ECONOMETRIC METHODS II: TIME SERIES

LECTURE NOTES ON THE KALMAN FILTER

KRISTOFFER P. NIMARK

The Kalman Filter

We will be concerned with state space systems of the form

$$X_t = A_t X_{t-1} + C_t \mathbf{u}_t \quad (0.1)$$

$$Z_t = D_t X_t + \mathbf{v}_t \quad (0.2)$$

where X_t is an $n \times 1$ vector of random variables, $\mathbf{u}_t$ is an $m \times 1$ vector of i.i.d. shocks with unit variance, i.e. $E\left[\mathbf{u}_t \mathbf{u}_{t+s}'\right] = I$ if $s = 0$ and $\mathbf{0}$ otherwise. A and C are ($n \times n$ and $n \times m$ respectively) coefficient matrices. Z_t is an $(l \times 1)$ vector of observables and D_t is an $(l \times n)$ selector matrix that combines elements of the state X_t into observable variables and $\mathbf{v}_t$ is an $(l \times 1)$ vector of measurement errors with covariance Σ_{vv}.

Given a system of the form (0.1) - (0.2) , the Kalman filter recursively computes estimates of X_t conditional on the history of observations $Z_t, Z_{t-1}, ...Z_0$ and an initial estimate (or prior) $X_{0|0}$ with variance $P_{0|0}$.

The form of the filter is

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1}\right) \quad (0.3)$$

and the task is thus to find the Kalman gain K so that the estimates $X_{t|t}$ are in some sense "optimal".

Date: June 1, 2010
I thank Loic Berger for drawing my attention to several typos in an earlier version of these notes.

1. A Simple Example

Let's say that we have a noisy measures z^1 of the unobservable process x so that

$$z_1 = x + v_1 \tag{1.1}$$

$$v_1 \sim N(0, \sigma_1^2) \tag{1.2}$$

Since the signal is unbiased, the minimum variance estimate $E\left[x \mid z^1\right] \equiv \widehat{x}$ of x is simply given by

$$\widehat{x} = z_1 \tag{1.3}$$

and its variance is equal to the variance of the noise

$$E\left[\widehat{x} - x\right]^2 = \sigma_1^2 \tag{1.4}$$

Now, let's say we have an second measure z_2 of x so that

$$z_2 = x + v_2 \tag{1.5}$$

$$v_2 \sim N(0, \sigma_2^2) \tag{1.6}$$

With two measures it makes intuitive sense that we should be able to get a better estimate of x than what we could get with a single signal. How can we combine the information in the two signals to find the a minimum variance estimate of x? If we restrict ourselves to linear estimators of the form

$$\widehat{x} = (1 - a)\, z_1 + a z_2 \tag{1.7}$$

we can simply minimize

$$E\left[(1 - a)\, z_1 + a z_2 - x\right]^2 \tag{1.8}$$

with respect to a. Rewrite (1.8) as

$$\begin{aligned} & E\left[(1-a)\left(x+v_1\right)+a\left(x+v_2\right)-x\right]^2 \\ = & E\left[(1-a) v_1+a v_2\right]^2 \\ = & \sigma_1^2-2 a \sigma_1^2+a^2 \sigma_1^2+a^2 \sigma_2^2 \end{aligned} \tag{1.9}$$

where the third line follows from the fact that v^1 and v^2 are uncorrelated so all expected cross terms are zero. Differentiate w.r.t. a and set equal to zero

$$-2\sigma_1^2+2a\sigma_1^2+2a\sigma_2^2=0 \tag{1.10}$$

and solve for a

$$a=\sigma_1^2/(\sigma_1^2+\sigma_2^2) \tag{1.11}$$

The minimum variance estimate is then given by

$$\widehat{x}=\frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}z_1+\frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}z_2 \tag{1.12}$$

with conditional variance

$$E\left[\widehat{x}-x\right]^2=\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1} \tag{1.13}$$

It is clear from (1.12) that less weight will be put on a more noisy signal. The expression for the variance (1.13) of the estimate also shows that the precision of the estimate increases when a second observation is available as long as the noise variance is finite. In the next section we will show that finding a is analogous to finding the Kalman gain for a system with a scalar state and a scalar signal.

