AUTONOMOUS
INTELLIGENCE GROUP

LECTURE 7 — STATE ESTIMATION (III)

INTRODUCTION TO AUTONOMOUS SYSTEMS · MODULE III

Topics: Kalman optimality · Joseph form · Riccati & observability · Steady-state filter · Sensor fusion

AGENDA

§	TOPIC	KEY IDEAS
4 (cont.)	**The Kalman Filter**	Optimality · Joseph form · Riccati · observability
5	**Sensor Fusion**	Precisions add · multi-rate fusion · failure modes

> *Lecture 6 covered §3 (Gaussian machinery) and §4 up to the Kalman filter algorithm.*
> *§6–§7 — EKF / UKF, particle filter, belief to policy — are in Lecture 8.*

LITERATURE

TOPIC	REFERENCE
Bayes, Gaussian and particle filters	Thrun, Burgard & Fox — *Probabilistic Robotics* (2005), Ch. 2–4
Rigorous filtering and smoothing	Särkkä — *Bayesian Filtering and Smoothing* (2013)
Kalman filter theory, Riccati, steady state	Anderson & Moore — *Optimal Filtering* (1979), Ch. 3–5
Original paper	Kalman — *A New Approach to Linear Filtering and Prediction Problems* (1960)

§4 (CONT.) · THE KALMAN FILTER

WHAT THE ALGORITHM GUARANTEES

RECAP — THE ALGORITHM

Algorithm (Kalman filter).
Input: $b_0 = \mathcal{N}(\mu_0, \Sigma_0)$; model A, B, H, Q, R; stream $(a_0, o_1), (a_1, o_2), \ldots$
For $t = 1, 2, \ldots$

1. **Predict:** $\bar{\mu}_t = A\mu_{t-1} + Ba_{t-1}, \quad \bar{\Sigma}_t = A\Sigma_{t-1}A^\top + Q$
2. **Update:** $S_t = H\bar{\Sigma}_t H^\top + R,\ K_t = \bar{\Sigma}_t H^\top S_t^{-1},\ \mu_t = \bar{\mu}_t + K_t(o_t - H\bar{\mu}_t),\ \Sigma_t = (I - K_t H)\bar{\Sigma}_t$

Output: $b_t = \mathcal{N}(\mu_t, \Sigma_t)$.

Cost per step: $O(n^3 + d^3)$ time, $O(n^2 + d^2)$ memory, **independent of** t. n^3 from $A\Sigma A^\top$, d^3 from inverting S_t. Only the current (μ_t, Σ_t) is stored.

Zero innovation: the mean does not move, but the covariance still shrinks — seeing the expected value is still information.

EXACTNESS AND OPTIMALITY

Theorem. Under the linear Gaussian model:

1. $b_t = \mathcal{N}(\mu_t, \Sigma_t)$ is the **exact** posterior $p(x_t \mid o_{1:t}, a_{0:t-1})$.
2. Among **all** estimators $\hat{x} = g(o_{1:t}, a_{0:t-1})$ — linear or not, hand-designed or a neural network — μ_t has the smallest MSE $\mathbb{E}\|x_t - \hat{x}\|^2$, equal to $\operatorname{tr} \Sigma_t$.
3. Without Gaussianity (same means/covariances, uncorrelated; actions known), μ_t is still the best **affine** estimator $\hat{x} = c + \sum_{s \le t} M_s o_s$, with MSE $\operatorname{tr} \Sigma_t$.

Part 1 holds by construction: predict and update are the *exact* Bayes filter steps; induction from b_0.

PROOF OF OPTIMALITY (PART 2)

Let $\hat{x} = g(\mathcal{D}_t)$ be any estimator.

