SOCN Course on Sequential Monte Carlo Methods: Hand-Assignments (due November 6th), Dieter Verbeke

1. (a) Since $\pi(x) = \frac{\tilde{\pi}(x)}{Z_\pi}$ is a pdf the integral of $\pi(x)$ over its support is equal to one:

$$Z_\pi = \int \tilde{\pi}(x)\mathrm{d}x = \int \frac{\tilde{\pi}(x)}{q(x)} q(x)\mathrm{d}x = \mathbb{E}_q\left[\frac{\tilde{\pi}(x)}{q(x)}\right] \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\tilde{\pi}(X^i)}{q(X^i)}, \tag{1}$$

where X^i are samples from the proposal distribution with pdf $q(x)$. The approximation on the right-hand side of (1) is a vanilla Monte Carlo estimator of the integral expression for Z_π. Therefore, the properties of such estimators apply and we may conclude that the estimator for Z_π is unbiased.

(b) We straightforwardly calculate the asymptotic variance of the normalization constant estimator by plugging the given target $\pi(x) \sim \mathcal{N}(x|0,1)$ and proposal distribution $q(x) \sim \mathcal{N}(x|0,\lambda^{-1})$ into the general expression

$$\int \frac{\pi(x)^2}{q(x)}\mathrm{d}x - 1 = \frac{1}{\sqrt{2\pi}\lambda}\int e^{-\frac{1}{2}(2-\lambda)x^2}\mathrm{d}x - 1. \tag{2}$$

The integral on the right-hand side of (2) is bounded for $(2-\lambda) > 0$. Restricting λ to be strictly positive, we find $\lambda \in]0,2[$, and consequently $\lambda^{-1} \in]1/2, +\infty[$. This means the variance of the proposal distribution must be finite and at the same time greater than a critical value. The critical value $\lambda = 2$ results in a constant integrand $\frac{\pi(x)^2}{q(x)}$ and leads to an unbounded integral over the (infinite) support. We also see that when the proposal distribution $q(x)$ matches the target distribution $\pi(x)$ exactly, i.e. when $\lambda = 1$, the asymptotic variance is equal to zero.

2. (a) The relevant probability density functions of the model are

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t|Ax_{t-1}, Q) \tag{3}$$
$$p(y_t|x_t) = \mathcal{N}(x_t|Cx_t, R) \tag{4}$$

with parameters $A = 0.7, C = 0.5$ and covariance matrices $Q = R = 0.1$.

(b) Figure 1 shows a simulation of the state trajectory and the Kalman filter estimate of the mean of $p(x_t|y_{1:T})$. In general one cannot expect the mean of the particle filter to be closer to the actual, simulated, trajectory: the Kalman filter provides the analytical solution to the optimal filtering problem for the linear Gaussian state space model.

(c) Figure 2 shows the Kalman and BPF estimate of the mean and variance of $p(x_t|y_{1:T})$ for $N = 1000$ particles. For a correct implementation of the particle filter, one can expect the error between Kalman and particle filter to vanish as $N \to \infty$.

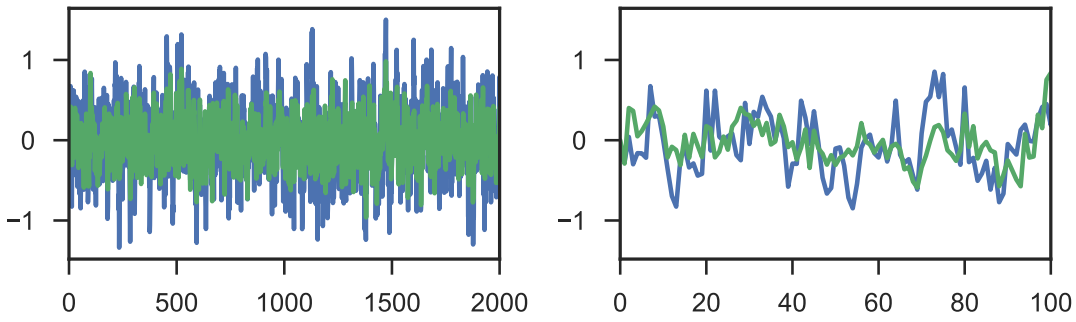

Figure 1: Simulation of the state trajectory in blue, Kalman filter estimate of the mean of $p(x_t|y_{1:T})$ in green. The panel on the right zooms in on the first 100 data points.