2. The Scalar Filter

Consider the process

$$x_t = \rho x_t + u_t \tag{2.1}$$

$$z_t = x_t + v_t \tag{2.2}$$

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}\right) \tag{2.3}$$

We want to form an estimate of x_t conditional on $z^t = \{z_t, z_{t-1,\ldots,}z_0\}$. In addition to the knowledge of the state space system (2.1) - (2.3) we have a "prior" belief about the initial value of the state x_0 so that

$$x_{0|0} = \overline{x}_0 \tag{2.4}$$

$$E\left(\overline{x}_0 - x_0\right)^2 = p_0 \tag{2.5}$$

With this information we can form an estimate of the state in period 1. Using the state transition equation we get

$$E\left[x_1 \mid x_{0|0}\right] = \rho x_{0|0} \tag{2.6}$$

We can call this the prior estimate of x_t in period 1 and denote it $x_{1|0}$. The variance of the prior estimate then is

$$E\left(x_{1|0} - x_1\right)^2 = \rho^2 p_0 + \sigma_u^2 \tag{2.7}$$

where the first term, $\rho^2 p_0$, is the uncertainty from period 0 carried over to period 1 and the second term, σ_u^2, is just the uncertainty in period 0 about the period 1 innovation to x_t. We can denote this uncertainty $p_{1|0}$ so that

$$p_{1|0} = \rho^2 p_0 + \sigma_u^2 \tag{2.8}$$

In period 1 we can also observe the period 1 signal z_1. The information in z_1 can be combined with the information in the prior in exactly the same way as we combined the two signals in the previous section. The optimal weight k_1 in

$$x_{1|1} = (1 - k)x_{1|0} + k_1 z_1 \tag{2.9}$$

is thus given by

$$k_1 = \frac{p_{1|0}}{p_{1|0} + \sigma_v^2} \tag{2.10}$$

and the period 1 posterior error covariance $p_{1|1}$ then is

$$p_{1|1} = \left(\frac{1}{p_{1|0}} + \frac{1}{\sigma_v^2} \right)^{-1} \tag{2.11}$$

or equivalently

$$p_{1|1} = p_{1|0} - p_{1|0}^2 (p_{1|0} + \sigma_v^2)^{-1} \tag{2.12}$$

We can again propagate the posterior error variance $p_{1|1}$ one step forward to get the next period prior variance $p_{2|1}$

$$p_{2|1} = \rho^2 p_{1|1} + \sigma_u^2 \tag{2.13}$$

or

$$p_{2|1} = \rho^2 \left(p_{1|0} - p_{1|0}^2 (p_{1|0} + \sigma_v^2)^{-1} \right) + \sigma_u^2 \tag{2.14}$$

By an induction type argument, we can find a general difference equation for the evolution of prior error variances

$$p_{t|t-1} = \rho^2 \left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right) + \sigma_u^2 \tag{2.15}$$

where

$$p_{t|t-s} = E\left[x_{t|t-s} - x_t \right]^2 \tag{2.16}$$

The associated period t Kalman gain is then given by

$$k_t = p_{t|t-1}(p_{t|t-1} + \sigma_v^2)^{-1}$$

There are two things worth noting about the difference equation for the prior error variances:

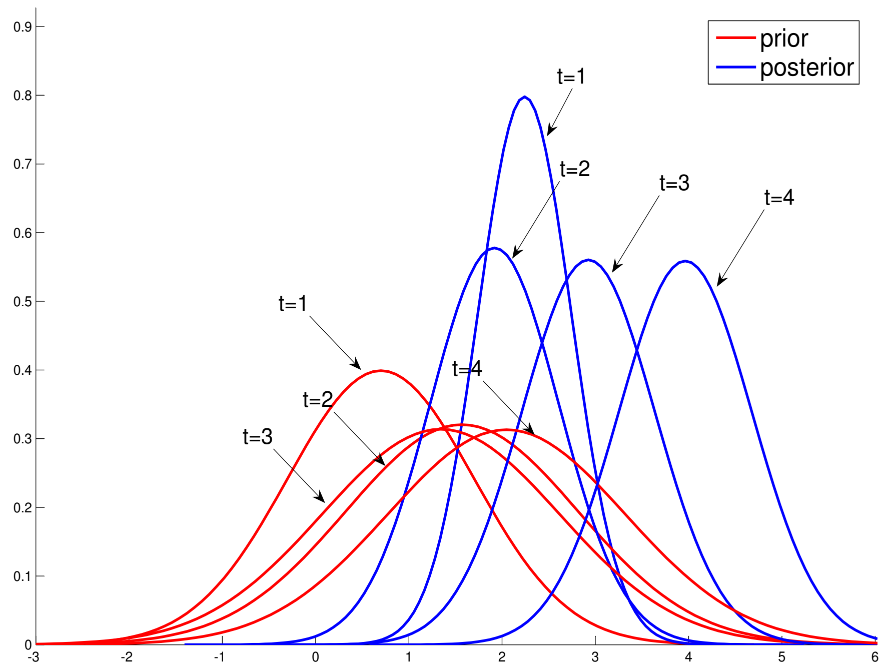

FIGURE 1. Propagation of prior and posterior distributions: $\overline{x}_0 = 1, p_0 = 1, \sigma_u^2 = 1, \sigma_v^2 = 1,\ z^t = \begin{bmatrix} 3.4 & 2.2 & 4.2 & 5.5 \end{bmatrix}$

(1) (a) The prior error variance is bounded both from above and below so that $\sigma_u^2 \leq p_{t|t-1} \leq 1/(1-\rho^2)\sigma_u^2$

(b) For $0 \leq |\rho| < 1$ (2.15) is a contraction

The upper bound in (a) is given by the optimality of the filter and that we cannot do worse than making the unconditional mean our estimate of x_t for all t. The error then is just the variance of x_t, or $1/(1-\rho^2)\sigma_u^2$. The lower bound is given by that the future is inherently uncertain as long as there are innovations in the x_t process, so even with a perfect estimate of x_{t-1}, x_t will still not be known with certainty.

