Closing the Gap: Online Probabilistic Discrepancy Models

Aaron Danielson

September 6, 2025

1 Data and Definitions

Let $\mathcal{D}_t = \{(s_j, y_j)\}_{j=1}^t$ where $s_j \in \mathbb{R}^n$ is the vector of simulated values and $y_j \in \mathbb{R}^n$ is the vector of observed values at time j. Define the per-time discrepancy

$$z_j \;=\; y_j - s_j \in \mathbb{R}^n.$$

For random variables constrained to a simplex (e.g., team-minutes per quarter), the discrepancy across players sums to zero at each time t:

$$\mathbf{1}^\top z_t \;=\; 0.$$

We enforce this exactly by working in the $(n-1)$-dimensional zero-sum subspace. Let

$$\Pi \;\equiv\; I_n - \tfrac{1}{n}\mathbf{1}\mathbf{1}^\top.$$

Compute a thin QR (or any orthonormal basis) of Π to obtain $C \in \mathbb{R}^{n\times(n-1)}$ satisfying

$$C^\top C = I_{n-1}, \qquad C^\top \mathbf{1} = 0, \qquad CC^\top = \Pi.$$

Any zero-sum vector $x \in \mathbb{R}^n$ can be represented as $x = C\eta$ with a unique $\eta \in \mathbb{R}^{n-1}$.

2 Probability Model

Our goal is to learn the posterior distribution of the discrepancies z_t given the observed data $\mathcal{D}_t$. We learn this distribution by modeling the observed discrepancy as a function of a lower-dimensional latent state

$$z_t = C\eta_t + \epsilon_t$$

where $\epsilon_t \sim \mathcal{N}(0, R)$ where R is the covariance matrix of the dimensions of the observation noise and η_t is the $(n-1)$-dimensional vector of latent discrepancies. That is, we observe the discrepancies z_t confounded by noise. Our sim is to learn the trajectory of latent discrepancy parameters over time. The n-dimensional discrepancy is recovered by premultiplication with C The predicted latent discrepancy at time $t+1$ as a function of the history of data observed at time t evolves according to

$$\eta_{t+1|t} = A\eta_t + \omega_t$$

where $\eta_{t+1|t}$ denotes the value of the lower-dimensional latent discrepancy at $t+1$ given data up to time t, $A \in \mathbb{R}^{(n-1)\times(n-1)}$ is the state transition matrix, and $\omega_t \sim \mathcal{N}(0, Q)$ with $Q \succeq 0$ process noise. As an example, if we assume $A = I_{n-1}$, i.e., a random walk model, then

$$\eta_{t+1|t} = \eta_t + \omega_t.$$

The filter maintains the posterior mean and covariance $(\eta_{t|t}, \Sigma_{t|t})$, where $\Sigma_{t|t} = \mathrm{Cov}(\eta_t \mid \mathcal{D}_t)$.

3 Algorithm Details

3.1 Initialization

- Build C from a QR of $\Pi = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$.
- Set $\eta_{0|0} = 0 \in \mathbb{R}^{n-1}$ and $\Sigma_{0|0} = \tau^2 I_{n-1}$ with a large τ.
- Set A (e.g., I), $Q_0 = \sigma_Q^2 I_{n-1}$, $R_0 = \sigma_R^2 I_n$.

3.2 Prediction

$$\eta_{t|t-1} = A\,\eta_{t-1|t-1}, \qquad \Sigma_{t|t-1} = A\,\Sigma_{t-1|t-1}A^\top + Q_{t-1}.$$

3.3 Update

Given (s_t, y_t), form $z_t = y_t - s_t$; if needed, project $z_t \leftarrow \Pi z_t$.

$$\begin{aligned}
v_t &= z_t - C\,\eta_{t|t-1} && \text{(innovation)}\\
S_t &= C\,\Sigma_{t|t-1}C^\top + R_{t-1} && \text{(innovation covariance)}\\
K_t &= \Sigma_{t|t-1}C^\top S_t^{-1} && \text{(Kalman gain)}\\
\eta_{t|t} &= \eta_{t|t-1} + K_t v_t && \text{(state update)}\\
\Sigma_{t|t} &= (I - K_t C)\,\Sigma_{t|t-1} && \text{(covariance update)}
\end{aligned}$$

