Factor Graphs for State Estimation

From Kalman Filters to Modern Optimization

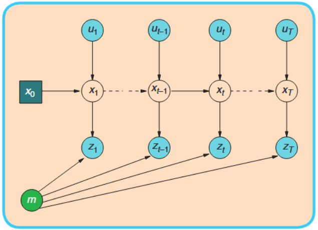
Zirui Zhang

Cheng Kar-Shun Robotics Institute
The Hong Kong University of Science and Technology

October 9, 2025

Markov Chain

The classic Kalman Filter corresponds to a simple **Markov Chain**:



Core Assumptions:

- **Markov Property:** The current state x_k depends *only* on the immediate previous state x_{k-1} and input u_k.
- **Conditional Independence:** Given the current state x_k, the observation y_k is independent of all other states and observations.

Scenario 1: Spatio-Temporal Constraints

A robot revisits a location, observing the same landmark l at two different times, k_1 and k_2:

$$\underset{\text{obs. } l}{x_{k_1}} \rightarrow x_{k_1+1} \rightarrow \cdots \rightarrow x_{k_2-1} \rightarrow \underset{\text{obs. } l}{x_{k_2}}$$

These two observations create a **direct constraint** between pose x_{k_1} and pose x_{k_2}. In the Kalman Filter, this connection is not direct.

Scenario 2: Physical Constraints

When tracking multiple objects, they may be subject to physical interaction constraints:

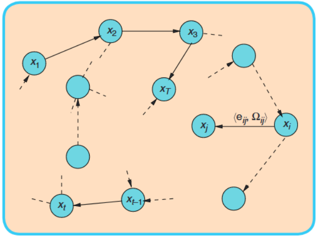
$$\text{Object A: } x_0^A \to \underbrace{x_1^A \to x_2^A}_{\text{interact with B}} \to \cdots$$

$$\text{Object B: } x_0^B \to \underbrace{x_1^B \to x_2^B}_{\text{interact with A}} \to \cdots$$

The Kalman Filter, designed for a single Markov chain, cannot natively represent this cross-object dependency.

Dynamic Bayesian Networks

Dynamic Bayesian Networks (DBNs) provide a more flexible framework than a simple Markov chain for representing probabilistic dependencies across time.



Extensions:

1. **Long-Term Dependencies:** States can depend on earlier states.
2. **Inter-Variable Links:** Variables within a time slice can be connected.
3. **Hierarchical States:** States can have sub-states with dependencies.

Dynamic Bayesian Networks (cont.)

A factor graph is a **bipartite graph** consisting of two types of nodes:

- **Variable Nodes:** Represent the unknown quantities we wish to estimate.
- **Factor Nodes:** Represent a **constraint** or a **measurement** on the set of variables they are connected to.

Goal: Find the most probable assignment of the variables that maximizes the product of all factors.

$$\mathbf{X}^* = \arg\max_{\mathbf{X}} \prod_i f_i(\mathscr{X}_i)$$

Under Gaussian assumptions, this becomes a nonlinear least-squares problem:

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \sum_i ||h_i(\mathscr{X}_i) - z_i||^2_{\mathbf{S}_i}$$

The Dual Information Form

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_{\alpha} \\ \boldsymbol{\mu}_{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\alpha\alpha} & \boldsymbol{\Sigma}_{\alpha\beta} \\ \boldsymbol{\Sigma}_{\beta\alpha} & \boldsymbol{\Sigma}_{\beta\beta} \end{bmatrix}\right) = \mathcal{N}^{-1}\left(\begin{bmatrix} \boldsymbol{\eta}_{\alpha} \\ \boldsymbol{\eta}_{\beta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Lambda}_{\alpha\alpha} & \boldsymbol{\Lambda}_{\alpha\beta} \\ \boldsymbol{\Lambda}_{\beta\alpha} & \boldsymbol{\Lambda}_{\beta\beta} \end{bmatrix}\right)$$

Operation	Covariance Form	Information Form
Marginalization	$\boldsymbol{\mu} = \boldsymbol{\mu}_{\alpha}$ $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\alpha\alpha}$	$\boldsymbol{\eta} = \boldsymbol{\eta}_{\alpha} - \boldsymbol{\Lambda}_{\alpha\beta}\boldsymbol{\Lambda}_{\beta\beta}^{-1}\boldsymbol{\eta}_{\beta}$ $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_{\alpha\alpha} - \boldsymbol{\Lambda}_{\alpha\beta}\boldsymbol{\Lambda}_{\beta\beta}^{-1}\boldsymbol{\Lambda}_{\beta\alpha}$
Conditioning	$\boldsymbol{\mu}' = \boldsymbol{\mu}_{\alpha} + \boldsymbol{\Sigma}_{\alpha\beta}\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_{\beta})$ $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}_{\alpha\alpha} - \boldsymbol{\Sigma}_{\alpha\beta}\boldsymbol{\Sigma}_{\beta\beta}^{-1}\boldsymbol{\Sigma}_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_{\alpha} - \boldsymbol{\Lambda}_{\alpha\beta}\boldsymbol{\beta}$ $\boldsymbol{\Lambda}' = \boldsymbol{\Lambda}_{\alpha\alpha}$