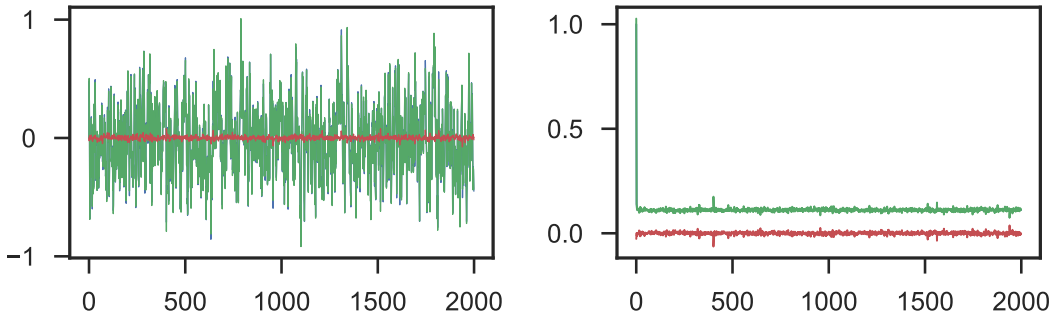

Figure 2: Implementation of the Kalman filter (blue, mostly hidden) and the bootstrap particle filter (green) with $N = 1000$ particles. The left panel shows the estimate of the mean of $p(x_t|y_{1:T})$, the right panel the estimate of the variance. The difference between Kalman and particle filter (red) vanishes as $N \to \infty$.

Figure 3 illustrates the influence of the number of particles on the average absolute difference between the mean and variance estimates of $p(x_t|y_{1:T})$ of the Kalman on the one hand, and the bootstrap and fully adapted (cf. infra) particle filters on the other hand. The error measures are defined as follows:

$$V_m = \frac{1}{T}\sum_{i=1}^{T} |x_{KF} - x_{PF}| \tag{5a}$$

$$V_{\sigma^2} = \frac{1}{T}\sum_{i=1}^{T} \left|\sigma^2_{KF} - \sigma^2_{PF}\right|. \tag{5b}$$

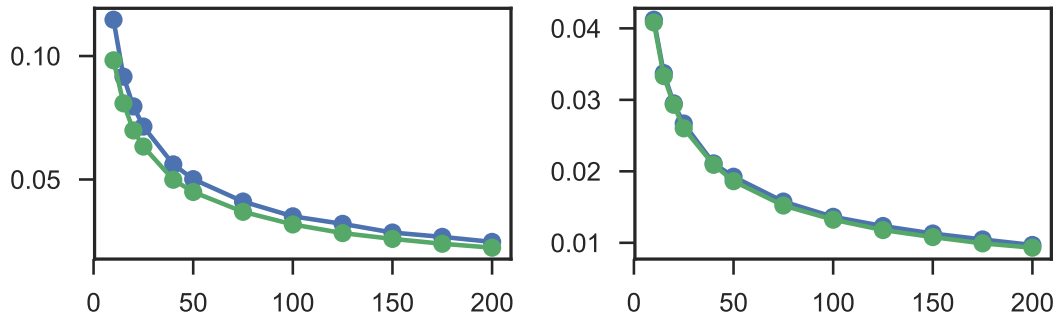

Figure 3: Influence of the number of particles N on the error measures V_m (left) and V_{σ^2} (right) defined in (5). For every N, ten simulations of the particle filter were obtained. Their corresponding error measures were calculated and averaged before plotting them. Both BPF (blue) and FAPF (green) show the same trend. The mean estimate of the FAPF is visibly closer to the (optimal) Kalman filter estimate.

