A Statistical Theory of the Kalman Filter

Vivak Patel

Department of Statistics, University of Chicago

SIAM UQ, Lausanne, Switzerland
April 8, 2016

Outline

I. Kalman Filter

II. Single State Estimation

III. Selective Assimilation

IV. Incremental Estimator

V. Numerical Experiment

VI. Conclusions

Kalman Filter

Let $\{\chi_t : t \in \mathbb{N}\}$ be a sequence of states in $\mathbb{R}^d$ related by the stochastic difference equation

$$\chi_{t+1} = D_t \chi_t + \eta_t$$

where $D_t \in \mathbb{R}^{d \times d}$ and $\eta_t \sim \mathcal{N}(0, Q_M)$.

Our **objective** is to estimate χ_t given D_t and a sequence of noisy observations $\{Z_t : t \in \mathbb{N}\}$ related to χ_t

$$Z_t = O_t \chi_t + \xi_t$$

where $O_t \in \mathbb{R}^{d \times d'}$ and $\xi_t \sim \mathcal{N}(0, Q_O)$

The **Kalman Filter** estimates χ_{t+1} and its covariance $\mathcal{C}_{t+1}$ from estimates $\hat{\chi}_t$ and C_t using:

$$\begin{aligned}
\hat{\chi}_{t+1} &= \arg\min \left\{ \|Z_{t+1} - O_{t+1}\chi\|^2_{Q_O^{-1}} + \|\chi - D_t \hat{\chi}_t\|^2_{(D_t C_t D_t^T + Q_M)^{-1}} \right\} \\
C_{t+1}^{-1} &= (D_t C_t D_t^T + Q_M)^{-1} + O_{t+1}^T Q_0^{-1} O_{t+1}
\end{aligned}$$

Kalman Filter

Observability/Controllability

- Kalman, 1960
- Kalman & Bucy, 1961

Exponential Convergence (Fixed State, Deterministic Observations)

- Johnstone, Johnson, Bitmead & Anderson, 1982
- Bittanti, Bolzern & Campi, 1990
- Parkum, Poulsen & Holst, 1992
- Cao & Schwarz, 2003

Nonexpansive in Specialized Metrics

- Bougerol, 1993
- Atar & Zeitouni, 1997
- Le Gland, et. al., 2004
- Carli & Sepulchre, 2015

Kalman Filter

Questions important to statisticians:

1. **Q:** How many observations are needed (with stochastic dynamics and noisy observations) to estimate a single state "well"?
   A: With probability approaching 1, objective decays like d/k.

2. **Q:** Does the covariance estimate actually estimate the asymptotic covariance of the parameter estimate?
   A: Yes, it does.

3. **Q:** How important are the *apriori* parameter and covariance estimate?
   A: ...

Kalman Filter

Questions important to numerical optimizer:

1. **Q:** Can we do "better" than the Kalman Filter?
   A: No.

2. **Q:** What is the rate of convergence of the Kalman Filter?
   A: Sublinear to a single state. Presumably, we cannot do better.

3. **Q:** Do we need to carry around the covariance estimate from state to state?
   A: ...

4. **Q:** Can we parallelize computations over relevant dimensions?
   A: ...

Kalman Filter

Questions important to UQ community (not in other categories):

1. **Q:** How well does the KF "perform" when the dynamics are just stable (e.v. of 1) or unstable (e.v. > 1)?
   A: ...

2. **Q:** How well does the KF "perform" when deterministic dynamics are misspecified?
   A: ...

3. **Q:** How well does the KF "perform" when model noise is misspecified/correlated from state to state?
   A: ...

4. **Q:** How well does the (E)KF "perform" when the observation function is misspecified?

5. **Q:** How well does the KF "perform" when the observation model noise is misspecified/correlated from state to state?

Single State Estimation

To get to statistical notation:

- Initialization:

$$\beta_0 = D_t \hat{\chi}_t \qquad M_0 = D_t C_t D_t^T + Q_M$$

- Observations, Errors, Covariates:

$$\begin{aligned} &Y_k \text{ is the k}^{th} \text{ component of } Z_t \\ &\epsilon_k \text{ is the k}^{th} \text{ component of } \xi_t \\ &X_k \text{ is the k}^{th} \text{ row of } O_t \end{aligned}$$

- True Parameter, Data Model:

$$\beta^* = \chi_{t+1} \qquad Y_k = X_k^T \beta^* + \epsilon_k$$

Selective Assimilation

Question: Suppose $\hat{\beta}_k$ is the estimator of β^* after k observations are assimilated. How quickly does $\hat{\beta}_k$ converge to β^*?