The significance of (b) is that the difference equation converges to a unique number, p_∞. To see this, note that we can rewrite (2.15) as

$$p_{t|t-1} = \rho^2 \left(\frac{1}{p_{t-1|t-2}} + \frac{1}{\sigma_v^2} \right)^{-1} + \sigma_u^2 \tag{2.17}$$

which satisfies Blackwell's sufficient conditions of discounting and monotonicity (see Ljungquist and Sargent 2005, p1010).

Figure 1 plots the prior and posterior densities of $x_{t|t}$ and illustrates the convergence of both the prior and posterior error variances. The parameters in (2.1) - (2.3) was set to $\rho = .8$ and $p_0 = \sigma_u^2 = \sigma_v^2 = \overline{x}_0 = 1$ and the history of observations $\{z_t\}_{t=1}^4 = \left\{ \begin{array}{cccc} 3.4, & 2.2, & 4.2 & 5.5 \end{array} \right\}$. One can see that the dispersion of the distributions converges quite quickly. It is also worth noting that the dispersion of the distributions at each point is independent of the actual realizations of z_t, though the location of the distribution is not.

3. The Discrete Time Kalman Filter

In this section we derive the formulas of the multivariate filter. We first restate the assumptions about the form of the filter and the initial conditions, and for the purposes of this section, it is convenient to first assume that the shock processes are Gaussian. As noted in the introduction, we will be concerned with systems of the form

$$X_t = A_{t-1} X_{t-1} + C_t \mathbf{u}_t \tag{3.1}$$

$$Z_t = D_t X_t + \mathbf{v}_t \tag{3.2}$$

where X_t is an $n \times 1$ vector of random variables, $\mathbf{u}_t$ is an $m \times 1$ vector of i.i.d. Gaussian shocks with unit variance, i.e. $E\left[\mathbf{u}_t \mathbf{u}_{t+s}'\right] = I$ if $s = 0$ and $\mathbf{0}$ otherwise. A and C are ($n \times n$ and $n \times m$ respectively) coefficient matrices. Z_t is an $(l \times 1)$ vector of observables and D_t is an $(l \times n)$ selector matrix that combines elements of the state X_t into observable variables and $\mathbf{v}_t$ is an $(l \times 1)$ vector of Gaussian measurement errors with covariance $\Sigma_{vv,t}$.

Given a system of the form (3.1) - (3.2) , the Kalman filter recursively computes estimates of X_t conditional on the history of observations $Z_t, Z_{t-1}, ...Z_0$ and an initial estimate (or prior) $X_{0|0}$ with variance $P_{0|0}$ defined as

$$E\left(X_{0|0} - X_0\right)\left(X_{0|0} - X_0\right)' = P_{0|0}$$

We further assume that $X_{0|0}$ is uncorrelated with the shock processes $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$.

3.1. **A Recursive Derivation using the Gaussian Errors Assumption.** First, we will follow a method of deriving the filter that has been described as "simple and direct but to some degree uninspiring".[1]

First, note that given the assumption above, X_1 and Z_1 are conditionally joint normally distributed random variables

$$\begin{bmatrix} X_1 \\ Z_1 \end{bmatrix} \sim N\left(\begin{bmatrix} A_0 X_{0|0} \\ D_1 A_0 X_{0|0} \end{bmatrix}, \begin{bmatrix} P_{1|0} & P'_{1|0} D'_1 \\ D_1 P_{1|0} & D_1 P'_{1|0} D'_1 + \Sigma_{vv,1} \end{bmatrix}\right) \tag{3.3}$$

where

$$P_{1|0} = A P_{0|0} A' + C_1 C'_1 \tag{3.4}$$

the entries in the covariance matrix (3.3) can be found by evaluating $E\left(X_1 - A_0 X_{0|0}\right)\left(X_1 - A_0 X_{0|0}\right)'$, $E\left(Z_1 - D_1 A_0 X_{0|0}\right)\left(Z_1 - D_1 A_0 X_{0|0}\right)'$ and $E\left(Z_1 - D_t A_t X_{0|0}\right)\left(X_1 - A_t X_{0|0}\right)'$ conditional on $X_{0|0}$ and $P_{0|0}$. We know that an conditional minimum variance estimate of jointly normally distributed variables are given by

$$E\left[X_1 \mid Z_1\right] = A_0 X_{0|0} + P'_{1|0} D_1 \left(D_1 P'_{1|0} D'_1 + \Sigma_{vv}\right)^{-1}\left(Z_1 - D_1 A_0 X_{0|0}\right) \tag{3.5}$$