1. **Tower property:** $\mathbb{E}\|x_t - \hat{x}\|^2 = \mathbb{E}_{\mathcal{D}_t}\Big[\,\mathbb{E}_{x_t|\mathcal{D}_t}\|x_t - \hat{x}\|^2\Big]$. Inside, $\hat{x}$ is a constant.
2. **Inner expectation.** Split $x_t - \hat{x} = (x_t - \mu_t) + (\mu_t - \hat{x})$:

$$\mathbb{E}\big[\|x_t - \hat{x}\|^2 \mid \mathcal{D}_t\big] = \mathbb{E}\big[\|x_t - \mu_t\|^2 \mid \mathcal{D}_t\big] + 2(\mu_t - \hat{x})^\top \underbrace{\mathbb{E}[x_t - \mu_t \mid \mathcal{D}_t]}_{=0} + \|\mu_t - \hat{x}\|^2.$$

3. **First term** $= \operatorname{tr}\operatorname{Cov}(x_t \mid \mathcal{D}_t) = \operatorname{tr}\Sigma_t$.
4. **Average over data.** Σ_t does not depend on the observed values (Property 1, §4):

$$\mathbb{E}\|x_t - \hat{x}\|^2 = \operatorname{tr}\Sigma_t + \mathbb{E}\|\mu_t - \hat{x}\|^2 \;\geq\; \operatorname{tr}\Sigma_t,$$

with equality iff $\hat{x} = \mu_t$ a.s. $\square$

PART 3 — BEST AFFINE ESTIMATOR, ANY NOISE

Idea. The best affine estimator depends on **means and covariances alone** — the same with or without Gaussianity.

1. By the regression remark (Lec 6, §3), with $y = (o_1, \dots, o_t)$:

$$\hat{x}^* = \mathbb{E}[x_t] + \mathrm{Cov}(x_t, y)\,\mathrm{Cov}(y)^{-1}\big(y - \mathbb{E}[y]\big).$$

2. x_t and y are built from x_0 and the noises by additions and fixed matrices → their moments depend only on (μ_0, Σ_0), $(0, Q)$, $(0, R)$.
3. In the Gaussian model, μ_t beats every estimator *and* is affine in y (gains don't involve data) → it is $\hat{x}^*$. ■

What is lost without Gaussianity. A *nonlinear* estimator can beat μ_t (e.g. a particle filter under heavy tails or multimodality). Σ_t is then the error covariance, not a posterior.

REMARK — WHEN THE PLAIN UPDATE FAILS

In floating point, the plain form $\Sigma_t = (I - K_t H)\bar{\Sigma}_t$ can lose symmetry or positive definiteness.

- **Why.** A rounding error δK enters at first order, $\delta\Sigma = -\delta K\, H\bar{\Sigma}_t$: not symmetric, any sign.
- **When.** Precise sensors ($R \ll H\bar{\Sigma}_t H^\top$): $\Sigma_t = \bar{\Sigma}_t - K_t H\bar{\Sigma}_t$ is a small difference of large matrices → cancellation, small eigenvalues go negative. Errors also accumulate over many steps.

Example (float32). $\bar{\Sigma} = H = 1$, $R = 10^{-8}$:

$$K = \frac{\bar{\Sigma}H}{H\bar{\Sigma}H + R} = \frac{1}{1 + 10^{-8}}.$$

Float32 precision is $\approx 10^{-7}$, so $1 + 10^{-8}$ rounds to $1 \to K = 1 \to \Sigma_t = (1 - 1) \cdot 1 = 0$. **False certainty**; the true value is $\approx 10^{-8}$.

REMARK — JOSEPH FORM

$$J(K) = (I - KH)\,\bar{\Sigma}_t\,(I - KH)^\top + KRK^\top, \qquad \Sigma_t = J(K_t).$$

A sum of two PSD terms → stays PSD under rounding. Example: $(1-1)^2 \cdot 1 + 1 \cdot 10^{-8} = 10^{-8}$ ✓.

- **Equals the standard form** for the Kalman gain: $(I - K_t H)\bar{\Sigma}_t H^\top = K_t R$, so the extra terms cancel.
- **Correct for any gain** K. Σ_t must be the covariance of the *actual* error of the estimate. If $\mu_t = \bar{\mu}_t + K(o_t - H\bar{\mu}_t)$ was computed with some K (exact or rounded), then with $o_t = Hx_t + v_t$: $x_t - \mu_t = (I - KH)(x_t - \bar{\mu}_t) - Kv_t$. Independent terms with covariances $\bar{\Sigma}_t$, R → the true error covariance is $J(K)$ **with that same** K.

