8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.2, June 6-8, 2007, Cancún, Mexico

INCREMENTAL NONLINEAR PARAMETER ESTIMATION IN DYNAMIC SYSTEMS

Heinz A. Preisig

Dept of Chemical Engineering, NTNU, 7491 Trondheim, Norway

Abstract: A Bayes-theorem-based interpretation of the Kalman filter is used to define the best choice of parameters in a nonlinear-in-parameters model, which is dynamic and linear in input and state.
Copyright ©2007 IFAC

Keywords: Nonlinear recursive parameter estimation, modelling

1. ORIGIN OF THIS WORK

The following text stood at the beginning of this project:

Parameter estimation by non-linear regression can often be a difficult and time consuming process, particularly in the early stages of an investigation when relatively little data is available and parameter estimates may be highly correlated.

However, in these circumstances, which may arise for example in the early stages of process development, it may be more important to have a general impression of the feasible range of parameters rather than a precise location of the best value. If the development has some further objective in mind it is important to know how sensitive the value of this objective is to changes in parameter values within the feasible region.

A method should be developed for displaying on a grid of values of the parameters, regions in which the parameters are likely to occur on the basis of current data. This display should be updated as new experimental information becomes available. It should be possible to display on the grid the values of an objective function which depends upon the parameters being estimated.

When a simple grid representation has been obtained, the possibility of improvements may be considered by e.g. automatically modifying the grid or applying interpolation procedures. [1]

This text stimulated looking into finding a recursive procedure for non-linear parameter identification. The procedure should be robust and apply to dynamic plants, as the group was deeply involved in batch process research. The text points towards process design, but the problem is certainly at least equally relevant to modelling for process control and may also extend to secondary applications like adaptive control.

The problem should be viewed from a probabilistic point of view: how does the probability distribution (objective function) change as more information becomes available. Besides these requirements, an additional one was added, namely the ability of the procedure to handle stochastic inputs, as it was observed that the input to the batch processes we were looking at displayed such characteristics. The nonlinearity in the parameter space was a must, as almost all models exhibit this property whilst linearity in state and input seemed little of a limitation.

[1] late David W T Rippin (internal report 1977)

1.1 Framework

The setting stimulated the use of a probabilistic based method that can handle input variations as well as measurement variations and is of recursive nature. This settles the choice quite obviously to be a Kalman filter and since one deals with an experiment to experiment kind of environment and/or experimental data taken by a data acquisition system, the filter would operate in the discrete time domain. Since the Kalman filter can also be derived from probabilistic arguments utilizing the Bias theorem, all one needs to do is to extract this information from the filter. The problem statement also states that the parameter space may be discretised. Though that may not be a requirement, it may be an essential feature for implementing such theory in the form of a computing procedure.

The project was executed in the seventies and a recent literature search revealed that the subject was essentially untouched since that time when one slowly discovered the richness of such problems. The theories that were consequently developed found - for various reasons - only a few applications and most of them were not particularly successful as too little experience with the techniques were accumulated. This, in turn, had the effect that the application domain pulled back. Consequently also implementation and theory developments stopped and the body of knowledge necessary to transfer the theoretical results into industry did not accumulate sufficiently. From today's point of view these failures seem not enough of a good reason to throw away all the knowledge that was accumulated in this field during this period, which probably is the main motivation to write this paper and bring this nearly forgotten subject up to the surface again.

2. THE APPROACH

The problem formulation asks for an iterative updating of parameters and a probabilistic criterion for selecting the "best" parameters most likely in a discretised parameter space. The latter induces the thought of using an estimator that is based on the Bias theorem. It was also thought that stochastic variations in the input of the plant are of importance, as mentioned, which lead to look into Kalman filtering. The idea follows largely a publication by Lainiotis (Lainiotis, 1976).

2.1 Set-up

- **Model :** The dynamic part of the model is:

$$\underline{\mathbf{x}}(k+1) := \underline{\underline{\mathbf{F}}}(k+1,k,\underline{\theta})\,\underline{\mathbf{x}}(k) + \\ + \underline{\underline{\mathbf{G}}}(k+1,k,\underline{\theta})\,\underline{\mathbf{w}}(k)\,.$$

The state is thereby nonlinearly affected by the defined parameter vector $\underline{\theta}$ and the white input noise is coloured by the matrix $\underline{\underline{\mathbf{G}}}$.