[1]Anderson and Moore (2005).

so that K_1 must be

$$K_1 = P'_{1|0} D_1 \left(D_1 P'_{1|0} D'_1 + \Sigma_{vv,1} \right)^{-1}$$

The posterior estimate $X_{1|1}$ has the conditional covariance

$$\begin{aligned} & E\left(E\left[X_1 \mid Z_1, X_{0|0}\right] - X_1\right)\left(E\left[X_1 \mid Z_1, X_{0|0}\right] - X_1\right)' && (3.6) \\ = \; & P_{1|0} - P'_{1|0} D'_1 \left(D_1 P'_{1|0} D'_1 + \Sigma_{vv} \right)^{-1} D_1 P_{1|0} && (3.7) \\ = \; & P_{1|1} && (3.8) \end{aligned}$$

To find the prior error covariance for period 2, we can propagate (3.8) forward

$$P_{2|1} = A_1 P_{1|1} A'_1 + CC' \qquad (3.9)$$

or

$$P_{2|1} = A_1 \left(P_{1|0} - P_{1|0} D'_1 \left(D_1 P_{1|0} D'_1 + \Sigma_{vv} \right)^{-1} D_1 P_{1|0} \right) A'_1 + C_2 C'_2 \qquad (3.10)$$

Equation (3.10) can be iterated forward in order to compute the prior error covariances for any period. Note again that the error covariances are independent of the observations, and could in principle be computed before the filter is run. For arbitrary t, we then have the main formulas needed in order to run the filter

$$\begin{aligned} P_{t|t-1} &= A_{t-1} \left(P_{t-1|t-2} - P_{t-1|t-2} D'_{t-1} \left(D_{t-1} P_{t-1|t-2} D'_{t-1} + \Sigma_{vv,t-1} \right)^{-1} D_{t-1} P_{t-1|t-2} \right) A'_1 \\ & \quad + C_t C'_t && (3.11) \\ K_t &= P_{t|t-1} D'_t \left(D_t P_{t|t-1} D'_t + \Sigma_{vv,t} \right)^{-1} && (3.12) \end{aligned}$$

3.2. **A Gram-Schmidt Orthogonalization Approach.** An alternative way to derive the Kalman filter is to use a Gram-Schmidt orthogonalization of the sequence of observable variables $\{Z_t\}_{t=1}^{T}$. This approach uses the projection theorem directly and also makes it clearer why the recursive form of the Kalman filter is equivalent to re-estimating the unobservable state in each period using the complete history of observations.

As before, we are given an initial prior about the state in period 0 which we will denote $X_{0|0}$ with variance $P_{0|0}$. Just like before, we also want to find a filter of the form

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1} \right) \tag{3.13}$$

that is optimal in some sense. Above, where we assumed Gaussian errors, we used results about jointly distributed Gaussian variables to derive the filter and under the Gaussian assumption, the Kalman filter is the minimum variance estimator of the unobservable state. In this section we drop the Gaussian assumption and show that the Kalman filter is the best *linear* estimator of X_t (in the minimum variance sense) regardless of the distribution of the errors.

3.2.1. *A Linear Minimum Variance Estimator.* We first state the general period t problem:

$$\min_{\alpha} E \left[X_t - \sum_{j=1}^{t} \alpha_j Z_{t-j+1} - \alpha_0 X_{0|0} \right] \left[X_t - \sum_{j=1}^{t} \alpha_j Z_{t-j+1} - \alpha_0 X_{0|0} \right]' \tag{3.14}$$

Our task is to find the coefficients (the αs) in (3.14). That is, we want to find the linear projection of X_t on the history of observables $Z_t, Z_{t-1}, ...Z_1$ and our prior $X_{0|0}$. From the projection theorem, we know that this will be given by the linear combination $\sum_{j=1}^{t} \alpha_j Z_{t-j+1} - \alpha_0 X_{0|0}$ that gives errors that are orthogonal to $Z_t, Z_{t-1}, ...Z_1$ and our prior $X_{0|0}$, that is, find the αs so that

$$\left(X_t - \sum_{j=1}^{t} \alpha_j Z_{t-j+1} - \alpha_0 X_{0|0} \right) \perp \{Z_j\}_{j=1}^{t} \tag{3.15}$$

and

$$\left(X_t - \sum_{j=1}^{t} \alpha_j Z_{t-j+1} - \alpha_0 X_{0|0} \right) \perp X_{0|0} \tag{3.16}$$

hold. We could of course just compute this directly as

$$\begin{aligned} & P\left(X_t \mid Z_t, Z_{t-1}, ...Z_1, X_{0|0}\right) && (3.17) \\ = \; & EX_t \left[Z_t' \; Z_{t-1}' \quad Z_1' \;\; X_{0|0}'\right]' \times \\ & \left(E\left[Z_t' \; Z_{t-1}' \quad Z_1' \;\; X_{0|0}'\right]\left[Z_t' \; Z_{t-1}' \quad Z_1' \;\; X_{0|0}'\right]'\right)^{-1}\left[Z_t' \; Z_{t-1}' \quad Z_1' \;\; X_{0|0}'\right]' \end{aligned}$$

However, this would not be very convenient as we would need compute the covariances of vectors that increase in size as time passes and the dimension of the history of observations grows. Instead, we will use a Gram-Schmidt orthogonalization and a result about projections on uncorrelated variables to derive the recursive form (3.13) of the filter. Hopefully, it will also be clear why the recursive formulation is equivalent to evaluating the expression (3.17).

