

Deriving the Discrete-Time Kalman Filter from Finite-Horizon Batch Estimation

Maïen Hamed

Abstract

The Kalman filter is typically presented as a sequence of Bayesian conditioning steps. While compact, that viewpoint can obscure where the recursion comes from algebraically and what global object it is implicitly factorizing. This note makes that structure explicit by centering the derivation on a single finite-horizon quantity: the inverse stacked innovation covariance $\mathbf{G}_{1:k}$. Viewing filtering through $\mathbf{G}_{1:k}$ clarifies how innovations are coupled across time by the dynamics, how conditioning and sensitivity appear in innovation space, and how fixed-window and smoothing variants correspond to retaining more of the same batch structure.

Concretely, state estimation over $1:k$ is posed as a constrained quadratic program in the stacked process disturbances and measurement noises. Eliminating the primal variables in the KKT system yields

$$\mathbf{G}_{1:k} = \left(\mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T + \mathbf{R}_{1:k} \right)^{-1}.$$

Extending the horizon from $k-1$ to k induces a structured block recursion for $\mathbf{G}_{1:k}$; the Schur complement of the appended block is exactly the per-step innovation covariance $\mathbf{S}_k$, and the Kalman gain appears as the corresponding block factor. Gaussianity is required only to interpret the quadratic program as a MAP estimator; all subsequent steps are deterministic linear algebra.

1 Introduction

The discrete-time Kalman filter is the canonical estimator for linear state-space models with additive Gaussian disturbances. It is most often introduced as a recursive Bayesian update, which concisely motivates the filter equations but can obscure the global algebraic structure from which the recursion originates.

This manuscript derives the Kalman filter by deriving, analyzing, and exploiting a single batch object: the inverse innovation covariance over a finite horizon. State estimation over

a horizon $1:k$ is formulated as a constrained quadratic program in the stacked process disturbances and measurement noises. Gaussian modeling assumptions enter only to justify a quadratic negative log-posterior (or equivalently a weighted least-norm) objective. From that point onward, the development is deterministic and linear-algebraic.

This paper is intended for readers already comfortable with linear systems, numerical linear algebra, and constrained optimization. It is not an introductory treatment of Kalman filtering. The objective is to make the structural relationship between the finite-horizon batch estimator and the standard recursion explicit.

By lifting the dynamics and measurement equations over the horizon, the stacked measurement residuals are expressed as an affine function of the stacked disturbances. The associated Karush-Kuhn-Tucker (KKT) conditions yield a structured saddle-point system. Eliminating the primal variables produces the inverse batch innovation matrix

$$\mathbf{G}_{1:k} = \left(\mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T + \mathbf{R}_{1:k} \right)^{-1},$$

which fully encodes the second-order structure of the finite-horizon estimation problem. This matrix defines the metric in innovation space, couples residuals across time through the dynamics, and maps the stacked residuals to the optimal dual variables and state corrections.

The central technical result of the paper is that $\mathbf{G}_{1:k}$ admits a structured one-step horizon-extension recursion. By embedding the $(k-1)$ -horizon quantities into the k -horizon space using canonical expansion and selection operators and applying block inversion identities, the Schur complement of the appended block appears as the per-step innovation covariance

$$\mathbf{S}_k = \mathbf{C}_k \mathbf{P}_{k|1:k-1} \mathbf{C}_k^T + \mathbf{R}_k.$$

The Kalman gain and the standard mean and covariance update equations arise directly as block factors in this recursion. From this viewpoint, the Kalman filter is a compressed representation of the horizon-extension properties of $\mathbf{G}_{1:k}$.

Emphasizing $\mathbf{G}_{1:k}$ exposes structure that is not visible in the stepwise recursion alone. Its eigenstructure characterizes conditioning of the stacked innovation covariance and reveals which linear combinations of innovations are weakly or strongly constrained by the model. Its blocks admit interpretation as cross-time innovation couplings induced by dynamics and process noise. Its associated dual variables quantify sensitivity of the optimum to perturbations of the measurement constraints. Finally, its batch form supports offline analysis, numerical experimentation, and fixed-window variants where partial parallelism is beneficial.

The remainder of the paper is organized as follows. Section 2 introduces the linear time-varying state-space model and the finite-horizon estimation viewpoint. Section 3 constructs the lifted operators and stacked equations. Section 4 formulates the constrained quadratic program, derives the KKT system, and identifies $\mathbf{G}_{1:k}$ as the central batch object. Section 5 derives the horizon-extension recursion for $\mathbf{G}_{1:k}$ and identifies $\mathbf{S}_k$ as a Schur complement. Finally, Section 6 shows how the standard Kalman filter equations emerge as a direct consequence of this recursion.

Relation to existing derivations. Classical presentations of the Kalman filter derive the recursion either from sequential Bayesian conditioning, from the information filter viewpoint, or from batch least-squares formulations followed by algebraic elimination [1, 2, 3]. The present derivation is closest in spirit to batch and information-form approaches, but differs in emphasis. Rather than eliminating state variables directly or working with information matrices at each step, the analysis centers on the inverse innovation covariance over a finite horizon, $\mathbf{G}_{1:k}$, and studies its horizon-extension structure. The standard Kalman recursion appears as a consequence of the block factorization of this single batch object.

Contributions. The derivation is organized around identifying the inverse stacked innovation covariance $\mathbf{G}_{1:k}$ as the central batch object produced by KKT elimination, and showing that the standard Kalman recursion is the horizon-by-horizon block factorization of $\mathbf{G}_{1:k}^{-1}$. In particular, the per-step innovation covariance $\mathbf{S}_k$ arises as a Schur complement under horizon extension, and the Kalman gain appears as the corresponding block factor mapping the newest innovation into the state correction.

Terminology (stacked innovation covariance). In this manuscript, $\mathbf{G}_{1:k}$ denotes the *inverse* of the stacked innovation covariance, while $\mathbf{G}_{1:k}^{-1}$ is the stacked innovation covariance itself. The term innovation covariance without qualification is reserved for the single-step quantity $\mathbf{S}_k$.

Notation

Table 1 summarizes the principal symbols used throughout the paper. Stacked quantities follow the index convention $i:j$. Lifted operators are block-structured with dimensions shown where they materially reduce ambiguity.