(d) The state and measurement model give immediate access to the following pdf's:

$$p(x_t|x_{t-1}) = \mathcal{N}(x_t|Ax_{t-1}, Q) \tag{6}$$
$$p(y_t|x_t, x_{t-1}) = p(y_t|x_t) = \mathcal{N}(y_t|Cx_t, R), \tag{7}$$

while for the locally optimal proposals we require the pdfs

$$\nu_{t-1} = p(y_t|x_{t-1}) \tag{8}$$
$$q(x_t|x_{t-1}, y_t) = p(x_t|x_{t-1}, y_t). \tag{9}$$

The pdfs (8) and (9) follow from (6) and (7), and the properties of the multivariate Gaussian distribution. Using Theorem 9 in [1] we can express the joint probability density function as

$$p(x_t, y_t|x_{t-1}) = \mathcal{N}\left(\begin{pmatrix} x_t \\ y_t \end{pmatrix} \middle| \begin{pmatrix} Ax_{t-1} \\ CAx_{t-1} \end{pmatrix}, \Sigma\right) \tag{10a}$$

with

$$\Sigma = \begin{pmatrix} Q & QC^T \\ CQ & R + CQC^T \end{pmatrix}. \tag{10b}$$

From Theorem 8 in [1] it follows that

$$p(x_t|x_{t-1}, y_t) = \mathcal{N}\left(x_t \middle| \mu_{x|y}, \Sigma_{x|y}\right) \tag{11a}$$

with

$$\mu_{x|y} = Ax_{t-1} + QC^T(R + CQC^T)^{-1}(y_t - CAx_{t-1}), \tag{11b}$$
$$\Sigma_{x|y} = Q - QC^T(R + CQC^T)^{-1}CQ \tag{11c}$$

Using the marginalization property of the multivariate Gaussian distribution (cf. Theorem 7 in [1]) on the joint pdf (10) we may write

$$p(y_t|x_{t-1}) = \mathcal{N}\left(y_t \middle| CAx_{t-1}, R + CQC^T\right). \tag{12}$$

Figure 4 shows the Kalman and FAPF estimate of the mean and variance of $p(x_t|y_{1:T})$ for $N = 1000$ particles.

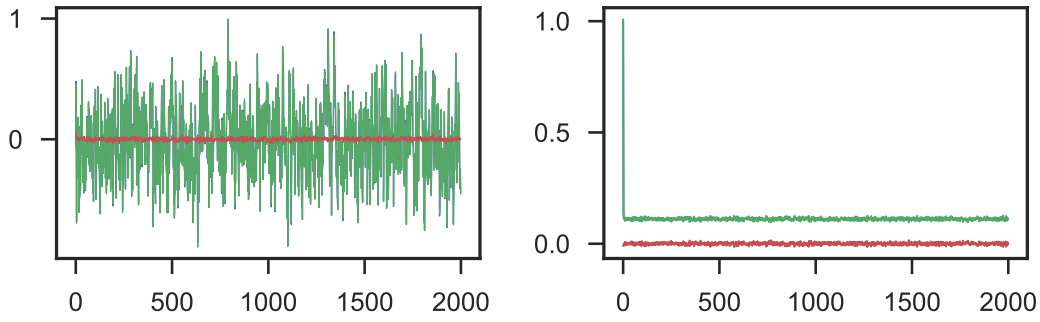

Figure 4: Implementation of the Kalman filter (blue, mostly hidden) and the fully adapted particle filter (green) with $N = 1000$ particles. The left panel shows the estimate of the mean of $p(x_t|y_{1:T})$, the right panel the estimate of the variance. The difference between Kalman and particle filter (red) vanishes as $N \to \infty$.

Figure 3 also presents a comparison between bootstrap particle filter and fully adapted particle filter. The fully adapted particle filter with the same number of particles is noticeably closer to the (optimal) Kalman filter estimate than the bootstrap particle filter. The explanation for this difference should be sought in the way the latest observation y_t enters the particle filter algorithm. In the BPF the latest observation y_t is not taken into account when simulating particles in the propagation step. In the FAPF the latest observation y_t *is* taken into account in the resampling and propagation step. When simulating particles $\{x^i\}_{i=1}^N$ it is preferable to include all available information to increase the probability of producing samples in the most relevant parts of the state space.

