

## Filtro de Kalman (T7)

Neste enunciado, primeiramente revisaremos conceitos essenciais de probabilidade e estatística. Após, veremos as regras básicas para implementação de um estimador de estado ótimo, o denominado Filtro de Kalman. Por fim, as tarefas do trabalho serão apresentadas.

### Probabilidade e Estatística

Variáveis aleatórias ou estocásticas são aquelas que apresentam alguma incerteza em sua definição, diferentemente do caso determinístico onde seu valor é conhecido com exatidão [1]. Um exemplo clássico de processo estocástico é o ruído de um sensor, onde não é possível prever com certeza como se dará sua evolução temporal. Apesar de não ser possível definir de forma precisa qual é o valor de uma variável aleatória  $x$ , podemos calcular qual é a sua probabilidade de apresentar um valor dentro de uma determinada faixa. Para isto, podemos aplicar a seguinte regra:

$$\text{Probability}\{a \leq x \leq b\} = \int_a^b p(x)dx, \quad (1)$$

onde  $p(x)$  representa a chamada função de densidade de probabilidade que descreve a natureza probabilística da variável aleatória  $x$ . Grande parte dos fenômenos aleatórios naturais podem ser modelados por funções de densidade gaussianas, conforme veremos adiante. Contudo, uma variável aleatória em geral pode assumir qualquer função de densidade  $p(x)$ , desde que esta função apresente a seguinte propriedade:

$$\int_{-\infty}^{\infty} p(x)dx = 1,$$

o que garante que a probabilidade de  $x$  assumir qualquer valor é 100%.

Para analisar uma variável aleatória, costuma-se utilizar certas medidas estatísticas, sendo as mais importantes delas a média e a variância. Na verdade, todas as medidas estatísticas são originárias de uma definição primordial chamada de valor esperado ou expectância:

$$E\{f(x)\} \triangleq \int_{-\infty}^{\infty} f(x)p(x)dx.$$

Esta operação pode ser interpretada como a melhor estimativa para o valor de uma função  $f(x)$  com base na informação da densidade de probabilidade  $p(x)$ . A estatística de média de uma variável é obtida pela aplicação da operação de valor esperado em  $x$  diretamente, ou seja:  $E\{x\}$ . Na literatura, costuma-se atribuir o símbolo  $\mu_x$  para indicar esta estatística fundamental:

$$\mu_x \triangleq E\{x\}.$$

Além da média, outra estatística fundamental de uma variável aleatória é a sua variância  $\sigma_x^2$ , a qual possui a seguinte definição:

$$\sigma_x^2 \triangleq \text{var}\{x\} \triangleq E\{(x - \mu_x)^2\}.$$

Quanto maior for a variância de determinada variável, maior será a sua variabilidade entorno do ponto médio. Outra medida relacionada é o desvio padrão  $\sigma_x$ , que é simplesmente a raiz quadrada da variância. A interpretação do desvio padrão é muito clara: representa a média dos desvios em relação à média.

Neste trabalho, iremos abordar a estimação simultânea de diversas variáveis aleatórias  $x_1, \dots, x_n$ , que irão representar os estados dinâmicos do processo controlado. Ao trabalhar com múltiplas variáveis, é interessante compactarmos a notação através do uso de vetores:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Assim como no casos monovariável, também existem funções de densidade de probabilidade para variáveis vetoriais. É possível calcular a probabilidade do vetor  $x$  estar localizado em uma determinada região através da mesma operação de integração mostrada em (1), porém expandida para múltiplas variáveis.

No caso multivariável, além das medidas de média e variância, podemos também definir a estatística de covariância, que representa o grau de interdependência entre pares de variáveis. A fórmula para o cálculo da covariância entre duas variáveis  $x_i$  e  $x_j$  é conforme abaixo:

$$\text{cov}\{x_i, x_j\} \triangleq E\{(x_i - \mu_i)(x_j - \mu_j)\}.$$

Observe que a estatística de covariância torna-se igual à tradicional variância quando aplicada em relação a mesma variável:

$$\text{cov}\{x_i, x_i\} = E\{(x_i - \mu_i)^2\} = \text{var}\{x_i\}.$$

Para analisar covariância cruzada entre todas as variáveis internas de um vetor aleatório  $x$ , bem com a variância individual de cada variável  $x_1$  até  $x_n$ , utiliza-se a chamada matriz de covariância  $\Sigma_x$ , que possui a seguinte definição:

$$\Sigma_x \triangleq E\{(x - \mu_x)(x - \mu_x)^T\} = \begin{bmatrix} \text{var}\{x_1\} & \text{cov}\{x_1, x_2\} & \dots & \text{cov}\{x_1, x_n\} \\ \text{cov}\{x_2, x_1\} & \text{var}\{x_2\} & \dots & \text{cov}\{x_2, x_n\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}\{x_n, x_1\} & \text{cov}\{x_n, x_2\} & \dots & \text{var}\{x_n\} \end{bmatrix}.$$