2 Linear Dynamical Model and Estimation Objective

2.1 Linear State Space Model

Consider the linear stochastic system with state $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and measurement $\mathbf{y}_k \in \mathbb{R}^{n_y}$ which, given the input sequence $\mathbf{u}_k \in \mathbb{R}^{n_u}$, are given by

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k, \quad (1)$$

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k. \quad (2)$$

Symbol	Meaning
$\mathbf{x}_k \in \mathbb{R}^{n_x}$	State at time k
$\mathbf{y}_k \in \mathbb{R}^{n_y}$	Measurement at time k
$\mathbf{u}_k \in \mathbb{R}^{n_u}$	Known input at time k
$\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}_k$	Dynamics, input, and measurement matrices
$\mathbf{w}_k, \mathbf{v}_k$	Process disturbance and measurement noise
$\mathbf{Q}_k, \mathbf{R}_k$	Covariances of $\mathbf{w}_k$ and $\mathbf{v}_k$
$\bar{\mathbf{x}}_{1:k}$	Open-loop mean state stack (no measurement correction)
$\Phi_{i,j}$	State transition from time j to i
$\mathbf{A}_{1:k} \in \mathbb{R}^{kn_x \times kn_x}$	Lifted transition operator over $1:k$ (block lower triangular)
$\mathbf{C}_{1:k} \in \mathbb{R}^{kn_y \times kn_x}$	Lifted measurement operator over $1:k$ (block diagonal)
$\mathbf{Q}_{0:k-1}, \mathbf{R}_{1:k}$	Lifted covariances (block diagonal)
$\hat{\mathbf{y}}_{1:k}$	Stacked measurement realizations
$\bar{\mathbf{r}}_{1:k}$	Open-loop residual stack: $\hat{\mathbf{y}}_{1:k} - \mathbf{C}_{1:k}\bar{\mathbf{x}}_{1:k}$
$\mathbf{w}_{0:k-1 1:k}, \mathbf{v}_{1:k 1:k}$	MAP disturbance and noise stacks consistent with $\hat{\mathbf{y}}_{1:k}$
$\boldsymbol{\xi}_{1:k 1:k}$	Stacked state correction: $\boldsymbol{\xi}_{1:k 1:k} = \mathbf{A}_{1:k}\mathbf{w}_{0:k-1 1:k}$
$\mathbf{x}_{k 1:k}$	Posterior state estimate using measurements $1:k$
$\mathbf{x}_{k 1:k-1}$	Prior state estimate using measurements $1:k-1$
$\mathbf{P}_{k 1:k}$	Posterior state covariance
$\mathbf{P}_{k 1:k-1}$	Prior state covariance
$\mathbf{S}_k$	Innovation covariance: $\mathbf{C}_k\mathbf{P}_{k 1:k-1}\mathbf{C}_k^T + \mathbf{R}_k$
$\mathbf{K}_k$	Kalman gain: $\mathbf{K}_k = \mathbf{P}_{k 1:k-1}\mathbf{C}_k^T\mathbf{S}_k^{-1}$
$\mathbf{r}_{k 1:k-1}$	Innovation: $\hat{\mathbf{y}}_k - \mathbf{C}_k\mathbf{x}_{k 1:k-1}$
$\mathbf{G}_{1:k}$	Inverse batch innovation matrix: $\left(\mathbf{C}_{1:k}\mathbf{A}_{1:k}\mathbf{Q}_{0:k-1}\mathbf{A}_{1:k}^T\mathbf{C}_{1:k}^T + \mathbf{R}_{1:k}\right)^{-1}$
$\mathbf{M}_{k 1:k} \in \mathbb{R}^{n_x \times kn_y}$	Residual-to-state gain map: $\mathbf{x}_{k 1:k} = \bar{\mathbf{x}}_k + \mathbf{M}_{k 1:k}\bar{\mathbf{r}}_{1:k}$
$\mathcal{E}_{n,k}, \mathcal{J}_{n,k}$	Canonical expansion and last-block selection operators

Table 1: Summary of notation used in the paper.

The process and measurement disturbances satisfy

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}_{n_x}, \mathbf{Q}_k), \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}_{n_y}, \mathbf{R}_k), \quad (3)$$

and are mutually uncorrelated across time.

The initial state is modeled as

$$\mathbf{x}_1 = \bar{\mathbf{x}}_1 + \mathbf{w}_0, \quad \mathbf{w}_0 \sim \mathcal{N}(\mathbf{0}_{n_x}, \mathbf{Q}_0). \quad (4)$$

Remark on indexing. Many Kalman filter presentations place the prior on $\mathbf{x}_0$. In this work, states are indexed starting at $k = 1$, with $\mathbf{x}_1$ modeled via $\bar{\mathbf{x}}_1$ and $\mathbf{Q}_0$. This choice reflects a 1-based indexing convention adopted throughout to maintain consistency with matrix indexing in the lifted formulations. Reindexing time by one step yields an equivalent formulation with identical recursions and estimators.

Probabilistic interpretation. Given measurements $\hat{\mathbf{y}}_{1:k}$, the unknown disturbance and noise stacks $(\mathbf{w}_{0:k-1}, \mathbf{v}_{1:k})$ are treated as latent random variables with Gaussian priors induced by $\mathbf{Q}_{0:k-1}$ and $\mathbf{R}_{1:k}$. The quadratic program derived below is the maximum a posteriori (MAP) estimator of $(\mathbf{w}_{0:k-1}, \mathbf{v}_{1:k})$ subject to the stacked measurement constraints. Because the priors are Gaussian and independent, the negative log-posterior is quadratic. The ensuing algebraic recursions follow from linear optimality conditions and Schur complements.

The goal is to estimate $\mathbf{x}_k$ sequentially from measurements $\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_k$. Rather than postulating a recursion a priori, we first construct a finite-horizon estimation problem whose solution is algebraically equivalent to the filter. The next section introduces a lifted representation over $1:k$ that makes the dependence of all measurements on the stacked disturbances explicit.

Assumptions on noise covariances. Throughout the paper, it is assumed that $\mathbf{Q}_k \succ 0$ and $\mathbf{R}_k \succ 0$ for all relevant indices. This guarantees uniqueness of the finite-horizon optimizer and invertibility of the batch innovation matrix $\mathbf{G}_{1:k}$. These assumptions are used implicitly in later Schur complement arguments.

All derivations are stated for time-varying systems. Time-invariant models are recovered as a special case by setting $\mathbf{A}_k = \mathbf{A}$, $\mathbf{C}_k = \mathbf{C}$, and constant covariances.

3 Finite Horizon Optimization Formulation

Having defined the linear dynamical model, we now reformulate estimation over a finite horizon as a deterministic optimization problem. The purpose of this section is to construct a lifted representation in which all measurements over $1:k$ depend linearly on the stacked disturbances and noises. This representation enables the estimation problem to be posed as a constrained quadratic program with explicit algebraic structure.

