Appendices

A. Storvik's filter as a Kalman filter

Let us consider the following model.

$$\begin{aligned} x_t &= \mathbf{A}x_{t-1} + v_t,\ v_t \sim N(0, \mathbf{Q}) \\ y_t &= \mathbf{H}x_t + w_t,\ w_t \sim N(0, \mathbf{R}) \end{aligned} \tag{17}$$

We will call the MMSE estimate Kalman filter returns as $x_{t|t} = \mathbb{E}[x_t \mid y_{0:t}]$ and the variance $\mathbf{P}_{t|t} = cov(x_t \mid y_{0:t})$. Then the update for the conditional mean estimate is as follows.

$$\begin{aligned} &x_{t|t} = \mathbf{A}x_{t-1|t-1} \\ &+ \underbrace{\mathbf{P}_{t|t-1}\mathbf{H}^T(\mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R})^{-1}}_{\mathbf{K}_t}(y_t - \mathbf{H}\mathbf{A}x_{t-1|t-1}) \end{aligned}$$

where as for the estimation covariance

$$\begin{aligned} \mathbf{P}_{t|t-1} &= \mathbf{A}\mathbf{P}_{t-1|t-1}\mathbf{A}^T + \mathbf{Q} \\ \mathbf{P}_{t|t} &= (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1} \end{aligned} \tag{18}$$

Matching the terms above to the updates in equation 6, one will obtain a linear model for which the transition matrix is $\mathbf{A} = \mathbf{I}$, the observation matrix is $\mathbf{H} = \mathbf{F}_t$, the state noise covariance matrix is $\mathbf{Q} = \mathbf{0}$, and the observation noise covariance matrix is $\mathbf{R} = \mathbf{Q}$

B. Proof of theorem 1

Let us assume that $x \in \mathbb{R}^d$,$\theta \in \mathbb{R}^p$ and $f_\theta(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is a vector valued function parameterized by θ. Moreover, due to the assumption of separability $f_\theta(x_{t-1}) = l(x_{t-1})^T h(\theta)$, where we assume that $l(\cdot) : \mathbb{R}^d \to \mathbb{R}^{m\times d}$ and $h(\cdot) : \mathbb{R}^p \to \mathbb{R}^m$ and m is an arbitrary constant. The stochastic perturbance will have the log-polynomial density $p(v_t) \propto \exp(\mathbf{\Lambda}_1 v_t + v_t^T \mathbf{\Lambda}_2 v_t + \dots)$ Let us analyze the case of $p(v_t) \propto \exp(\mathbf{\Lambda}_1 v_t + v_t^T \mathbf{\Lambda}_2 v_t)$, for mathematical simplicity.

Proof.

$$\begin{aligned} &\log p(\theta \mid x_{0:T}) \propto \log p(\theta) + \sum_{t=1}^{T} \log p(x_t \mid x_{t-1}, \theta) \\ &\propto \log p(\theta) + \sum_{t=1}^{T} \mathbf{\Lambda}_1 \left(x_t - l(x_{t-1})^T h(\theta)\right) + \\ &\left(x_t - l(x_{t-1})^T h(\theta)\right)^T \mathbf{\Lambda}_2 \left(x_t - l(x_{t-1})^T h(\theta)\right) \\ &\propto \log p(\theta) + \underbrace{\left(\sum_{t=1}^{T} -(\mathbf{\Lambda}_1 + 2x_t^T\mathbf{\Lambda}_2) l(x_{t-1})^T\right)}_{\mathbf{S}_1} h(\theta) \\ &+ h^T(\theta) \underbrace{\left(\sum_{t=1}^{T} l(x_{t-1})\mathbf{\Lambda}_2 l^T(x_{t-1})\right)}_{\mathbf{S}_2} h(\theta) + \text{constants} \end{aligned}$$

Therefore, sufficient statistics ($\mathbf{S}_1 \in \mathbb{R}^{1\times m}$ and $\mathbf{S}_2 \in \mathbb{R}^{m\times m}$) exist. The analysis can be generalized for higher-order terms in v_t in similar fashion. □

