Uncertainty in Kalman–Bucy Filtering

SAMUEL N. COHEN
Mathematical Institute
University of Oxford

Research Supported by:
The Oxford–Man Institute for Quantitative Finance
The Oxford–Nie Financial Big Data Laboratory

Based on joint work with ANDREW ALLAN

UNIVERSITY OF OXFORD
Mathematical Institute

Oxford Mathematics

Uncertain filtering

I was born not knowing and have had only a little time to change that here and there. —Feynman

Filtering is a common problem in time series and statistics.

- Kalman filters for tracking space shuttles.
- Hidden Markov models for speech recognition.
- Estimation of ARMA processes in basic time series.

Essentially, we have a process X which we cannot see, which affects the state of a process Y, which we can see.

- Almost always, X and (X, Y) are Markov processes in 'nice' spaces.

Using Bayes' theorem and a probabilistic model, we calculate the distribution of X given Y.

A basic problem in practice is that the probabilistic model is often unknown.

- The filter dynamically incorporates new observations of Y, but requires the joint dynamics of X and Y as inputs.
- Practically, these dynamics are approximated statistically, and the filter run using this approximation.
- Errors in the dynamics are not necessarily 'small', but often may be quantifiable.

We will consider how to incorporate our uncertainty in the underlying model into the filter.

- We will work with a conceptually rigorous description of our uncertainty, which leads to efficient calculations.

Suppose our 'signal' X is a scalar diffusion satisfying

$$dX_t = \alpha_t X_t dt + \sqrt{\beta_t} dB_t, \qquad X_0 = x$$

and Y is an observation process satisfying

$$dY_t = c_t X_t dt + dW_t, \qquad Y_0 = 0.$$

- B, W are independent Brownian motions
- α, β, c_t are locally bounded processes, which may depend (continuously and non-anticipatively) on the path of Y
- $\mathcal{Y}_t = \sigma(Y_s; s \leq t)$ is the observation filtration.

Our aim is to calculate the distribution of $X_t | \mathcal{Y}_t$.

Suppose $X_0 = x \sim N(q_0, R_0)$. As our problem is Gaussian, one can show that

$$X_t | \mathcal{Y}_t \sim N(q_t, R_t)$$

where

$$\begin{aligned} dR_t &= (\beta_t + 2\alpha_t R_t - c_t^2 R_t^2)dt \\ dq_t &= \alpha_t q_t dt + c_t R_t dV_t \\ dV_t &= dY_t - c_t q_t dt \end{aligned}$$

This is the Kalman–Bucy filter, and is easily implemented, assuming we know α, β, c, q_0 and R_0.

In practice, we don't know the parameters, so it is difficult to implement this filter

- Equivalently, as these define the generators of the probability measure on our space, we can say that we do not know what the measure should be.
- This is a form of 'Knightian' uncertainty – developed initially by Von Kries, Keynes, Knight, Meinong, Nitsche (among others)
- An axiomatic approach to dealing with these problems is given by 'nonlinear expectations'

Given a space $(\Omega, \mathcal{F}, P)$, a convex expectation is a map $\mathcal{E} : L^\infty(\mathcal{F}) \to \mathbb{R}$ such that

- Monotonicity: If $\xi \geq \xi'$ P-a.s. then $\mathcal{E}(\xi) \geq \mathcal{E}(\xi')$,
- Constant triviality and equivariance: For any $k \in \mathbb{R}$, $\mathcal{E}(k) = k$ and $\mathcal{E}(\xi + k) = \mathcal{E}(\xi) + k$,
- Convexity: For any ξ, ξ', any $\lambda \in [0, 1]$,

$$\mathcal{E}(\lambda\xi + (1 - \lambda)\xi') \leq \lambda\mathcal{E}(\xi) + (1 - \lambda)\mathcal{E}(\xi').$$

The measure P acts as a 'reference' measure.