3.1 Expanded Measurement at Time k

Expanding the state recursion, the k th measurement satisfies

$$\begin{aligned}
\mathbf{y}_k &= \mathbf{C}_k \mathbf{A}_{k-1} \mathbf{A}_{k-2} \cdots \mathbf{A}_1 \bar{\mathbf{x}}_1 \\
&\quad + \mathbf{C}_k \mathbf{A}_{k-1} \mathbf{A}_{k-2} \cdots \mathbf{A}_1 \mathbf{w}_0 \\
&\quad + \mathbf{C}_k \mathbf{A}_{k-1} \mathbf{A}_{k-2} \cdots \mathbf{A}_2 (\mathbf{B}_1 \mathbf{u}_1 + \mathbf{w}_1) \\
&\quad + \mathbf{C}_k \mathbf{A}_{k-1} \mathbf{A}_{k-2} \cdots \mathbf{A}_2 (\mathbf{B}_2 \mathbf{u}_2 + \mathbf{w}_2) \\
&\quad + \cdots \\
&\quad + \mathbf{C}_k \mathbf{A}_{k-1} (\mathbf{B}_{k-2} \mathbf{u}_{k-2} + \mathbf{w}_{k-2}) \\
&\quad + \mathbf{C}_k (\mathbf{B}_{k-1} \mathbf{u}_{k-1} + \mathbf{w}_{k-1}) \\
&\quad + \mathbf{v}_k.
\end{aligned} \tag{5}$$

This expansion motivates the lifted representation used below. Each measurement $\mathbf{y}_k$ is an affine function of the initial condition, the input sequence, and the stacked disturbances $\mathbf{w}_{0:k-1}$, with coefficients given by products of the state transition matrices.

3.2 Stacked Dynamics and Measurement Equations

Define the state transition operator from time j to time i by

$$\Phi_{i,j} := \begin{cases} \mathbf{A}_{i-1} \mathbf{A}_{i-2} \cdots \mathbf{A}_j, & i > j, \\ \mathbf{I}_{n_x}, & i = j, \\ \mathbf{0}_{n_x \times n_x}, & i < j. \end{cases} \tag{6}$$

We extend this notation to block index ranges. For index intervals $i:j$ and $k:l$, define

$$\Phi_{i:j, k:l} := \begin{bmatrix} \Phi_{i,k} & \Phi_{i,k+1} & \cdots & \Phi_{i,l} \\ \Phi_{i+1,k} & \Phi_{i+1,k+1} & \cdots & \Phi_{i+1,l} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{j,k} & \Phi_{j,k+1} & \cdots & \Phi_{j,l} \end{bmatrix} \in \mathbb{R}^{(j-i+1)n_x \times (l-k+1)n_x}. \tag{7}$$

Singleton ranges are identified with single indices; for example, $\Phi_{i,1:k} \equiv \Phi_{i:i,1:k}$ and $\Phi_{1:k,j} \equiv \Phi_{1:k,j:j}$.

Lifted Transition Matrix Using this convention, the batch transition matrix over the horizon $1:k$ is defined as

$$\mathbf{A}_{1:k} := \Phi_{1:k,1:k} \in \mathbb{R}^{kn_x \times kn_x}. \tag{8}$$

By construction, $\mathbf{A}_{1:k}$ is block lower triangular.

3.3 Stacked Variables and Covariances

Define the stacked state, measurement, disturbance, and noise vectors

$$\mathbf{x}_{1:k} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_k \end{bmatrix}, \quad \mathbf{y}_{1:k} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{bmatrix}, \quad \mathbf{w}_{0:k-1} = \begin{bmatrix} \mathbf{w}_0 \\ \vdots \\ \mathbf{w}_{k-1} \end{bmatrix}, \quad \mathbf{v}_{1:k} = \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k \end{bmatrix}. \quad (9)$$

Introduce the block-diagonal matrices

$$\mathbf{C}_{1:k} := \text{blkdiag}(\mathbf{C}_1, \dots, \mathbf{C}_k) \in \mathbb{R}^{kn_y \times kn_x}, \quad (10)$$

$$\mathbf{Q}_{0:k-1} := \text{blkdiag}(\mathbf{Q}_0, \dots, \mathbf{Q}_{k-1}) \in \mathbb{R}^{kn_x \times kn_x}, \quad (11)$$

$$\mathbf{R}_{1:k} := \text{blkdiag}(\mathbf{R}_1, \dots, \mathbf{R}_k) \in \mathbb{R}^{kn_y \times kn_y}. \quad (12)$$

3.4 Open-Loop Mean and Input Lifting

Let

$$\mathbf{u}_{1:k-1} = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{k-1} \end{bmatrix} \quad (13)$$

denote the stacked input sequence and define the time-shifted subdiagonal batch input operator $\mathbf{B}_{1:k-1}^{(\text{prev})} \in \mathbb{R}^{kn_x \times (k-1)n_u}$ for $k \geq 2$ by

$$\mathbf{B}_{1:k-1}^{(\text{prev})} := \begin{bmatrix} \mathbf{0}_{n_x \times n_u} & \mathbf{0}_{n_x \times n_u} & \cdots & \mathbf{0}_{n_x \times n_u} \\ \mathbf{B}_1 & \mathbf{0}_{n_x \times n_u} & \cdots & \mathbf{0}_{n_x \times n_u} \\ \mathbf{0}_{n_x \times n_u} & \mathbf{B}_2 & \cdots & \mathbf{0}_{n_x \times n_u} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n_x \times n_u} & \mathbf{0}_{n_x \times n_u} & \cdots & \mathbf{B}_{k-1} \end{bmatrix}. \quad (14)$$

Define the open-loop state sequence mean

$$\bar{\mathbf{x}}_{1:k} := \begin{cases} \bar{\mathbf{x}}_1, & k = 1, \\ \Phi_{1:k,1} \bar{\mathbf{x}}_1 + \mathbf{A}_{1:k} \mathbf{B}_{1:k-1}^{(\text{prev})} \mathbf{u}_{1:k-1}, & k \geq 2, \end{cases} \quad (15)$$

which is the mean state stack obtained by setting $\mathbf{w}_{0:k-1} = \mathbf{0}$ and $\mathbf{v}_{1:k} = \mathbf{0}$.

Indexing sanity check (example $k = 3$). From (1), the open-loop means satisfy

$$\begin{aligned}\bar{\mathbf{x}}_2 &= \mathbf{A}_1 \bar{\mathbf{x}}_1 + \mathbf{B}_1 \mathbf{u}_1, \\ \bar{\mathbf{x}}_3 &= \mathbf{A}_2 \bar{\mathbf{x}}_2 + \mathbf{B}_2 \mathbf{u}_2 = \mathbf{A}_2 \mathbf{A}_1 \bar{\mathbf{x}}_1 + \mathbf{A}_2 \mathbf{B}_1 \mathbf{u}_1 + \mathbf{B}_2 \mathbf{u}_2,\end{aligned}$$

which matches (15) with the time-shifted input operator.