REMARK — GAIN ERRORS: PLAIN VS. JOSEPH

Rounding gives $K = K_t + \delta K$. Substitute into $J(K) = \bar{\Sigma}_t - KH\bar{\Sigma}_t - \bar{\Sigma}_t H^\top K^\top + K S_t K^\top$ and group by powers of δK:

$$J(K_t + \delta K) = \underbrace{J(K_t)}_{\Sigma_t} + \delta K \underbrace{(S_t K_t^\top - H\bar{\Sigma}_t)}_{=0} + \underbrace{(K_t S_t - \bar{\Sigma}_t H^\top)}_{=0}\delta K^\top + \delta K\, S_t\, \delta K^\top,$$

both brackets vanish since $K_t S_t = \bar{\Sigma}_t H^\top$. Plain form: $\big(I - (K_t + \delta K)H\big)\bar{\Sigma}_t = \Sigma_t - \delta K\, H\bar{\Sigma}_t$. Hence

$$\text{Joseph: } \Sigma_t + \delta K\, S_t\, \delta K^\top, \qquad \text{plain: } \Sigma_t - \delta K\, S_t K_t^\top.$$

- **Joseph error:** second order, symmetric, PSD → it only ever *overstates* uncertainty. (Also why K_t minimizes J.)
- **Plain error:** first order and non-symmetric (its transpose $K_t S_t\, \delta K^\top$ differs, since δK is unrelated to K_t) → not a covariance. Scalar: $K = 1 + 10^{-7}$ → variance -10^{-7}.
- **Size:** $\delta K \sim 10^{-7}$ → Joseph error $\sim 10^{-14}$, plain error $\sim 10^{-7}$.
- **Why it matters:** Σ_t feeds the next step → S_{t+1} can turn indefinite, the gain goes wrong, errors compound → divergence.

Takeaway. The plain form is exact only in exact arithmetic. In code, use the Joseph form.

PROPERTIES — THE COVARIANCE IGNORES THE DATA

Write $P_t = \bar{\Sigma}_t$, $S_t = HP_tH^\top + R$.

Property 1. Σ_t, $\bar{\Sigma}_t$, S_t, K_t depend only on the model (A, H, Q, R, Σ_0), never on $o_{1:t}$ or the actions. The whole gain sequence can be computed **before the system runs**.

Proof. Induction: the covariance lines use only model matrices and Σ_{t-1}; data enter only the mean. □

Definition (Riccati equation). An equation for a symmetric matrix P:

$$P_{\text{new}} = F(P), \qquad F(P) = APA^\top + Q - APH^\top \left(HPH^\top + R\right)^{-1} HPA^\top.$$

Riccati recursion: $P_{t+1} = F(P_t)$. **DARE:** $P = F(P)$. (After Riccati's $\dot{y} = q + 2ay + by^2$ — quadratic in the unknown.)

Property 2. $P_{t+1} = F(P_t)$: one filter cycle on the covariance ($K_t = P_tH^\top S_t^{-1}$),

$$P_t \xrightarrow{\text{update}} \Sigma_t = P_t - P_tH^\top S_t^{-1} HP_t \xrightarrow{\text{predict}} P_{t+1} = A\Sigma_t A^\top + Q = F(P_t).$$

The Riccati map **is** update + predict for the covariance.

MODES, OBSERVABILITY, DETECTABILITY

Definition (Mode). An eigenvector $Av = \lambda v$. **Stable** if $|\lambda| < 1$, **unstable** if $|\lambda| \geq 1$, **unobservable** if $Hv = 0$ (then $o_t = H\lambda^t v = 0$ forever).

Definition. (A, H) is

- **observable** if $\operatorname{rank} \begin{pmatrix} H \\ HA \\ \vdots \\ HA^{n-1} \end{pmatrix} = n$ (Lec 2) — equivalently, no mode is unobservable;
- **detectable** if every unobservable mode is stable — the unseen part decays by itself.

EXAMPLE — OBSERVABLE VS. DETECTABLE

$$A = \begin{pmatrix} 1.1 & 0 \\ 0 & 0.5 \end{pmatrix}, \qquad \det(A - \lambda I) = (1.1 - \lambda)(0.5 - \lambda) = 0.$$

- $\lambda_1 = 1.1$, eigenvector $v_1 = e_1$: $|\lambda_1| \geq 1$ → **unstable**.
- $\lambda_2 = 0.5$, eigenvector $v_2 = e_2$: $|\lambda_2| < 1$ → **stable**.

