1 Linear filtering theory

Introduction

The objective of this chapter is to derive the Kalman filter in discrete time as well as in continuous time.

There are several ways to derive these formulas, especially in discrete time where an elementary direct approach is feasible. A more systematic way to proceed is to use the connection between the optimal filter and the least square or maximum likelihood estimate. This approach is convenient in continuous time.

1.1 Filtering theory in discrete time

1.1.1 The model

We consider the following dynamic system in discrete time:

$$x_{k+1} = F_k x_k + f_k + w_k, \quad k = 0, 1, \ldots, N-1,$$
$$x_0 = \xi \tag{1.1.1}$$
$$y_k = H_k x_k + h_k + b_k, \quad k = 0, 1, \ldots, N-1 \tag{1.1.2}$$

where

$$x_k \in R^n, \quad y_k \in R^m; \tag{1.1.3}$$

F_k, H_k are matrices with appropriate sizes; f_k is a sequence of vectors in R^n, h_k a sequence of vectors in R^m.

ξ is a random variable with gaussian probability law, with mean x_0 and covariance matrix P_0 (1.1.4)

w_k, b_k are random gaussian variables with mean 0 and covariance matrices Q_k, P_k respectively; the matrix R_k is positive definite. The variables, ξ, w_k, b_k are mutually independent.

Let us consider the sequence of σ-algebras,

$$\mathcal{Y}^k = \sigma(y_0, \ldots, y_{k-1}), \quad k = 1, \ldots, N.$$

The problem is to compute

$$\hat{x}_N = E[x_N | \mathcal{Y}^N]. \tag{1.1.5}$$

1.1.2 The best linear estimate

A linear estimate is defined as follows:

$$\mathcal{F}_S = \bar{x}_N + \sum_{k=0}^{N-1} S_k (y_k - \bar{y}_k)$$

where $S_1, \ldots, S_N$ are matrices in $L(R^m; R^n)$ defining the filter $\mathcal{F}$, and $\bar{x}_N$, $\bar{y}_k$ represent the means of x_N, y_k respectively. The set $S_1, \ldots, S_N$ is represented globally by S.

Note that the sequences $\bar{x}_k$, $\bar{y}_k$ are defined by the formulas

$$\bar{x}_{k+1} = F_k \bar{x}_k + f_k, \quad k = 0, \ldots, N-1 \tag{1.1.6}$$

$$\bar{y}_k = H_k \bar{x}_k + h_k.$$

The best linear filter is obtained by choosing S in order to minimize the functional

$$\mathcal{L}(S) = E(x_N - \mathcal{F}_S)^*(x_N - \mathcal{F}_S) \tag{1.1.7}$$

where * denotes the transpose.

We need some notation. Let Λ_{kl} denote the correlation matrix of the process x_k, namely

$$\Lambda_{kl} \in L(R^n; R^n)$$

and

$$\Lambda_{kl} = E(x_k - \bar{x}_k)(x_l - \bar{x}_l)^*.$$

Note that $\Lambda_{kl}^* = \Lambda_{lk}$, and if $M \in L(R^n; R^n)$

$$E(x_k - \bar{x}_k)^* M (x_l - \bar{x}_l) = \text{tr } \Lambda_{lk} M$$

we then deduce easily that

$$\begin{aligned}\mathcal{L}(S) =& \text{tr } \Lambda_{NN} + \sum_{k=0}^{N-1} \text{tr } R_k S_k^* S_k \\ &+ \sum_{k,l=0}^{N-1} \text{tr } \Lambda_{lk} H_k^* S_k^* S_l H_l \\ &- 2 \sum_{k=0}^{N-1} \text{tr } \Lambda_{kN} S_k H_k.\end{aligned} \tag{1.1.8}$$

We then deduce easily

Proposition 1.1.1 *There exists a unique S minimizing the functional $\mathcal{L}(S)$.*