## Processos Estocásticos Normais

A distribuição de probabilidade normal, também chamada de gaussiana, é uma das mais utilizadas para modelar fenômenos estocásticos naturais. Para denotar que uma determinada variável escalar  $x$ , apresenta uma distribuição probabilística normal, costuma-se utilizar a seguinte notação:

$$x \sim \mathcal{N}(\mu_x, \sigma_x^2),$$

onde  $\mu_x$  e  $\sigma_x^2$  denotam respectivamente a média e variância da variável. Neste caso monovariável, a função de densidade de probabilidade gaussiana tem o seguinte formato:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right).$$

Observe que esta função é naturalmente definida de acordo com os parâmetros de média e variância. De acordo com a regra de integração mostrada anteriormente, podemos calcular as seguintes probabilidades para diferentes faixas de valores da variável normal  $x$  entorno de sua média:

$$\begin{aligned} \text{Probability}\{x \text{ between } \mu_x \pm \sigma_x\} &\approx 68.2\% \\ \text{Probability}\{x \text{ between } \mu_x \pm 2\sigma_x\} &\approx 95.4\% \\ \text{Probability}\{x \text{ between } \mu_x \pm 3\sigma_x\} &\approx 99.6\% \end{aligned}.$$

A distribuição gaussiana também apresenta uma versão multivariável para modelar vetores aleatórios normais:

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_x|}} \exp \left( -\frac{(x - \mu_x)^T \Sigma^{-1} (x - \mu)}{2} \right).$$

Neste caso multivariável, a função de densidade de probabilidade é definida pelo vetor de média  $\mu_x$ , que tem a mesma dimensão de  $x$ , e pela matriz de covariância  $\Sigma_x$ , que tem dimensão  $n$  por  $n$ .

## Filtro de Kalman

O Filtro de Kalman, na literatura conhecido como *Kalman Filter* (KF), é amplamente utilizado no processo de estimação dos estados de um sistema dinâmico com base na informação proveniente de sensores ruidosos [2]. Como veremos, o KF trata-se de uma abordagem ótima para solução deste problema, pois leva em conta de forma criteriosa as propriedades estatísticas do modelo do sistema e do ruído das medições. Este estimador funciona em dois passos distintos: primeiramente realiza-se a predição dos estados do sistema com base no seu modelo e, posteriormente, aplica-se correções com base nas medidas coletadas.

A formulação do KF tradicional baseia-se em um modelo discreto linear sujeito à sinais aleatórios gaussianos:

$$\begin{cases} x_k = A_k x_{k-1} + B_k u_{k-1} + w_k \\ y_k = C_k x_k + v_k \end{cases}, \quad \begin{cases} w_k \sim \mathcal{N}(0, Q_k) \\ v_k \sim \mathcal{N}(0, R_k) \end{cases}. \quad (2)$$

Os elementos mostrados no modelo acima têm o seguinte significado:

- $k$ : índice da amostra  $(0, 1, 2, 3, 4, \dots)$ ;
- $x_k$  vetor de estados do processo;
- $u_k$ : vetor entradas de controle do processo;
- $y_k$ : vetor de medidas do processo;
- $w_k$ : sinal aleatório normal referente à incerteza do modelo;
- $v_k$ : sinal aleatório normal referente ao ruído dos sensores;

- $A_k, B_k, C_k$ : matrizes referentes ao modelo nominal do sistema;
- $Q_k, R_k$ : matrizes de covariância referentes respectivamente à incerteza do modelo e ao ruído dos sensores.

Para implementação do algoritmo do Filtro de Kalman, é necessário que o projetista tenha o conhecimento do modelo nominal da planta  $A_k, B_k, C_k$  bem como dos parâmetros estatísticos  $Q_k, R_k$  que modelam a aleatoriedade do sistema. O modelo nominal  $A_k, B_k, C_k$  pode ser obtido pela modelagem física do sistema ou pela identificação a partir de ensaios. Caso o projetista tenha somente um modelo contínuo do processo, se faz necessário realizar a sua discretização. Caso modelo da planta seja não-linear, deve-se aplicar o Filtro de Kalman Estendido ao invés do tradicional. Mais adiante iremos tratar sobre este caso geral.

