Appendices

A. Storvik’s filter as a Kalman filter

Let us consider the following model.

\[ x_t = Ax_{t-1} + v_t, \quad v_t \sim N(0, Q) \]
\[ y_t = Hx_t + w_t, \quad w_t \sim N(0, R) \]  \hspace{1cm} (17)

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

\[
x_{t|t} = Ax_{t-1|t-1} + P_{t|t-1}H^T(HP_{t-1|t-1}H^T + R)^{-1}(y_t - HAx_{t-1|t-1})
\]

where as for the estimation covariance

\[
P_{t|t-1} = AP_{t-1|t-1}A^T + Q
\]
\[
P_{t|t} = (I - K_tH)P_{t|t-1}
\]  \hspace{1cm} (18)

Matching the terms above to the updates in equation 6, one will obtain a linear model for which the transition matrix is \( A = I \), the observation matrix is \( H = F_t \), the state noise covariance matrix is \( Q = 0 \), and the observation noise covariance matrix is \( R = Q \).

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 \( \theta \). Moreover, due to the assumption of separability \( f_\theta(x_{t-1}) = l(x_{t-1})^T h(\theta) \), where we assume that \( l(\cdot) : \mathbb{R}^d \to \mathbb{R}^m \) and \( h(\cdot) : \mathbb{R}^p \to \mathbb{R}^m \) and \( m \) is an arbitrary constant. The stochastic perturbation will have the log-polynomial density \( p(v_t) \propto \exp(A_1 v_t + v_t^T A_2 v_t + \ldots) \).

Let us analyze the case of \( p(v_t) \propto \exp(A_1 v_t + v_t^T A_2 v_t) \), for mathematical simplicity.

\[ \log p(\theta | x_{0:T}) \propto \log p(\theta) + \sum_{t=1}^{T} \log p(x_t | x_{t-1}, \theta) \]

\[ \propto \log p(\theta) + \sum_{t=1}^{T} \Lambda_1 (x_t - l(x_{t-1})^T h(\theta)) + \]

\[ (x_t - l(x_{t-1})^T h(\theta))^T A_2 (x_t - l(x_{t-1})^T h(\theta)) \]

\[ \propto \log p(\theta) + \sum_{t=1}^{T} -(\Lambda_1 + 2x_t^T A_2 l(x_{t-1})^T h(\theta)) \]

\[ + h^T(\theta) \left( \sum_{t=1}^{T} l(x_{t-1}) A_2^T x_{t-1} \right) h(\theta) + \text{constants} \]

Therefore, sufficient statistics \((S_1 \in \mathbb{R}^{1 \times m} \text{ and } S_2 \in \mathbb{R}^{m \times m}) \) 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 \( I = (x - a, + b) \) 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

\[ \forall y \in I, \quad |f^{(k+1)}(y)| \leq U \]

We may then bound

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

We define the following terms

\[ \epsilon = U \frac{a^{M+1}}{(M+1)!} \]

\[ Z = \int \exp(S(x))dx \]

\[ \hat{Z} = \int \exp(P(x))dx \]
Since $\exp(\cdot)$ is monotone and increasing and $|S(x) - P(x)| \leq \epsilon$, we can derive tight bounds relating $Z$ and $\bar{Z}$.

$$Z = \int_I \exp(S(x)) \, dx \leq \int_I \exp(P(x) + \epsilon) \, dx = \bar{Z} \exp(\epsilon)$$

$$Z = \int_I \exp(S(x)) \, dx \geq \int_I \exp(P(x) - \epsilon) \, dx = \bar{Z} \exp(-\epsilon)$$

Proof.

\[
D_{KL}(p||\hat{p}) = \int_I \ln \left( \frac{p(x)}{\hat{p}(x)} \right) p(x) \, dx = \int_I \left( S(x) - P(x) + \ln(Z) - \ln(Z) \right) p(x) \, dx \\
\leq \int_I |S(x) - P(x)| \, p(x) \, dx + \int_I |\ln(\bar{Z}) - \ln(Z)| \, p(x) \, dx \\
\leq 2\epsilon \times \frac{\sigma^{M+1}}{(M+1)!} \approx \frac{1}{\sqrt{2\pi(M+1)!}} \left( \frac{\sigma}{M+1} \right)^{M+1}
\]

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.

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

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

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

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. \qed

E. Proof of theorem 4

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

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

\[
D_{KL}(p_T||\hat{p}_T) = \int_{S_\eta} \left( \frac{1}{\sigma^2} \sum_{k=1}^{T} \epsilon_k (x_k - \hat{f}_\theta(x_{k-1})) \right) p_T(\theta \mid x_{0:T}) \, d\theta
\]

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 $\theta_c$)

\[
\lim_{T \to \infty} D_{KL}(p_T||\hat{p}_T) = \frac{1}{\sigma^2} \lim_{T \to \infty} \sum_{k=1}^{T} \epsilon_k \int_{S_\eta} (x_k - \hat{f}_\theta(x_{k-1})) p_T(\theta \mid x_{0:T}) \, d\theta
\]

\[
= \frac{1}{\sigma^2} \lim_{T \to \infty} \sum_{k=1}^{T} \epsilon_k \left( x_k - \sum_{i=0}^{M} H_i(x_{k-1}) \frac{d\theta}{(\theta - \theta_c)^i} \right)
\]

\[
= \frac{1}{\sigma^2} \lim_{T \to \infty} \sum_{k=1}^{T} \epsilon_k \left( x_k - \sum_{i=0}^{M} H_i(x_{k-1})(\theta^* - \theta_c)^i \right)
\]

If the center of the Taylor approximation $\theta_c$ is the true parameter value $\theta^*$, we can show that

\[
\lim_{T \to \infty} D_{KL}(p_T||\hat{p}_T) = \frac{1}{\sigma^2} \lim_{T \to \infty} \sum_{k=1}^{T} \epsilon_k (x_k - \hat{f}_\theta(x_{k-1}))
\]

\[
= \frac{1}{\sigma^2} \lim_{T \to \infty} \sum_{k=1}^{T} \epsilon_k v_k = 0
\]
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 $\theta_c$, 

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

(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 $M \to \infty$, $D_{KL}(p||\hat{p}) \to 0$. \qed