Adaptive R (innovation matching). With forgetting factor $\lambda \in (0,1)$,

$$\widehat{R}_t^{\text{raw}} = (1-\lambda)R_{t-1} + \lambda\big(v_t v_t^\top - C\,\Sigma_{t|t-1}C^\top\big).$$

Optionally shrink to ensure PSD:

$$R_t = (1-\alpha)\,\widehat{R}_t^{\text{raw}} + \alpha\,\frac{\operatorname{tr}(\widehat{R}_t^{\text{raw}})}{n}\,I_n, \qquad \alpha \in [0,1).$$

Adaptive Q (state residual matching). With forgetting factor $\gamma \in (0,1)$, define the latent residual

$$u_t = \eta_{t|t} - A\,\eta_{t-1|t-1}, \qquad Q_t = (1-\gamma)Q_{t-1} + \gamma\,(u_t u_t^\top).$$

3.4 Algorithm Pseudocode

Algorithm 1 Online Zero-Sum Subspace Kalman Filter with Adaptive R and Q

Require: Streams $\{(s_t, y_t)\}_{t\ge 1}$; step sizes $\lambda, \gamma \in (0,1)$; shrinkage $\alpha \in [0,1)$

Construct $\Pi \leftarrow I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$; obtain $C \in \mathbb{R}^{n\times(n-1)}$ with $C^\top C = I$, $C^\top \mathbf{1} = 0$, $CC^\top = \Pi$
Init $\eta_{0|0} \leftarrow 0$, $\Sigma_{0|0} \leftarrow \tau^2 I$, $A \leftarrow I$, $Q_0 \leftarrow \sigma_Q^2 I$, $R_0 \leftarrow \sigma_R^2 I$
for $t = 1, 2, \ldots$ **do**
  $z_t \leftarrow y_t - s_t$ ▷ Optionally: $z_t \leftarrow \Pi z_t$ to enforce zero-sum
  Predict: $\eta_{t|t-1} \leftarrow A\,\eta_{t-1|t-1}$; $\Sigma_{t|t-1} \leftarrow A\,\Sigma_{t-1|t-1}A^\top + Q_{t-1}$
  Innovation: $v_t \leftarrow z_t - C\,\eta_{t|t-1}$; $S_t \leftarrow C\,\Sigma_{t|t-1}C^\top + R_{t-1}$
  Gain: $K_t \leftarrow \Sigma_{t|t-1}C^\top S_t^{-1}$
  Update: $\eta_{t|t} \leftarrow \eta_{t|t-1} + K_t v_t$; $\Sigma_{t|t} \leftarrow (I - K_t C)\Sigma_{t|t-1}$
  Adapt R: $\widehat{R}_t^{\text{raw}} \leftarrow (1-\lambda)R_{t-1} + \lambda(v_t v_t^\top - C\Sigma_{t|t-1}C^\top)$
  $R_t \leftarrow (1-\alpha)\widehat{R}_t^{\text{raw}} + \alpha\,\frac{\operatorname{tr}(\widehat{R}_t^{\text{raw}})}{n}\,I$
  Adapt Q: $u_t \leftarrow \eta_{t|t} - A\,\eta_{t-1|t-1}$; $Q_t \leftarrow (1-\gamma)Q_{t-1} + \gamma(u_t u_t^\top)$
  (Optional output) $\hat{\delta}_{t|t} \leftarrow C\,\eta_{t|t}$ ▷ Zero-sum by construction
end for

Algorithm 2 Online Zero-Sum Kalman Filter with Adaptive R and Q

Require: Streams $\{(s_t, y_t)\}_{t\geq 1}$ with $s_t, y_t \in \mathbb{R}^n$; step sizes $\lambda_R, \gamma_Q \in (0,1)$; shrinkage $\alpha \in [0,1)$

Construct zero-sum basis: Form $C \in \mathbb{R}^{n\times(n-1)}$ s.t. $C^\top C = I_{n-1}$ and $C^\top \mathbf{1} = 0$ (e.g., QR of $I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$).