H	He_1	He_2	UNOBSERVABLE MODE	VERDICT
$(0 \;\; 1)$	0	1	e_1, unstable	**not detectable**
$(1 \;\; 0)$	1	0	e_2, stable	**detectable, not observable**

Rank check for $H = (1 \;\; 0)$: $\begin{pmatrix} H \\ HA \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1.1 & 0 \end{pmatrix}$, rank $1 < 2$ → not observable, as the mode test says.

NOISE MUST REACH THE UNSTABLE MODES

Definition (Stabilisable). For the controlled system $x_t = Ax_{t-1} + Bu_t$, (A, B) is stabilisable if some $u_t = -Kx_{t-1}$ makes $A - BK$ stable.
Popov–Belevitch–Hautus (PBH) test (without proof): for every left eigenvector $u^\top A = \lambda u^\top$ with $|\lambda| \geq 1$, $u^\top B \neq 0$.
Meaning: $z_t = u^\top x_t$ is the coordinate of mode λ, and $z_t = \lambda z_{t-1} + (u^\top B)\, u_t$. If $u^\top B = 0$, the input never reaches this mode, so $z_t = \lambda^t z_0$ cannot be stabilised. The test: the input reaches every unstable mode.

Definition (Noise excites the unstable modes). Write $w_t = Q^{1/2}\varepsilon_t$, $\mathrm{Cov}(\varepsilon_t) = I$, so $x_t = Ax_{t-1} + Q^{1/2}\varepsilon_t$: noise enters like an input with $B = Q^{1/2}$. The noise **excites every unstable mode** if $(A, Q^{1/2})$ is stabilisable: $u^\top Q u > 0$ for every left eigenvector u with $|\lambda| \geq 1$.

Meaning: each unstable mode coordinate $z_t = u^\top x_t = \lambda z_{t-1} + u^\top w_t$ gets a fresh random kick of variance $u^\top Q u > 0$ every step. Nothing is controlled here: ε_t is not chosen, there is no feedback. We reuse only the PBH condition "the input reaches every unstable mode", with noise as the input.

LONG-RUN BEHAVIOUR

Property 3 (Detectability keeps uncertainty bounded).
(a) If (A, H) is detectable, P_t stays bounded for every Σ_0.
(b) An unobservable unstable mode that receives noise makes the uncertainty grow without bound.

The structure of (A, H) decides whether the filter works at all; Q and R decide how well.

Property 4 (Steady state). If (A, H) is detectable and noise excites every unstable mode, then $P_t \to P_\infty$ for every Σ_0, the unique PSD solution of the DARE. Gains converge to $K_\infty = P_\infty H^\top (H P_\infty H^\top + R)^{-1}$, and $A(I - K_\infty H)$ is stable.

Properties 3–4 without proof: Anderson & Moore (1979).

STEADY-STATE KALMAN FILTER

By Property 1 the gains need no data; by Property 4 they converge, $K_t \to K_\infty$. So solve the DARE **offline** and run only the mean recursion with a constant gain:

$$\mu_t = \bar{\mu}_t + K_\infty\big(o_t - H\bar{\mu}_t\big), \qquad \bar{\mu}_t = A\mu_{t-1} + Ba_{t-1}.$$

- **Cost:** matrix–vector products only, $O(n^2 + nd)$ per step — no covariance update, no matrix inverse (vs. $O(n^3 + d^3)$).
- **Memory and timing:** one stored $n \times d$ matrix; fixed, predictable runtime → microcontrollers, fixed-point; common in embedded systems.
- **Numerics:** no covariance is propagated, so the rounding problems of the Joseph-form slides cannot occur.

STEADY-STATE KALMAN FILTER — STABILITY AND PRICE

Error dynamics. Let $e_t = x_t - \mu_t$. Subtracting the prediction from $x_t = Ax_{t-1} + Ba_{t-1} + w_t$ gives $x_t - \bar{\mu}_t = Ae_{t-1} + w_t$; the innovation is $o_t - H\bar{\mu}_t = H(x_t - \bar{\mu}_t) + v_t$. So

$$e_t = (I - K_\infty H)(Ae_{t-1} + w_t) - K_\infty v_t = Me_{t-1} + \text{noise}, \qquad M = (I - K_\infty H)A.$$