General Assumptions:

- ϵ_k are independent, $\mathbb{E}\left[\epsilon_k | X_k\right] = 0$ and $\sup_k \mathbb{E}\left[\epsilon_k^2 | X_k\right] < \infty$
- X_k explore the entire space $\mathbb{R}^d$ regularly

(Iterative) Batch Estimator:

$$\hat{\beta}_k = \arg\min \left\{ \sum_{j=1}^{k} \frac{1}{\sigma_j^2} (Y_j - X_j^T \beta)^2 + \|\beta - \beta_0\|_{M_0^{-1}}^2 \right\}$$

If $\beta_0 \in \mathbb{R}^d$ and $M_0 \succ 0$ and conditioned on $X_1, \ldots, X_k$ then $\hat{\beta}_k \to \beta^*$ almost surely. However:

- σ_j^2 are never known *a priori*. Requires iteration.
- Batch estimator must be recomputed if k is incremented.

Incremental Estimator

Kalman Filter (Single State)

$$\hat{\beta}_{k+1} = \arg\min \left\{ \frac{1}{\gamma_k^2} (Y_{k+1} - X_{k+1}^T \beta)^2 + \left\| \beta - \hat{\beta}_k \right\|_{M_k^{-1}}^2 \right\}$$
$$M_{k+1}^{-1} = M_k^{-1} + \frac{1}{\gamma_k^2} X_{k+1} X_{k+1}^T$$

Assumptions:

- ϵ_k are independent, $\mathbb{E}\left[\epsilon_k | X_k\right] = 0$ and $\sup_k \mathbb{E}\left[\epsilon_k^2 | X_k\right] < \infty$
- $X_1, X_2, \ldots$ are independent, identically distributed with finite second moment.
- $\mathbb{P}\left[|X_1^T v| = 0\right] < 1$ for all $v \in \mathbb{R}^d$ s.t. $\|v\|_2 = 1$.

Incremental Estimator

Tuning Parameter Requirements:

$$0 < \delta^2 \leq \inf_k \gamma_k^2 \leq \sup_k \gamma_k^2 \leq \Delta^2 < \infty$$

$$\beta_0 \text{ is arbitrary and } M_0 \succ 0$$

Theorem 1: Conditioned on $X_1, X_2, \ldots, X_k$, $\|\beta_k - \beta^*\| \to 0$ almost surely.

Theorem 2: Let

$$\mathcal{M}_k = \mathbb{E}\left[\left(\hat{\beta}_k - \beta^*\right) \left(\hat{\beta}_k - \beta^*\right)^T \middle| X_1, \ldots, X_k \right]$$

If $\sigma_j^2 = \sigma^2 \; \forall j \in \mathbb{N}$ then for all $\epsilon > 0$ asymptotically almost surely

$$\frac{1-\epsilon}{\Delta^2} M_k \preceq \frac{1}{\sigma^2} \mathcal{M}_k \preceq \frac{1+\epsilon}{\delta^2} M_k$$

Numerical Experiment

Data Set

- Source: Public Use File from Center of Medicare and Medicaid Services.
- Described health care expenses, type of visit, patient demographics.
- Contained $N = 2.8$ million records (too big to fit in my computer's 8GB memory).

Linear Model

- Response: Cost of visit.
- Predictors: Patient's gender, Type of Facility, Patient's age.
- Intercept term was included, giving $p = 31$ unknown variables.