Initialize: $\eta_{0|0} \leftarrow 0 \in \mathbb{R}^{n-1}$; $\Sigma_{0|0} \leftarrow \tau^2 I_{n-1}$ with large τ.

Set models: $A \in \mathbb{R}^{(n-1)\times(n-1)}$ (e.g., $A = I_{n-1}$); $Q_0 \leftarrow \sigma_Q^2 I_{n-1}$; $R_0 \leftarrow \sigma_R^2 I_n$.

for $t = 1, 2, \dots$ **do**

  Form discrepancy: $z_t \leftarrow y_t - s_t \in \mathbb{R}^n$ ▷ $\mathbf{1}^\top z_t = 0$ should hold (or project: $z_t \leftarrow (I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top) z_t$)

  Predict (time update):

    $\eta_{t|t-1} \leftarrow A\, \eta_{t-1|t-1}$

    $\Sigma_{t|t-1} \leftarrow A\, \Sigma_{t-1|t-1} A^\top + Q_{t-1}$

  Innovation (measurement prediction):

    $v_t \leftarrow z_t - C\, \eta_{t|t-1}$ ▷ innovation in player space

    $S_t \leftarrow C\, \Sigma_{t|t-1}\, C^\top + R_{t-1}$ ▷ innovation covariance

  Kalman gain:

    $K_t \leftarrow \Sigma_{t|t-1}\, C^\top\, S_t^{-1}$

  Measurement update:

    $\eta_{t|t} \leftarrow \eta_{t|t-1} + K_t\, v_t$

    $\Sigma_{t|t} \leftarrow (I_{n-1} - K_t C)\, \Sigma_{t|t-1}$

  Adaptive R (innovation matching):

    $\widehat{R}_t^{\text{raw}} \leftarrow (1-\lambda_R)\, R_{t-1} + \lambda_R \big(v_t v_t^\top - C\, \Sigma_{t|t-1}\, C^\top \big)$

    $R_t \leftarrow (1-\alpha)\, \widehat{R}_t^{\text{raw}} + \alpha\, \frac{\operatorname{tr}(\widehat{R}_t^{\text{raw}})}{n}\, I_n$ ▷ optional shrinkage

  Adaptive Q (state residual matching):

    $u_t \leftarrow \eta_{t|t} - A\, \eta_{t-1|t-1}$ ▷ latent residual

    $Q_t \leftarrow (1-\gamma_Q)\, Q_{t-1} + \gamma_Q\, (u_t u_t^\top)$

  (Optional) Output player discrepancies: $\hat{\delta}_{t|t} \leftarrow C\, \eta_{t|t}$ ▷ $\mathbf{1}^\top \hat{\delta}_{t|t} = 0$ by construction

end for

4 Posterior Predictive Distribution

Suppose we have observed data up to time t and updated the latent state and covariance $(\eta_{t|t}, \Sigma_{t|t})$ as well as the noise covariances Q_t and R_t. We are interested in the predictive distribution of the next discrepancy z_{t+1}.

Latent state prediction

The latent discrepancy one step ahead follows the state transition model:

$$\eta_{t+1|t} \;=\; A\, \eta_{t|t}, \qquad \Sigma_{t+1|t} \;=\; A\, \Sigma_{t|t} A^\top + Q_t.$$

Thus

$$\eta_{t+1} \,\big|\, \mathcal{D}_t \;\sim\; \mathcal{N}(\eta_{t+1|t}, \Sigma_{t+1|t}).$$

Observation prediction

The observed discrepancy at time $t+1$ is related to the latent state by

$$z_{t+1} = C\, \eta_{t+1} + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \mathcal{N}(0, R_t).$$