(e) Figures 5-7 illustrate the path degeneracy problem for the the fully adapted particle filter with $N = 100$ particles and different resampling strategies. The particle genealogy is not always depicted for the full time horizon T, in order to improve readibility.

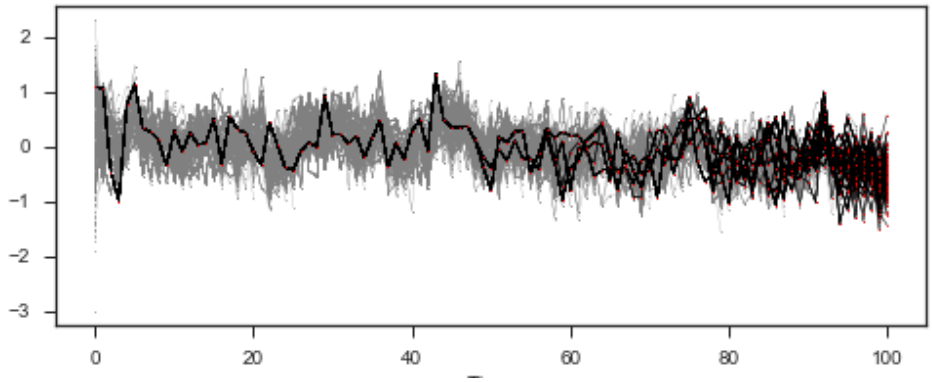

Figure 5: Particle genealogy from time t=100 back to time t=0 for the FAPF with multinomial resampling. The particle degeneracy phenomenon is evident.

Systematic resampling most strongly suppresses path degeneracy. Adaptive resampling mitigates path degeneracy as well. In the latter case resampling is not done at each iteration, only when the effective sample size (ESS) is below some treshold. As a result the stepwise reduction of the number of unique particles is slowed down. However, path degeneracy is bound to manifest itself eventually.

3. (a) Let ϕ be the unknown, while $\sigma = 0.16$ and $\beta = 0.7$. After an initial exploration of the parameter ϕ on a coarse grid ranging from 0 to 1.0 with steps of 0.1, the search was refined. In Figure 8 ϕ varies between 0.905 and 1 with steps of 0.01. For each grid point 10 estimates of the log-likelihood are computed with the bootstrap particle filter. The result is visualized in a box-plot.

   (b) Let the variance parameters σ^2 and β^2 be unknown, while $\phi = 0.985$. We consider a Bayesian setting and put inverse Gaussian priors on the parameters:

$$\sigma^2 \sim \mathcal{IG}(a = 0.01, b = 0.01) \tag{13}$$
$$\beta^2 \sim \mathcal{IG}(a = 0.01, b = 0.01), \tag{14}$$

where the pdf of the inverse gamma distribution with parameters (a,b) is given by

$$\mathcal{IG}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\left(-\frac{b}{x}\right). \tag{15}$$

The priors are only supported on the positive real line. If a negative value is proposed for any of the parameters, the proposal is rejected without having to compute a likelihood estimate.

We use a particle Metropolis Hastings (PMH) algorithm to find the posterior distribution $p(\sigma^2, \beta^2|y_{1:T})$ using a Gaussian random walk proposal for $\theta =$