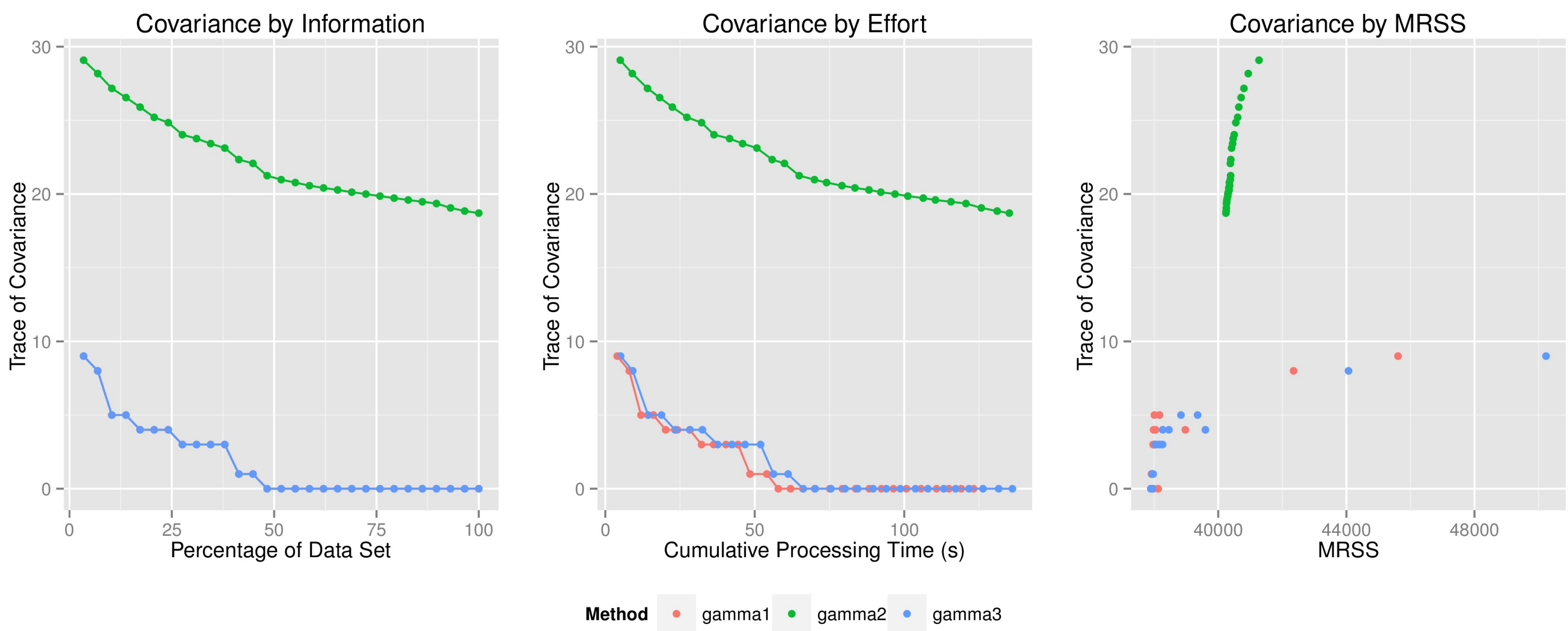
Tuning Parameter

- $\gamma_1^2(k) = \frac{1}{k}$
- $\gamma_2 = 37000$
- $\gamma_3 = 0.001$

Numerical Experiment

Convergence of Estimated Covariance.

- Recall: $\frac{1-\epsilon}{\max_k \gamma_k^2} M_k \prec \frac{1}{\sigma^2}\mathcal{M}_k \prec \frac{1+\epsilon}{\min_k \gamma_k^2} M_k$.
- Estimated covariance tracks well with mean residuals squared.



Conclusions

1. **Q:** How many observations are needed (with stochastic dynamics and noisy observations) to estimate a single state "well"?
   A: (The trace of) M_k is sufficient for determining convergence.

2. **Q:** Does the covariance estimate actually estimate the asymptotic covariance of the parameter estimate?
   A: Yes. $\frac{1-\epsilon}{\max_k \gamma_k^2} M_k \prec \frac{1}{\sigma^2} \mathcal{M}_k \prec \frac{1+\epsilon}{\min_k \gamma_k^2} M_k$

3. **Q:** Can we converge faster than the Kalman Filter?
   A: For a single state and given that M_k is estimating $\mathcal{M}_k$, no. We will incrementally invert the hessian of the objective. In fact, we show that the conditioning has no impact on the rate of convergence.

Thank You

Acknowledgements

- Mihai Anitescu
- Gabriel Terejanu & Haiyan Cheng
- University of Chicago, Department of Statistics & SIAM Travel Award

Slides
This slide deck can be found at galton.uchicago.edu/~vpatel.

Reference

Patel, Vivak. "Kalman-based Stochastic Gradient Method with Stop Condition and Insensitivity to Conditioning." arXiv preprint arXiv:1512.01139 (2015).