C. Proof of theorem 2

Proposition 1. *Let $S(x)$ be a $M+1$ times differentiable function and $P(x)$ its order M Taylor approximation. Let $I = (x-a, x+a)$ be an open interval around x. Let $R(x)$ be the remainder function, so that $S(x) = P(x) + R(x)$. Suppose there exists constant U such that*

$$\forall y \in I, \quad \left|f^{(k+1)}(y)\right| \le U$$

We may then bound

$$\forall y \in I, \quad |R(y)| \le U \frac{a^{M+1}}{(M+1)!}$$

We define the following terms

$$\begin{aligned} \epsilon &= U \frac{a^{M+1}}{(M+1)!} \\ Z &= \int_I \exp(S(x))dx \\ \hat{Z} &= \int_I \exp(P(x))dx \end{aligned}$$

Since $\exp(\cdot)$ is monotone and increasing and $|S(x) - P(x)| \leq \epsilon$, we can derive tight bounds relating Z and $\widehat{Z}$.

$$Z = \int_I \exp(S(x))dx \leq \int_I \exp(P(x) + \epsilon)dx$$
$$= \hat{Z}\exp(\epsilon)$$
$$Z = \int_I \exp(S(x))dx \geq \int_I \exp(P(x) - \epsilon)dx$$
$$= \hat{Z}\exp(-\epsilon)$$

Proof.

$$D_{KL}(p||\hat{p}) = \int_I \ln\left(\frac{p(x)}{\hat{p}(x)}\right)p(x)dx$$
$$= \int_I \Big(S(x) - P(x) + \ln(\hat{Z}) - \ln(Z)\Big)\, p(x)dx$$
$$\leq \int_I |S(x) - P(x)|\, p(x)dx$$
$$+ \int_I \Big|\ln(\hat{Z}) - \ln(Z)\Big|\, p(x)dx$$
$$\leq 2\epsilon \propto \frac{a^{M+1}}{(M+1)!} \approx \frac{1}{\sqrt{2\pi(M+1)!}}\left(\frac{ae}{M+1}\right)^{M+1}$$

where the last approximation follows from Stirling's approximation. Therefore, $D_{KL}(p||\hat{p}) \to 0$ as $M \to \infty$. □

D. Proof of theorem 3

Proof.

$$\log \hat{p}(\theta \mid x_{0:T}) = \log\left(p(\theta)\prod_{k=0}^{T}\hat{p}(x_k|x_{k-1},\theta)\right)$$
$$= \log p(\theta) + \sum_{k=0}^{T}\log \hat{p}(x_k \mid x_{k-1},\theta)$$

We can calculate the form of $\log \hat{p}(x_k \mid x_{k-1},\theta)$ explicitly.

$$\log \hat{p}(x_k \mid x_{k-1},\theta) = \log \mathcal{N}(\hat{f}(x_{k-1},\theta),\sigma^2)$$
$$= -\log(\sigma\sqrt{2\pi}) - \frac{(x_k - \hat{f}(x_{k-1},\theta))^2}{2\sigma^2}$$
$$= -\log(\sigma\sqrt{2\pi}) - \frac{x_k^2 - 2x_k\hat{f}(x_{k-1},\theta) + \hat{f}(x_{k-1},\theta)^2}{2\sigma^2}$$
$$= -\log(\sigma\sqrt{2\pi}) - \frac{x_k^2}{2\sigma^2} - \frac{\sum_{i=0}^{M} x_k H^i(x_{k-1})\theta^i}{\sigma^2}$$
$$+ \frac{\sum_{i=0}^{2M} J^i_{x_{k-1}}\theta^i}{2\sigma^2}$$