- **Unroll:** $e_t = M^t e_0 + \sum_{s=1}^{t} M^{t-s}\,(\text{noise}_s)$. A wrong start enters **only** through $M^t e_0$.
- **M is stable:** XY and YX have the same eigenvalues ($XYv = \lambda v \Rightarrow YX(Yv) = \lambda Yv$), and $A(I - K_\infty H)$ is stable by Property 4. So $M^t e_0 \to 0$ geometrically: the initial error is forgotten, and only the noise-driven part, with covariance Σ_∞, remains.
- **Price — the transient:** early on, $P_t \gg P_\infty$ (vague prior), so the optimal $K_t > K_\infty$ trusts the first measurements more. The constant K_∞ under-weights them and clings to the bad prior. Example (a): error variance 4.70 vs. optimal 2.93 at $t = 1$ (+60%), +17% at $t = 3$, <1% by $t = 6$.
- **Also:** needs a time-invariant A, H, Q, R; no online Σ_t (but P_∞ is known).

EXAMPLE (A) — WATCHING THE GAIN CONVERGE

$x_t = x_{t-1} + a_{t-1}\Delta t + w_t$, GPS $o_t = x_t + v_t$: $A = H = 1, Q = 1, R = 4$, vague prior $\Sigma_0 = 10$.

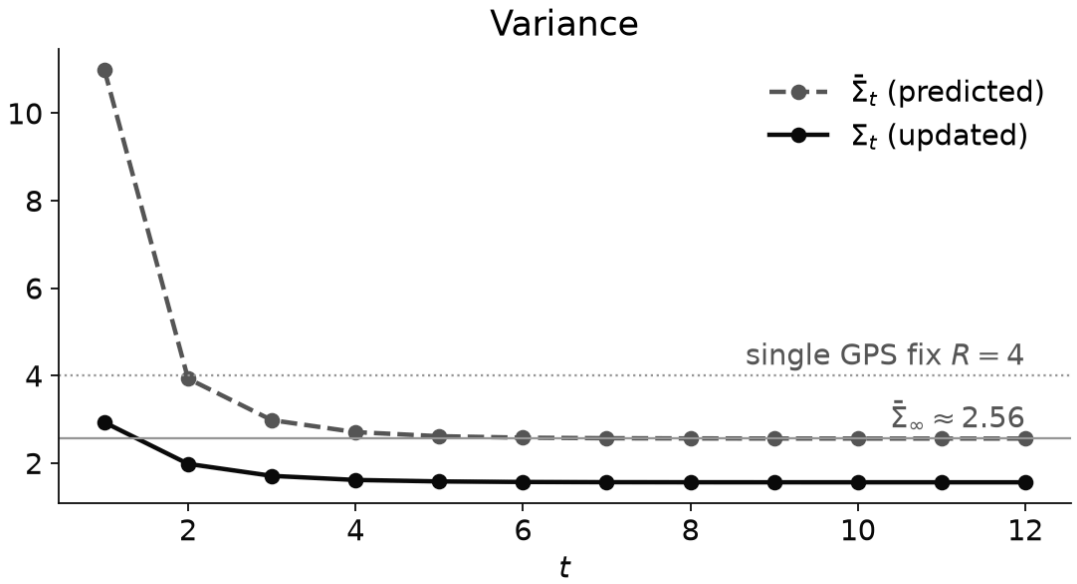
t	$\bar{\Sigma}_t$	K_t	Σ_t
1	11.00	0.733	2.93
2	3.93	0.496	1.98
3	2.98	0.427	1.71
∞	2.56	0.390	1.56

