Advanced State-Space Inference with Applications to Frozen-Rate Estimation

Conrad Voigt

June 11, 2026

Abstract

This review investigates the problem of recursive, covariance-driven estimation of a latent Gaussian process from noisy observations through four works that span more than six decades of methodological development. We begin with Kalman's (1960) reformulation of the Wiener filtering problem in terms of conditional expectations and orthogonal projections, which yields the recursive linear filter and the matrix Riccati equation that govern essentially all subsequent state-space inference. We then review Hartikainen and Särkkä (2010), who show that a large class of temporal Gaussian process regression models admit an exact linear state-space representation, reducing the cubic cost of direct Gaussian process regression to linear cost and handling missing and irregularly sampled data naturally through Kalman filtering and Rauch–Tung–Striebel smoothing. Moving from the temporal to the high-dimensional spatial setting, we review Al-Ghattas et al. (2023), who establish non-asymptotic error bounds for thresholded covariance operator estimators and prove an exponential improvement in sample complexity in the small-lengthscale regime, with direct consequences for localization in ensemble Kalman filters. Finally, we review Fang and Gu (2024), whose *inverse* Kalman filter computes exact covariance–vector products in linear time, enabling scalable solution of the inverse problems that arise when latent functions must be recovered from indirectly observed, noise-corrupted data. We close by sketching how this body of theory maps onto a concrete quantitative finance problem: estimating "frozen" European interest rates after the European market close but while U.S. rates continue to trade, a task central to the consistent valuation of cross-currency swaps.

1 Introduction

The estimation of an unobserved, temporally or spatially correlated quantity from partial and noisy measurements is a central problem in applied statistics, engineering, and the physical and financial sciences. Three structurally distinct but mathematically allied formulations recur: *filtering*, in which the latent quantity is estimated at the present instant; *smoothing*, in which it is estimated at a past instant using all available data; and *prediction*, in which it is estimated at a future instant. Kalman [1] unified these under the single heading of *estimation* and provided a recursive algorithm whose computational and conceptual elegance has made it foundational. The defining feature of the Kalman approach is that all inference reduces to the propagation of first and second moments—means and covariances—through linear dynamics, with the covariance of the estimation error evolving according to a deterministic, nonlinear matrix difference equation.

This review examines four papers that, read together, exhibit a coherent arc. The unifying object is a latent Gaussian process whose covariance structure must be estimated, propagated, or inverted, and the unifying machinery is the Kalman recursion and its descendants. We organize the discussion so that each paper addresses a limitation exposed by its predecessor:

1. *The intractability of the Wiener-Hopf integral.* Kalman [1] establishes the recursive linear filter and the Riccati equation, the baseline against which all later work is measured.

2. *The cubic cost scaling of Gaussian processes.* Hartikainen and Särkkä [2] connect the Kalman recursion to Gaussian process regression, showing that temporal Gaussian processes with widely used covariance functions are linear state-space models, and that smoothing handles time gaps in the data exactly.

3. *The failure of sample covariance in high dimensions.* Al-Ghattas et al. [3] confront the estimation of the covariance operator itself in high (indeed infinite) dimension, proving that sparsity-exploiting thresholded estimators dramatically reduce the sample size required to track high-dimensional correlation, with application to ensemble Kalman filters.

4. *The difficulty of exact inverse operations.* Fang and Gu [4] reverse the direction of the Kalman recursion to compute covariance–vector products and covariance inverses in linear time, enabling recovery of latent functions from indirectly observed data.

Throughout, we maintain a fixed notation (Section 2) and, where the original papers differ, we translate their symbols into it. An interesting application, developed in detail in Section 8, concerns the pricing of cross-currency swaps. Such instruments depend simultaneously on a U.S. dollar interest-rate curve (the secured overnight financing rate, SOFR) and a euro-area curve (the euro short-term rate, ESTR). Because the European market closes several hours before the U.S. market, an end-of-day U.S. valuation must rely on *stale*, or "frozen," European observations. The latent quantity to be estimated—where the European curve would have settled at the U.S. close—is precisely an unobserved Gaussian process to be recovered from correlated, still-trading observables. We emphasize that the review of the methodological literature is the primary objective; the financial application is illustrative and deliberately schematic.

2 Notation and the Linear-Gaussian State-Space Model

We fix a discrete time index $k = 0, 1, 2, \dots$ and consider a latent *state* $\boldsymbol{x}_k \in \mathbb{R}^n$ and an *observation* $\boldsymbol{y}_k \in \mathbb{R}^p$ related by the linear-Gaussian state-space model