Proof

The set of S is clearly a finite dimensional vector space. We equip it with the scalar product

$$(S, \tilde{S}) = \sum_k \text{tr} S_k^* \tilde{S}_k. \tag{1.1.9}$$

The functional $\mathcal{L}(S)$ is a quadratic form and

$$\sum_{k,l} \text{tr} \Lambda_{lk} H_k^* S_k^* S_l H_l \geq 0$$

$$\begin{aligned}\sum_k \text{tr} R_k S_k^* S_k =& \sum_{k=0}^{N-1} \sum_{h=1}^{n} \sum_{i,j=1}^{m} R_{k,ij} S_{k,hi} S_{k,hj} \\ \geq& \alpha \sum_{k=0}^{N-1} \sum_{h=1}^{n} \sum_{i=1}^{m} (S_{k,hi})^2 \\ =& \alpha \|S\|^2.\end{aligned}$$

Therefore $\mathcal{L}(S)$ has a unique minimum $\hat{S}$ and $\hat{S}$ is uniquely defined by the equation

$$\begin{aligned} &\sum_k \operatorname{tr}(\hat{S}_k S_k R_k + R_k \hat{S}_k^* S_k) \\ &+ \sum_{k,l} \operatorname{tr}(\Lambda_{lk} H_k^* \hat{S}_k^* S_l H_l + H_l^* \hat{S}_l^* S_k H_k \Lambda_{kl}) \\ &- 2 \sum_k \operatorname{tr} \Lambda_{kN} S_k H_k = 0, \quad \forall S. \end{aligned} \tag{1.1.10}$$

■

In fact the best linear estimate is also the best possible estimate, by virtue of the gaussian properties. We can state the following:

Proposition 1.1.2 *We have*

$$\hat{x}_N = \mathcal{F}_{\hat{S}} = \bar{x}_N + \sum_{k=0}^{N-1} \hat{S}_k (y_k - \bar{y}_k). \tag{1.1.11}$$

Proof

Define

$$\epsilon_N = x_N - \mathcal{F}_{\hat{S}};$$

then from the definition (1.1.7) one has

$$\sum_k E[\epsilon_N^* S_k y_k + y_k^* S_k^* \epsilon_N] = 0 \quad \forall S_0, \dots, S_{N-1}.$$

Since $S_0, \dots, S_{N-1}$ are arbitrary, it follows that ϵ_N^* is not correlated with $y_0, \dots, y_{N-1}$ thus is also independent of them. Hence the desired result. ■

1.1.3 The discrete time Kalman filter

In equation (1.1.10) the explicit relations defining the process x_k do not play any role. On the other hand the best estimate $\hat{S}$ depends on N and there is no recursive property as N varies.

A recursive form of the filter can be obtained using the explicit relations (1.1.1), and this form is called the Kalman filter, since the seminal papers of Kalman (Kalman 1960; Kalman and Bucy, 1961).

Let us set

$$P_N = E(x_N - \mathcal{F}_{\hat{S}})(x_N^* - F_{\hat{S}}^*) = E\epsilon_N \epsilon_N^* \tag{1.1.12}$$

which represents the covariance of the error. Clearly one has

$$\mathcal{L}(\hat{S}) = \operatorname{tr} P_N. \tag{1.1.13}$$

Let us also define

$$\begin{aligned} \hat{x}_N^+ &= E[x_N | y^{N+1}] \\ &= E[x_N | y_0, \dots, y_N] \end{aligned} \tag{1.1.14}$$

$$\epsilon_N^+ = x_N - \hat{x}_N^+ \tag{1.1.15}$$

$$P_N^+ = E\epsilon_N^+ (\epsilon_N^+)^*. \tag{1.1.16}$$

We shall prove the following formulae.

Theorem 1.1.1 *For the model (1.1.1), (1.1.2) the following formulae hold:*

$$\hat{x}_{N+1} = F_N \hat{x}_N^+ + f_N \tag{1.1.17}$$
$$\hat{x}_N^+ = \hat{x}_N + P_N H_N^*(R_N + H_N P_N H_N^*)^{-1}(y_N - H_N \hat{x}_N - h_N)$$

$$\begin{aligned} P_{N+1} &= Q_N + F_N P_N^+ F_N^* \\ P_N^+ &= P_N - P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} H_N P_N \\ \hat{x}_0 &= x_0 \\ P_0 &= P_0. \end{aligned} \tag{1.1.18}$$

The proof requires the concept of innovation, which we develop now.