- $\gamma \log E[\exp(\xi/\gamma)]$ and $E_Q[\cdot]$, for $Q \sim P$, are examples
- If $\mathcal{E}$ is a convex nonlinear expectation, then $\rho(\xi) = \mathcal{E}(-\xi)$ is a convex risk measure.

We can describe a convex expectation through its dual:

- Suppose $\mathcal{E}$ is lower semicontinuous: If $\xi_n \uparrow 0$ then $\mathcal{E}(\xi_n) \uparrow 0$.
- Then $\mathcal{E}$ has the representation

$$\mathcal{E}(\xi) = \sup_{Q\sim P} \left\{ E_Q[\xi] - \mathcal{Z}(Q) \right\}$$

  where the supremum is taken over probability measures equivalent to P (assuming $\mathcal{Z}(Q) < \infty$ for some $Q \sim P$).
- Knowing $\mathcal{Z}$ is enough to allow us to calculate $\mathcal{E}$ for every random variable

This convex expectation can then be used to give robust estimates of many quantities, by minimizing a loss functional, e.g.

$$\arg\inf_{\hat{\xi}\in\mathbb{R}} \mathcal{E}(\|\xi - \hat{\xi}\|^2 \,|\mathcal{Y}_t),$$

for any random variable ξ.

This formulation is also natural when looking at 'robust control' problems, where we are interested in minimizing the maximum expected cost (among 'reasonable' models of costs).

We still need to choose a penalty $\mathcal{Z}$. This should come from our uncertainty over the inputs $(\alpha, \beta, c, q_0, R_0)$.

To do this, we need to answer two questions:

- Do we have a fixed penalty over the inputs, but then evolve using standard filtering, or
- do we try and learn the values of the inputs, and so modify our penalty as we observe new data?

and

- Are the values of α, β, c fixed through time, or
- do the values of the parameters vary dynamically?

Today, we will look at a fixed penalty, with dynamic parameters. We will also assume c is known.

Statistical uncertainty and expectations

Those who ignore statistics are condemned to reinvent it. —Efron (attrib.)

- For initial parameters, we have a penalty $\kappa_0(q_0, R_0)$.
- For a given process (α_t, β_t), we have a penalty

$$\int_{[0,T]} \gamma_t(\alpha_t, \beta_t)dt$$

- κ_0 and γ are known functions, which may depend on Y non-anticipatively, and have minimal value 0.
- A natural choice is given by the negative log-likelihood from precalibration

Given this, for fixed parameters k, k', we define

$$\mathcal{Z}(Q^{(q_0,R_0,\alpha,\beta)}) := \Big(\frac{1}{k}\Big(\kappa_0(q_0, R_0) + \int_{[0,T]} \gamma_t(\alpha_t, \beta_t)dt\Big)\Big)^{k'}$$

Looking only at the current state...

When my information changes, I alter my conclusions. What do you do, sir? —Keynes (attrib.)

For filtering, we don't want to consider the measure on $(\Omega, \mathcal{F})$, but only the law of $X_t|\mathcal{Y}_t$

- For every model under consideration, the law is given by $X_t|\mathcal{Y}_t \sim N(q_t, R_t)$, for some q, R.
- Applying the same logic, we can look for a penalty on q, R, to obtain

$$\mathcal{E}(f(X_t)|\mathcal{Y}_t) = \sup_{q,R}\left\{ \int f(x)\phi_{q,R}(x)dx - \Big(\frac{1}{k}\kappa_t(q,R)\Big)^{k'}\right\}$$

  where κ is to be determined.
- If we can calculate κ, then we can calculate the convex expectation of any function of the current state.

Finding κ through control

When my information changes, I alter my conclusions. What do you do, sir? —Keynes (attrib.)