3.2.2. *Gram-Schmidt Orthogonalization in* $\mathbb{R}^m$. The columns of a matrix can be orthogonalized using the so called Gram-Schmidt procedure. It consists of simply taking the first column as it is, subtracting the projection of the second column on the first from the second column and use the result as the "new" second column. The projection of the third column on the first two are then subtracted from the third column and the result replaces the third column and so on.

Let the matrix $Y \;\; (m \times n)$ have columns $\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_n$.

$$Y = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix} \tag{3.18}$$

We want to construct a new matrix $\widetilde{Y}$ with the same column space as Y that has orthogonal columns so that $\widetilde{Y}\widetilde{Y}'$ is a diagonal matrix. The first column can be chosen arbitrarily so we might as well keep the first column of Y as it is. The second column should be orthogonal to the first. One way of doing this is to subtract the projection of $\mathbf{y}_2$ on $\mathbf{y}_1$ from $\mathbf{y}_2$ and define a new column vector $\widetilde{\mathbf{y}}_2$

$$\widetilde{\mathbf{y}}_2 = \mathbf{y}_2 - \mathbf{y}_1\left(\mathbf{y}_1'\mathbf{y}_1\right)^{-1}\mathbf{y}_1'\mathbf{y}_2 \tag{3.19}$$

or

$$\widetilde{\mathbf{y}}_2 = \left(I - \mathcal{P}_{y_1}\right)\mathbf{y}_2 \tag{3.20}$$

and then subtract the projection of $\mathbf{y}_3$ on $[\mathbf{y}_1 \;\; \mathbf{y}_2]$ from $\mathbf{y}_3$ to construct $\widetilde{\mathbf{y}}_3$ and so on. Since we can reconstruct the original matrix using a linear combinations of the orthogonal columns we know that the two matrices share the same column space. Below, we will use the same procedure but with the inner product defined as $E(XY')$ rather than $\sum_{i=1}^{n} x_i y_i$ to construct the *innovations sequence* that span the same space as the complete history of observable variables.

3.2.3. *Projections on uncorrelated variables.* The usefulness of the Gram-Schmidt orthogonalization process for deriving the recursive form of the Kalman filter stems from the following result about projections on uncorrelated variables.

Let $\left\{\widetilde{Z}_j\right\}_{j=1}^{t}$ denote a sequence of uncorrelated mean zero variables so that

$$E\left[\widetilde{Z}_t \widetilde{Z}_{t-s}\right] = 0 : s \neq 0 \tag{3.21}$$

then

$$P\left(X_t | \widetilde{Z}_t, \widetilde{Z}_{t-1}, ...\widetilde{Z}_1\right) = \mathcal{P}\left(X_t | \widetilde{Z}_t\right) + \mathcal{P}\left(X_t | \widetilde{Z}_{t-1}\right) + ... + \mathcal{P}\left(X_t | \widetilde{Z}_1\right) \tag{3.22}$$

(To see why, just write out the projection formula. If the variables that we project on are orthogonal, the inverse will be taken of a diagonal matrix.)

3.2.4. *A recursive formulation via projections on the innovation sequence.* As before, we will start from the first period problem of how to optimally combine the information in the prior $X_{0|0}$ and the signal Z_1. Use that

$$Z_1 = D_1 A_0 X_0 + D_1 C \mathbf{u}_1 + \mathbf{v}_1 \tag{3.23}$$

and that we know that $\mathbf{u}_t$ and $\mathbf{v}_t$ are orthogonal to $X_{0|0}$ to first find the optimal projection of Z_1 on $X_{0|0}$

$$Z_{1|0} = D_1 A_0 X_{0|0} \tag{3.24}$$

We can then define the period 1 innovation $\widetilde{Z}_1$ in Z_1 as

$$\widetilde{Z}_1 = Z_1 - Z_{1|0} \tag{3.25}$$

. By result (3.22) above we know that

$$\mathcal{P}\left(X_1|\widetilde{Z}_1, X_{0|0}\right) = \mathcal{P}\left(X_1|\widetilde{Z}_1\right) + \mathcal{P}\left(X_1|X_{0|0}\right) \tag{3.26}$$

since $\widetilde{Z}_1 \perp X_{0|0}$ and $\mathcal{P}\left(X_1|X_{0|0}\right) = D_1 A_0 X_{0|0}$. From the projection theorem, we know that we should look for a K_1 such that the inner product of the projection error and $\widetilde{Z}_1$ are zero

$$\left\langle X_1 - K_1\widetilde{Z}_1, \widetilde{Z}_1\right\rangle = 0 \tag{3.27}$$