The static part, representing the link between the state and the measurement (output):

$$\underline{\mathbf{y}}(k) := \underline{\underline{\mathbf{H}}}(k,\underline{\theta})\,\underline{\mathbf{x}}(k) + \underline{\mathbf{v}}(k); \quad \theta \in \Theta\,,$$

is again nonlinearly dependent on the parameters, whilst a simple random error is added.

- **Required properties :** We request the two stochastic inputs $\underline{\mathbf{w}}$ and $\underline{\mathbf{v}}$ to be independent and Gauss-normal distributed:

$$\mathbf{E}\left[\underline{\mathbf{w}}(k)\right] := 0; \quad \mathbf{E}\left[\underline{\mathbf{w}}(k)\underline{\mathbf{w}}(k)^T\right] := \underline{\underline{\mathbf{Q}}}(k,\underline{\theta})\delta_{k,j}\,,$$
$$\mathbf{E}\left[\underline{\mathbf{v}}(k)\right] := 0; \quad \mathbf{E}\left[\underline{\mathbf{v}}(k)\underline{\mathbf{v}}(k)^T\right] := \underline{\underline{\mathbf{R}}}(k,\underline{\theta})\delta_{k,j}\,.$$

The variance-covariance matrices may be a function of the parameters too. The two signals shall be independent of each other and of the initial state

$$\mathbf{E}\left[\{\underline{\mathbf{w}}(k) - \mathbf{E}\left[\underline{\mathbf{w}}(k)\right]\}\,\underline{\mathbf{v}}^T(k)\right] := 0\,,$$
$$\mathbf{E}\left[\{\underline{\mathbf{x}}(0) - \mathbf{E}\left[\underline{\mathbf{x}}(0)\right]\}\,\underline{\mathbf{v}}^T(k)\right] := 0\,.$$

- **Initial conditions :** For the initial conditions, the initial state is Gauss normal distributed with a mean $\hat{\underline{\mathbf{x}}}(0|0,\underline{\theta})$ and the variance $\underline{\underline{\mathbf{P}}}(0|0,\underline{\theta})$.

- **Measurements :** The procedure will use a sequence of measurements $\lambda_k := \{\underline{\mathbf{y}}(0),\ldots,\underline{\mathbf{y}}(k)\}$ and the parameters are in the space Θ.

- **Parameter estimate :** In the continuous domain the parameters are weighted over the defined domain:

$$\hat{\underline{\theta}} := \int_{\Theta} \underline{\theta}\;\underline{\mathbf{p}}_k(\underline{\theta}|\lambda_k)\;\mathrm{d}\underline{\theta}\,.$$

With $\underline{\mathbf{p}}_k(\underline{\theta}|\lambda_k)$ being the conditional probability density function of the parameters given the observations λ_k.

- **State estimate :** is simply the expectation after having "seen" a measurement history λ_k

$$\hat{\underline{\mathbf{x}}}(k|k,\underline{\theta}) := \mathbf{E}\left[\underline{\mathbf{x}}(k|\lambda_k,\underline{\theta}\right]\,.$$

- **Kalman filter :** Based on these definitions and assumptions, the Kalman filter can be derived ((Jazwinski, 1970)):

$$\hat{\underline{\mathbf{x}}}(k|k,\underline{\theta}) := \hat{\underline{\mathbf{x}}}(k|k-1,\underline{\theta}) + \underline{\underline{\mathbf{K}}}(k,\underline{\theta})\,\hat{\underline{\mathbf{e}}}(k|k-1,\underline{\theta})\,,$$
$$\underline{\underline{\mathbf{P}}}(k|k,\underline{\theta}) := \left[\underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{K}}}(k,\underline{\theta})\,\underline{\underline{\mathbf{H}}}(k,\underline{\theta})\right]\,\underline{\underline{\mathbf{P}}}(k|k-1,\underline{\theta})$$

with:

$$
\begin{aligned}
\hat{\underline{\mathbf{x}}}(k|k-1,\underline{\theta}) &:= \underline{\underline{\mathbf{F}}}(k,k-1,\underline{\theta})\,\hat{\underline{\mathbf{x}}}(k-1|k-1,\underline{\theta})\,,\\
\underline{\mathbf{e}}(k|k-1,\underline{\theta}) &:= \underline{\mathbf{y}}(k) - \underline{\underline{\mathbf{H}}}(k,\underline{\theta})\,\underline{\mathbf{x}}(k,k-1,\underline{\theta})\\
\underline{\underline{\mathbf{K}}}(k,\underline{\theta}) &:= \underline{\underline{\mathbf{P}}}(k|k-1,\underline{\theta})\,\underline{\underline{\mathbf{H}}}(k|k-1,\underline{\theta})\\
&\quad\ \underline{\underline{\mathbf{L}}}(k|k-1,\underline{\theta})\,,\\
\underline{\underline{\mathbf{L}}}(k|k-1,\underline{\theta}) &:= [\underline{\underline{\mathbf{H}}}(k,\underline{\theta})\,\underline{\underline{\mathbf{P}}}(k|k-1,\underline{\theta})\,\underline{\underline{\mathbf{H}}}(k,\underline{\theta}) +\\
&\quad\ + \underline{\underline{\mathbf{R}}}(k,\underline{\theta})]^{-1}\,,\\
\underline{\underline{\mathbf{P}}}(k|k-1,\underline{\theta}) &:= \underline{\underline{\mathbf{F}}}(k,k-1,\underline{\theta})\,\underline{\underline{\mathbf{P}}}(k-1|k-1,\underline{\theta})\\
&\quad\ \underline{\underline{\mathbf{F}}}^{T}(k,k-1,\underline{\theta}) +\\
&\quad\ + \underline{\underline{\mathbf{G}}}(k,k-1,\underline{\theta})\,\underline{\underline{\mathbf{Q}}}(k-1,\underline{\theta})\\
&\quad\ \underline{\underline{\mathbf{G}}}^{T}(k,k-1,\underline{\theta})\,.
\end{aligned}
$$

The first of the above block of equation is simply the state propagation according to the model, namely an autonomous process, in this case. The second equation gives the output error, namely the difference between the measurements and the predicted output. The $\underline{\underline{\mathbf{K}}}$ is the Kalman gain, the gain in the loop feeding the error back into the estimator for forcing the state to converge. The $\underline{\underline{\mathbf{L}}}^{-1}$ matrix is the cumulative variance in the error and, whilst it is usually not lifted out as a special feature of the Kalman filter, it is essential for our purpose as it is the basis for the "confidence" measure later being derived. Finally the variance in the state estimate is captured in $\underline{\underline{\mathbf{P}}}$.

- **Bayes Theorem :**

$$
\begin{aligned}
p(\underline{\theta}|\lambda_k) &:= \frac{M(k|k-1|\underline{\theta})\,p(\underline{\theta}|\lambda_{k-1})}{\int_\Theta M(k|k-1|\underline{\theta})\,p(\underline{\theta}|\lambda_{k-1})\,d\underline{\theta}}\,,\\
M(k|k-1|\underline{\theta}) &:= |\underline{\underline{\mathbf{L}}}(k|k-1,\underline{\theta}|^{0.5}\,e^{N(k)}\,,\\
N(k) &:= -\frac{1}{2}\,||\hat{\underline{\mathbf{e}}}(k|k-1,\underline{\theta})||^2_{\underline{\underline{\mathbf{L}}}(k|k-1,\underline{\theta})}\,.
\end{aligned}
$$

The Bayes theorem forms a recursive structure which updates the probability of the parameters in the parameter space, which is exactly what was asked for.