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_k \, \boldsymbol{x}_{k-1} + \boldsymbol{w}_k, \qquad \boldsymbol{w}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_k), \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{H}_k \, \boldsymbol{x}_k + \boldsymbol{v}_k, \qquad \boldsymbol{v}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}_k), \tag{2}$$

where $\boldsymbol{\Phi}_k \in \mathbb{R}^{n \times n}$ is the *state-transition matrix*, $\boldsymbol{H}_k \in \mathbb{R}^{p \times n}$ is the *observation matrix*, $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ are the process- and observation-noise covariances, and the noise sequences $\{\boldsymbol{w}_k\}$, $\{\boldsymbol{v}_k\}$ are mutually independent, zero-mean, and white. The initial state satisfies $\boldsymbol{x}_0 \sim \mathcal{N}(\boldsymbol{m}_0, \boldsymbol{P}_0)$. We write $\hat{\boldsymbol{x}}_{k|j} := \mathbb{E}[\boldsymbol{x}_k \mid \boldsymbol{y}_0, \dots, \boldsymbol{y}_j]$ for the conditional mean given observations up to time j, and $\boldsymbol{P}_{k|j} := \mathrm{Cov}[\boldsymbol{x}_k \mid \boldsymbol{y}_0, \dots, \boldsymbol{y}_j]$ for the corresponding error covariance. The cases $j = k$, $j = k-1$, and $j < k$ correspond to filtering, one-step prediction, and smoothing, respectively.

The four papers reviewed adopt different but reconcilable conventions. Kalman [1] writes the transition as $\boldsymbol{\Phi}(t{+}1; t)$ and the observation matrix as $M(t)$, and—importantly—folds measurement noise into an auxiliary state component rather than introducing an explicit $\boldsymbol{R}_k$. Hartikainen and Särkkä [2] use a continuous-time generator $\boldsymbol{F}$ and, after discretization, a transition matrix $\boldsymbol{A}_{k-1} = \exp(\boldsymbol{F}\,\Delta t_k)$; their observation matrix is $\boldsymbol{H} = (1, 0, \dots, 0)$. Fang and Gu [4] use $\boldsymbol{G}_t$ for the transition, $\boldsymbol{F}_t$ $(1 \times q)$ for the observation, $\boldsymbol{W}_t$ for the process noise, and $\boldsymbol{V}_t$ for the observation noise. Al-Ghattas et al. [3] work in function space, replacing the covariance *matrix* by a covariance *operator* $\mathcal{C}$ and the observation matrix by a bounded linear observation operator $\boldsymbol{A}$. We flag these correspondences as each paper is introduced.

3 Kalman (1960): Recursive Linear Filtering and Prediction

3.1 Reframing the Wiener problem

By 1960 the optimal estimation of a stationary random signal corrupted by noise had been solved by Wiener through the Wiener–Hopf integral equation, whose practical use was limited by four difficulties that Kalman [1] enumerates: the optimal filter is specified only through its impulse response and is awkward to synthesize; numerical determination of that response is ill-suited to machine computation; nonstationary and growing-memory generalizations each require fresh, difficult derivations; and the underlying assumptions are obscured by the analysis. Kalman's remedy is to abandon the impulse-response viewpoint in favor of the *state-transition* viewpoint, describing linear systems by first-order difference equations of the form (1) and treating the estimation problem through conditional distributions and orthogonal projections.

The conceptual core is a pair of results. First, for any loss function that is positive, nondecreasing, and symmetric, and under symmetry and convexity conditions on the posterior, the estimator minimizing expected loss is the conditional expectation $\hat{\boldsymbol{x}}(t_1 \mid t) = \mathbb{E}[\boldsymbol{x}(t_1) \mid \boldsymbol{y}(t_0), \dots, \boldsymbol{y}(t)]$; under quadratic loss this holds without the distributional restrictions. Second, when the processes are Gaussian (or when attention is restricted to linear estimators under quadratic loss), this conditional expectation coincides with the *orthogonal projection* of $\boldsymbol{x}(t_1)$ onto the linear span $Y(t)$ of the observations. The geometry of Hilbert space thus replaces the analysis of integral equations: the optimal estimate is the point of $Y(t)$ nearest $\boldsymbol{x}(t_1)$, and the estimation error is orthogonal to every observation. Crucially, only first- and second-order moments enter, so "no other statistical data are needed."

3.2 The recursive estimator and the Riccati equation

The decisive computational consequence is recursion. Let $\boldsymbol{\nu}_k := \boldsymbol{y}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k-1}$ denote the *innovation*—the component of the new observation orthogonal to the past—which Kalman shows is a white-noise sequence. Writing the optimal one-step predictor in Kalman's combined form,

$$\hat{\boldsymbol{x}}_{k+1|k} \;=\; \boldsymbol{\Phi}_k^{*}\, \hat{\boldsymbol{x}}_{k|k-1} \;+\; \boldsymbol{\Delta}_k^{*}\, \boldsymbol{y}_k, \tag{3}$$

the gain and effective transition matrices are

$$\boldsymbol{\Delta}_k^{*} = \boldsymbol{\Phi}_k\, \boldsymbol{P}_k\, \boldsymbol{H}_k^{\top} \big(\boldsymbol{H}_k\, \boldsymbol{P}_k\, \boldsymbol{H}_k^{\top}\big)^{-1}, \tag{4}$$

$$\boldsymbol{\Phi}_k^{*} = \boldsymbol{\Phi}_k - \boldsymbol{\Delta}_k^{*}\, \boldsymbol{H}_k. \tag{5}$$