The stacked dynamics and measurement equations are then

$$\mathbf{x}_{1:k} = \bar{\mathbf{x}}_{1:k} + \mathbf{A}_{1:k} \mathbf{w}_{0:k-1}, \quad (16)$$

$$\mathbf{y}_{1:k} = \mathbf{C}_{1:k} \bar{\mathbf{x}}_{1:k} + \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{w}_{0:k-1} + \mathbf{v}_{1:k}. \quad (17)$$

4 Optimality Conditions and Normal Equations

Using the lifted measurement equation, estimation over the horizon $1:k$ can be posed as a constrained quadratic optimization problem. This section derives the Karush-Kuhn-Tucker conditions of that problem and identifies the inverse batch innovation matrix $\mathbf{G}_{1:k}$ as the central algebraic object governing the solution.

4.1 Estimation Objective

At time k , given the measurement realizations $(\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_k)$, define

$$\hat{\mathbf{y}}_{1:k} := \begin{bmatrix} \hat{\mathbf{y}}_1 \\ \vdots \\ \hat{\mathbf{y}}_k \end{bmatrix}. \quad (18)$$

Consider disturbance and noise stacks $(\mathbf{w}_{0:k-1|1:k}, \mathbf{v}_{1:k|1:k})$ consistent with these measurements, such that

$$\hat{\mathbf{y}}_{1:k} = \mathbf{C}_{1:k} \bar{\mathbf{x}}_{1:k} + \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{w}_{0:k-1|1:k} + \mathbf{v}_{1:k|1:k}. \quad (19)$$

The corresponding state estimate is

$$\mathbf{x}_{1:k|1:k} = \bar{\mathbf{x}}_{1:k} + \mathbf{A}_{1:k} \mathbf{w}_{0:k-1|1:k}. \quad (20)$$

Define the stacked open-loop residual

$$\bar{\mathbf{r}}_{1:k} := \hat{\mathbf{y}}_{1:k} - \mathbf{C}_{1:k} \bar{\mathbf{x}}_{1:k}, \quad (21)$$

which substitutes into (19) to give the stacked constraint

$$\mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{w}_{0:k-1|1:k} + \mathbf{v}_{1:k|1:k} = \bar{\mathbf{r}}_{1:k}. \quad (22)$$

For convenience, define the stacked correction sequence

$$\boldsymbol{\xi}_{1:k|1:k} := \mathbf{A}_{1:k} \mathbf{w}_{0:k-1|1:k} \in \mathbb{R}^{kn_x}, \quad (23)$$

so that

$$\mathbf{x}_{1:k|1:k} = \bar{\mathbf{x}}_{1:k} + \boldsymbol{\xi}_{1:k|1:k}. \quad (24)$$

4.2 KKT System

Assumptions for uniqueness and invertibility. Assume $\mathbf{Q}_i \succ 0$ for $i = 0, \dots, k-1$ and $\mathbf{R}_i \succ 0$ for $i = 1, \dots, k$. Then $\mathbf{Q}_{0:k-1}$ and $\mathbf{R}_{1:k}$ are symmetric positive definite, the KKT matrix is nonsingular, the optimizer is unique, and $\mathbf{G}_{1:k}$ exists. Extensions to semidefinite covariances require additional rank conditions and are not treated here.

Remark on semidefinite noise models. If $\mathbf{Q}_k$ or $\mathbf{R}_k$ are only positive semidefinite, the KKT system may remain solvable under appropriate rank conditions on the lifted operators, but $\mathbf{G}_{1:k}$ may fail to exist as a full inverse. In such cases, the analysis proceeds using generalized inverses and constrained subspace projections. These extensions are not pursued here.

The MAP problem is the constrained minimization

$$\min_{\mathbf{w}_{0:k-1}, \mathbf{v}_{1:k}} J(\mathbf{w}_{0:k-1}, \mathbf{v}_{1:k}) := \frac{1}{2} \mathbf{w}_{0:k-1}^T \mathbf{Q}_{0:k-1}^{-1} \mathbf{w}_{0:k-1} + \frac{1}{2} \mathbf{v}_{1:k}^T \mathbf{R}_{1:k}^{-1} \mathbf{v}_{1:k}, \quad (25)$$

subject to (22).

Introduce Lagrange multipliers $\bar{\boldsymbol{\lambda}}_i \in \mathbb{R}^{n_y}$ for each measurement constraint and stack them as

$$\bar{\boldsymbol{\lambda}}_{1:k} = \begin{bmatrix} \bar{\boldsymbol{\lambda}}_1 \\ \vdots \\ \bar{\boldsymbol{\lambda}}_k \end{bmatrix} \in \mathbb{R}^{kn_y}. \quad (26)$$

The KKT conditions yield the saddle-point linear system

$$\begin{bmatrix} \mathbf{Q}_{0:k-1}^{-1} & \mathbf{0}_{kn_x \times kn_y} & \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \\ \mathbf{0}_{kn_y \times kn_x} & \mathbf{R}_{1:k}^{-1} & \mathbf{I}_{kn_y} \\ \mathbf{C}_{1:k} \mathbf{A}_{1:k} & \mathbf{I}_{kn_y} & \mathbf{0}_{kn_y \times kn_y} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{w}}_{0:k-1|1:k} \\ \bar{\mathbf{v}}_{1:k|1:k} \\ \bar{\boldsymbol{\lambda}}_{1:k|1:k} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{kn_x} \\ \mathbf{0}_{kn_y} \\ \bar{\mathbf{r}}_{1:k} \end{bmatrix}. \quad (27)$$

Define the inverse batch innovation matrix

$$\mathbf{G}_{1:k} := \left(\mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T + \mathbf{R}_{1:k} \right)^{-1} \in \mathbb{R}^{kn_y \times kn_y}, \quad (28)$$

so that elimination yields

$$\bar{\mathbf{w}}_{0:k-1|1:k} = \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \mathbf{G}_{1:k} \bar{\mathbf{r}}_{1:k}, \quad (29a)$$

$$\bar{\mathbf{v}}_{1:k|1:k} = \mathbf{R}_{1:k} \mathbf{G}_{1:k} \bar{\mathbf{r}}_{1:k}, \quad (29b)$$

$$\bar{\lambda}_{1:k|1:k} = \mathbf{G}_{1:k} \bar{\mathbf{r}}_{1:k}. \quad (29c)$$

Weighted projection viewpoint. The solution $(\bar{\mathbf{w}}_{0:k-1|1:k}, \bar{\mathbf{v}}_{1:k|1:k})$ is the projection of $(\mathbf{0}, \mathbf{0})$ onto the affine constraint set defined by (22) in the norm induced by the weighting matrix $\text{blkdiag}(\mathbf{Q}_{0:k-1}^{-1}, \mathbf{R}_{1:k}^{-1})$.

