SLAM Summer School 2006

Supplement to the Information Filter Practical

M. Walter, Massachusetts Institute of Technology
mwalter@mit.edu

1 Background

Kalman Filter SLAM algorithms represent Gaussian distributions in terms of a covariance matrix, Σ, and mean vector, μ. The information (canonical) form is an alternative parametrization that describes the Gaussian by the information matrix, Λ, and information vector, η. The two representations are related according to:

$$\Lambda = \Sigma^{-1} \quad \Lambda\mu = \eta \tag{1}$$

You can easily derive this relationship by playing with the quadratic term within the Gaussian exponential, i.e. $(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$.

One of the major limitations of the Kalman Filter is that we must keep track of the covariance matrix[1]. This matrix is dense as we show in the left-hand side of Figure 1 and storing it requires an amount of memory that is quadratic in the size of the map. Secondly, the computational expense of the KF update step is $\mathcal{O}(n^2)$, where n is the number of states. Together, these costs restrict KF SLAM algorithms to relatively small maps, on the order of hundreds of features.

On the other hand, the corresponding information matrix, shown on the right in Figure 1 tends to be "relatively" sparse. Some of the terms are large while the majority are, in proportion, very small. If these terms were actually zero, a careful implementation of the IF can shed most of the computational and memory burden [1].

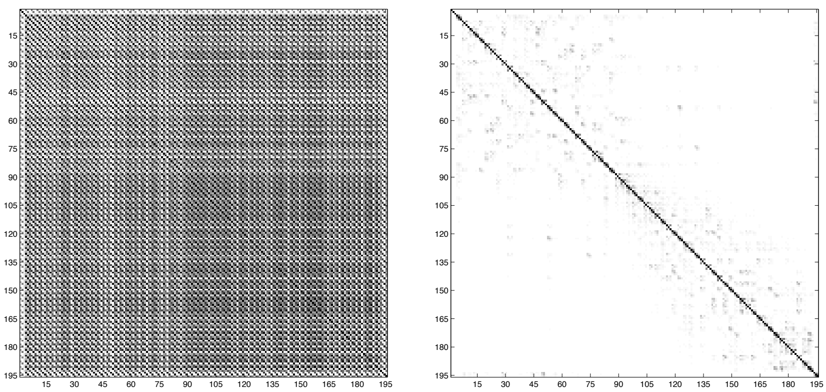

Figure 1: A comparison of a typical normalized covariance matrix (left) and the corresponding normalized information matrix (right). Darker shades correspond to larger magnitudes.

[1] Of course, the covariance matrix is also a big help since it allows us to take advantage of the correlation between the states to improve our estimates when we run the update step.

1.1 Interpretation as a Graphical Model

A particularly useful property of the canonical form is that it explicitly describes the structure of the undirected graphical model associated with the distribution. More specifically, nonzero off-diagonal terms in the information matrix indicate the existence (and strength) of links between the corresponding variables in the Gaussian Markov Random Field (GMRF) [2]. For states that do not "share information" within the matrix, there is no edge between the corresponding nodes in the graph. The schematic on the left in Figure 2 shows an example in which each feature shares information only with the robot pose and, in turn, the only edges are those to the robot.

Since an undirected graph encodes the conditional independence between states [3], those that are not directly connected according to the information matrix are conditionally independent of one another given their immediate neighbors. This characteristic of the information form is particularly useful in understanding many of the issues related to IF SLAM.

2 Duality Between the Standard and Information Forms

Gaussian-based algorithms, like any other Bayesian approach to the SLAM problem, rely upon marginalization and conditioning to keep track of the state distribution. The time projection step can be thought of, in part, as a marginalization while measurement updates condition the state on new observations.

It is fairly easy to show that the standard (covariance) and information forms are duals of one another when it comes to the math involved in marginalization and conditioning. As we show in Table 1, the structure of the marginalization step for the canonical form is very similar to the conditioning step for the covariance form and vice-versa.