Here $\boldsymbol{P}_k$ is the covariance of the prediction error $\tilde{\boldsymbol{x}}_{k|k-1}$, and the innovation covariance $\boldsymbol{H}_k \boldsymbol{P}_k \boldsymbol{H}_k^{\top}$ plays the role usually assigned to $\boldsymbol{H}_k \boldsymbol{P}_k \boldsymbol{H}_k^{\top} + \boldsymbol{R}_k$ in the modern formulation; the difference reflects Kalman's modeling of measurement noise as a state variable. Eliminating $\boldsymbol{\Delta}^{*}$ and $\boldsymbol{\Phi}^{*}$ yields the celebrated matrix Riccati difference equation for the error covariance,

$$\boldsymbol{P}_{k+1} = \boldsymbol{\Phi}_k \Big[\boldsymbol{P}_k - \boldsymbol{P}_k \boldsymbol{H}_k^{\top} \big(\boldsymbol{H}_k \boldsymbol{P}_k \boldsymbol{H}_k^{\top}\big)^{-1} \boldsymbol{H}_k \boldsymbol{P}_k\Big] \boldsymbol{\Phi}_k^{\top} + \boldsymbol{Q}_k, \tag{6}$$

which is integrated forward from a specified $\boldsymbol{P}_{t_0}$. Equation (6) is the structural analogue of the Wiener–Hopf equation; but unlike the latter it is solved by simple forward iteration, and once $\boldsymbol{P}_k$ is known the filter coefficients (4)–(5) follow "without further calculations." Because $\boldsymbol{P}_k$ evolves deterministically, it can be computed offline, a point we return to in the application.

3.3 Duality and significance

Kalman further establishes (his Theorem 4) a duality between the filtering problem and the noise-free optimal regulator problem of control theory: the two reduce to the same Riccati-type equation under a transposition and time reversal of the system matrices. This duality, made explicit for the first time, connected estimation to the emerging state-space theory of control and remains a standard pedagogical and theoretical device. For our purposes the essential inheritances are three: (i) optimal estimation is recursive and requires only second-order statistics; (ii) the error covariance obeys a deterministic Riccati recursion; and (iii) the innovations form a white sequence, so the optimal filter is a feedback system driven by prediction residuals. Every subsequent paper in this review builds on at least one of these inheritances.

4 Hartikainen and Särkkä (2010): Temporal Gaussian Processes as State-Space Models

4.1 Motivation: the cubic cost of Gaussian process regression

Gaussian process (GP) regression [6] supposes that an unknown function f is a priori a zero-mean GP, $f \sim \mathcal{GP}(0, k(\cdot,\cdot;\boldsymbol{\theta}))$, observed through noisy data $y_i = f(t_i) + \epsilon_i$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2_{\text{noise}})$. The posterior at a test input t_* is Gaussian with mean and variance

$$\mu_{\mathcal{GP}}(t_*) = \boldsymbol{k}_{*,f}\, (\boldsymbol{K} + \sigma^2_{\text{noise}} \boldsymbol{I})^{-1} \boldsymbol{y}, \tag{7}$$

$$\sigma^2_{\mathcal{GP}}(t_*) = k(t_*, t_*) - \boldsymbol{k}_{*,f}\, (\boldsymbol{K} + \sigma^2_{\text{noise}} \boldsymbol{I})^{-1} \boldsymbol{k}_{*,f}^{\top}, \tag{8}$$

where $\boldsymbol{K}$ is the $n \times n$ Gram matrix $[\boldsymbol{K}]_{ij} = k(t_i, t_j)$. The matrix inversion costs $O(n^3)$ time and $O(n^2)$ memory, which is prohibitive for the long data records typical of time series. Hartikainen and Särkkä [2] observe that for one-dimensional inputs and stationary covariance functions this cost can be reduced to $O(n)$ by recasting the regression as Kalman filtering and Rauch–Tung–Striebel (RTS) smoothing of a linear state-space model. As they note, this state-space view of GP inference is, in spirit, exactly the reformulation that Kalman's 1960 article performed for the filtering problem of Wiener; the present paper carries the idea into the machine-learning context.

4.2 From covariance functions to stochastic differential equations

The key device is to represent the process $f(t)$ as the output of a linear time-invariant stochastic differential equation (SDE) driven by white noise $w(t)$ of spectral density q,

$$\frac{\mathrm{d}^m f}{\mathrm{d}t^m} + a_{m-1}\frac{\mathrm{d}^{m-1} f}{\mathrm{d}t^{m-1}} + \cdots + a_0 f \;=\; w(t), \tag{9}$$

equivalently, in first-order vector form, the continuous-time analogue of (1),

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = \boldsymbol{F}\,\boldsymbol{x}(t) + \boldsymbol{L}\,w(t), \qquad f(t) = \boldsymbol{H}\,\boldsymbol{x}(t), \tag{10}$$

where the state stacks f and its first $m-1$ derivatives, $\boldsymbol{H} = (1, 0, \ldots, 0)$, and $\boldsymbol{F}$, $\boldsymbol{L}$ take companion form. Taking a formal Fourier transform of (10) gives the power spectral density

$$S(\omega) = \boldsymbol{H}(\boldsymbol{F} + i\omega \boldsymbol{I})^{-1}\boldsymbol{L}\,q\,\boldsymbol{L}^{\top}(\boldsymbol{F} - i\omega \boldsymbol{I})^{-\top}\boldsymbol{H}^{\top}, \tag{11}$$

which is a rational function of ω^2. The construction is therefore *exact* for any stationary covariance function whose spectral density is rational, and the stationary state covariance $\boldsymbol{P}_\infty$—needed to initialize the filter so as to reproduce the exact GP—solves the algebraic Lyapunov equation $\boldsymbol{F}\boldsymbol{P}_\infty + \boldsymbol{P}_\infty\boldsymbol{F}^{\top} + \boldsymbol{L}q\boldsymbol{L}^{\top} = \boldsymbol{0}$.