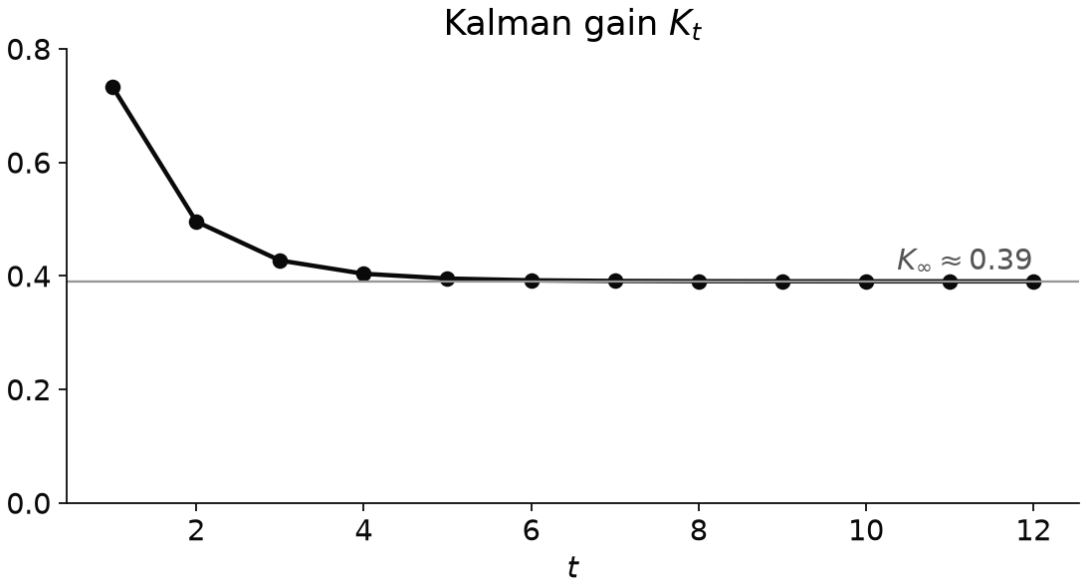
Scalar DARE: $\bar{\Sigma}^2 = Q\bar{\Sigma} + QR$

$$\bar{\Sigma}_\infty = \tfrac{1}{2}\left(1 + \sqrt{17}\right) \approx 2.56$$

- Gain starts high (trust data over a vague prior), settles fast.
- $\sqrt{1.56} \approx 1.25$ m **below** the 2 m of one GPS fix: the model averages over time.
- No number depends on the measurements.

EXAMPLE (A) — PICTURED





EXAMPLE (B) — 2D TRACKING, UNMEASURED VELOCITY

State $x = (p_x, p_y, v_x, v_y)^\top$, constant-velocity model:

$$A = \begin{pmatrix} I_2 & \Delta t\, I_2 \\ 0 & I_2 \end{pmatrix}, \qquad Q = q \begin{pmatrix} \frac{\Delta t^3}{3} I_2 & \frac{\Delta t^2}{2} I_2 \\ \frac{\Delta t^2}{2} I_2 & \Delta t\, I_2 \end{pmatrix}, \qquad H = \begin{pmatrix} I_2 & 0 \end{pmatrix}.$$

Q = exact discretisation of white-noise acceleration with intensity q.

Observability. $\begin{pmatrix} H \\ HA \end{pmatrix} = \begin{pmatrix} I_2 & 0 \\ I_2 & \Delta t\, I_2 \end{pmatrix}$ has rank 4 → velocity is observable though never measured.

The filter recovers it through the position–velocity correlation in $\bar{\Sigma}_t$ (Lec 6, §4.4).

> **Common confusion:** "K_t is a weight in $[0, 1]$." Only in the scalar case. In general K_t is $n \times d$ and corrects states that are never measured.

This is the setting of the Kalman implementation task in Assignment 2.

§5 · SENSOR FUSION

EACH SENSOR IS ANOTHER LIKELIHOOD FACTOR

STATIC FUSION — SETUP

Real systems combine GPS (absolute, noisy, slow), wheel odometry (smooth, drifting), IMU (fast, biased), cameras. In the Bayesian framework **no new theory is needed**.

Prior $x \sim \mathcal{N}(\mu_0, \Sigma_0)$; k sensors:

$$o^{(j)} = H_j x + v^{(j)}, \qquad v^{(j)} \sim \mathcal{N}(0, R_j), \qquad j = 1, \dots, k,$$

noises **independent of each other and of x**.

Why the likelihoods multiply. Given x, each reading is random only through its own noise → **conditionally independent given x**:

$$p\big(o^{(1)}, \dots, o^{(k)} \mid x\big) = \prod_{j=1}^{k} \mathcal{N}\big(o^{(j)};\, H_j x,\, R_j\big), \qquad p(x \mid o^{(1:k)}) \propto p(x) \prod_{j} p(o^{(j)} \mid x).$$

The readings are *not* unconditionally independent — they share the cause x (Lec 4).