The posterior covariance of $(\mathbf{w}_{0:k-1|1:k}, \mathbf{v}_{1:k|1:k})$ is

$$\text{Var} \left(\begin{bmatrix} \mathbf{w}_{0:k-1|1:k} \\ \mathbf{v}_{1:k|1:k} \end{bmatrix} \right) = \begin{bmatrix} \mathbf{Q}_{0:k-1} & \mathbf{0}_{kn_x \times kn_y} \\ \mathbf{0}_{kn_y \times kn_x} & \mathbf{R}_{1:k} \end{bmatrix} - \begin{bmatrix} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \\ \mathbf{R}_{1:k} \end{bmatrix} \mathbf{G}_{1:k} \begin{bmatrix} \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} & \mathbf{R}_{1:k} \end{bmatrix}. \quad (30)$$

Therefore

$$\mathbf{Q}_{0:k-1|1:k} := \text{Var}(\mathbf{w}_{0:k-1|1:k}) = \mathbf{Q}_{0:k-1} - \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \mathbf{G}_{1:k} \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1}, \quad (31)$$

and the posterior state covariance is

$$\mathbf{P}_{1:k|1:k} := \text{Var}(\mathbf{x}_{1:k|1:k}) = \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1|1:k} \mathbf{A}_{1:k}^T \in \mathbb{R}^{kn_x \times kn_x}. \quad (32)$$

In particular,

$$\begin{aligned} \mathbf{P}_{k|1:k} &= \Phi_{k,1:k} \mathbf{Q}_{0:k-1|1:k} \Phi_{k,1:k}^T \\ &= \Phi_{k,1:k} \left(\mathbf{Q}_{0:k-1} - \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \mathbf{G}_{1:k} \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \right) \Phi_{k,1:k}^T \in \mathbb{R}^{n_x \times n_x}. \end{aligned} \quad (33)$$

4.3 Interpretation and Uses of the Batch Innovation Matrix

The inverse batch innovation matrix

$$\mathbf{G}_{1:k} = \left(\mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T + \mathbf{R}_{1:k} \right)^{-1}$$

is the central second-order object in the finite-horizon estimator. The interpretations below motivate why $\mathbf{G}_{1:k}$ is useful to analyze, independent of whether one computes it explicitly.

Innovation-space geometry. The quadratic form $\bar{\mathbf{r}}_{1:k}^T \mathbf{G}_{1:k} \bar{\mathbf{r}}_{1:k}$ defines the squared norm of the open-loop residual $\bar{\mathbf{r}}_{1:k}$ in the metric induced jointly by the dynamics, process covariance, and measurement covariance. The MAP estimate corresponds to a weighted projection induced by $\text{blkdiag}(\mathbf{Q}_{0:k-1}^{-1}, \mathbf{R}_{1:k}^{-1})$.

Spectral structure and conditioning. The eigenstructure of $\mathbf{G}_{1:k}^{-1}$, which is the stacked innovation covariance, characterizes conditioning of the batch residual model. Large eigenvalues of $\mathbf{G}_{1:k}^{-1}$ correspond to innovation modes with high variance under the model (often dominated by accumulated process uncertainty), while small eigenvalues correspond to modes tightly constrained by measurement noise and dynamics.

Cross-time coupling of residuals. Off-diagonal blocks of $\mathbf{G}_{1:k}$ quantify how residual components at different times are coupled through dynamics and process noise.

Dual variables as sensitivity measures. The multipliers satisfy $\bar{\boldsymbol{\lambda}}_{1:k|1:k} = \mathbf{G}_{1:k} \bar{\mathbf{r}}_{1:k}$. Under standard convex duality, these multipliers quantify sensitivity of the optimum value to perturbations of the measurement constraints.

Batch estimation and smoothing. Solving the full KKT system yields the MAP disturbance and noise stacks, and hence the full MAP state stack $\mathbf{x}_{1:k|1:k}$. Extracting earlier states $\mathbf{x}_{j|1:k}$ for $j < k$ corresponds to fixed-interval smoothing. The filter recursion arises when only the last marginal is propagated.

Relation to information filters and smoothing. Because $\mathbf{G}_{1:k}^{-1}$ is the covariance of the stacked innovations, its block structure is closely related to information-form and square-root representations of the Kalman filter. Retaining all blocks corresponds to a fixed-interval batch estimator and yields Rauch–Tung–Striebel smoothing when earlier state marginals are extracted. The present formulation makes these connections explicit through Schur complements rather than through backward recursions.

5 Horizon Expansion and Recursive Structure

The batch formulas in Section 4 are global in time: they couple all measurements through $\mathbf{G}_{1:k}$. Filtering requires an update rule that reuses computations from the previous horizon. In this section we derive a horizon-extension recursion for $\mathbf{G}_{1:k}$ and for the gain quantities that map residuals to state corrections. The key step is a block decomposition of $\mathbf{G}_{1:k}^{-1}$ and evaluation of the associated Schur complement.

Roadmap for the block structure. In what follows, the lifted matrices are partitioned so that the upper-left block corresponds to the horizon $1:k-1$ and the lower-right block to the new measurement at time k . Expansion operators embed $(k-1)$ -horizon objects into the k -horizon space, while selection operators extract the newly appended components. With this structure in place, the Kalman filter appears as a direct consequence of Schur complements.

5.1 Expansion and Selection Operators

For each $k \geq 2$, define the canonical block expansion operator for block-dimension n

$$\mathcal{E}_{n,k} := \begin{bmatrix} \mathbf{I}_{(k-1)n} \\ \mathbf{0}_{n \times (k-1)n} \end{bmatrix} \in \mathbb{R}^{kn \times (k-1)n}, \quad (34)$$

and the last-block selection operator

$$\mathcal{J}_{n,k} := \begin{bmatrix} \mathbf{0}_{n \times (k-1)n} & \mathbf{I}_n \end{bmatrix} \in \mathbb{R}^{n \times kn}. \quad (35)$$

These satisfy

$$\mathcal{E}_{n,k}^T \mathcal{E}_{n,k} = \mathbf{I}_{(k-1)n}, \quad (36a)$$

$$\mathcal{J}_{n,k} \mathcal{E}_{n,k} = \mathbf{0}_{n \times (k-1)n}, \quad (36b)$$

$$\mathcal{J}_{n,k} \mathcal{J}_{n,k}^T = \mathbf{I}_n, \quad (36c)$$

$$\mathcal{E}_{n,k} \mathcal{E}_{n,k}^T + \mathcal{J}_{n,k}^T \mathcal{J}_{n,k} = \mathbf{I}_{kn}. \quad (36d)$$

We note that

$$\Phi_{k,1:k} = \Phi_{k,1:k-1} \mathcal{E}_{n_x,k}^T + \mathcal{J}_{n_x,k} = \mathbf{A}_{k-1} \Phi_{k-1,1:k-1} \mathcal{E}_{n_x,k}^T + \mathcal{J}_{n_x,k}, \quad (37)$$

and

$$\mathbf{A}_{1:k} = \mathcal{E}_{n_x,k} \mathbf{A}_{1:k-1} \mathcal{E}_{n_x,k}^T + \mathcal{J}_{n_x,k}^T \Phi_{k,1:k}. \quad (38)$$

5.2 Batch MAP Matrix and the Schur Complement

This subsection derives the block structure of $\mathbf{G}_{1:k}^{-1}$ under horizon extension and identifies $\mathbf{S}_k$ as a Schur complement. The goal is to express $\mathbf{G}_{1:k}$ and the residual-to-state map $\mathbf{M}_{k|1:k}$ in terms of $(k-1)$ -horizon quantities.