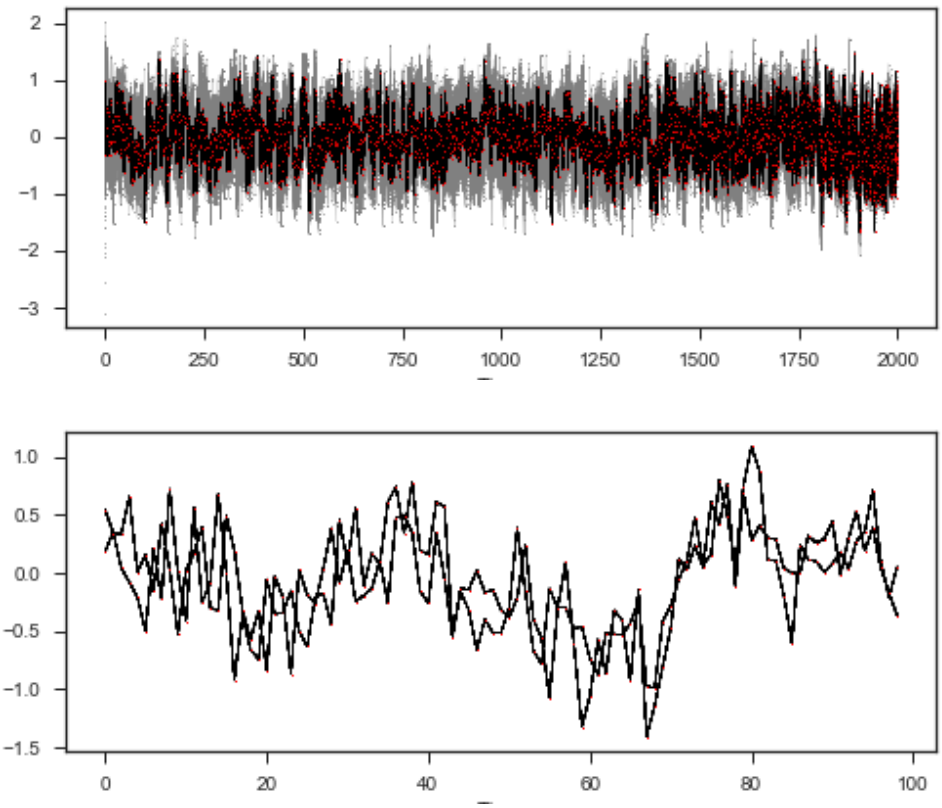

Figure 6: Particle genealogy from time t=2000 back to time t=0 for the FAPF with systematic resampling. The "active paths" in the zone from t=0 to t=100 are highlighted in the lower panel. Path degeneracy is clearly suppressed.

(σ^2, β^2), more specifically

$$q(\theta' \,|\, \theta[k-1]) = \mathcal{N}\left(\theta' | \, \theta[k-1], \Sigma\right) \tag{16}$$

with

$$\theta[0] = \begin{bmatrix} 0.2^2 \\ 0.9^2 \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} 0.02^4 & 0 \\ 0 & 0.09^4 \end{bmatrix}. \tag{17}$$

Note that in the expression for the acceptance probability the contribution of the proposal distribution $q(\cdot)$ cancels since it is symmetric in (θ', θ).

Figure 9 shows the marginal distributions of the variance parameters for a PMH chain with $M = 15000$ iterations, burn-in of $M_0 = 2500$ iterations and $N = 500$ particles. These settings are loosely based on the settings in [2] which calibrated a related stochastic volatility model to very similar data. Note that in [2] the proposal and prior density were defined w.r.t. the square root of the variance parameter (i.e. σ) while in this assignment it is defined w.r.t. the