Quanto aos parâmetros estatísticos, a matriz de covariância  $R_k$  das medições pode ser obtida por um simples experimento prático de ensaio dos sensores, ou em alguns casos, pela consulta das especificações técnicas fornecidas pelo fabricante do sensor. Por outro lado, a matriz de covariância  $Q_k$  do modelo do sistema é mais difícil de ser determinada de maneira puramente metodológica. Em geral, costuma-se definir esta matriz de forma intuitiva, utilizando fatores de penalidade que determinam a confiabilidade/incerteza de cada equações do modelo dinâmico do processo. Por exemplo, podemos definir esta matriz na seguinte forma diagonal  $Q_k = \text{diag}\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ , onde  $\epsilon_i$  são escalares positivos. Quanto maior for a incerteza da  $i$ -ésima equação do modelo, maior deverá ser o valor atribuído a  $\epsilon_i$ .

Para representar uma estimativa do estado real  $x_k$  do sistema considerando as medidas desde  $k = 0$  até  $k = \ell$ , vamos utilizar a seguinte notação:

$$\hat{x}_{k|\ell} \triangleq E\{x_k | y_\ell, y_{\ell-1}, \dots, y_1, y_0\} ,$$

que representa o valor esperado de  $x_k$  dada a informação estatística proveniente dos sensores. Para cada instante  $k$ , a estimativa dita “em tempo real” considera a informação dos sensores desde 0 até  $k$ , ou seja:  $\hat{x}_{k|k}$ .

A solução ótima do problema de estimação, conforme proposto por Rudolf Kalman [2], requer a estimação conjunta da matriz de covariância do erro de estimação juntamente com a própria estimativa do estado. Usualmente, utiliza-se a seguinte notação para denotar esta estimativa da matriz de covariância das estimativas:

$$P_{k|\ell} \triangleq E\{(\hat{x}_{k|\ell} - x_k)(\hat{x}_{k|\ell} - x_k)^\top\} .$$

Com base no modelo estocástico do processo apresentado em (2), o objetivo do Filtro de Kalman é apresentar rotinas para o cálculo recursivo de  $\hat{x}_{k|k}$  de modo que esta estimativa seja a melhor possível. A otimalidade do Filtro de Kalman está associada a minimização da seguinte função custo:

$$J = \sum_{k=0}^{\infty} E\{(\hat{x}_{k|k} - x_k)^\top (\hat{x}_{k|k} - x_k)\} = \sum_{k=0}^{\infty} \text{trace}(P_{k|k}), \quad (3)$$

que representa o somatório para todo  $k$  do erro quadrático (variância) entre o estado real  $x_k$  e a sua estimativa em tempo real  $\hat{x}_{k|k}$ . Note que esta função objetivo pode ser expressada de forma equivalente como o somatório do traço da matriz de covariância  $P_{k|k}$ , lembrando que o traço representa a soma dos elementos da diagonal principal de uma matriz.

A implementação do algoritmo ótimo de Kalman que minimiza função de custo (3) é conforme organizado no Algoritmo 1. A dedução matemática para estas equações pode ser consultada em [3].

Para rodar iterativamente o KF mostrado no Algoritmo 1, é necessário inicializar as variáveis  $\hat{x}_{0|0}$  e  $P_{0|0}$ , que representam respectivamente a estimativa inicial dos estados e sua respectiva matriz de covariância. Mesmo que a inicialização  $\hat{x}_{0|0}$  esteja bem discrepante em relação ao estado real  $x_k$  do processo, o KF é capaz de corrigir esta estimativa com o passar do tempo. No entanto, a rapidez com que o Filtro de Kalman converge está ligado à incerteza configurada na matriz  $P_{0|0}$ . Quanto mais próximo de zero estiverem os autovalores de  $P_{0|0}$ , maior será a confiança atribuída na condição inicial. Quanto se há muita incerteza sobre a condição inicial do sistema, recomenda-se utilizar uma matriz  $P_{0|0}$  com autovalores elevados.

Algoritmo 1: Filtro de Kalman tradicional.