For $k \geq 2$, the matrix $\mathbf{G}_{1:k}$ can be updated from $\mathbf{G}_{1:k-1}$ using $\mathbf{Q}_{k-1}$, $\mathbf{A}_{k-1}$, $\mathbf{C}_k$, and $\mathbf{R}_k$ through block inversion identities (Appendix A).

First, note that

$$\mathbf{R}_{1:k} = \mathcal{E}_{n_y,k} \mathbf{R}_{1:k-1} \mathcal{E}_{n_y,k}^T + \mathcal{J}_{n_y,k}^T \mathbf{R}_k \mathcal{J}_{n_y,k} = \begin{bmatrix} \mathbf{R}_{1:k-1} & \mathbf{0}_{(k-1)n_y \times n_y} \\ \mathbf{0}_{n_y \times (k-1)n_y} & \mathbf{R}_k \end{bmatrix}, \quad (39)$$

and

$$\begin{aligned} \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T &= \mathcal{E}_{n_y,k} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathcal{E}_{n_y,k}^T \\ &\quad + \mathcal{J}_{n_y,k}^T \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathcal{E}_{n_y,k}^T \\ &\quad + \mathcal{E}_{n_y,k} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T \mathcal{J}_{n_y,k} \\ &\quad + \mathcal{J}_{n_y,k}^T \mathbf{C}_k \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \Phi_{k,1:k}^T \mathbf{C}_k^T \mathcal{J}_{n_y,k} \\ &= \begin{bmatrix} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T & \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T \\ \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T & \mathbf{C}_k \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \Phi_{k,1:k}^T \mathbf{C}_k^T \end{bmatrix}. \end{aligned} \quad (40)$$

Substituting into (28) yields the block form

$$\begin{aligned} \mathbf{G}_{1:k}^{-1} &= \mathcal{E}_{n_y,k} \mathbf{G}_{1:k-1}^{-1} \mathcal{E}_{n_y,k}^T \\ &\quad + \mathcal{J}_{n_y,k}^T \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathcal{E}_{n_y,k}^T \\ &\quad + \mathcal{E}_{n_y,k} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T \mathcal{J}_{n_y,k} \\ &\quad + \mathcal{J}_{n_y,k}^T \left(\mathbf{R}_k + \mathbf{C}_k \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \Phi_{k,1:k}^T \mathbf{C}_k^T \mathcal{J}_{n_y,k} \right) \\ &= \begin{bmatrix} \mathbf{G}_{1:k-1}^{-1} & \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T \\ \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T & \mathbf{C}_k \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \Phi_{k,1:k}^T \mathbf{C}_k^T + \mathbf{R}_k \end{bmatrix}. \end{aligned} \quad (41)$$

Identification of the Schur complement with $\mathbf{S}_k$. Let $\mathbf{P}_{k-1|1:k-1}$ denote the posterior covariance at step $k-1$, and define the predicted covariance

$$\mathbf{P}_{k|1:k-1} := \mathbf{A}_{k-1} \mathbf{P}_{k-1|1:k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}. \quad (42)$$

Applying the block inverse formula (63) to (41), and using (33), the Schur complement of the $(k-1)$ -horizon block $\mathbf{G}_{1:k-1}^{-1}$ is

$$\begin{aligned} \mathbf{S}_k &:= \mathbf{C}_k \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \Phi_{k,1:k}^T \mathbf{C}_k^T + \mathbf{R}_k \\ &\quad - \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathbf{G}_{1:k-1}^{-1} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T \\ &= \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T + \mathbf{C}_k \mathbf{Q}_{k-1} \mathbf{C}_k^T + \mathbf{R}_k \\ &\quad - \mathbf{C}_k \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathbf{G}_{1:k-1}^{-1} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k,1:k-1}^T \mathbf{C}_k^T \\ &= \mathbf{C}_k \Phi_{k,1:k-1} \left(\mathbf{Q}_{0:k-2} - \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathbf{G}_{1:k-1}^{-1} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \right) \Phi_{k,1:k-1}^T \mathbf{C}_k^T \\ &\quad + \mathbf{C}_k \mathbf{Q}_{k-1} \mathbf{C}_k^T + \mathbf{R}_k \\ &= \mathbf{C}_k \left(\mathbf{A}_{k-1} \mathbf{P}_{k-1|1:k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \right) \mathbf{C}_k^T + \mathbf{R}_k \\ &= \mathbf{C}_k \mathbf{P}_{k|1:k-1} \mathbf{C}_k^T + \mathbf{R}_k. \end{aligned} \quad (43)$$

which is the standard per-step innovation covariance.

5.3 Residual-to-State Map and Horizon Extension

Objective of this subsection. This subsection derives a recursive expression for the residual-to-state map $\mathbf{M}_{k|1:k}$, showing how the Kalman gain emerges as the factor that maps the newest innovation into the state correction.

Define the gain mapping residuals up to step k to the correction at step k by

$$\mathbf{M}_{k|1:k} := \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \mathbf{G}_{1:k} \in \mathbb{R}^{n_x \times k n_y}, \quad (44)$$

so that

$$\bar{\mathbf{x}}_{k|1:k} = \bar{\mathbf{x}}_k + \mathbf{M}_{k|1:k} \bar{\mathbf{r}}_{1:k}. \quad (45)$$

Applying (63) to (41) yields the horizon-extension recursion

$$\mathbf{G}_{1:k} = \begin{bmatrix} \mathbf{G}_{1:k-1} + \mathbf{M}_{k-1|1:k-1}^T \mathbf{A}_{k-1}^T \mathbf{C}_k^T \mathbf{S}_k^{-1} \mathbf{C}_k \mathbf{A}_{k-1} \mathbf{M}_{k-1|1:k-1} & -\mathbf{M}_{k-1|1:k-1}^T \mathbf{A}_{k-1}^T \mathbf{C}_k^T \mathbf{S}_k^{-1} \\ -\mathbf{S}_k^{-1} \mathbf{C}_k \mathbf{A}_{k-1} \mathbf{M}_{k-1|1:k-1} & \mathbf{S}_k^{-1} \end{bmatrix}. \quad (46)$$