Matérn family. For the Whittle–Matérn covariance with half-integer smoothness $\nu = p + \tfrac{1}{2}$ and $\lambda := \sqrt{2\nu}/\ell$, the spectral density is

$$S(\omega) \propto (\lambda^2 + \omega^2)^{-(p+1)}, \tag{12}$$

which is rational, so the SDE representation is exact with state dimension $m = p + 1$. The stable transfer function is $\mathcal{H}(i\omega) = (\lambda + i\omega)^{-(p+1)}$, and, for example, the $p = 1$ case ($\nu = 3/2$) yields

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \begin{pmatrix} 0 & 1 \\ -\lambda^2 & -2\lambda \end{pmatrix}\boldsymbol{x} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} w(t). \tag{13}$$

Squared-exponential covariance. The squared-exponential (SE) covariance $k(\tau) = \sigma^2 \exp(-\tau^2/2\ell^2)$ corresponds to an infinitely differentiable process; its spectral density, $S(\omega) = \sigma^2\sqrt{\pi/\kappa}\,\exp(-\omega^2/4\kappa)$ with $\kappa = 1/2\ell^2$, is *not* rational, so no finite-dimensional Markov representation is exact. Hartikainen and Särkkä recover a finite model by expanding $1/S(\omega)$ in a Taylor series in ω^2 and truncating at order N, yielding a rational approximant whose order controls accuracy. They report that an order-$N = 6$ approximation is practically indistinguishable from the exact SE covariance, while lower orders trade fidelity near the origin against fidelity in the tails.

4.3 Inference, time gaps, and hyperparameters

Discretizing (10) over (possibly nonuniform) sampling intervals Δt_k produces the discrete model (1) with $\boldsymbol{\Phi}_k = \exp(\boldsymbol{F}\,\Delta t_k)$ and an analytically computable $\boldsymbol{Q}_k$. The posterior over the full state trajectory is then obtained exactly by the Kalman filter followed by the RTS smoother [8], at cost $O(n\,m^3)$ and $O(n\,m^2)$; since m is a small constant (typically below ten), the practical scaling is $O(n)$. The authors document this empirically, performing inference for $n = 10^7$ observations in roughly a hundred seconds, where the brute-force GP is limited to a few thousand points.

Two features of this reformulation matter directly for the financial application. First, *missing data are handled by simply skipping the update step*: when no observation arrives at time k, the filter propagates the predictive distribution forward via (1) alone, and the smoother subsequently fills in the latent trajectory using observations on both sides of the gap. Irregularly spaced data are accommodated by recomputing $\boldsymbol{\Phi}_k$ and $\boldsymbol{Q}_k$ between observations. Second, the one-step predictive likelihood factors $p(\boldsymbol{y}_k \mid \boldsymbol{y}_{0:k-1}, \boldsymbol{\theta})$ are produced as byproducts of the filter, so the marginal likelihood $p(\boldsymbol{y}_{0:n} \mid \boldsymbol{\theta}) = \prod_k p(\boldsymbol{y}_k \mid \boldsymbol{y}_{0:k-1}, \boldsymbol{\theta})$ —and hence hyperparameter learning by type-II maximum likelihood—requires only a filtering pass, with no smoothing. The paper thus delivers a non-parametric learning algorithm whose cost is linear in the data while retaining the exact GP posterior for the Matérn family.

5 Al-Ghattas et al. (2023): Sparse Covariance Operators and Ensemble Kalman Filters

5.1 From covariance matrices to covariance operators

The previous two papers take the covariance structure as known and focus on inference of the latent state. In high-dimensional and operational settings, however, the covariance must itself be estimated from a limited ensemble, and estimation error in the covariance propagates into every downstream Kalman update. Al-Ghattas et al. [3] address this problem in function space. Let $u, u_1, \ldots, u_N$ be independent, centered, almost-surely continuous Gaussian random fields on $D = [0,1]^d$ with covariance kernel $k(x,x') = \mathbb{E}[u(x)u(x')]$ and covariance operator $\mathcal{C}$ acting on $L^2(D)$ by $(\mathcal{C}\psi)(\cdot) = \int_D k(\cdot,x')\psi(x')\,\mathrm{d}x'$. The sample covariance function and its *thresholded* counterpart are

$$\hat{k}(x,x') = \frac{1}{N}\sum_{n=1}^{N} u_n(x)u_n(x'), \qquad \hat{k}_{\rho_N}(x,x') = \hat{k}(x,x')\,\mathbf{1}\Big\{\Big|\hat{k}(x,x')\Big| \geq \rho_N\Big\}, \tag{14}$$

the latter setting to zero all entries below a data-driven threshold ρ_N. Sparsity is formalized through an approximate ℓ_q condition generalizing the row-wise sparsity used for covariance matrices: for some $q \in (0,1)$,

$$\sup_{x\in D}\left(\int_D \big|k(x,x')\big|^q\,\mathrm{d}x'\right)^{1/q} \leq R_q. \tag{15}$$