Using this expansion, we calculate

$$\log \hat{p}(\theta \mid x_{0:T}) = \log p(\theta) + \sum_{k=0}^{T}\log \hat{p}(x_k \mid x_{k-1},\theta)$$
$$= \log p(\theta) - (T+1)\log(\sigma\sqrt{2\pi})$$
$$- \frac{1}{2\sigma^2}\left(\sum_{k=0}^{T} x_k^2\right) - T(\theta)^T\eta(x_0,\ldots,x_T)$$

where we expand $T(\theta)^T\eta(x_0,\ldots,x_T)$ as in 3. The form for $\log \hat{p}(\theta \mid x_{0:T})$ is in the exponential family. □

E. Proof of theorem 4

Proof. Assume that function f has bounded derivatives and bounded support I. Then the maximum error satisfies $\left|f_\theta(x_{k-1}) - \hat{f}_\theta(x_{k-1})\right| \leq \epsilon_k$. It follows that $\hat{f}_\theta(x_{k-1})^2 - f_\theta(x_{k-1})^2 = -\epsilon_k^2 - 2\hat{f}_\theta(x_{k-1})\epsilon_k \approx -2\hat{f}_\theta(x_{k-1})\epsilon_k$.

Then the KL-divergence between the real posterior and the approximated posterior satisfies the following formula.

$$D_{KL}(p_T||\hat{p}_T) \tag{19}$$
$$= \int_{\mathcal{S}_\theta}\left(\frac{1}{\sigma^2}\sum_{k=1}^{T}\epsilon_k(x_k - \hat{f}_\theta(x_{k-1}))\right)p_T(\theta|x_{0:T})d\theta$$

Moreover, recall that as $T \to \infty$ the posterior shrinks to $\delta(\theta - \theta^*)$ by the assumption of identifiability. Then we can rewrite the KL-divergence as (assuming Taylor approximation centered around θ_c)

$$\lim_{T\to\infty} D_{KL}(p_T||\hat{p}_T) \tag{20}$$
$$= \frac{1}{\sigma^2}\lim_{T\to\infty}\sum_{k=1}^{T}\epsilon_k\int_{\mathcal{S}_\theta}(x_k - \hat{f}_\theta(x_{k-1}))p_T(\theta|x_{0:T})d\theta$$
$$= \frac{1}{\sigma^2}\lim_{T\to\infty}\sum_{k=1}^{T}\epsilon_k\cdot \tag{21}$$
$$\left(x_k - \sum_{i=0}^{M}H^i(x_{k-1})\int_{\mathcal{S}_\theta}(\theta-\theta_c)^i p(\theta|x_{0:T})d\theta\right)$$
$$= \frac{1}{\sigma^2}\lim_{T\to\infty}\sum_{k=1}^{T}\epsilon_k\left(x_k - \sum_{i=0}^{M}H^i(x_{k-1})(\theta^*-\theta_c)^i\right)$$

If the center of the Taylor approximation θ_c is the true parameter value θ^*, we can show that

$$\lim_{T\to\infty} D_{KL}(p_T||\hat{p}_T) = \frac{1}{\sigma^2}\lim_{T\to\infty}\sum_{k=1}^{T}\epsilon_k\left(x_k - f_{\theta^*}(x_{k-1})\right))$$
$$= \frac{1}{\sigma^2}\lim_{T\to\infty}\sum_{k=1}^{T}\epsilon_k v_k = 0 \tag{22}$$

where the final statement follows from law of large numbers. Thus, as $T \to \infty$, the Taylor approximation of any order will converge to the true posterior given that $\theta_c = \theta^*$. For an arbitrary center value θ_c,

$$D_{KL}(p_T || \hat{p}_T) = \frac{1}{\sigma^2} \sum_{k=1}^{T} \epsilon_k \left(x_k - \sum_{i=0}^{M} H^i(x_{k-1})(\theta^* - \theta_c)^i \right) \tag{23}$$

Notice that $\epsilon_k \propto \frac{1}{(M+1)!}$ (by our assumptions that f has bounded derivative and is supported on interval I) and $H^i(\cdot) \propto \frac{1}{M!}$. The inner summation will be bounded since $M! > a^M, \forall a \in \mathbb{R}$ as $M \to \infty$. Therefore, as $M \to \infty$, $D_{KL}(p || \hat{p}) \to 0$. $\square$