By the laws of linear–Gaussian models, the predictive distribution is

$$z_{t+1} \,\big|\, \mathcal{D}_t \;\sim\; \mathcal{N}\Big(C\, \eta_{t+1|t},\; C\, \Sigma_{t+1|t}\, C^\top + R_t \Big).$$

Interpretation

- The predictive *mean* is the projection of the one-step-ahead latent estimate back into the player space:

$$\mathbb{E}[z_{t+1} \mid \mathcal{D}_t] = C\,\eta_{t+1|t}.$$

- The predictive *covariance* accounts both for state uncertainty propagated forward ($\Sigma_{t+1|t}$) and observation noise (R_t):

$$\mathrm{Cov}[z_{t+1} \mid \mathcal{D}_t] = C\,\Sigma_{t+1|t}\,C^\top + R_t.$$

Hence the posterior predictive distribution is multivariate normal in the n-dimensional player space, centered at the projected latent prediction and respecting the zero-sum constraint in expectation.

5 Starter-Aware Observation Model (Immediate Effect)

Let $h_t \in \{0,1\}^n$ denote the per-player starter status at time t. Since our discrepancy $z_t = y_t - s_t$ is constrained to be zero-sum across players, we map h_t into the zero-sum subspace before using it:

$$\Pi \;=\; I_n - \tfrac{1}{n}\mathbf{1}\mathbf{1}^\top, \qquad \tilde{h}_t \;=\; \Pi h_t \in \mathbb{R}^n, \qquad u_t \;=\; C^\top \tilde{h}_t \in \mathbb{R}^{n-1}.$$

We model an *immediate* (time-t) mean shift from starter status in the observation:

$$z_t \;=\; C\,\eta_t \;+\; C\,\Gamma\,u_t \;+\; \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, R), \tag{1}$$

where $\Gamma \in \mathbb{R}^{(n-1)\times(n-1)}$ is an unknown matrix capturing how the projected starter pattern u_t shifts the discrepancy. The C on the left ensures that the shift lives in the zero-sum column space.

State model (unchanged).

$$\eta_{t+1} \;=\; A\,\eta_t \;+\; \omega_t, \qquad \omega_t \sim \mathcal{N}(0, Q).$$

Filter updates with starter-aware observation

Prediction is unchanged:

$$\eta_{t|t-1} = A\,\eta_{t-1|t-1}, \qquad \Sigma_{t|t-1} = A\,\Sigma_{t-1|t-1}A^\top + Q_{t-1}.$$

The *innovation* subtracts the instantaneous starter offset:

$$v_t = z_t - C\,\eta_{t|t-1} - C\,\Gamma\,u_t, \tag{2}$$

$$S_t = C\,\Sigma_{t|t-1}C^\top + R_{t-1}, \tag{3}$$

$$K_t = \Sigma_{t|t-1}C^\top S_t^{-1}. \tag{4}$$

Then

$$\eta_{t|t} = \eta_{t|t-1} + K_t v_t, \qquad \Sigma_{t|t} = (I - K_t C)\,\Sigma_{t|t-1}.$$

Learning Γ online (RLS in the subspace). Work in the subspace to avoid multiplying by C twice. Define the *subspace innovation*

$$\tilde{v}_t \;\equiv\; C^\top z_t - \eta_{t|t-1} \;\approx\; \Gamma\,u_t + \tilde{\varepsilon}_t, \quad \tilde{\varepsilon}_t \sim \mathcal{N}(0, \tilde{R}_t),$$

and run a recursive least squares (RLS) update with forgetting $\rho \in (0,1]$:

$$g_t \;=\; \frac{P_\Gamma u_t}{\rho + u_t^\top P_\Gamma u_t}, \qquad \Gamma \leftarrow \Gamma + \big(\tilde{v}_t - \Gamma u_t\big) g_t^\top, \qquad P_\Gamma \leftarrow \rho^{-1}\big(P_\Gamma - g_t u_t^\top P_\Gamma\big),$$

where $P_\Gamma \succ 0$ is the RLS covariance (initialize $P_\Gamma = \kappa I$ with large κ).