5.2 Non-asymptotic error bounds

The first main result bounds all moments of the operator-norm error of the thresholded estimator. With the threshold scaled to the expected supremum of the field, $\rho_N \asymp N^{-1/2}\,\mathbb{E}[\sup_{x\in D} u(x)]$, the estimator $\hat{\mathcal{C}}_{\rho_{\hat{N}}}$ satisfies, for every $p \geq 1$,

$$\Big(\mathbb{E}\,\Big\|\hat{\mathcal{C}}_{\rho_{\hat{N}}} - \mathcal{C}\Big\|^p\Big)^{1/p} \;\lesssim_p\; R_q^q\,\rho_N^{1-q} \;+\; \rho_N\,e^{-\frac{c}{p}N(\rho_N\wedge\rho_N^2)}. \tag{16}$$

The first term mirrors the classical finite-dimensional rate for ℓ_q-sparse covariance *matrices*; the second, governed only by the expected supremum of the field, is negligible in the regime of interest. A notable theoretical contribution is that (16) holds for all moment orders p and that the tuning parameter need not be tied to a confidence level, in contrast to the high-probability statements standard in the matrix literature.

5.3 The small-lengthscale regime and exponential improvement

The paper's most striking message concerns isotropic kernels $k_\lambda(x, x') = k_\lambda(|x - x'|)$ indexed by a correlation lengthscale λ. Treating the spatial dimension $d \in \{1, 2, 3\}$ as constant, the authors compare the standard sample covariance estimator against the thresholded estimator as $\lambda \to 0$. The sample covariance obeys, via the operator concentration results of Koltchinskii and Lounici [7], the relative-error scaling

$$\frac{\mathbb{E}\left\|\hat{\mathcal{C}} - \mathcal{C}\right\|}{\|\mathcal{C}\|} \asymp \sqrt{\frac{\lambda^{-d}}{N}} \vee \frac{\lambda^{-d}}{N}, \tag{17}$$

so that controlling the relative error requires $N \gtrsim \lambda^{-d}$ samples. By contrast, the thresholded estimator achieves

$$\frac{\mathbb{E}\left\|\hat{\mathcal{C}}_{\hat{\rho_N}} - \mathcal{C}\right\|}{\|\mathcal{C}\|} \leq c(q)\left(\frac{\log(\lambda^{-d})}{N}\right)^{\frac{1-q}{2}}, \tag{18}$$

requiring only $N \gtrsim \log(\lambda^{-d})$ samples—an exponential improvement in sample complexity. The quantity λ^{-d} thus plays, in the infinite-dimensional setting, exactly the role that the ambient dimension d_u plays in finite-dimensional covariance matrix estimation. The intuition, made precise through a covering-number argument and a heuristic mesh of $(1/\lambda)^d$ approximately uncorrelated points, is that a field of lengthscale λ has roughly λ^{-d} effective degrees of freedom; thresholding exploits the resulting spatial decay of correlations, whereas the unstructured sample covariance pays for every degree of freedom. Numerical experiments with squared-exponential and Matérn kernels in $d = 1, 2$ confirm that, with $N = 5\log(\lambda^{-d})$ samples, the thresholded relative error remains flat as λ decreases while the sample-covariance error diverges.

5.4 Application to ensemble Kalman filters

The theory's intended payoff is for the ensemble Kalman filter (EnKF) [5], a family of algorithms that represent forecast and analysis distributions by an ensemble of N particles—typically $N \sim 10^2$ against discretization levels of 10^9 in operational weather forecasting. The *mean-field* analysis update, which uses the true forecast covariance, is

$$v_n^\star = u_n + \mathcal{K}(\mathcal{C})\big(y - \boldsymbol{A}u_n - \eta_n\big), \qquad \mathcal{K}(\mathcal{C}) = \mathcal{C}\boldsymbol{A}^*(\boldsymbol{A}\mathcal{C}\boldsymbol{A}^* + \boldsymbol{\Gamma})^{-1}, \tag{19}$$

where $\boldsymbol{A}$ is a bounded observation operator, $\boldsymbol{\Gamma}$ the observation-noise covariance, and $\mathcal{K}$ the Kalman gain operator, the function-space analogue of (4). Practical filters replace $\mathcal{C}$ by either the sample estimator $\hat{\mathcal{C}}$ (yielding the stochastic, or perturbed-observation, EnKF) or the thresholded "localized" estimator $\hat{\mathcal{C}}_{\rho_N}$. The paper's Theorem 2.10 shows that, in the small-lengthscale regime with $N \gtrsim \log(\lambda^{-d})$, the localized update tracks the mean-field update with error $O\big((\log(\lambda^{-d})/N)^{(1-q)/2}\big)$, whereas the unlocalized update incurs the far larger $O\big(\sqrt{\lambda^{-d}/N}\big)$. This is, to the authors' knowledge, the first functional-setting justification of the long-standing empirical practice of covariance localization in EnKFs.