To compute $\mathbf{M}_{k|1:k}$ recursively from $\mathbf{M}_{k-1|1:k-1}$, define the open loop state covariance

$$\mathbf{P}_k := \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \Phi_{k,1:k}^T. \quad (47)$$

Substituting this into (44), (31), and (33) gives

$$\Phi_{k,1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T \mathbf{M}_{k|1:k}^T = \mathbf{P}_k - \mathbf{P}_{k|1:k}. \quad (48)$$

Next, note that

$$\begin{aligned} \Phi_{k,1:k} \mathbf{Q}_{0:k-1} \mathbf{A}_{1:k}^T \mathbf{C}_{1:k}^T &= \left(\Phi_{k,1:k-1} \mathcal{E}_{n_x,k}^T + \mathcal{J}_{n_x,k} \right) \\ &\quad \left(\mathcal{E}_{n_x,k} \mathbf{Q}_{0:k-2} \mathcal{E}_{n_x,k}^T + \mathcal{J}_{n_x,k}^T \mathbf{Q}_{k-1} \mathcal{J}_{n_x,k} \right) \\ &\quad \left(\mathcal{E}_{n_x,k} \mathbf{A}_{1:k-1}^T \mathcal{E}_{n_x,k}^T + \mathcal{E}_{n_x,k} \mathbf{A}_{k-1} \Phi_{k-1,1:k-1}^T \mathcal{J}_{n_x,k} + \mathcal{J}_{n_x,k}^T \mathcal{J}_{n_x,k} \right) \\ &\quad \left(\mathcal{E}_{n_x,k} \mathbf{C}_{1:k-1}^T \mathcal{E}_{n_y,k}^T + \mathcal{J}_{n_x,k}^T \mathbf{C}_k^T \mathcal{J}_{n_y,k} \right) \\ &= \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathcal{E}_{n_y,k}^T \\ &\quad + \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \Phi_{k-1,1:k-1}^T \mathbf{A}_{k-1}^T \mathbf{C}_k^T \mathcal{J}_{n_y,k} \\ &\quad + \mathbf{Q}_{k-1} \mathbf{C}_k^T \mathcal{J}_{n_y,k} \\ &= \Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \mathcal{E}_{n_y,k}^T + \left(\mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \right) \mathbf{C}_k^T \mathcal{J}_{n_y,k} \\ &= \left[\Phi_{k,1:k-1} \mathbf{Q}_{0:k-2} \mathbf{A}_{1:k-1}^T \mathbf{C}_{1:k-1}^T \quad \left(\mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \right) \mathbf{C}_k^T \right]. \quad (49) \end{aligned}$$

Then

$$\begin{aligned}
M_{k|1:k} &= \Phi_{k,1:k} Q_{0:k-1} A_{1:k}^T C_{1:k}^T G_{1:k} \\
&= \Phi_{k,1:k-1} Q_{0:k-2} A_{1:k-1}^T C_{1:k-1}^T G_{1:k-1} \mathcal{E}_{n_y,k}^T \\
&\quad + \Phi_{k,1:k-1} Q_{0:k-2} A_{1:k-1}^T C_{1:k-1}^T M_{k-1|1:k-1}^T A_{k-1}^T C_k^T S_k^{-1} C_k A_{k-1} M_{k-1|1:k-1} \mathcal{E}_{n_y,k}^T \\
&\quad - \left(A_{k-1} P_{k-1} A_{k-1}^T + Q_{k-1} \right) C_k^T S_k^{-1} C_k A_{k-1} M_{k-1|1:k-1} \mathcal{E}_{n_y,k}^T \\
&\quad - \Phi_{k,1:k-1} Q_{0:k-2} A_{1:k-1}^T C_{1:k-1}^T M_{k-1|1:k-1}^T A_{k-1}^T C_k^T S_k^{-1} \mathcal{J}_{n_y,k} \\
&\quad + \left(A_{k-1} P_{k-1} A_{k-1}^T + Q_{k-1} \right) C_k^T S_k^{-1} \mathcal{J}_{n_y,k} \\
&= A_{k-1} M_{k-1|1:k-1} \mathcal{E}_{n_y,k}^T \\
&\quad - \left(A_{k-1} P_{k-1|1:k-1} A_{k-1}^T + Q_{k-1} \right) C_k^T S_k^{-1} C_k A_{k-1} M_{k-1|1:k-1} \mathcal{E}_{n_y,k}^T \\
&\quad + \left(A_{k-1} P_{k-1|1:k-1} A_{k-1}^T + Q_{k-1} \right) C_k^T S_k^{-1} \mathcal{J}_{n_y,k} .
\end{aligned} \tag{50}$$

Define the Kalman gain

$$K_k := P_{k|1:k-1} C_k^T S_k^{-1}. \tag{51}$$

Then the residual-to-state recursion is

$$\begin{aligned}
M_{k|1:k} &= (I_{n_x} - K_k C_k) A_{k-1} M_{k-1|1:k-1} \mathcal{E}_{n_y,k}^T + K_k \mathcal{J}_{n_y,k} \\
&= \begin{bmatrix} (I_{n_x} - K_k C_k) A_{k-1} M_{k-1|1:k-1} & K_k \end{bmatrix} .
\end{aligned} \tag{52}$$

6 Kalman Filter Expressions

This section translates the horizon-extension relations into the conventional Kalman filter notation. The posterior mean is expressed as a prior plus a gain times the innovation, and the posterior covariance update is recovered in a standard symmetric form.

The open-loop residual $\bar{\mathbf{r}}_k$ at time k is $\bar{\mathbf{r}}_k = \hat{\mathbf{y}}_k - \mathbf{C}_k \bar{\mathbf{x}}_k$. The filtered innovation uses the predicted mean $\bar{\mathbf{x}}_{k|1:k-1}$ and is defined below.