Defining the inner product $\langle X, Y\rangle$ as $E\left(XY'\right)$ we get

$$E\left[\left(X_1 - K_1\widetilde{Z}_1\right)\widetilde{Z}_1'\right] = 0 \tag{3.28}$$

$$E\left[X_1\widetilde{Z}_1'\right] - K_1 E\left[\widetilde{Z}_1\widetilde{Z}_1'\right] = 0 \tag{3.29}$$

$$K_1 = E\left[X_1\widetilde{Z}_1'\right]\left(E\left[\widetilde{Z}_1\widetilde{Z}_1'\right]\right)^{-1} \tag{3.30}$$

We thus need to evaluate the two expectanal expressions in (). Before doing so it helps to define the state innovation

$$\widetilde{X}_1 = X_1 - X_{1|0} \tag{3.31}$$

that is, $\widetilde{X}_1$ is the one period error. The first expectation factor of K_1 in (3.27) can now be manipulated in the following way

$$E\left[X_1\widetilde{Z}_1'\right] = E\left(\widetilde{X}_1 + X_{1|0}\right)\widetilde{Z}_1' \tag{3.32}$$

$$= E\widetilde{X}_1\widetilde{Z}_1' \tag{3.33}$$

$$= E\widetilde{X}_1\left(\widetilde{X}_1'D' + E\mathbf{v}_t'\right) \tag{3.34}$$

$$= P_{1|0}D' \tag{3.35}$$

where the first equality uses the definition (3.31) the second uses that the innovation is orthogonal to $X_{1|0}$, the third equality uses the definition of the innovation (3.27). The last line uses the definition of the prior error covariance and that and that $E\left(\widetilde{X}_1\mathbf{v}_t'\right) = 0$ since $E\left(X_1\mathbf{v}_t'\right) = 0$ and $E\left(X_{1|0}\mathbf{v}_t'\right) = 0$.

Evaluating the second expectation factor

$$E\left[\widetilde{Z}_1\widetilde{Z}_1'\right] = E\left[\left(D_1\widetilde{X}_1 + \mathbf{v}_t\right)\left(D_1\widetilde{X}_1 + \mathbf{v}_t\right)'\right] \tag{3.36}$$

$$= D_1P_{1|0}D_1' + \Sigma_{vv} \tag{3.37}$$

together with (3.35) gives us what we need for the formula for K_1

$$K_1 = P_{1|0}D_1'\left(D_1P_{1|0}D_1' + \Sigma_{vv}\right)^{-1} \tag{3.38}$$

where we know that $P_{1|0} = A_0P_{0|0}A_0' + C_0C_0'$. We can add the projections of X_1 on $\widetilde{Z}_1$ and $X_{0|0}$ to get our linear minimum variance estimate $X_{1|1}$

$$X_{1|1} = \mathcal{P}\left(X_1|X_{0|0}\right) + \mathcal{P}\left(X_t|\widetilde{Z}_1\right) \tag{3.39}$$

$$= A_0X_{0|0} + K_1\widetilde{Z}_1 \tag{3.40}$$

Again, we can use the estimate $X_{1|1}$ to form a projection of X_2 on $X_{0|0}$ and $\widetilde{Z}_1$.

$$X_{2|1} = A_1 X_{1|1}$$

To see that projecting on the prior and the innovation is equivalent to projecting on the prior and the actual observation note that

$$\begin{aligned} X_{2|1} &= A_1 A_0 X_{0|0} + A_1 K_1 \widetilde{Z}_1 && (3.41) \\ &= A_1 A_0 X_{0|0} + A_1 K_1 \left(Z_1 - Z_{1|0} \right) && (3.42) \\ &= A_1 \left(A_0 - D A_0 \right) X_{0|0} + A_1 K_1 Z_1 && (3.43) \end{aligned}$$

Then, since we can always back out the history of observations as a linear combination of innovations, we know that the sequence of innovations and observable span the same space. (Of course, we also need to be able to get the innovations sequence as a linear combination of the observables sequence.) To find the period innovation $\widetilde{Z}_2$, that is the component of Z_2 that is orthogonal to $Z_{2|1}$, we again subtract the predicted component from the actual

$$\widetilde{Z}_2 = Z_2 - D_2 A_1 X_{1|1} \tag{3.44}$$

and the period 2 posterior estimate of X_2 is then given by

$$X_{2|2} = A_1 X_{1|1} + K_2 \widetilde{Z}_2 \tag{3.45}$$

with

$$K_2 = P_{2|1} D_1' \left(D_1 P_{2|1} D_1' + \Sigma_{vv} \right)^{-1} \tag{3.46}$$