---

**Informações de Entrada**

---

Medição atual dos sensores:  $y_k$

Sinal de controle prévio:  $u_{k-1}$

Estimativa prévia dos estados:  $\hat{x}_{k-1|k-1}$

Covariâncias das estimativas prévias:  $P_{k-1|k-1}$

Modelo atual do sistema:  $A_k, B_k, C_k$

Covariâncias atuais do sistema:  $Q_k, R_k$

---

---

**Etapas de predição**

---

1:  $\hat{x}_{k|k-1} = A_k \hat{x}_{k-1|k-1} + B_k u_{k-1}$

2:  $P_{k|k-1} = A_k P_{k-1|k-1} A_k^T + Q_k$

---

---

**Etapas de correção**

---

1:  $S_k = C_k P_{k|k-1} C_k^T + R_k$

2:  $L_k = P_{k|k-1} C_k^T S_k^{-1}$

3:  $\hat{y}_{k|k-1} = C_k \hat{x}_{k|k-1}$

4:  $\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k (y_k - \hat{y}_{k|k-1})$

5:  $P_{k|k} = P_{k|k-1} - L_k C_k P_{k|k-1}$

---

---

**Informações de Saída**

---

Estimativa atualizada dos estados:  $\hat{x}_{k|k}$

Covariâncias das estimativas atualizadas:  $P_{k|k}$

---

## Filtro de Kalman Estendido

O Filtro de Kalman Estendido, na literatura conhecido como *Extended Kalman Filter* (EKF), representa uma generalização do Filtro de Kalman tradicional para permitir o tratamento de sistemas não-lineares conforme abaixo:

$$\begin{cases} x_k = f_k(x_{k-1}, u_{k-1}) + w_k \\ y_k = h_k(x_k) + v_k \end{cases}, \quad \begin{cases} w_k \sim \mathcal{N}(0, Q_k) \\ v_k \sim \mathcal{N}(0, R_k) \end{cases}. \quad (4)$$

O EKF funciona de forma híbrida, pois tanto o modelo não-linear original quanto linearizações são utilizadas. O modelo não-linear do processo é aplicado para realizar a predição da evolução do sistema de  $\hat{x}_{k-1|k-1}$  para  $\hat{x}_{k|k-1}$ , bem como para fazer a predição  $\hat{y}_{k|k-1}$  da medida dos sensores. Já para realizar a propagação da covariância das estimativas, o EKF emprega funções matriciais jacobianas que representam a linearização ponto a ponto do sistema:

$$A_k(x_{k-1}, u_{k-1}) \triangleq \frac{\partial f_k(x_{k-1}, u_{k-1})}{\partial x_{k-1}}, \quad C_k(x_k) \triangleq \frac{\partial h_k(x_k)}{\partial x_k}. \quad (5)$$

O algoritmo completo do EKF pode ser conferido na Algoritmo 2, onde as funções jacobianas  $A_k(x_{k-1}, u_{k-1})$  e  $C_k(x_k)$  devem ser calculadas conforme em (5).

Observe, que o EKF requer que a linearização do sistema seja computada em tempo real com base no estado atual que o sistema se encontra. Este fato torna o EKF muito mais complexo de se projetar e implementar em comparação ao tradicional KF. Tendo em vista que o EKF baseia-se nesta linearizações, é notável que a propriedade de otimalidade vinda do KF não é mantida. Contudo, visto que o EKF atualiza constantemente as linearizações, o ponto de operação do sistema está sempre próximo ao ponto de linearização, portanto para atualizações pequenas entre amostras, EKF tende a apresentar um desempenho muito próximo do que seria o ótimo teórico.



## Algoritmo 2: Filtro de Kalman Estendido.

**Informações de Entrada**

---

Medição atual dos sensores:  $y_k$

Sinal de controle prévio:  $u_{k-1}$

Estimativa prévia dos estados:  $\hat{x}_{k-1|k-1}$

Covariâncias das estimativas prévias:  $P_{k-1|k-1}$

Modelo não-linear atual do sistema:  $f_k(x_{k-1}, u_{k-1}), h_k(x_k)$

Covariâncias atuais do sistema:  $Q_k, R_k$

---

**Etapa de predição**

---

1:  $\hat{x}_{k|k-1} = f_k(x_{k-1|k-1}, u_{k-1})$

2:  $P_{k|k-1} = A_k(x_{k-1|k-1}, u_{k-1}) P_{k-1|k-1} A_k(x_{k-1|k-1}, u_{k-1})^\top + Q_k$

---

**Etapa de correção**

---

1:  $S_k = C_k(\hat{x}_{k|k-1}) P_{k|k-1} C_k(\hat{x}_{k|k-1})^\top + R_k$

2:  $L_k = P_{k|k-1} C_k(\hat{x}_{k|k-1})^\top S_k^{-1}$

3:  $\hat{y}_{k|k-1} = h_k(\hat{x}_{k|k-1})$

4:  $\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k(y_k - \hat{y}_{k|k-1})$