1.1.4 Innovation process

Let us consider the process

$$\nu_N = y_N - (H_N \hat{x}_N + h_N), \tag{1.1.19}$$

then we have the following:

Proposition 1.1.3 *The process ν_N is a gaussian process such that*

$$\begin{aligned} E\nu_k &= 0 \\ E\nu_k \nu_l^* &= \delta_{kl}(R_k + H_k P_k H_k^*), \quad \forall k, l. \end{aligned} \tag{1.1.20}$$

Moreover ν_N is independent of $\mathcal{Y}^N$.

Proof

We have

$$\nu_k = H_k \epsilon_k + b_k, \quad k = 0, 1, \ldots \tag{1.1.21}$$

and ν_k is a gaussian variable such that $E\nu_k = 0$.

Moreover

$$E\nu_k \nu_k^* = E(H_k \epsilon_k + b_k)(\epsilon_k^* H_k^* + b_k^*).$$

Since b_k is independent of ϵ_k, we deduce immediately

$$E\nu_k \nu_k^* = R_k + H_k P_k H_k^*.$$

Moreover for $l = 0, \ldots, k-1$,

$$E[\nu_k y_l^* | \mathcal{Y}^k] = E[\nu_k | \mathcal{Y}^k] y_l^* = 0$$

hence also

$$E[\nu_k \nu_l^*] + E[\nu_k \hat{x}_l^*] H_l^* = 0.$$

Using

$$E[\nu_k \hat{x}_l^*] = E\{E[\nu_k | \mathcal{Y}^k] \hat{x}_l^*\} = 0$$

we deduce

$$E\nu_k \nu_l^* = 0, \quad \text{for } l = 0, \ldots, k-1.$$

The proof has been completed. ∎

1.1.5 Proof of Theorem 1.1.1

The first relation of (1.1.17) as well as the first relation of (1.1.18) are immediate.

Moreover using the innovation we can assert that

$$\hat{x}_N^+ = E[x_N | y_0, \ldots, y_{N-1}, \nu_N].$$

Using the fact that $\hat{x}_N^+$ is also the best linear estimate of x_N, given $y_0, \ldots, y_{N-1}, \nu_N$ (by analogy with Proposition 1.1.2), we can write the formula

$$\hat{x}_N^+ = \hat{x}_N + K_N \nu_N \tag{1.1.22}$$

where K_N is a gain factor to be determined. Note that the independence of ν_N from $y_0, \ldots, y_{N-1}$ has been used to derive (1.1.22). It remains to fix K_N, knowing that it minimizes the covariance of the error.

The error can be written as

$$\epsilon_N^+ = \epsilon_N - K_N \nu_N$$

hence the covariance $E\epsilon_N^+(\epsilon_N^+)^*$ is

$$P_N^+ = P_N + K_N(H_N P_N H_N^* + R_N)K_N^* - K_N E\nu_N \epsilon_N^* - E\epsilon_N \nu_N^* K_N.$$

But

$$\begin{aligned} E\nu_N\epsilon_N^* &= E y_N \epsilon_N^* \\ &= E(H_N x_N + h_N)\epsilon_N^* \\ &= E H_N(x_N - \hat{x}_N)\epsilon_N^* = H_N P_N. \end{aligned}$$

Therefore we have

$$\begin{aligned} P_N^+ &= P_N + K_N(H_N P_N H_N^* + R_N)K_N^* - K_N H_N P_N - P_N H_N^* K_N^* \\ &= P_N + [K_N - P_N H_N^*(H_N P_N H_N^* + R_N)^{-1}](H_N P_N H_N^* + R_N) \\ &\quad \times [K_N^* - (H_N P_N H_N^* + R_N)^{-1} H_N P_N] \\ &\quad - P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} H_N P_N \end{aligned}$$

by completing the square.

It follows immediately that the best value of K_N is

$$K_N = P_N H_N^*(H_N P_N H_N^* + R_N)^{-1}$$

and the second formulae in (1.1.17) and (1.1.18) follow. The proof is now complete. ∎

1.1.6 Least squares estimate

There is another approach to derive the Kalman filter, which is to some extent very intuitive but at the same time is more a lucky consequence of the linear gaussian character of the model than based on theoretical reasons.