Table 1: Summary of Marginalization and Conditioning Operations on a Gaussian Distribution Expressed in Covariance and Information Form

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

	MARGINALIZATION $p(\boldsymbol{\alpha}) = \int p(\boldsymbol{\alpha}, \boldsymbol{\beta})\, d\boldsymbol{\beta}$	CONDITIONING $p(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = p(\boldsymbol{\alpha}, \boldsymbol{\beta}) / p(\boldsymbol{\beta})$
COVARIANCE FORM	$\boldsymbol{\mu} = \boldsymbol{\mu}_\alpha$ $\Sigma = \Sigma_{\alpha\alpha}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$
INFORMATION FORM	$\boldsymbol{\eta} = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

3 Basic SLAM Steps

The information filter SLAM algorithm, just like the KF approach, iterates over the time projection and measurement update steps and adds new features to the map when they are observed. In this section,

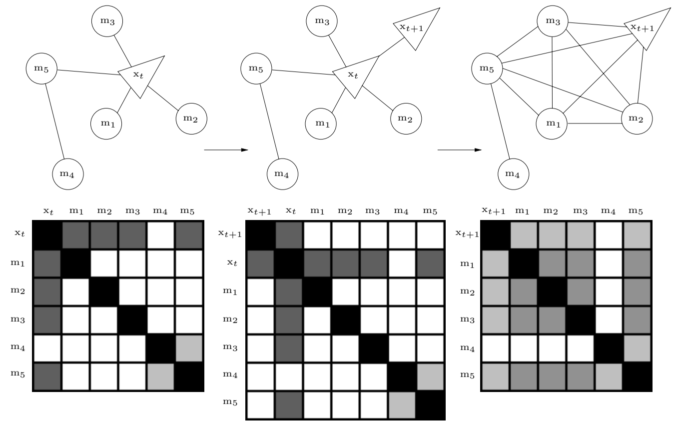

Figure 2: The time projection step of the information filter viewed as state augmentation followed by marginalization. The elements of the information matrix are shaded according to their magnitude with the strongest constraints shown as being darkest.

we take a quick look at how these steps are implemented in the IF.

3.1 Adding New Features to the Map

Suppose that our state vector, $\boldsymbol{\xi}_t = \begin{bmatrix}\mathbf{x}_t^\top \ \mathbf{M}^\top\end{bmatrix}^\top$, currently consists of the robot pose, $\mathbf{x}_t$ and a number of features, $\mathbf{M} = \{\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n\}$. The information parametrization of the Gaussian distribution over the state is

$$\Lambda_t = \begin{bmatrix}\Lambda_{x_t x_t} & \Lambda_{x_t M} \\ \Lambda_{M x_t} & \Lambda_{MM}\end{bmatrix} \qquad \boldsymbol{\eta}_t = \begin{bmatrix}\boldsymbol{\eta}_{x_t} \\ \boldsymbol{\eta}_M\end{bmatrix}.$$

The robot observes a new feature according to the model

$$\mathbf{m}_{n+1} = \mathrm{G}\mathbf{x}_t + \mathbf{v}_t$$

where $\mathbf{v}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{R}\big)$ denotes zero mean, white Gaussian noise. We add the new landmark to the end of the state vector, $\boldsymbol{\xi}_t = \begin{bmatrix}\mathbf{x}_t^\top \ \mathbf{M}^\top \ \mathbf{m}_{n+1}^\top\end{bmatrix}^\top$, and modify the information matrix and vector per (2).

$$\bar{\Lambda} = \begin{bmatrix}\big(\Lambda_{x_t x_t} + \mathrm{G}^\top \mathrm{R}^{-1}\mathrm{G}\big) & \Lambda_{x_t M} & -\mathrm{G}^\top \mathrm{R}^{-1} \\ \Lambda_{M x_t} & \Lambda_{MM} & 0 \\ -\mathrm{R}^{-1}\mathrm{G} & 0 & \mathrm{R}^{-1}\end{bmatrix} \qquad \bar{\boldsymbol{\eta}}_t = \begin{bmatrix}\boldsymbol{\eta}_{x_t} \\ \boldsymbol{\eta}_M \\ \mathbf{0}\end{bmatrix} \tag{2}$$

Notice that the new feature shares information only with the robot pose and not the rest of the map. This is consistent with the fact that, if we knew the robot pose, the previous map wouldn't tell us anything about the location of the new feature. In terms of the graphical model, the lone edge connected to $\mathbf{m}_{n+1}$ pairs it with the robot pose: the new feature is conditionally independent of the map, given the pose of the vehicle.