- We need κ to align with our general penalty $\mathcal{Z}$
- We only need consider the *smallest* penalty associated with q, R, that is

$$\kappa(q,R) = \inf_{\alpha,\beta}\Big\{\kappa_0(q_0,R_0) + \int_{[0,T]} \gamma_t(\alpha_t,\beta_t)dt \quad : \quad X_t|\mathcal{Y}_t \sim_{Q^{(q_0,R_0,\alpha,\beta)}} N(q,R)\Big\},$$

  all others will be ignored by the optimization.
- This yields a control problem for κ.
- The natural space for α_t, β_t is all nonanticipative functions of $(t, Y_{s\leq t})$. Standard dynamic programming yields an optimizer in terms of (t, q_t, R_t).

- We can now view our problem as a random control problem, forwards in time, for a *fixed* observed path Y.
- The dynamics are given by (for a control α, β)

$$\begin{aligned} dR_t &= (\beta_t + 2\alpha_t R_t - c_t^2 R_t^2)dt \\ dq_t &= \alpha_t q_t dt + c_t R_t dV_t \\ dV_t &= dY_t - c_t q_t dt \end{aligned}$$

- And we seek to minimize

$$\kappa_0(q_0, R_0) + \int_{[0,T]} \gamma_t(\alpha_t, \beta_t)dt$$

subject to the terminal condition $(q_T, R_T) = (q, R)$

- As we look at our problem pathwise, the stochastic integral is concerning.
- We assume c is known, so we can take the transformation

$$(q, R) \to w = \Big(\frac{q}{R} - \eta_t, \frac{1}{R}\Big)$$

where η is a fixed version of $\int c_s dY_s$

- We then find w has dynamics

$$\frac{dw}{dt} = f(w, t, \alpha_t, \beta_t) = \begin{pmatrix} -(w_1 + \eta_t)(\alpha + \beta w_2) \\ -\beta w_2^2 - 2\alpha w_2 + c_t^2 \end{pmatrix}$$

- These have the advantage of being absolutely continuous, but do not have nice growth.

- In our new variables, we can consider the HJB equation for our control
$$\frac{\partial \tilde{\kappa}}{\partial t} + H(w, t, \nabla \tilde{\kappa}) = 0$$
- The Hamiltonian is
$$H(w, t, p) = \sup_{(a,b)\in\mathbb{R}\times[0,\infty)} \left\{ f(w, t, a, b) \cdot p - \gamma(t, a, b) \right\}$$
and domain $(w, t) \in U \times (0, \infty)$, where $U = (\mathbb{R} \times (0, \infty))$.
- This is a first order control problem, as we are working pathwise.

- We shall assume c and η are locally bounded, and the coercivity condition that $\exists p > 1$ such that $\forall T$

$$\inf_{0\le t\le T} \frac{\gamma(t,a,b)}{|a|^p + b^p} \to \infty \quad \text{as } |a| + b \to \infty$$

- The initial penalty $\tilde{\kappa}_0$ is assumed finite-valued, bounded below, locally Lipschitz and explodes on the boundary of U.
- We will look for a solution $\tilde{\kappa}$ which satisfies these properties for each t

Theorem

The value function $\tilde{\kappa}$ is the unique locally Lipschitz viscosity solution to the HJB equation which explodes (uniformly on compacts in time) near the boundary of U.

- That $\tilde{\kappa}$ solves the HJB equation (in viscosity sense) is easy, from standard methods.
- Uniqueness comes from proving a comparison principle, through a modified 'doubling the variables' argument.
- We have infinite speed of propagation and explosive boundary conditions.
- Bounded controls cause the explosion to be faster.
- Coercivity helps, as we know the optimizer will not travel too far towards the boundary.

- For a fixed path, we then can consider methods to approximate the value function $\tilde{\kappa}$.
- We can then change variables to obtain κ, and hence our nonlinear expectation.
- This gives a 'robust' method to estimate expectations of functions of the current state of X (and all past values of Y, with only minor modifications).
- Explosions in $\tilde{\kappa}$ are natural, as they indicate 'impossible' models.