6 Fang and Gu (2024): Fast Covariance Operations via the Inverse Kalman Filter

6.1 Reversing the Kalman recursion

The three preceding papers move forward through the state-space model: they propagate states, or they estimate the covariance that drives the propagation. Fang and Gu [4] ask the reverse question. Given a dynamic linear model (DLM),

$$y_t = \boldsymbol{F}_t \boldsymbol{\theta}_t + v_t, \qquad v_t \sim \mathcal{N}(0, V_t), \tag{20}$$

$$\boldsymbol{\theta}_t = \boldsymbol{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{W}_t), \tag{21}$$

with $\boldsymbol{\theta}_t \in \mathbb{R}^q$ a q-dimensional latent state, let $\boldsymbol{\Sigma} = \mathrm{Cov}[\boldsymbol{y}_{1:N}] = \boldsymbol{L}\boldsymbol{L}^\top$ denote the induced $N \times N$ observation covariance with Cholesky factor $\boldsymbol{L}$. The classical Kalman filter already computes $\boldsymbol{L}^{-1}\boldsymbol{u}$ for any vector $\boldsymbol{u}$ in $O(q^3 N)$ operations—this is the whitening of observations into innovations, the discrete-time embodiment of Kalman's white-noise innovations sequence. The *inverse Kalman filter* (IKF) supplies the missing operations: it computes $\boldsymbol{L}\boldsymbol{u}$ and $\boldsymbol{L}^\top\boldsymbol{u}$, and hence the covariance–vector product $\boldsymbol{\Sigma}\boldsymbol{u} = \boldsymbol{L}(\boldsymbol{L}^\top\boldsymbol{u})$, also in $O(q^3 N)$ time and *without approximation*. Whereas direct multiplication by $\boldsymbol{\Sigma}$ costs $O(N^2)$, the IKF is linear in N; the authors report multiplying a covariance against a vector of dimension 10^6 in about two seconds.

The construction proceeds through two lemmas. The first computes $\tilde{\boldsymbol{x}} = \boldsymbol{L}^\top\boldsymbol{u}$ by a backward recursion whose coefficients are expressed entirely through Kalman-filter quantities—the predictive variance Q_t, the conditional state covariance B_t, and the Kalman gain $K_t = B_t \boldsymbol{F}_t^\top Q_t^{-1}$. The second computes $\boldsymbol{x} = \boldsymbol{L}\tilde{\boldsymbol{x}}$ by a forward recursion that, the authors emphasize, is precisely the reverse of the standard innovations computation—hence the name. Because the $O(q^3 N)$ cost is dominated by the one-time precomputation of the filter quantities, the per-iteration cost in an iterative solver is only $O(q^2 N)$. A robustness modification adds artificial noise V to stabilize the recursion when the predictive variance Q_t approaches zero, returning $\boldsymbol{\Sigma}\boldsymbol{u}$ exactly and independently of V.

6.2 Scalable inversion and the inverse problem of latent recovery

The IKF becomes powerful when paired with the conjugate gradient (CG) algorithm, which solves linear systems $\boldsymbol{\Sigma}_y \boldsymbol{z} = \boldsymbol{b}$ using only matrix–vector products. This addresses a structure that defeats conventional Gaussian process approximations: covariances of the form

$$\boldsymbol{\Sigma}_y = \sum_{j=1}^{J} \boldsymbol{A}_j \boldsymbol{\Sigma}_j \boldsymbol{A}_j^\top + \boldsymbol{\Lambda}, \tag{22}$$

where each $\boldsymbol{\Sigma}_j$ is a DLM-induced covariance, each $\boldsymbol{A}_j$ is sparse, and $\boldsymbol{\Lambda}$ is diagonal. Approximation methods that target the individual $\boldsymbol{\Sigma}_j$ cannot be applied to $\boldsymbol{\Sigma}_y$ directly, but the resulting IKF-CG algorithm computes the required products $\boldsymbol{\Sigma}_y\boldsymbol{u}$ exactly and in pseudolinear time, reducing the $O(\tilde{N}^3)$ cost of forming and inverting $\boldsymbol{\Sigma}_y$ to a scalable order.

The motivating inverse problem is the nonparametric recovery of particle interaction functions: each observed velocity is an additive, sparsely-weighted combination of latent interaction functions evaluated at many pairwise distances, so a single observation depends on multiple latent states—a dependence structure outside the reach of the standard filter. Modeling each interaction function as a GP with a Matérn kernel (half-integer smoothness with $\nu \le 5/2$, so $q \le 3$) and applying IKF-CG, the authors recover interaction laws from systems with $N \approx 3 \times 10^5$ velocity observations, including real cell-trajectory microscopy data, with predictive errors essentially identical to direct

computation but at a fraction of the cost. The relevance to our review is conceptual: the IKF supplies the computational engine for solving *inverse* problems in which the clean latent function is observed only through a noisy, linearly distorted, and structurally complex transformation.