3.2 Time Projection Step

It is intuitive to break the time prediction step into two processes: state augmentation and marginalization. We first add the new pose, $\mathbf{x_{t+1}}$, to the state vector, $\boldsymbol{\xi}_t \rightarrow \bar{\boldsymbol{\xi}}_{t+1} = \begin{bmatrix}\mathbf{x}_t^\top \ \underline{\mathbf{x}_{t+1}^\top} \ \mathbf{M}^\top\end{bmatrix}^\top$ where the motion of the robot is governed by the linear model

$$\mathbf{x}_{t+1} = \mathrm{F}\mathbf{x}_t + \mathbf{w}_t$$

where $\mathbf{w}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{Q}\big)$ is Gaussian noise. We can think of the new pose as a new feature and, as a result, the new canonical parametrization follows from (2),

$$p\left(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) = \mathcal{N}^{-1}\big(\hat{\boldsymbol{\eta}}_{t+1}, \hat{\Lambda}_{t+1}\big)$$

$$\hat{\Lambda}_{t+1} = \left[\begin{array}{c|cc} \left(\Lambda_{x_t x_t} + \mathrm{F}^\top \mathrm{Q}^{-1} \mathrm{F}\right) & -\mathrm{F}^\top \mathrm{Q}^{-1} & \Lambda_{x_t M} \\ \hline -\mathrm{Q}^{-1}\mathrm{F} & \mathrm{Q}^{-1} & 0 \\ \Lambda_{M x_t} & 0 & \Lambda_{MM} \end{array}\right] = \left[\begin{array}{c|c} \hat{\Lambda}_{t+1}^{11} & \hat{\Lambda}_{t+1}^{12} \\ \hline \hat{\Lambda}_{t+1}^{21} & \hat{\Lambda}_{t+1}^{22} \end{array}\right] \tag{3a}$$

$$\hat{\boldsymbol{\eta}}_{t+1} = \left[\begin{array}{c} \boldsymbol{\eta}_{x_t} \\ \hline \mathbf{0} \\ \boldsymbol{\eta}_M \end{array}\right] = \left[\begin{array}{c} \hat{\boldsymbol{\eta}}_{t+1}^{1} \\ \hline \hat{\boldsymbol{\eta}}_{t+1}^{2} \end{array}\right] \tag{3b}$$

As with a new feature, the new pose is linked with the old pose, but not with the map, as we indicate in the middle plot of Figure 2.

Next, we get rid of the old robot pose from the state by marginalizing over the distribution.

$$p\left(\mathbf{x}_{t+1}, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) = \int\limits_{\mathbf{x}_t} p\left(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) \partial \mathbf{x}_t$$

Looking back at Table 1, the corresponding information form follows if we let $\boldsymbol{\alpha} = \left[\mathbf{x}_{t+1}^\top \; \mathbf{M}^\top\right]^\top$ and $\boldsymbol{\beta} = \mathbf{x}_t$,

$$p\left(\boldsymbol{\xi}_{t+1} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) = \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_{t+1}, \bar{\Lambda}_{t+1}\big)$$

$$\bar{\Lambda}_{t+1} = \hat{\Lambda}_{t+1}^{22} - \hat{\Lambda}_{t+1}^{21} \left(\hat{\Lambda}_{t+1}^{11}\right)^{-1} \hat{\Lambda}_{t+1}^{12} \tag{4a}$$

$$\bar{\boldsymbol{\eta}}_{t+1} = \hat{\boldsymbol{\eta}}_{t+1}^{2} - \hat{\Lambda}_{t+1}^{21} \left(\hat{\Lambda}_{t+1}^{11}\right)^{-1} \hat{\boldsymbol{\eta}}_{t+1}^{1} \tag{4b}$$

Unlike the state augmentation component to the time prediction step, marginalization significantly changes the information matrix as we show in the right-most schematic of Figure 2. Within the graphical model, marginalization has added edges between the new pose and all landmarks that were previously linked to the old robot state. Additionally, these features are now fully connected. Elements in the information matrix that were zero (i.e. missing edges) are now filled in as, in essence, information that was contained in the old pose has been distributed among the remaining map and pose. At the same time, the shade of the matrix elements (darker shades imply larger magnitudes) suggests that existing constraints have become weaker. These effects are typical of the time projection step, which tends to populate the information matrix while decreasing the magnitude of existing terms.