- **Quantification :** Since $p(\underline{\theta}|\lambda_{k-1})$ is infinite dimensional, a discrete approximation is introduced (Lainiotis and Deshpande, 1974):

$$
p(\underline{\theta}|\lambda_{k-1}) := \sum_{i:=1}^{m} p_i(\underline{\theta}|\lambda_{k-1})\,\delta(\underline{\theta}-\underline{\theta}_i)\,.
$$

This yields a discrete representation of the Bayes theorem:

$$
p_i(\underline{\theta}_i,\lambda_{k-1}) := \frac{M(k|k-1|\underline{\theta}_i)\,p_i(\underline{\theta}_i,\lambda_{k-1})}{\sum_i^m M(k|k-1|\underline{\theta}_i)\,p_i(\underline{\theta}_i,\lambda_{k-1})}\,.
$$

The derivation of Bayes theorem is the result of applying repetitively the multiplication rule:

$$
p(A_j|B) := \frac{p(B,A_j)}{\sum_{i:=1}^m p(B,A_i)}\,.
$$

$p(B,A_j)$ is the marginal distribution $p(\underline{\theta},\underline{\mathbf{z}}_k) := M(k|k-1|\underline{\theta})$. Applying the multiplication rule repetitively:

$$
\ln p(\underline{\theta},\lambda_k) := \sum_{r:=1}^{k} \ln M(r|r-1|\underline{\theta})\,.
$$

The joint probability for the parameters is also quantified:

$$
p(\theta_{q,i}) := \sum_{i:=1}^{m} p(\theta_q,\lambda_k)\,\delta(\theta_q-\theta_{q,i})\,,
$$

with $\theta_{q,i}$ being the q-th component of the parameter vector of dimension r and the i-th quantification. The marginal distribution is then

$$
p(\theta_{a,i},\lambda_k) := \sum_{i:=1}^{m}\ \sum_{q:=1,q\neq a}^{r} p(\theta_{q,i},\lambda_k)\,,
$$

and using this in the Bayes theorem gives:

$$
p(\theta_{a,i}|\lambda_k) := \frac{p(\theta_{a,i},\lambda_k)}{\sum_{i:=1}^m p(\theta_{a,i},\lambda_k)}\,.
$$

And for the parameter:

$$
\hat{\theta}_a := \sum_{i:=1}^{m} p(\theta_{a,i}|\lambda_k)\,\theta_{a,i}\,.
$$

For the conditional probability at every grid point:

$$
p(\underline{\theta}|\lambda_k) := \frac{p(\underline{\theta},\lambda_k)}{\sum_{i:=1}^m p(\theta_{a,i},\lambda_k)}\,.
$$

3. COMPUTATIONAL EXPERIMENT

For the demonstration of this procedure, a nonlinear model was chosen that is relatively simple, namely an ISTR with two reactions running in parallel. Choosing a particular parameterisation on can cast the problem into the following form:

$$
\begin{aligned}
\underline{\mathbf{x}}(k) &:= \begin{bmatrix} \dfrac{1}{1+\theta_1} & 0 \\ 0 & \dfrac{1}{1+\theta_2} \end{bmatrix} \underline{\mathbf{x}}(k-1) + \underline{\mathbf{w}}(k-1)\,,\\
\underline{\mathbf{z}}(k) &:= \underline{\mathbf{x}}(k) + \underline{\mathbf{v}}(k)\,.
\end{aligned}
$$

For the initial conditions, parameters and the stochastic disturbances one chose:

$$
\begin{aligned}
\underline{\mathbf{x}}(0) &:= \begin{bmatrix} 1 \\ 1 \end{bmatrix},\\
\underline{\theta} &:= \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix},\\
w_i &:= \mathbf{N}(0,0.04) \quad ;i=1,2\,,\\
v_i &:= \mathbf{N}(0,0.04) \quad ;i=1,2\,.
\end{aligned}
$$

