

that I_t is surjective. Suppose $Z \in L^2(Y, t, d')$ so that $\|Z_n - Z\|_2 \rightarrow 0$ for some $Z_n \in \mathcal{E}(Y, t, d')$. By isometry we conclude that $H_n := I_t^{-1}(Z_n)$ is a Cauchy sequence converging to some $H \in L^2(X, t, d')$. Finally $\|I_t H - Z\| = \lim_{n \rightarrow \infty} \|Z_n - Z\| = 0$ and therefore $I_t(H) = Z$.

Equation (3.11) is evident in the case that γ is a simple function. For general $\gamma \in L_t(X, d')$, choose a sequence of simple $\gamma^{(n)}$ with $\|\gamma^{(n)} - \gamma\|_2 \rightarrow 0$. By (3.9) and Itô's isometry we have

$$\left\| \int_0^t \gamma^{(n)}(s) dX(s) - \int_0^t \gamma(s) dX(s) \right\|_2 \rightarrow 0, \quad \left\| \int_0^t \gamma^{(n)}(s) dY(s) - \int_0^t \gamma(s) dY(s) \right\|_2 \rightarrow 0$$

as $n \rightarrow \infty$. The claim then follows because the continuity of the operator I_t yields

$$I_t \int_0^t \gamma(s) dX(s) = \lim_{n \rightarrow \infty} I_t \int_0^t \gamma^{(n)}(s) dX(s) = \lim_{n \rightarrow \infty} \int_0^t \gamma^{(n)}(s) dY(s) = \int_0^t \gamma(s) dY(s). \quad \square$$

4 Filtering, smoothing, and prediction

This section is devoted to optimal linear filtering, prediction and smoothing of partially observed polynomial processes. We let either $I := \mathbb{N}$ or $I := \mathbb{R}_+$ and fix a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in I}, \mathbb{P})$ as well as an \mathbb{R}^d -valued adapted process $X = (X(t))_{t \in I}$. If $I = \mathbb{R}_+$, we assume $(\mathcal{F}_t)_{t \in I}$ to be right-continuous. Suppose that the components $X_{m+1}(t), \dots, X_d(t)$ are observable whereas $X_1(t), \dots, X_m(t)$ are not. We let the subscript o stand for the observable part of a vector $x \in \mathbb{R}^d$ and let $H := (\delta_{m+i,j})_{i=1, \dots, d-m; j=1, \dots, d}$, i.e. $x_o := Hx = (x_{m+1}, \dots, x_d)$. For $\Sigma \in \mathbb{R}^{d \times d}$ we set $\Sigma_{:,o} := \Sigma H^\top = \Sigma_{1:d, m+1:d}$, $\Sigma_{o,:} := H\Sigma = \Sigma_{m+1:d, 1:d}$ as well as $\Sigma_o := H\Sigma H^\top = \Sigma_{m+1:d, m+1:d}$. The subscript u standing for the unobservable part of a vector is treated in the same manner.

We suppose that $\mathbb{E}(\|X(t)\|^2) < \infty$ for $t \in I$ and consider the following general filtering problem for fixed $t \in I$. The goal is to minimise the mean square error $\mathbb{E}(\|X(t) - Y\|^2)$ over all random variables Y that are measurable with respect to the observable information

$$\mathcal{G}_t := \sigma(\{X_o(s) : s \in I, s \leq t\}). \quad (4.1)$$

We call the minimiser of (4.1) the optimal filter for X . Regardless of any specific model the optimal filter is then given by the conditional mean $\hat{X}(t, t) := \mathbb{E}(X(t) | \mathcal{G}_t)$.

4.1 Discrete-time linear filtering problems

Let $I = \mathbb{N}$. For Gaussian state space models, the *optimal filter* can be computed recursively:

Proposition 4.1 (Kálmán filter). *Suppose that X is a linear Gaussian state space model as in Definition 3.1 and set $C(t) := B(t)B(t)^\top$. Let $\hat{X}(0, -1) := \mathbb{E}(X(0))$, $\hat{\Sigma}(0, -1) := \text{Cov}(X(0))$ and*

$$\begin{aligned} \hat{X}(t+1, t) &:= a(t+1) + A(t+1)\hat{X}(t, t), \\ \hat{X}(t, t) &:= \hat{X}(t, t-1) + \hat{\Sigma}_{:,o}(t, t-1)\hat{\Sigma}_o(t, t-1)^+ (X_o(t) - \hat{X}_o(t, t-1)), \\ \hat{\Sigma}(t+1, t) &:= A(t+1)\hat{\Sigma}(t, t)A(t+1)^\top + C(t+1), \\ \hat{\Sigma}(t, t) &:= \hat{\Sigma}(t, t-1) - \hat{\Sigma}_{:,o}(t, t-1)\hat{\Sigma}_o(t, t-1)^+ \hat{\Sigma}_{o,:}(t, t-1) \end{aligned}$$