_{\infty,1}) = 0$.

Univariate Treatment of Multivariate Series

For the univariate treatment of multivariate series, we need to find $G_\theta(Z_t^*)$, $G_\theta(d_t^*)$, and $G_\theta(H_t^*)$ given $G_\theta(Z_t)$, $G_\theta(d_t)$, $G_\theta(H_t)$, and C_t, the univariate factorization matrix. To do so,

we solve for $G_\theta(H_t^*)$ and $G_\theta(C_t)$, noting that elements of $G_\theta(C_t)$ associated with the upper triangular elements of C_t must be zero while elements of $G_\theta(H_t^*)$ associated with off-diagonal elements of H_t^* must be zero:

$$
\begin{aligned}
G_\theta(H_t) &= G_\theta(C_t H_t^* C_t') \\
&= G_\theta(C_t)(H_t^* C_t' \otimes I_p) N_p + G_\theta(H_t^*)(C_t' \otimes C_t') \\
&= \begin{bmatrix} G_\theta(C_t) & G_\theta(H_t^*) \end{bmatrix} \begin{bmatrix} (H_t^* C_t' \otimes I_p) N_p \\ C_t' \otimes C_t' \end{bmatrix} = G_t W_t = \tilde{G}_t \tilde{W}_t
\end{aligned}
$$

where $\tilde{G}_t$ contains the nonzero columns of G_t and $\tilde{W}_t$ contains the corresponding rows of W_t. The product $G_\theta(H_t)\tilde{W}_t^+$ then provides an approximate solution for $\tilde{G}_t$ that can be reordered to solve for $G_\theta(C_t)$ and $G_\theta(H_t^*)$.

Having solved for $G_\theta(C_t)$, we can solve for the other gradients:

$$
\begin{aligned}
G_\theta(Z_t^*) &= -G_\theta(C_t)(C_t^{-1} \otimes C_t'^{-1})(Z_t \otimes I_p) + G_\theta(Z_t)(I_m \otimes C_t'^{-1}) \\
G_\theta(d_t^*) &= -G_\theta(C_t)(C_t^{-1} \otimes C_t'^{-1})(d_t \otimes I_p) + G_\theta(d_t)C_t'^{-1}
\end{aligned}
$$

3 Parameter Estimation

We are interested in the maximum likelihood estimate of a set of *structural parameters* (θ) given data $y_1, \ldots, y_n$ where elements of the *state space parameters* (Z_t, d_t, H_t, T_t, c_t, R_t, and Q_t) depend on θ. For computational simplicity, the state space parameters will be restricted such that each of their scalar elements must be a function of a single element of a vector of the *reduced form parameters* (ψ) which depend on θ. The reduced form parameters are unrestricted in how they may depend on θ, allowing for a rich specification of parameter restrictions. In most estimations, there will be the same number of structural parameters and reduced form parameters.

More formally, let $\theta \in \mathbb{R}^{n_\theta}$ and $\psi \in \mathbb{R}^{n_\psi}$ where $n_\theta \leq n_\psi$. Define a set of functions $\Psi_i : \mathbb{R}^{n_\theta} \to \mathbb{R}$ such that each element of ψ is a function of the structural parameters, $\psi_i = \Psi_i(\theta)$. Except in cases of cross-parameter restrictions, $n_\theta = n_\psi$ and $\Psi_i(\theta) = \theta_i$ so that $\psi = \theta$. In cases of cross-parameter restrictions, the user must define the Ψ_i functions for each restricted element. Also, unless initial values will be given as the vector θ (which is often inconvenient), the inverse function $\Psi^{-1} : \mathbb{R}^{n_\psi} \to \mathbb{R}^{n_\theta}$ must be specified.