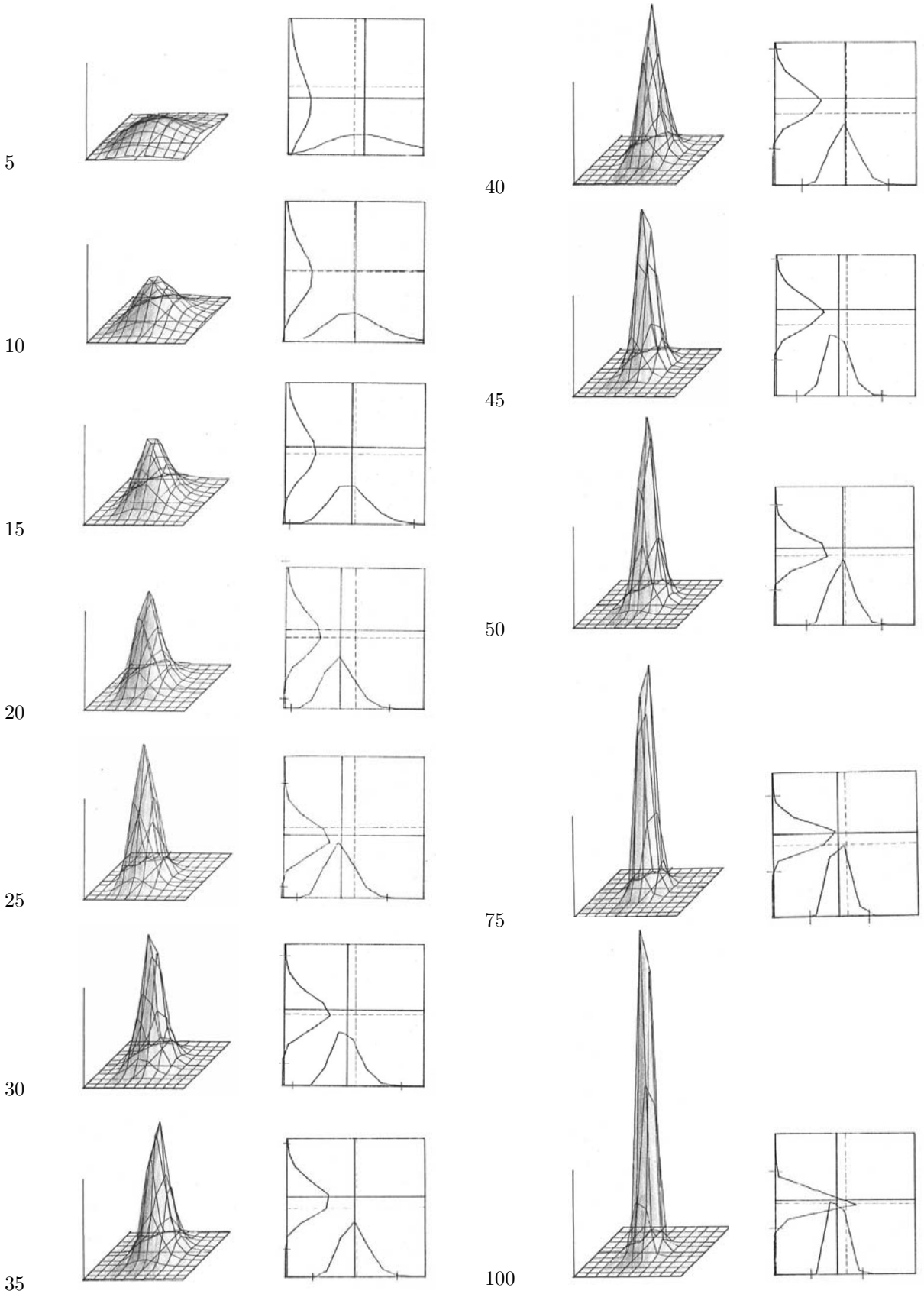

Figure 1: Shows snapshots of probability distributions in the two-parameter space and the marginal distributions for 5, 10, ..., 100 samples (explanations see figure 2).

3.1 Results

The attached figure1show snapshots of the development of the probability distribution in the parameter space as the number of experiments increases. In each column on the right the distribution is shown in a 3D-plot, on the left the marginal distribution functions for the two parameters. The currently best estimate is indicated with a full line and the nominal values are shown with dotted lines.

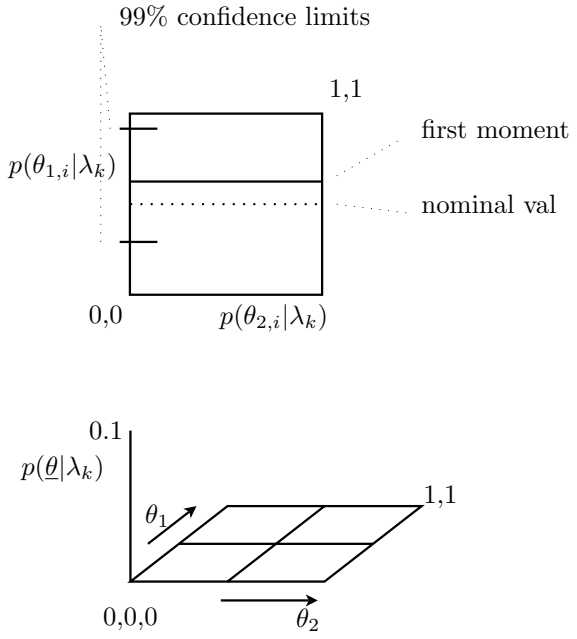

Figure 2 : Explanation to figure 1.