However, we have not yet derived an expression for the prior covariance matrices $P_{2|1}$. We can perhaps go directly to deriving a formula for $P_{t|t-1}$ without loosing anyone. We start by finding an expression for $P_{t|t}$. We can rewrite

$$X_{t|t} = K_t \widetilde{Z}_t + X_{t|t-1} \tag{3.47}$$

as

$$X_t - X_{t|t} + K_t \widetilde{Z}_t = X_t - X_{t|t-1} \tag{3.48}$$

by adding X_t to both sides and rearranging. Since the period t error $X_t - X_{t|t}$ is orthogonal to $\widetilde{Z}_t$ the variance of the right hand side must be equal to the sum of the variances of the terms on the left hand side. We thus have

$$P_{t|t} + K_t \left(DP_{t|t-1}D' + \Sigma_{vv} \right) K_t' = P_{t|t-1} \tag{3.49}$$

or by rearranging

$$\begin{aligned}
P_{t|t} &= P_{t|t-1} - K_t \left(DP_{t|t-1}D' + \Sigma_{vv} \right) K_t' && (3.50) \\
&= P_{t|t-1} - P_{t|t-1}D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right) && (3.51) \\
&\quad \times \left(P_{t|t-1} D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} \right)' && (3.52) \\
&= P_{t|t-1} - P_{t|t-1} D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} D_t P_{t|t-1} && (3.53)
\end{aligned}$$

It is then straight forward to show that

$$\begin{aligned}
P_{t+1|t} &= A_t P_{t|t} A_t' + CC' && (3.54) \\
&= A_t' \left(P_{t|t-1} - P_{t|t-1} D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} D_t P_{t|t-1} \right) A_t' + CC' && (3.55)
\end{aligned}$$

3.3. **Summing up.** For the state space system

$$\begin{aligned}
X_t &= A_t X_{t-1} + C_t \mathbf{u}_t && (3.56) \\
Z_t &= D_t X_t + \mathbf{v}_t && (3.57) \\
\begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix} &\sim N \left(\mathbf{0}, \begin{bmatrix} I_n & \mathbf{0}_{n \times l} \\ \mathbf{0}_{l \times n} & \Sigma_{vv} \end{bmatrix} \right) && (3.58)
\end{aligned}$$

we get the state estimate update equation

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1} \right) \tag{3.59}$$

The recursive formulas for the Kalman gain

$$K_t = P_{t|t-1} D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} \tag{3.60}$$

and the prior error covariance

$$\begin{aligned} P_{t+1|t} &= A_t \left(P_{t|t-1} - P_{t|t-1} D_{t1}' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} D_t P_{t|t-1} \right) A_t' \\ &\quad + C_{t+1} C_{t+1}' \end{aligned} \tag{3.61}$$

The innovation sequence can be computed recursively from the innovation representation

$$\widetilde{Z}_t = Z_t - D_t X_{t|t-1} \tag{3.62}$$

$$X_{t+1|t} = A_{t-1} X_{t|t-1} + A_{t-1} K_t \widetilde{Z}_t \tag{3.63}$$

4. The Time Invariant Filter

Under some conditions, the Kalman filter will be invariant, by which we mean that the Kalman gain matrix K_t is time invariant. Since K_t is a function of the prior error covariance matrix $P_{t|t-1}$, time invariance of K_t requires time invariance of $P_{t|t-1}$. A necessary but not sufficient condition for time invariance of the filter is that the associated state space system is time invariant. That is, the matrices A_t, C_t, D_t and the measurement error covariance Σ_{vv} should not depend on t. In this section we will therefore suppress the time subscripts on the coefficient matrices. However, time invariant matrices in the state space system is not enough. We also need to restrict the initial error covariance $P_{1|0}$ to be the solution to the

Riccati equation

$$P = A'\left(P - PD'\left(DPD' + \Sigma_{vv}\right)^{-1} DP\right) A' + CC' \tag{4.1}$$

We thus need to show that a solution exists which we will show by demonstrating that iterating on (3.61) yields a convergent sequence. As before, we can argue heuristically that we know that P is bounded from both above and below by

$$CC' \leq P \leq \Sigma_{XX} \tag{4.2}$$

where $\Sigma_{XX} = E\left(X_t - \mu_x\right)\left(X_t - \mu_x\right)'$. The lower bound come from the fact that even when the current state is known with certainty, that is $P_{t|t} = \mathbf{0}$, the future values of the state are still uncertain as long as there are non-zero variance innovations hitting the state. The upper bound comes from that we could always choose $K = 0$ and the variance of the estimates would then be the same as the unconditional variance of the state X_t. To show that the limit of iterating on (3.61) exists, we will show that for arbitrary initial $P_{1|0}$, the sequence $\left\{P_{t|t-1}\right\}_{t=1}^{\infty}$ is either monotonically increasing or decreasing. We do this by relying on the boundedness of P and a monotonicity argument about the sequence $\left\{P_{t|t-1}\right\}_{t=1}^{\infty}$.