The Dual Information Filter

	Kalman Filter	Information Filter
Prediction Step	$\boldsymbol{\mu}_{t\|t-1} = \mathbf{A}_t \boldsymbol{\mu}_{t-1}$ $\boldsymbol{\Sigma}_{t\|t-1} = \mathbf{A}_t \boldsymbol{\Sigma}_{t-1} \mathbf{A}_t^\top + \mathbf{Q}_t$	$\boldsymbol{\Lambda}_{t\|t-1} = (\mathbf{A}_t \boldsymbol{\Lambda}_{t-1}^{-1} \mathbf{A}_t^\top + \mathbf{Q}_t)^{-1}$ $\boldsymbol{\eta}_{t\|t-1} = \boldsymbol{\Lambda}_{t\|t-1} \mathbf{A}_t \boldsymbol{\Lambda}_{t-1}^{-1} \boldsymbol{\eta}_{t-1}$
Update Step	$\mathbf{K}_t = \boldsymbol{\Sigma}_{t\|t-1} \mathbf{H}_t^\top (\mathbf{H}_t \boldsymbol{\Sigma}_{t\|t-1} \mathbf{H}_t^\top + \mathbf{R}_t)^{-1}$ $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t\|t-1} + \mathbf{K}_t(\mathbf{z}_t - \mathbf{H}_t \boldsymbol{\mu}_{t\|t-1})$ $\boldsymbol{\Sigma}_t = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \boldsymbol{\Sigma}_{t\|t-1}$	$\boldsymbol{\Lambda}_t = \boldsymbol{\Lambda}_{t\|t-1} + \mathbf{H}_t^\top \mathbf{R}_t^{-1} \mathbf{H}_t$ $\boldsymbol{\eta}_t = \boldsymbol{\eta}_{t\|t-1} + \mathbf{H}_t^\top \mathbf{R}_t^{-1} \mathbf{z}_t$

Factor Graph with Information Form

The global nonlinear optimization problem

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \sum_i ||h_i(\mathscr{X}_i) - z_i||^2_{\mathbf{S}_i}$$

can be linearized at current estimation $\mathbf{X}_0$:

$$h_i(\mathscr{X}_i) \approx h_i(\mathscr{X}_{i,0}) + \mathbf{J}_i \cdot \delta\mathscr{X}_i$$

where:

- $\mathbf{J}_i = \frac{\partial h_i}{\partial \mathscr{X}_i}|_{\mathscr{X}_{i,0}}$ is the **Jacobian matrix** of measurement function h_i
- $\mathbf{r}_i = z_i - h_i(\mathscr{X}_{i,0})$ is the **residual vector**

Factor Graph with Information Form (cont.)

Local Information Form	Global Information Form
$\mathbf{\Lambda}_i = \mathbf{J}_i^\top \mathbf{S}_i^{-1} \mathbf{J}_i$ $\boldsymbol{\eta}_i = \mathbf{J}_i^\top \mathbf{S}_i^{-1} \mathbf{r}_i$	$\mathbf{\Lambda} = \sum_i \mathbf{A}_i^\top \mathbf{\Lambda}_i \mathbf{A}_i$ $\boldsymbol{\eta} = \sum_i \mathbf{A}_i^\top \boldsymbol{\eta}_i$

where $\mathbf{A}_i$ is the **selection matrix** that maps local variables $\mathcal{X}_i$ to the global state vector $\mathbf{X}$. The optimal update is then:

$$\mathbf{X}^* = \mathbf{X}_0 + (\mathbf{\Lambda})^{-1} \boldsymbol{\eta}$$

From Filtering to Smoothing

Kalman Filter, Information Filter, and Factor Graph are fundamentally solving the same problem: **state estimation under Gaussian assumptions**. They are probabilistically equivalent.

Despite their equivalence, FG-based smoothing dominates modern applications because:

1. It naturally encodes arbitrary constraints;
2. It exploits sparse structure for efficient solving;
3. It's batch-based update enabling non-linear optimization;
4. It corrects past states by future evidence.

Thank you for listening !

Zirui Zhang