5:  $P_{k|k} = P_{k|k-1} - L_k C_k(\hat{x}_{k|k-1}) P_{k|k-1}$

---

**Informações de Saída**

---

Estimativa atualizada dos estados:  $\hat{x}_{k|k}$

Covariâncias das estimativas atualizadas:  $P_{k|k}$

---

## Tarefas

O objetivo deste trabalho consiste em estimar o vetor de estados  $x$  do sistema carro-pêndulo, abordado nos trabalhos anteriores, com base no vetor de medidas  $y$  provenientes dos sensores ruidosos. Para obter a nota total neste trabalho, deve-se replicar o mesmo resultado obtido no T6, porém agora usando o feedback exclusivo do vetor de medidas  $y$ , sendo aqui proibido o uso da variável  $x$  dentro da função `control.m`. Vale destacar que a flag `noise` deve ser deixada como `true`, para que o ruído dos sensores esteja ativado na simulação. Para alcançar o objetivo requisitado, deve-se empregar o EKF, conforme a teoria preliminar apresentada.

O modelo estocástico das medições provenientes dos sensores do sistema carro-pêndulo é conforme abaixo:

$$y_k = \begin{bmatrix} z_k \\ \theta_k \end{bmatrix} + v_k ,$$

onde  $z_k$  e  $\theta_k$  representam respectivamente a posição real do carrinho e o ângulo real da haste em cada amostra  $k$  de controle. O sinal vetorial aleatório  $v_k \sim \mathcal{N}(0, R)$  modela o ruído branco dos sensores, que possui a seguinte estatística de covariância:

$$R = \begin{bmatrix} 10^{-4} & 0 \\ 0 & 10^{-4} \end{bmatrix} .$$

Já o modelo dinâmico nominal do processo a ser usado no EKF é conforme dado nas equações (11), (12) e (13) do enunciado do T5. **Observe que para programar o EKF, é necessário realizar a discretização deste modelo contínuo.** Para isso, o seguinte procedimento pode ser aplicado. **Considere um modelo dinâmico contínuo originalmente dado por:**

$$\dot{x} = f_c(x, u) .$$

Definindo  $T_k$  como o intervalo de amostragem do sistema de controle entre as amostras  $k-1$  e  $k$ , segue que o modelo não-linear discretizado do sistema é aproximadamente conforme:

$$x_k = f_k(x_{k-1}, u_{k-1}) \approx T_k f_c(x_{k-1}, u_{k-1}) + x_{k-1} .$$

Em nosso, caso o intervalo de amostragem é constante e igual a 1 milissegundo. Observe que o modelo fornecido não apresenta uma definição clara

quanto a entrada aleatória de incerteza  $w_k$ , conforme apresentado em (4). Neste trabalho, adotaremos uma modelagem empírica para esta variável considerando-a normal, ou seja  $w_k \sim \mathcal{N}(0, Q)$ , com a seguinte matriz de covariância:

$$Q = \alpha I,$$

onde  $\alpha$  é um escalar positivo pequeno, preferencialmente menor que 1, e  $I$  é uma matriz identidade 4 por 4. Quanto menor for o valor escolhido para  $\alpha$ , maior será a confiabilidade que o EKF dará para o modelo do sistema em relação às leituras dos sensores, o que fará com que a propagação do ruído dos sensores seja reduzida. Por outro lado, um valor de  $\alpha$  exageradamente baixo tornará o EKF pouco robusto à distúrbios e incertezas não modeladas, como o erro introduzido pela discretização aproximada que foi sugerida anteriormente. Portanto, escolha de  $\epsilon$  está associada a uma relação de compromisso entre a filtragem do ruído e a robustez do estimador.

No seu relatório, lembre de colocar os desenvolvimentos matemáticos realizados durante o projeto, a lógica completa para programação do EKF desenvolvido. Caso não obtenha 100% de êxito no projeto, procure indicar as possíveis causas dos problemas observados.

## Referências

- [1] Thrun, S., Burgard, W. and Fox, D. “Probabilistic Robotics,” The MIT Press, Cambridge, Massachusetts.
- [2] Kalman, R. E. “A New Approach to Linear Filtering and Prediction Problems,” *Transaction of the ASME Journal of Basic Engineering*, pp. 35-45 (March 1960).
- [3] Anderson B.D.O., Moore, J.B. “Optimal Filtering,” Prentice Hall, New Jersey, 1979.