PRECISIONS ADD

Proposition (Information fusion).

$$\Sigma^{-1} = \Sigma_0^{-1} + \sum_{j=1}^{k} H_j^\top R_j^{-1} H_j, \qquad \Sigma^{-1}\mu = \Sigma_0^{-1}\mu_0 + \sum_{j=1}^{k} H_j^\top R_j^{-1} o^{(j)}.$$

Proof. Stack: $o = Hx + v$, $H = \begin{pmatrix} H_1 \\ \vdots \\ H_k \end{pmatrix}$, $R = \mathrm{blockdiag}(R_1, \dots, R_k)$ by independence. Information form (Lec 6, §3.4), and $R^{-1} = \mathrm{blockdiag}(R_j^{-1})$:

$$H^\top R^{-1} H = \sum_j H_j^\top R_j^{-1} H_j, \qquad H^\top R^{-1} o = \sum_j H_j^\top R_j^{-1} o^{(j)}. \qquad \square$$

Correlated noises → off-diagonal blocks in R → cross terms. **Independence is what makes each contribution separate.**

SCALAR EXAMPLE

Scalar example. GPS 10.0 m (var 4), UWB 11.0 m (var 1), flat prior:

$$\sigma^{-2} = \tfrac{1}{4} + 1 = 1.25 \Rightarrow \sigma^2 = 0.8, \qquad \mu = 0.8\left(\tfrac{10}{4} + \tfrac{11}{1}\right) = 10.8\ \mathrm{m}.$$

- The fused estimate is better than either sensor alone ($0.8 < 1$).
- It lies closer to the more precise sensor: weight 0.8 on UWB against 0.2 on GPS.

> Information adds and precisions add — each sensor weighted by its own precision R_j^{-1}.

WHAT IF GPS AND UWB SHARE A COMMON ERROR SOURCE?

TWO TYPES OF MEASUREMENT

State $x = (p, v)$; prior from a predict step, so correlated:

$$\mu_0 = \begin{pmatrix} 10 \\ 2 \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \Sigma_0^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad \Sigma_0^{-1}\mu_0 = \begin{pmatrix} 8 \\ -6 \end{pmatrix}.$$

GPS: $H_1 = (1 \;\; 0)$, $R_1 = 2$, $o^{(1)} = 12$. **Speedometer:** $H_2 = (0 \;\; 1)$, $R_2 = 1$, $o^{(2)} = 1$.

$$\Sigma^{-1} = \Sigma_0^{-1} + \mathrm{diag}(\tfrac{1}{2}, 0) + \mathrm{diag}(0, 1) = \begin{pmatrix} 1.5 & -1 \\ -1 & 3 \end{pmatrix}, \quad \Sigma^{-1}\mu = \begin{pmatrix} 14 \\ -5 \end{pmatrix} \Rightarrow \mu \approx \begin{pmatrix} 10.57 \\ 1.86 \end{pmatrix}.$$

	$\hat{p}$	$\hat{v}$	$\mathrm{Var}(p)$	$\mathrm{Var}(v)$
prior	10	2	2	1
GPS only	11.00	2.50	1.00	0.75
speedometer only	9.50	1.50	1.50	0.50
both	10.57	1.86	0.86	0.43

WHAT THE TABLE SHOWS

- **Each sensor changes both components.** GPS measures only position, yet moves velocity ($2 \rightarrow 2.5$) and shrinks its variance ($1 \rightarrow 0.75$) — through the prior correlation. With diagonal Σ_0, each sensor would affect only its own component.
- **The sensors complement each other.** GPS: "further than predicted" → velocity up. Speedometer: "slower than predicted" → velocity down. The fused estimate balances both and is more certain than either alone in both components.

Shared error source: the fused variance is too optimistic — **double counting**.