Adaptive R and Q (unchanged). Use innovation matching for R with the *starter-corrected* innovation (**??**):

$$\widehat{R}_t^{\text{raw}} = (1-\lambda)R_{t-1} + \lambda\big(v_t v_t^\top - C\,\Sigma_{t|t-1}C^\top\big),$$

optionally shrunk toward a scaled identity to keep PSD; and process matching for Q with $u_t^{(\text{state})} = \eta_{t|t} - A\eta_{t-1|t-1}$:

$$Q_t = (1-\gamma)Q_{t-1} + \gamma\left(u_t^{(\text{state})} u_t^{(\text{state})\top}\right).$$

Posterior predictive with starters. Conditioned on $\mathcal{D}_t$ and h_{t+1} (hence u_{t+1}), the one-step predictive law is

$$z_{t+1} \mid \mathcal{D}_t, h_{t+1} \;\sim\; \mathcal{N}\Big(C\,\eta_{t+1|t} + C\,\Gamma\,u_{t+1},\;\; C\,\Sigma_{t+1|t}C^\top + R_t\Big).$$

Algorithm 3 Starter-Aware Observation (Immediate Effect)

Require: Streams $\{(s_t, y_t, h_t)\}$; step sizes λ, γ; RLS forgetting ρ; shrinkage α

Precompute $\Pi = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, basis C with $C^\top C = I$, $C^\top \mathbf{1} = 0$

Init $\eta_{0|0} = 0$, $\Sigma_{0|0} = \tau^2 I$, $A = I$, $Q_0 = \sigma_Q^2 I$, $R_0 = \sigma_R^2 I$, $\Gamma = 0$, $P_\Gamma = \kappa I$

for $t = 1, 2, \ldots$ **do**

  $z_t \leftarrow y_t - s_t;\quad \tilde{h}_t \leftarrow \Pi h_t;\quad u_t \leftarrow C^\top \tilde{h}_t$

  Predict: $\eta_{t|t-1} \leftarrow A\eta_{t-1|t-1};\; \Sigma_{t|t-1} \leftarrow A\Sigma_{t-1|t-1}A^\top + Q_{t-1}$

  Innovation: $v_t \leftarrow z_t - C\eta_{t|t-1} - C\Gamma u_t;\; S_t \leftarrow C\Sigma_{t|t-1}C^\top + R_{t-1}$

  Gain/Update: $K_t \leftarrow \Sigma_{t|t-1}C^\top S_t^{-1};\; \eta_{t|t} \leftarrow \eta_{t|t-1} + K_t v_t;\; \Sigma_{t|t} \leftarrow (I - K_t C)\Sigma_{t|t-1}$

  RLS for Γ: $\tilde{v}_t \leftarrow C^\top z_t - \eta_{t|t-1};\; g_t \leftarrow \frac{P_\Gamma u_t}{\rho + u_t^\top P_\Gamma u_t};\; \Gamma \leftarrow \Gamma + (\tilde{v}_t - \Gamma u_t)g_t^\top;\; P_\Gamma \leftarrow \rho^{-1}(P_\Gamma - g_t u_t^\top P_\Gamma)$

  Adapt R: $\widehat{R}^{raw} \leftarrow (1-\lambda)R_{t-1} + \lambda(v_t v_t^\top - C\Sigma_{t|t-1}C^\top);\; R_t \leftarrow (1-\alpha)\widehat{R}^{raw} + \alpha\frac{\mathrm{tr}(\widehat{R}^{raw})}{n} I$

  Adapt Q: $u_t^{state} \leftarrow \eta_{t|t} - A\eta_{t-1|t-1};\; Q_t \leftarrow (1-\gamma)Q_{t-1} + \gamma(u_t^{state} u_t^{state\top})$

  (Optional) Output: $\hat{\delta}_{t|t} \leftarrow C\eta_{t|t} + C\Gamma u_t$ ▷ instantaneous starter-adjusted mean

end for