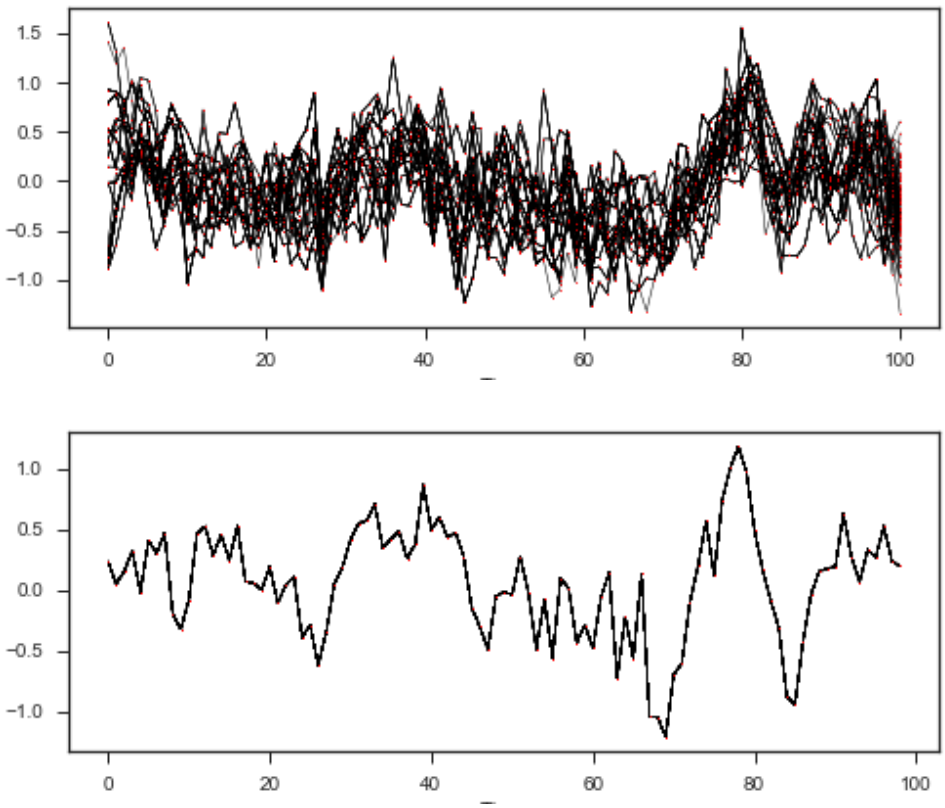

Figure 7: Adaptive resampling (here with ESS treshold of $N/2 = 50$) in the FAPF mitigates path degeneracy. Particle genealogy from time t=100 back to time t=0 in the upper panel. Particle genealogy from time t=2000 back to time t=0 in the lower panel. Note that only the interval between $t = 0$ and $t = 100$ is plotted and the "dead" particles are removed to highlight path degeneracy.

variance parameters (i.e. σ^2 and β^2). In general one should have a look at convergence and mixing diagnostics to assess the quality of the PMH estimates. Choice of the prior, initialization, tuning of the proposal distribution and the number of particles all have an impact on the obtained result. However, for the sake of brevity, the above topics were not further addressed in this exercise. Nonetheless, it seems the choice of identical priors (15) is questionable. The mode of the inverse gamma prior is $b/(1+a) \approx 0.01$ for both parameters. The modes of the posterior distributions are clearly (even orders of magnitude) apart from each other. As a means of sanity check, the log-likelihood was inspected as a function of one parameter while the other was fixed to the mean of the PMH samples. Figure 10 confirms that the PMH estimates are reasonable.

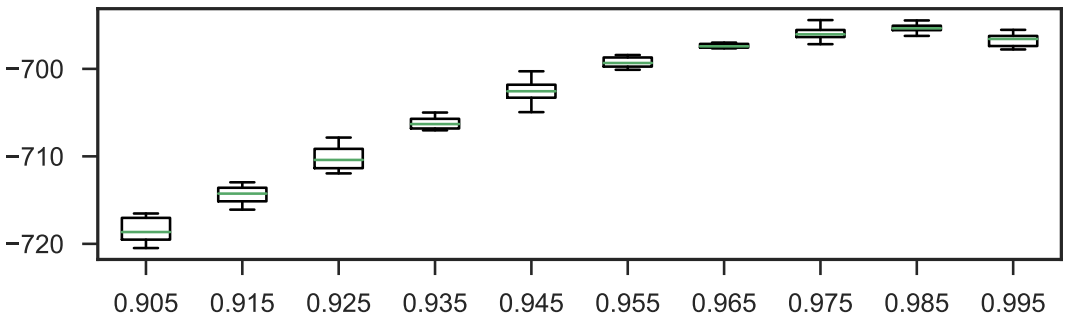

Figure 8: Estimate of the log-likelihood in function of the parameter ϕ while $\sigma = 0.16$ and $\beta = 0.7$.