FUSION IN THE KALMAN FILTER

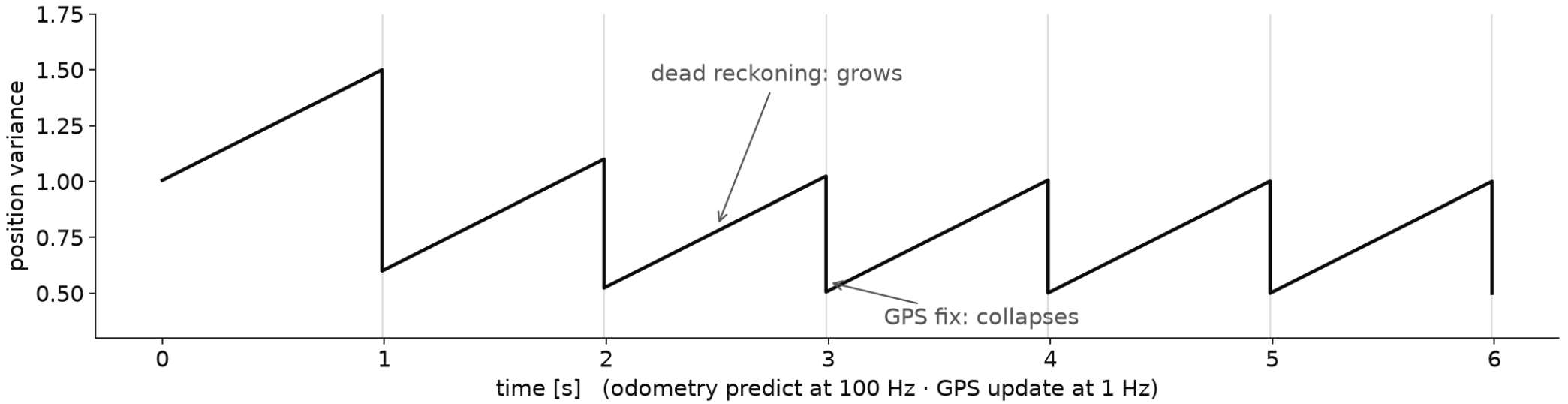
Two equivalent ways for simultaneous measurements:

1. **Stacked update.** $o_t = \begin{pmatrix} o_t^{(1)} \\ o_t^{(2)} \end{pmatrix}$, $H = \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}$, $R = \mathrm{blockdiag}(R_1, R_2)$, one update.
2. **Sequential updates.** Update with sensor 1, use the result as prior for sensor 2. Same answer: Bayes absorbs independent likelihoods in any order.

Multi-rate, asynchronous fusion. Predict at the highest needed rate; when *any* sensor reports, update with its H_j, R_j. A silent sensor needs no special handling: the filter keeps predicting and the covariance grows until the next fix.

EXAMPLE — GPS + ODOMETRY

Odometry (velocity, or equivalently the action a_t) at 100 Hz; GPS (position) at 1 Hz.



Between fixes the filter integrates odometry and the variance grows (dead reckoning); at each fix it collapses. **Smooth like odometry, drift-free like GPS** — neither sensor achieves both alone.

WHEN FUSION IS NOT JUST MULTIPLICATION

Fusion relies on (1) a known model per sensor, (2) errors independent given the state, (3) a shared, correctly defined state.

SITUATION	PROBLEM	WHAT IS USED INSTEAD
Correlated errors, correlation unknown	Double counting → overconfident	Covariance intersection
Fusing *estimates* (track-to-track)	Each estimate already contains a prior	Remove common information first
Unknown data association	Likelihood is a mixture over assignments	JPDA, multiple-hypothesis tracking
Outliers, sensor faults	Gaussian trusts a wild reading fully	Innovation gating, heavy-tailed likelihoods
Unmodelled bias (IMU)	Bias is not white noise	State augmentation
Learned fusion (camera + lidar)	No explicit $p(o \mid x)$	End-to-end architectures (Lec 9)
Complementary filter	Heuristic frequency weighting	Often a steady-state KF in disguise

SUMMARY

1. Linear Gaussian model → the Kalman filter is the **exact posterior** and the **minimum-MSE** estimator; for any noise it is still the **best affine** estimator.
2. In code, update the covariance in **Joseph form**: it stays PSD and is robust to gain errors.
3. The covariance **ignores the data**: $P_{t+1} = F(P_t)$, the Riccati map.
4. (A, H) detectable + noise reaches every unstable mode → $P_t \to P_\infty$ (stabilising DARE solution) → **steady-state Kalman filter**.
5. Sensor fusion: independent sensors are independent likelihood factors. Gaussian: **precisions add**.

WHAT'S NEXT

Lecture 8 — State Estimation (IV)

- Everything here needed linear f, h and Gaussian noise
- §6 Keep the Gaussian, approximate the model: **EKF / UKF**
- §6 Drop the Gaussian: the **particle filter**
- §7 The belief as a sufficient statistic: estimator + controller