3.3 Measurement Update Step

Assume that the robot observes landmarks on an individual basis according to the linear measurement model,

$$\mathbf{z}_t = \mathrm{H}\boldsymbol{\xi}_t + \mathbf{v}_t$$

that, again, includes additive Gaussian noise, $\mathbf{v}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{R}\big)$. We us Bayes' rule to incorporate this information into the posterior, $p\left(\boldsymbol{\xi}_t \mid \mathbf{z}^{t-1}, \mathbf{u}^t\right) = \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_t, \bar{\Lambda}_t\big)$, via

$$p\left(\boldsymbol{\xi}_t \mid \mathbf{z}^t, \mathbf{u}^t\right) \propto p\left(\mathbf{z}_t \mid \boldsymbol{\xi}_t\right) \, p\left(\boldsymbol{\xi}_t \mid \mathbf{z}^{t-1}, \mathbf{u}^t\right). \tag{5}$$

This is equivalent to first adding $\mathbf{z}_t$ to the state vector and then conditioning on the measurement (per Table 1). The resulting information filter update step is:

$$p\left(\boldsymbol{\xi}_t \mid \mathbf{z}^t, \mathbf{u}^t\right) = \mathcal{N}^{-1}\left(\boldsymbol{\eta}_t, \Lambda_t\right)$$

$$\Lambda_t = \bar{\Lambda}_t + \mathrm{H}^\top \mathrm{R}^{-1} \mathrm{H} \tag{6a}$$

$$\boldsymbol{\eta}_t = \bar{\boldsymbol{\eta}}_t + \mathrm{H}^\top \mathrm{R}^{-1} \mathbf{z}_t \tag{6b}$$

Each row of the matrix, H, corresponds to a different measurement and is sparse, with non-zero values only at positions corresponding to the robot pose and the feature that is observed. As a result the $\mathrm{H}^\top \mathrm{R}^{-1} \mathrm{H}$ term is also sparse and modifies off-diagonal constraints between the vehicle and observed features, as well as their diagonal elements. In essence, we are adding "new information" as we strengthen links between the robot and map.

4 Conclusion

This description of feature-based SLAM information filters will, hopefully, provide some insight into some of the interesting characteristics of the canonical form and its application to SLAM. It is, by no means, complete since there are a lot of important issues that we haven't covered (e.g. nonlinear models, mean recovery[2], data association, etc.). For a more detailed discussion of SLAM information filters, including the scalability issues, [1, 4, 5] are good references.

References

[1] S. Thrun, Y. Liu, D. Koller, A.Y. Ng, Z. Ghahramani, and H. Durrant-Whyte. Simultaneous localization and mapping with sparse extended information filters. *International Journal of Robotics Research*, 23(7-8):693–716, July-August 2004.

[2] T.P. Speed and H.T. Kiiveri. Gaussian Markov distributions over finite graphs. *Annals of Statistics*, 14(1):138–150, March 1986.

[3] M. Jordan. Graphical models. *Statistical Science*, 19(1):140–155, 2004.

[4] R. Eustice, H. Singh, and J. Leonard. Exactly sparse delayed state filters. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2428–2435, Barcelona, Spain, April 2005.

[5] M. Walter, R. Eustice, and J. Leonard. A provably consistent method for imposing exact sparsity in feature-based SLAM information filters. In *Proceedings of the 12th International Symposium of Robotics Research (ISRR)*, San Francisco, CA, October 2005.

[2]Remember, we do not have an explicit representation for the mean, but can recover it from the information vector and matrix via (1).