7 Synthesis

Paper	**What is known**	**What is estimated**	**Key contribution**
Kalman (1960)	$\boldsymbol{\Phi}, \boldsymbol{H}, \boldsymbol{Q}, \boldsymbol{R}$	latent state $\boldsymbol{x}_k$	recursive filter; Riccati equation; duality
Hartikainen & Särkkä (2010)	GP kernel (Matérn / SE)	latent function $f(t)$	GP regression as exact state-space model; $O(n)$ cost
Al-Ghattas et al. (2023)	sparsity level R_q, lengthscale λ	covariance operator $\mathcal{C}$	thresholded estimator; exponential sample-complexity gain; EnKF localization
Fang & Gu (2024)	DLM parameters $\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{W}_t$	$\boldsymbol{\Sigma u}$, $\boldsymbol{\Sigma}_y^{-1}$, latent functions	inverse Kalman filter; $O(q^3 N)$ covariance products; scalable inversion

Table 1: The four reviewed papers, organized by what each treats as known versus estimated, all anchored to the linear-Gaussian state-space model of Section 2.

Read in sequence, the four papers describe a single object from four vantage points. Kalman [1] provides the recursion and the Riccati equation governing the error covariance, establishing that optimal linear estimation needs only second-order moments. Hartikainen and Särkkä [2] reveal that a broad class of Gaussian process priors *are* state-space models, so that the Kalman recursion delivers exact GP inference at linear cost and absorbs missing and irregular data through its smoothing pass. Al-Ghattas et al. [3] confront the empirical reality that the covariance itself is estimated from finite ensembles, and show that sparsity-aware thresholding restores statistical tractability in high dimension—the same covariance that Kalman assumed known and that Hartikainen and Särkkä encoded in an SDE. Fang and Gu [4] close the loop by making the covariance operations themselves—multiplication and, through CG, inversion—scalable, and by reversing the filter to attack inverse problems. Table 1 summarizes the correspondence.

A useful way to see the unity is through the error covariance. Kalman's $\boldsymbol{P}_k$ is propagated; Hartikainen and Särkkä initialize it at the stationary $\boldsymbol{P}_\infty$ to match a prescribed kernel; Al-Ghattas et al. estimate the operator $\mathcal{C}$ that generates it; and Fang and Gu make arithmetic with that operator cheap. Each paper relaxes an assumption the previous one made tacitly.

8 Application: Frozen-Rate Estimation for Cross-Currency Swaps

We now sketch how the reviewed methods bear on a problem in fixed-income pricing. This section is intentionally a secondary, illustrative companion to the literature review above; we indicate where each method applies rather than developing a full empirical study.

8.1 The problem: time-zone asymmetry in cross-currency valuation

A cross-currency basis swap exchanges floating payments referenced to two currencies—say a USD leg indexed to SOFR and a EUR leg indexed to ESTR—together with a basis spread that clears the cross-currency funding market [9]. Its value at a given valuation time depends jointly on the USD discount and projection curve and the EUR curve, across many tenors. The operational difficulty is temporal: the European cash and derivatives markets close several hours before the U.S. market. When a U.S. desk marks its book at the U.S. close, the European inputs are *frozen*—the last quotes observed before the European close—while the USD curve has continued to move. Using stale European levels against a fresh USD curve produces a distorted basis and a contaminated end-of-day profit-and-loss (PnL). The estimation task is to infer where the European curve *would* have settled at the U.S. close, using the still-trading, correlated observables.

8.2 Baseline: a state-space model of the curves (Kalman)

Following Section 3, let the hidden state $\boldsymbol{x}_k$ collect the (unobserved, at the U.S. close) European curve factors together with the contemporaneous USD factors, where k indexes intraday time steps. The dynamics (1) encode the joint mean-reverting evolution of the two curves and the basis, with $\boldsymbol{\Phi}_k$ and $\boldsymbol{Q}_k$ estimated from the overlapping hours during which both markets trade. The observations (2) are the live, still-trading instruments: the SOFR curve, USD-listed proxies for European risk (e.g. EUR FX forwards and EUR rate futures that trade in U.S. hours), and the cross-currency basis quotes themselves. During European trading hours the European curve is directly observed and the filter behaves conventionally; the European curve is the noisy observable and the analyst simply tracks it. After the European close, that observation channel goes dark, and the European curve becomes a hidden state to be estimated recursively from the remaining channels via (3)–(6). Because the error covariance $\boldsymbol{P}_k$ evolves deterministically, the desk can precompute the uncertainty attached to a frozen-rate mark before any data arrive—a direct operational use of Kalman's observation that $\boldsymbol{P}_k$ is data-independent.

8.3 Smoothing across the frozen window (Hartikainen & Särkkä)

The frozen interval—European close to U.S. close—is exactly a window of *missing observations* on the European channel. Section 4 shows that such gaps are handled by skipping the update step and propagating the predictive distribution, with the smoother reconstructing the latent trajectory using information on both sides of the gap. Modeling the EUR–USD basis spread as a temporal Gaussian process with a Matérn covariance, and exploiting the exact SDE representation, the basis becomes a low-dimensional linear state-space model amenable to the Rauch–Tung–Striebel smoother. Concretely, the smoother estimates the trajectory of the basis spread through the frozen hours and produces a consistent terminal estimate at the U.S. close, with calibrated uncertainty; the next morning, when the European market reopens, the newly observed level enters as an observation that the same smoother uses to revise the overnight trajectory retrospectively. The linear-time scaling is practically relevant because the historical record used to estimate the basis dynamics is long and sampled irregularly across holidays and half-days, both of which the method accommodates by recomputing the transition and noise matrices between observations. Hyperparameters of the basis kernel (lengthscale, magnitude, smoothness) can be fit by the marginal likelihood assembled from the filter's one-step predictive densities.