Filtration consistency

Probability does not exist —de Finetti

Can we define $\mathcal{E}(\phi(X_T)|\mathcal{Y}_t)$ for $T \neq t$?

If we are looking towards decision making and control, this should satisfy

- Relevance: $\mathcal{E}(I_A\phi(X_T)|\mathcal{Y}_t) = I_A\mathcal{E}(\phi(X_T)|\mathcal{Y}_t)$ for all $A \in \mathcal{Y}_t$
- Recursivity: $\mathcal{E}(\mathcal{E}(\phi(X_T)|\mathcal{Y}_t)|\mathcal{Y}_s) = \mathcal{E}(\phi(X_T)|\mathcal{Y}_s)$ for $s < t$.

Generally, this is not the case for the expectation we have constructed with $\mathcal{Z}$.

- The difficulty is that our uncertainty does not agree with the arrow of time.

Filtration consistency: Pasting?

Probability does not exist —de Finetti

- To enforce dynamic-consistency, we can consider recursively defining $\mathcal{E}$.
- If $\mathcal{E}$ is dynamically consistent, this will not have any effect.
- In discrete time, this generates a variation of the conditional expectation

$$\vec{\mathcal{E}}(\phi(X_T)|\mathcal{Y}_t) := \xi_t = \mathcal{E}(\xi_{t+1}|\mathcal{Y}_t)$$

- In continuous time, we expect that this will satisfy a BSDE in the observation filtration.
- BSDEs are a natural way to encode changes of measure, particularly in a non-Markov or infinite dimensional setting.

Filtration consistency: BSDEs

Probability does not exist —de Finetti

- In the observation filtration, for a fixed reference measure P we have the dynamics

$$dY_s = c_s q_s ds + dV_s$$

- The process V is the innovation process, and has the predictable representation property (even though it may not generate the filtration).
- We note that, in the observation filtration, our uncertainty only comes in through the state estimate q. (Different parameters also affect V, but not in law.)
- Not knowing q corresponds (again) to a lack of knowledge of the law of Y.

Filtration consistency: BSDEs

Probability does not exist —de Finetti

For a formal derivation of the BSDE, we take

$$\begin{aligned}\xi_{t+h} - \xi_t &= \xi_{t+h} - E_P[\xi_{t+h}|\mathcal{Y}_t] + E_P[\xi_{t+h}|\mathcal{Y}_t] - \xi_t \\ &\approx Z_t(V_{t+h} - V_t) - (\xi_t - E_P[\xi_{t+h}|\mathcal{Y}_t])\end{aligned}$$

this suggests the drift term

$$\begin{aligned}&h^{-1}(\xi_t - E_P[\xi_{t+h}|\mathcal{Y}_t]) \\ &= h^{-1}(\mathcal{E}(\xi_{t+h}|\mathcal{Y}_t) - E_P[\xi_{t+h}|\mathcal{Y}_t]) \\ &= \sup_{q,R}\left(Z_t E_Q\left[\frac{W_{t+h} - W_t}{h}\right] - h^{-1}\left(\frac{1}{k}\kappa(q,R)\right)^{k'}\right) \\ &= \sup_{q,R}\left(Z_t c_t(q - q_t^*) - h^{-1}\left(\frac{1}{k}\kappa(q,R)\right)^{k'}\right)\end{aligned}$$

where q^* is the filter state under the reference measure P.