Additionally define a set of functions $\tau_{X_{i,j}} : \mathbb{R} \to \mathbb{R}$ for $X \in \{Z_t, d_t, H_t, T_t, c_t, R_t, Q_t\}$ such that each element of the parameter matrices to be estimated is a transformation of an element of ψ. Common $\tau_{X_{i,j}}$ transformations include the identity, exponential, negative exponential, and logistic transformations to allow for an unbounded $\psi \in \mathbb{R}^{n_\psi}$ while maintaining bounds on the parameter matrices. When estimating models with mixed-frequency data, all state space parameters shared between the high-frequency and low-frequency states depend on the same reduced form parameters simply by using different $\tau_{X_{i,j}}$ functions. In almost all cases, the specification of the $\tau_{X_{i,j}}$ functions will be done automatically to account for state space parameter bounds and accumulator definitions. User definitions of $\tau_{X_{i,j}}$ should be rare.

The gradients of the state space parameters with respect to the reduced form parameters, $G_\psi(X)$, are easily computed according to the derivatives of $\tau_{X_{i,j}}$. The gradients of the state space parameters with respect to θ simply require the chain rule, $G_\theta(X) = G_\theta(\psi) G_\psi(X)$. Note that in cases where $\psi \neq \theta$, $G_\theta(\psi)$ must be provided by the user.

With the likelihood and gradient of the structural parameters defined, the optimization can be performed using common gradient ascent methods available in the Matlab Optimization Toolbox. In practice, it has been observed that further improvements in the likelihood are possible using simplex based optimization methods the gradient ascent method has converged. Due to this, the two methods are repeated until neither is able to improve the likelihood using standard convergence criteria.

A   Gradient Derivation

Denote $\partial\left[\text{vec}(A)'\right]/\partial\theta$ by $G_\theta(A)$, where for an $m \times n$ matrix A and a θ of length n_θ, $G_\theta(A)$ will be a $n_\theta \times mn$ matrix of partial derivatives. Several identities from [3] are needed in computing the gradient (assume matrices are of dimensions such that the expression on the left-hand side exists):

(a) $G_\theta(A+B) = G_\theta(A) + G_\theta(B)$

(b) $G_\theta(AB) = G_\theta(A)(B \otimes I_{p_A}) + G_\theta(B)(I_{q_B} \otimes A')$

(c) $G_\theta(A') = G_\theta(A)K_{p_A q_A}$

(d) $G_\theta(ACA') = G_\theta(A)(CA' \otimes I_{p_A})N_{p_A} + G_\theta(C)(A' \otimes A')$

(e) $G_\theta(D^{-1}) = -G_\theta(D)(D^{-1} \otimes D'^{-1})$

(f) $G_\theta(\log|D|) = G_\theta(D)\ \text{vec}(D'^{-1})$

where C is symmetric, D is nonsingular, K_{mn} is a commutation matrix of size mn, and $N_m = I_{m^2} + K_{mm}$. For any given matrix M is of dimension $p_M \times q_M$. Note that $G_\theta(AA')$ can easily be computed from (d) by treating C as the identity matrix. Also note that $A \otimes I_1 = A$ and $\log|B| = \log B$ if B is a scalar.

The chain rule for the operator $G_\theta(\cdot)$ is simply $G_\theta(A) = G_\theta(\psi)G_\psi(A)$ where A is a function of ψ which is a function of θ. This can be shown by noting that

$$\frac{\partial A_{i,j}}{\partial\theta_k} = \sum_{l=1}^{n_\psi} \frac{\partial A_{i,j}}{\partial\psi_l}\frac{\partial\psi_l}{\partial\theta_k}$$

which defines the appropriate element of the matrix product $G_\theta(\psi)G_\psi(A)$.

References

[1] James Durbin and Siem Jan Koopman. *Time Series Analysis by State Space Methods: Second Edition (Oxford Statistical Science Series)*. Oxford University Press, 2012.

[2] S. J. Koopman and J. Durbin. Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, (3):281–296, 2000.

[3] Daisuke Nagakura. Exact gradient vector of loglikelihood for linear gaussian state space model. 2013.