The idea consists in looking at ξ, w_k in (1.1.1) as decision variables to be chosen in order to minimize the cost

$$\begin{aligned} K(\xi,w) =&(\xi - x_0)^* P_0^{-1}(\xi - x_0) + \sum_{k=0}^{N-1} w_k^* Q_k^{-1} w_k \\ &+ \sum_{k=0}^{N-1} [y_k - (H_k x_k + h_k)]^* R_k^{-1} [y_k - (H_k x_k + h_k)]. \end{aligned} \tag{1.1.23}$$

In (1.1.23) the quantities $y_0, \dots, y_{N-1}$ are considered as given. This is a control problem in which the state equations are (1.1.1) and ξ, w_k are the control variables. The functional K is called a likelihood function. This is because we can express

$$\text{Prob}(\xi = \xi_0, w_k = w_{k_0}, y_k = y_{k_0}) = C\exp\left(-\frac{1}{2} K_0(\xi_0, w_0)\right) \tag{1.1.24}$$

in which $K_0(\xi, w)$ represents the expression (1.1.23) with $y_k = y_{k_0}$. Therefore if we minimize (1.1.23) we maximize the conditional probability of ξ, w given y. The control problem (1.1.83) is solved by standard techniques. Denoting the optimal controls by $\hat\xi$, $\hat w$ and the corresponding optimal state by $\hat p$, we have the relations

$$\begin{aligned} \hat p_{k+1} =&F_k \hat p_k + f_k + \hat w_k, \quad k = 0, \dots, N-1 \\ \hat p_0 =&\hat\xi \end{aligned} \tag{1.1.25}$$

and

$$(\hat\xi - \bar x_0)^* P_0^{-1} \xi + \sum_{k=0}^{N-1} \hat w_k^* Q_k^{-1} w_k - \sum_{k=0}^{N-1} (y_k - H_k \hat p_k - h_k)^* R_k^{-1} H_k \tilde x_k = 0 \tag{1.1.26}$$

for any ξ, w and $\tilde x_k$ given by

$$\begin{aligned} \tilde x_{k+1} =&F_k \tilde x_k + w_k \\ \tilde x_0 =&\xi. \end{aligned} \tag{1.1.27}$$

Introducing the adjoint variable $\hat q_k$ as the solutions of

$$\begin{aligned} \hat q_k =&F_k^* \hat q_{k+1} - H_k^* R_k^{-1}(y_k - H_k \hat p_k - h_k) \\ \hat q_N =&0 \end{aligned} \tag{1.1.28}$$

we deduce from (1.1.26) that

$$(\hat\xi - x_0) P_0^{-1} \xi + \sum_{k=0}^{N-1} \hat w_k^* Q_k^{-1} w_k + \sum_{k=0}^{N-1} (\hat q_k^* - \hat q_{k+1}^* F_k) \tilde x_k = 0$$

and using (1.1.27)

$$(\hat\xi - x_0) P_0^{-1} \xi + \sum_{k=0}^{N-1} \hat w_k^* Q_k^{-1} w_k + \sum_{k=0}^{N-1} \hat q_k^* \tilde x_k - \sum_{k=0}^{N-1} \hat q_{k+1}^* \tilde x_{k+1} + \sum_{k=0}^{N-1} \hat q_{k+1}^* w_k = 0$$

or finally

$$[(\hat{\xi} - x_0)^* P_0^{-1} + \hat{q}_0^*]\xi + \sum_{k=0}^{N-1} (\hat{w}_k^* Q_k^{-1} + \hat{q}_{k+1}^*) w_k = 0$$

which implies

$$\begin{aligned} \hat{\xi} &= x_0 - P_0 \hat{q}_0, \\ \hat{w}_k &= - Q_k \hat{q}_{k+1} \end{aligned}$$

which used in (1.1.25) yields

$$\begin{aligned} \hat{p}_{k+1} &= F_k \hat{p}_k + f_k - Q_k \hat{q}_{k+1} \\ \hat{p}_0 &= x_0 - P_0 \hat{q}_0. \end{aligned} \tag{1.1.29}$$

Let us check that the pair $\hat{p}_k$, $\hat{q}_k$ satisfies the following affine relation

$$\hat{p}_k = r_k - \Sigma_k \hat{q}_k \tag{1.1.30}$$

where r_k, Σ_k are to be determined. Using (1.1.30) in (1.1.28), (1.1.29) yields

$$\begin{aligned} r_{k+1} - \Sigma_{k+1} \hat{q}_{k+1} &= F_k (r_k - \Sigma_k \hat{q}_k) - Q_k \hat{q}_{k+1} + f_k \\ (I + H_k^* R_k^{-1} H_k \Sigma_k) \hat{q}_k &= F_k^* \hat{q}_{k+1} - H_k^* R_k^{-1} (y_k - H_k r_k - h_k). \end{aligned}$$