4.0.1. *Monotonicity of* $\left\{P_{t|t-1}\right\}_{t=1}^{\infty}$. A sequence of covariance matrices $P_{t|t-1}$are said to be increasing monotonically if

$$P_{1|0} \leq P_{2|1} \leq ... \leq P_{t|t-1} : \forall \;\; t \tag{4.3}$$

where the larger than or equal sign means that $P_{t+1|t} - P_{t|t-1}$ is a positive semi definite matrix. A monotonically decreasing sequence is defined analogously, but with the larger than or equal signs replaced by less than or equal signs.

We start by defining the prior state error $\widetilde{X}_t$

$$\widetilde{X}_t = X_t - X_{t|t-1} \tag{4.4}$$

We then have

$$\widetilde{X}_{t+1} = (A - AKD)\, \widetilde{X}_t - K_t \mathbf{v}_{t+1} + C\mathbf{u}_{t+1} \tag{4.5}$$

To see why, just plug the definition (4.4) into (4.5)

$$X_{t+1} - X_{t+1|t} = (A - AK_tD)\left(X_t - X_{t|t-1}\right) - AK_t\mathbf{v}_{t+1} + C\mathbf{u}_{t+1} \tag{4.6}$$

The variance of (4.5) can then be written as

$$P_{t+1|t} = (A - AK_tD)\, P_{t|t-1}\,(A - AK_tD_t)' + AK_t\mathbf{\Sigma}_{vv}K_t'A' + CC' \tag{4.7}$$

which we can use to demonstrate monotonicity of $\left\{P_{t|t-1}\right\}_{t=1}^{\infty}$.

We will define two sequences of $\left\{P_{t|t-1}\right\}_{t=1}^{\infty}$ and $\left\{\widehat{P}_{t|t-1}\right\}_{t=1}^{\infty}$ which only differ in the values of their initial conditions. We also have the associated Kalman gain sequences $\{K_t\}_{t=1}^{\infty}$ and $\left\{\widehat{K}_t\right\}_{t=1}^{\infty}$. We have that

$$P_{0|-1} = 0 \tag{4.8}$$

$$\widehat{P}_{1|0} = 0 \tag{4.9}$$

so we know that for any t

$$P_{t|t-1} = \widehat{P}_{t+1|t} \tag{4.10}$$

To prove that $\left\{\widehat{P}_{t|t-1}\right\}_{t=1}^{\infty}$ is monotonically increasing we will rely on the optimality properties of the filter:

$$\begin{aligned} P_{t+1|t} &= \min_{K_1^*}\left[(A - AK_t^*D)\, P_{t|t-1}\,(A - AK_t^*D)' + AK_t^*\mathbf{\Sigma}_{vv}K_t^{*\prime}A' + CC'\right] && (4.11)\\ &= \left[(A - AK_tD)\, P_{t|t-1}\,(A - AK_tD)' + AK_t\mathbf{\Sigma}_{vv}K_t'A' + CC'\right] && (4.12)\end{aligned}$$

where K_0 is the Kalman gain that minimizes the prior error covariance. We thus have

$$\begin{aligned}
P_{t+1|t} &= (A - AK_tD)\, P_{t|t-1}\, (A - AK_tD)' + AK_t\mathbf{\Sigma}_{vv}K_t'A' + CC' && (4.13)\\
&\geq (A - AK_tD)\, \widehat{P}_{t|t-1}\, (A - AK_tD)' + AK_t\mathbf{\Sigma}_{vv}K_t'A' + CC' && (4.14)\\
&\geq \min_{K_1^*}\left[(A - AK_0^*D)\, \widehat{P}_{0|-1}\, (A - AK_0^*D)' + AK_0^*\mathbf{\Sigma}_{vv}K_0^{*\prime}A' + CC'\right] && (4.15)\\
&= \left(A - A\widehat{K}_0D\right) \widehat{P}_{0|-1} \left(A - A\widehat{K}_0D\right)' + A\widehat{K}_0\mathbf{\Sigma}_{vv}'\widehat{K}_0'A' + CC' && (4.16)\\
&= \widehat{P}_{t+1|t} && (4.17)
\end{aligned}$$

so for arbitrary t we have that

$$P_{t+1|t} \geq \widehat{P}_{t+1|t}$$

but since $P_{t|t-1} = \widehat{P}_{t+1|t}$ we know that $\widehat{P}_{t|t-1} \leq \widehat{P}_{t+1|t}$, that is, $\left\{\widehat{P}_{t|t-1}\right\}_{t=1}^{\infty}$ is an increasing sequence starting from $\widehat{P}_{1|0} = 0$. Together with boundedness, this implies that the limit exists. A symmetric argument can be made for monotonically decreasing sequence starting with some upper bound as initial condition.