6.1 Mean update

Define the prior state and innovation at step k based on measurements up to $k-1$ as

$$\bar{\mathbf{x}}_{k|1:k-1} := A_{k-1} \bar{\mathbf{x}}_{k-1|1:k-1} + B_{k-1} \mathbf{u}_{k-1}, \tag{53}$$

$$\bar{\mathbf{r}}_{k|1:k-1} := \hat{\mathbf{y}}_k - C_k \bar{\mathbf{x}}_{k|1:k-1}. \tag{54}$$

Remark (open-loop residual versus innovation). Using $\mathbf{x}_{k|1:k-1} = \bar{\mathbf{x}}_k + \mathbf{A}_{k-1}\boldsymbol{\xi}_{k-1|1:k-1}$, one obtains the explicit relationship

$$\bar{\mathbf{r}}_{k|1:k-1} = \bar{\mathbf{r}}_k - \mathbf{C}_k \mathbf{A}_{k-1} \bar{\boldsymbol{\xi}}_{k-1|1:k-1}. \quad (55)$$

Then the posterior mean update is

$$\bar{\mathbf{x}}_{k|1:k} = \bar{\mathbf{x}}_{k|1:k-1} + \mathbf{K}_k \bar{\mathbf{r}}_{k|1:k-1}. \quad (56)$$

6.2 Covariance update

Substituting into equation (48) transposed gives

$$\begin{aligned} \mathbf{P}_k - \mathbf{P}_{k|1:k} &= \mathbf{M}_{k|1:k} \mathbf{C}_{1:k} \mathbf{A}_{1:k} \mathbf{Q}_{0:k-1} \boldsymbol{\Phi}_{k,1:k}^T \\ &= \left[(\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) \mathbf{A}_{k-1} \mathbf{M}_{k-1|1:k-1} \quad \mathbf{K}_k \right] \begin{bmatrix} \mathbf{C}_{1:k-1} \mathbf{A}_{1:k-1} \mathbf{Q}_{0:k-2} \boldsymbol{\Phi}_{k-1,1:k-1}^T \\ \mathbf{C}_k \mathbf{P}_k \end{bmatrix} \\ &= (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) \mathbf{A}_{k-1} (\mathbf{P}_{k-1} - \mathbf{P}_{k-1|1:k-1}) \mathbf{A}_{k-1}^T + \mathbf{K}_k \mathbf{C}_k \mathbf{P}_k, \end{aligned} \quad (57)$$

so that

$$\begin{aligned} \mathbf{P}_{k|1:k} &= \mathbf{P}_k - (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) \mathbf{A}_{k-1} (\mathbf{P}_{k-1} - \mathbf{P}_{k-1|1:k-1}) \mathbf{A}_{k-1}^T - \mathbf{K}_k \mathbf{C}_k \mathbf{P}_k \\ &= (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) (\mathbf{A}_{k-1} \mathbf{P}_{k-1|1:k-1} \mathbf{A}_{k-1}^T + \mathbf{P}_k - \mathbf{A}_{k-1} \mathbf{P}_{k-1} \mathbf{A}_{k-1}^T) \\ &= (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) (\mathbf{A}_{k-1} \mathbf{P}_{k-1|1:k-1} \mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}) \\ &= (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) \mathbf{P}_{k|1:k-1}. \end{aligned} \quad (58)$$

Substituting $\mathbf{K}_k = \mathbf{P}_{k|1:k-1} \mathbf{C}_k^T \mathbf{S}_k^{-1}$ also yields the standard symmetric update

$$\mathbf{P}_{k|1:k} = \mathbf{P}_{k|1:k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T. \quad (59)$$

For numerical robustness, one may use the Joseph form

$$\mathbf{P}_{k|1:k} = (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k) \mathbf{P}_{k|1:k-1} (\mathbf{I}_{n_x} - \mathbf{K}_k \mathbf{C}_k)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T. \quad (60)$$

7 Discussion and Conclusion

This manuscript derived the discrete-time Kalman filter by centering the analysis on the batch innovation matrix $\mathbf{G}_{1:k}$ arising from a finite-horizon quadratic estimation problem. After the modeling step that yields a quadratic negative log-posterior, all results follow from

deterministic linear algebra: the KKT system, elimination of primal variables, and Schur complement identities under horizon extension.

The matrix $\mathbf{G}_{1:k}$ encodes the complete second-order structure relating the stacked measurement residuals to optimal disturbance corrections, dual variables, and posterior state estimates. Its inverse is the covariance of the stacked innovations, including cross-time correlations induced by dynamics and process noise.

Extending the horizon from $k - 1$ to k yields a structured block recursion for $\mathbf{G}_{1:k}$. The Schur complement of the newly appended block is the per-step innovation covariance $\mathbf{S}_k = \mathbf{C}_k \mathbf{P}_{k|1:k-1} \mathbf{C}_k^T + \mathbf{R}_k$, and the Kalman gain appears as the corresponding block factor. From this perspective, the Kalman filter recursion is the horizon-by-horizon factorization of a single structured batch object.

This formulation clarifies which assumptions are essential. The algebraic recursions require only a quadratic objective and positive definiteness of the covariance blocks. Gaussianity is required only to interpret the optimizer as a MAP estimator. Modifying the objective yields alternative estimators governed by the same lifting and horizon-extension principles, even when the resulting recursion is no longer linear.

Computational implications. The batch derivation shows that filtering corresponds to a particular elimination ordering of a structured linear system. In fixed-lag or fixed-window settings, the same KKT system admits factorizations that can be reused across horizon shifts, yielding efficient sliding-window solvers. Sparsity and block-banded structure are properties of the KKT system (and of equivalent information-form normal equations in the state variables), not of $\mathbf{G}_{1:k}^{-1}$ in general. Consequently, practical algorithms exploit structured factorizations of the KKT matrix or square-root / information-form representations, rather than explicitly forming $\mathbf{G}_{1:k}$.

The batch viewpoint also exposes extensions that are less visible from the stepwise recursion alone. Retaining earlier marginals yields smoothing. Spectral analysis of $\mathbf{G}_{1:k}^{-1}$ reveals conditioning of stacked innovation modes. The dual variables quantify sensitivity to measurement constraints. In fixed-window and iteratively linearized settings, the same block structure and Schur complements govern factorization reuse and update strategies.

A 2×2 block matrix inverse (Schur complement)

Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad (61)$$

where $\mathbf{A} \in \mathbb{R}^{m \times m}$ and $\mathbf{D} \in \mathbb{R}^{n \times n}$.

A.1 Inverse using the Schur complement of A

Assume $\mathbf{A}$ is invertible and define the Schur complement

$$\mathbf{S}_A := \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}. \quad (62)$$

If $\mathbf{S}_A$ is invertible, then $\mathbf{M}$ is invertible and

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{S}_A^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{S}_A^{-1} \\ -\mathbf{S}_A^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{S}_A^{-1} \end{bmatrix}. \quad (63)$$

A.2 Inverse using the Schur complement of D

Assume $\mathbf{D}$ is invertible and define the Schur complement

$$\mathbf{S}_D := \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}. \quad (64)$$

If $\mathbf{S}_D$ is invertible, then

$$\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{S}_D^{-1} & -\mathbf{S}_D^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{S}_D^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{S}_D^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}. \quad (65)$$

A.3 Woodbury matrix identity

Comparing the top-left blocks of (63) and (65) gives the Woodbury matrix identity:

$$\left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (66)$$

References

- [1] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [2] P. S. Maybeck, *Stochastic Models, Estimation, and Control, Volume 1*, Academic Press, New York, 1979.
- [3] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice–Hall, Englewood Cliffs, NJ, 1979.