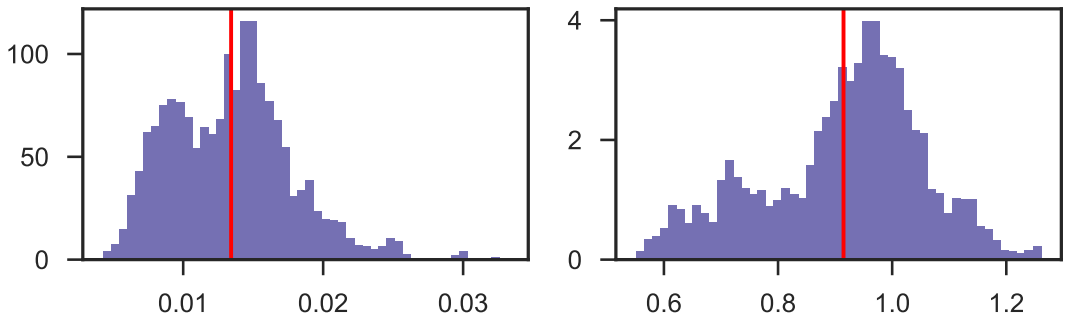

Figure 9: Marginal distributions (normalized histograms) of the variance parameters σ (left) and β (right). The red vertical lines indicate the average estimates: $\hat{\sigma}^2 = 0.0134$ and $\hat{\beta}^2 = 0.9147$.

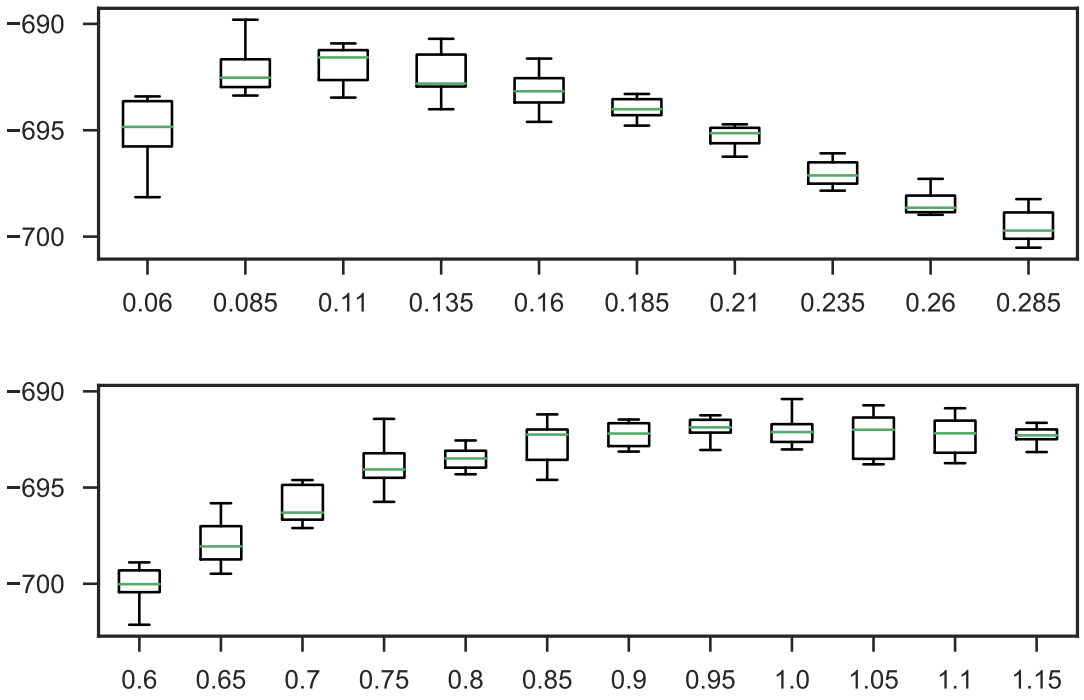

Figure 10: Log-likelihood in function of σ for $\beta = \hat{\beta} = 0.9534$ in the upper panel. Log-likelihood in function of β for $\sigma = \hat{\sigma} = 0.1143$.

References

[1] T. B. Schön and F. Lindsten, "Learning of dynamical systems - particle filters and markov chain methods," 2017, Lecture notes.

[2] J. Dahlin and T. B. Schön, "Getting Started with Particle Metropolis-Hastings for Inference in Nonlinear Dynamical Models," *ArXiv e-prints*, Nov. 2015.