- This shows that, without proper scaling, unless $k' = \infty$ our uncertainty over future values will vanish with recursion.
- Rescaling κ with $h^{1/k'}$, we (formally) obtain a nontrivial BSDE for $\vec{\mathcal{E}}(\phi(X_T)|\mathcal{Y}_t) := \xi_t$,

$$d\xi_t = -g(Z_t)dt + Z_t dV_t; \qquad \xi_T = \mathcal{E}(\phi(X_T)|\mathcal{Y}_T)$$

where V is the (reference) innovations process and

$$g(z) = \sup_{q,R} \Big(Z_t c_t(q - q_t^*) - \big(k^{-1}\kappa(q,R)\big)^{k'} \Big).$$

- For bounded ϕ, existence holds via monotonicity
- More generally, minimal supersolutions exist by convexity

Filtration consistency: BSDEs

Probability does not exist —de Finetti

- By construction, our expectation $\vec{\mathcal{E}}$ is $\mathcal{Y}$-consistent.
- There is an interesting 'double counting' of the uncertainty in our definition of $\vec{\mathcal{E}}$.
- We include both our uncertainty in the state $X_T|\mathcal{Y}_T$, but our uncertainty at $\mathcal{Y}_t$ about the value we will give at $\mathcal{Y}_s$ for $s > t$. Given c is known, this double counting only appears through q_t.
- This added conservatism is natural from a dynamic risk management perspective.
- One can use this to give value functions for optimal control problems with uncertain observation.

Multidimensional Problems

Behold, these things pass away, that others may replace them —Augustine of Hippo

- In this presentation today, we have only considered scalar processes.
- In the Kalman–Bucy setting, this can be relaxed, provided we have some non-degeneracy of our observations (the filtered variance matrix must remain invertible for all models)
- Appropriate coercivity assumptions follow, even though the domain of the control problem defining κ becomes more delicate (R now lives on the space of positive definite matrices).
- That $\tilde{\kappa}$ solves the viscosity equation follows as before.
- This allows a very wide range of classical time-series to be considered.

Uncertainty in the signal-observation link

Behold, these things pass away, that others may replace them —Augustine of Hippo

- We would like to allow c to also be uncertain.
- In discrete time, this poses no difficulty.
- In continuous time, we have problems due to the pathwise formulation of our control problem.
- We have been considering using various methods to overcome this. (Suggestions welcome!)

Non-Kalman–Bucy settings

Behold, these things pass away, that others may replace them —Augustine of Hippo

University of Oxford
Mathematical Institute

- In the setting where X is a finite-state Markov chain, analogous equations can be found (using the Wonham filter)
- We expect the same results will hold, *mutatis mutandis*.
- Observations with jumps do not change the problem significantly in this setting.
- In general, we need to consider optimal control with the Zakai equation generating the state process.
- We expect this to pose some interesting problems in stochastic control.

Learning and uncertainty

Behold, these things pass away, that others may replace them —Augustine of Hippo

- Today we have considered an uncertain-prior model, which does not include learning
- In discrete time, we have an analogous theory which includes learning, through a change in the dynamics of κ.
- A derivation of our problem from likelihood theory adds a term $c_t q_t dY_t - (c_t q_t)^2 dt$ to the dynamics of κ.
- Given c is known, this formally leads to κ solving a stochastic HJB equation (up to a translation).
- This allows us to combine statistical estimation of our model with filtering, in a consistent manner.

Numerical methods

Behold, these things pass away, that others may replace them —Augustine of Hippo

- Given our existence/uniqueness result for the HJB equation, we are well placed to apply numerical methods. Some tweaking is needed to use the exploding boundary.
- Basic experiments suggest that, for well-calibrated models, a form of central limit theorem applies. In particular, κ is smooth enough to be well approximated by a quadratic, at least around its minimal value.
- Making this approximation rigorous (through approximation of control problems) will give fast approximations to this estimation problem.
- There are also numerical questions surrounding calculation of $\vec{\mathcal{E}}$, as the state variable κ is infinite dimensional.

- Using the expectation $\vec{\mathcal{E}}$ in a control problem leads to an interesting class of problems, combining filtering and control with uncertainty.
- Understanding how this changes behaviour is of interest, in particular in what circumstances it leads to 'exploratory' behaviour.
- Giving a rigorous mathematical basis for the value of information as also potentially possible in this framework.