8.4 Multi-tenor covariance and joint tracking (Al-Ghattas et al.)

A swap is not a single rate but a curve across many tenors, and the joint US–EU state can easily span dozens of factors. During the overlapping open hours, the model must estimate the high-dimensional cross-covariance between the U.S. and European curves from a limited number of intraday ensemble snapshots—precisely the regime of Section 5. Two features of the theory transfer. First, curve covariances exhibit approximate sparsity: correlations decay with separation in tenor space, so the cross-tenor covariance is well approximated by an ℓ_q-sparse structure as in (15), and thresholding the off-diagonal (long-tenor-separation) entries suppresses spurious sampling noise. Second, an ensemble Kalman filter with a *localized* (thresholded) covariance can track the joint curve state with an ensemble size growing only logarithmically in the effective dimension, rather than linearly—the practical content of Theorem 2.10. In curve terms, localization prevents a short-dated European tenor from being spuriously updated by a long-dated U.S. tenor merely because the finite ensemble happened to show a sample correlation between them. The result is a more stable estimate of the cross-currency covariance that anchors the frozen-rate inference of the preceding subsections.

8.5 Back-calculating the clean basis (Fang & Gu)

In practice one frequently observes the *distorted* end-of-day quantity—the contaminated portfolio PnL, or a set of stale composite marks—and must back out the clean, implied basis adjustments consistent with the still-trading observables. This is an inverse problem: the observed PnL is a sparse, linear functional of the latent curve increments, plus noise, and recovering the latent increments requires repeatedly solving large linear systems whose covariance has exactly the additive, sparsely-weighted form of (22), with each $\mathbf{\Sigma}_j$ a DLM-induced curve covariance, each $\boldsymbol{A}_j$ the sparse map from curve factors to instrument sensitivities, and $\mathbf{\Lambda}$ the quote-level noise. Across many tenors and many intraday time points, direct inversion is infeasible. The IKF-CG algorithm of Section 6 supplies exact covariance–vector products in time linear in the number of time points and cubic only in the small latent dimension, so the conjugate gradient solver can recover the implied clean basis adjustments at scale. Conceptually, this reverses the forward filter: rather than predicting the distorted observation from the clean state, it reconstructs the clean, active-quote-consistent state from the distorted observation.

8.6 Summary of the application

The four methods compose into a coherent pipeline for frozen-rate estimation: a state-space model defines the latent European curve and its observable U.S. counterparts (Kalman); a temporal-GP smoother bridges the frozen window with calibrated uncertainty (Hartikainen & Särkkä); a localized, thresholded covariance stabilizes the high-dimensional cross-currency correlation during overlapping hours (Al-Ghattas et al.); and an inverse solver recovers the clean basis from distorted end-of-day data (Fang & Gu). We stress that this is a conceptual mapping; a deployment would require careful curve construction, attention to discount and projection bases, and validation against realized reopening levels.

9 Conclusion

This review followed a single methodological thread across four works. The recursive linear filter and matrix Riccati equation of Kalman [1] furnish the structural backbone: optimal estimation

as the propagation of conditional means and covariances, with a deterministic error-covariance recursion and a white innovations sequence. Hartikainen and Särkkä [2] showed that temporal Gaussian process regression is, for the Matérn family exactly and for the squared-exponential family approximately, a linear state-space model, so that the Kalman filter and RTS smoother deliver exact GP inference at $O(n)$ cost while naturally handling missing and irregular data. Al-Ghattas et al. [3] confronted the estimation of the covariance operator that drives these recursions, proving that sparsity-exploiting thresholded estimators achieve an exponential reduction in sample complexity in the small-lengthscale regime and thereby explaining the empirical success of localization in ensemble Kalman filters. Fang and Gu [4] reversed the filter to compute covariance products and inverses in linear time, enabling the solution of structurally complex inverse problems. The accompanying application to cross-currency swap pricing illustrated how the same statistical ideas—state-space modeling, smoothing across data gaps, sparse covariance estimation, and scalable inversion—combine to address the estimation of frozen European rates against still-trading U.S. rates. Taken together, the four papers show that the careful recursive management of second-order structure is sufficient, and remarkably scalable, machinery for learning hidden processes from imperfect observations.

References

[1] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[2] J. Hartikainen and S. Särkkä. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 379–384, 2010.

[3] O. Al-Ghattas, J. Chen, D. Sanz-Alonso, and N. Waniorek. Covariance operator estimation: Sparsity, lengthscale, and ensemble Kalman filters. *arXiv preprint arXiv:2310.16933*, 2023.

[4] X. Fang and M. Gu. The inverse Kalman filter. *arXiv preprint arXiv:2407.10089*, 2024.

[5] G. Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2009.

[6] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[7] V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133, 2017.

[8] S. Särkkä. *Bayesian Filtering and Smoothing*. Cambridge University Press, 2013.

[9] D. Brigo and F. Mercurio. *Interest Rate Models—Theory and Practice: With Smile, Inflation and Credit*. Springer, 2006.