Assuming for a while that Σ_k is positive definite (it is also assumed symmetric), we deduce

$$\begin{aligned} r_{k+1} - \Sigma_{k+1} \hat{q}_{k+1} = &F_k r_k - F_k (\Sigma_k^{-1} + H_k^* R_k^{-1} H_k)^{-1} \\ &\times [F_k^* \hat{q}_{k+1} - H_k^* R_k^{-1} (y_k - H_k r_k - h_k)] - Q_k \hat{q}_{k+1} + f_k. \end{aligned}$$

Identifying terms so that this relation holds whatever the value of $\hat{q}_{k+1}$, we write

$$\begin{aligned} \Sigma_{k+1} &= F_k (\Sigma_k^{-1} + H_k^* R_k^{-1} H_k)^{-1} F_k^* + Q_k \\ r_{k+1} &= F_k r_k + f_k + F_k (\Sigma_k^{-1} + H_k^* R_k^{-1} H_k)^{-1} H_k^* R_k^{-1} (y_k - H_k r_k - h_k). \end{aligned} \tag{1.1.31}$$

By virtue of the initial condition appearing in (1.1.29), we take

$$r_0 = x_0, \quad \Sigma_0 = P_0.$$

We note the algebraic relations

$$\begin{aligned} H_k (\Sigma_k^{-1} + H_k^* R_k^{-1} H_k)^{-1} &= R_k (H_k \Sigma_k H_k^* + R_k)^{-1} H_k \Sigma_k \\ \Sigma_k - \Sigma_k H_k^* (R_k + H_k \Sigma_k H_k^*)^{-1} H_k \Sigma_k &= (\Sigma_k^{-1} + H_k^* R_k^{-1} H_k)^{-1} \end{aligned} \tag{1.1.32}$$

and hence we can rewrite (1.1.31) as follows:

$$\begin{aligned} \Sigma_{k+1} &= F_k \Sigma_k F_k^* + Q_k - F_k \Sigma_k H_k^* (R_k + H_k \Sigma_k H_k^*)^{-1} H_k \Sigma_k F_k^* \\ r_{k+1} &= F_k r_k + f_k + F_k \Sigma_k H_k^* (H_k \Sigma_k H_k^* + R_k)^{-1} (y_k - H_k r_k - h_k). \end{aligned} \tag{1.1.33}$$

Comparing these formulae with those of Theorem 1.1, we deduce easily that

$$\Sigma_k = P_k, \quad r_k = \hat{x}_k.$$

Now from (1.1.30) we see

$$\hat{p}_N = r_N = \hat{x}_N$$

which proves that the optimal state at time N, corresponding to minimizing the likelihood function, coincides with the best estimate, the conditional mean.

This fact is not general. The maximum likelihood estimate does not coincides with the conditional mean in non linear, non gaussian situations. It remains an interesting suboptimal estimate to consider in all cases, in view of its nice computational features.

To illustrate further the interest of the maximum likelihood approach to guess intuitively the form of a desired estimate, let us suppose that we want to compute $\hat{x}_N$. By analogy with (1.1.23) we introduce this time the likelihood function

$$\begin{aligned} K^+(\xi, w) =&(\xi - x_0)^* P_0^{-1}(\xi - x_0) + \sum_{k=0}^{N-1} w_k^* Q_k^{-1} w_k \\ &+ \sum_{k=0}^{N}(y_k - H_k x_k - h_k)^* R_k^{-1}(y_k - H_k x_k - h_k). \end{aligned} \tag{1.1.34}$$

Denoting again (to save notation) by $\hat{\xi}$, $\hat{w}_k$, $\hat{r}_k$ the corresponding optimal control and state, we get

$$\begin{aligned} \hat{p}_{k+1} =&F_k \hat{p}_k + f_k + \hat{w}_k, \quad k = 0, \dots, N-1 \\ \hat{p}_0 =&\hat{\xi} \end{aligned} \tag{1.1.35}$$

and

$$(\hat{\xi} - \bar{x}_0)^* P_0^{-1}\xi + \sum_{k=0}^{N-1} \hat{w}_k^* Q_k^{-1} w_k - \sum_{k=0}^{N}(y_k - H_k \hat{p}_k - h_k)^* R_k^{-1} H_k \tilde{x}_k = 0 \tag{1.1.36}$$

$$\begin{aligned} \tilde{x}_{k+1} =&F_k \tilde{x}_k + w_k \\ \tilde{x}_0 =&\xi. \end{aligned}$$