One can clearly see on how the probability builds up a sharper and sharper peak converging to the true parameter values. The estimate is also clearly unbiased. The original work compares these results with a simple minimal sum of squares error measure on the same discretised parameter domain. In that case the procedure builds up a bias as expected, which is due to the neglected input error , as this estimator is not un-biased.

4. CONCLUSIONS

The recursive discretised parameter estimation procedures, as they were to a large extent introduced in the 70ties, are excellently suited for recursive nonlinear parameter estimation. This procedure could indeed be used as originally suggested by David W T Rippin in the design of processes, but equally well in batch-to-batch model-based planning and control or other adaptive schemes.

5. REFERENCES

Jazwinski, A H (1970). *Stochastic Processes and Filter Theory*. Academic Press. New York.

Lainiotis, Demetrios G (1976). Partitioning: A unifying framework for adaptive systems, i: Estimation. *Proceedings of the IEEE* **64**(8), 1126–1143.

Lainiotis, Demetrios G and J G Deshpande (1974). Parameter estimimation using splines. *Information Sciences* **7**, 291–315.

6. APPENDIX

This appendix derives the variance of the output error. To ease the writing, we reduce the problem to a scalar plant and introduce the notation: 11 for $k|k,\underline{\theta}$, 10 for $k|k-1,\underline{\theta}$, 00 for $k-1|k-1,\underline{\theta}$, 1 for $k|\underline{\theta}$ and 0 for $k-1|\underline{\theta}$. For the variance of w we use q and for the variance of v we use r.

The model equations are simplified to:

$$x_1 := f_{10}\, x_0 + w_0\,, \tag{1}$$

$$y_1 := h_1\, x_1 + v_1\,. \tag{2}$$

It is to be shown that

$$\mathbf{E}\left[(y_1 - \mathbf{E}\left[y_1\right]^2)\right] := h_1^2\, p_{10} + r_1\,.$$

Proof:

$$\begin{aligned}
\mathbf{E}\left[\left(y_1 - \mathbf{E}\left[y_1\right]^2\right)\right] :=& \\
:=& \mathbf{E}\left[((h_1\, x_1 + v_1) - \mathbf{E}\left[(h_1\, x_1 + v_1)\right])^2\right]\,, \\
:=& \mathbf{E}\left[((h_1\, x_1 - \mathbf{E}\left[h_1\, x_1\right]) + (v_1 - \mathbf{E}\left[v_1\right]))^2\right]\,, \\
:=& \mathbf{E}\left[(h_1\,(x_1 - \mathbf{E}\left[x_1\right]) + v_1)^2\right]\,, \\
:=& h_1^2\, \mathbf{E}\left[(x_1 - \mathbf{E}\left[x_1\right])^2\right] + \\
& +2\, h_1\, \mathbf{E}\left[x_1 - \mathbf{E}\left[x_1\right]\right]\, \mathbf{E}\left[v_1\right] + \mathbf{E}\left[v_1^2\right]\,, \\
& h_1^2\, \mathbf{E}\left[(x_1 - \mathbf{E}\left[x_1\right])^2\right] + r_1\,.
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\left[\left(x_1 - \mathbf{E}\left[x_1\right]^2\right)\right] :=& \\
:=& \mathbf{E}\left[((f_0\, x_0 + w_0) - \mathbf{E}\left[(f_0\, x_0 + v_0)\right])^2\right]\,, \\
:=& f_0^2\, \mathbf{E}\left[(x_1 - \mathbf{E}\left[x_1\right])^2\right] + q_0\,, \\
:=& f_0^2\, p_{00} + q_0\,, \\
:=& p_{10}
\end{aligned}$$

Which, when substituted, gives the desired result.