Defining $\hat{q}_k$ by

$$\begin{aligned} \hat{q}_k =&F_k^* \hat{q}_{k+1} - H_k^* R_k^{-1}(y_k - H_k \hat{p}_k - h_k), \quad k = 0, \dots, N \\ \hat{q}_{N+1} =&0 \end{aligned} \tag{1.1.37}$$

we deduce from (1.1.36)

$$(\hat{\xi} - \bar{x}_0)^* P_0^{-1}\xi + \sum_{k=0}^{N-1} \hat{w}_k^* Q_k^{-1} w_k + \sum_{k=0}^{N}(\hat{q}_k^* - \hat{q}_{k+1}^* F_k)\tilde{x}_k = 0$$

that is

$$(\hat{\xi} - \bar{x}_0)^* P_0^{-1}\xi + \sum_{k=0}^{N-1} \hat{w}_k^* Q_k^{-1} w_k + \hat{q}_0^* \xi + \sum_{k=0}^{N-1} \hat{q}_{k+1}^* w_k = 0, \quad \forall \xi, \forall w_k$$

from which $\hat{\xi}$, $\hat{w}_k$ follow. We thus get

$$\begin{aligned} \hat{p}_{k+1} =&F_k \hat{p}_k + f_k - Q_k \hat{q}_{k+1}, \quad k = 0, \dots, N-1 \\ \hat{p}_0 =&x_0 - P_0 \hat{q}_0. \end{aligned} \tag{1.1.38}$$

Note that (1.1.37) can be written as

$$\begin{aligned}\hat{q}_k =&F_k^*\hat{q}_{k+1} - H_k^* R_k^{-1}(y_k - H_k\hat{p}_k - h_k), \quad k = 0,\dots,N-1 \\ \hat{q}_N =& - H_N^* R_N^{-1}(y_N - H_N\hat{p}_N - h_N).\end{aligned} \tag{1.1.39}$$

One can easily check that

$$\hat{p}_k = \hat{x}_k - P_k\hat{q}_k$$

hence in particular, recalling that $\hat{p}_N = \hat{x}_N^+$, we deduce

$$\hat{x}_N^+ = \hat{x}_N + P_N H_N^* R_N^{-1}(y_N - H_N\hat{x}_N^+ - h_N)$$

hence

$$\hat{x}_N^+ = \hat{x}_N + (P_N^{-1} + H_N^* R_N^{-1} H_N)^{-1} H_N^* R_N^{-1}(y_N - H_N\hat{x}_N - h_N)$$

i.e.

$$\hat{x}_N^+ = \hat{x}_N + P_N H_N^*(H_N P_N H_N^* + R_N)^{-1}(y_N - H_N\hat{x}_N - h_N). \tag{1.1.40}$$

We can in particular note the relation

$$\hat{x}_N = F_{N-1}\hat{x}_{N-1}^+ + f_{N-1}. \tag{1.1.41}$$

It also follows from (1.1.40) that defining

$$\epsilon_N^+ = x_N - \hat{x}_N^+$$

$$P_N^+ = E\epsilon_N^+(\epsilon_N^+)^*$$

we have

$$\begin{aligned}\epsilon_N^+ = [I - P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} H_N]\epsilon_N \\ -P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} b_N\end{aligned}$$

hence

$$\begin{aligned}P_N^* =&[I - P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} H_N]P_N[I - H_N^*(H_N P_N H_N^* + R_N)^{-1} H_N P_N] \\ &+ P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} R_N (H_N P_N H_N^* + R_N)^{-1} H_N P_N.\end{aligned}$$

Simplifying, we deduce

$$P_N^+ = P_N - P_N H_N^*(H_N P_N H_N^* + R_N)^{-1} H_N P_N. \tag{1.1.42}$$

We recover in this manner all the formulae of Theorem 1.1.1.