1 Appendices

1.1 Inverted Pendulum Dynamics Derivation

We can find the state space equations for the Inverted Pendulum using d'Alembert forces. Firstly we define the distance and velocity vectors to the important points:

$$\begin{aligned}
\boldsymbol{r}_P &= x\boldsymbol{i} \\
\boldsymbol{r}_{B_1/P} &= Lsin\theta\boldsymbol{i} + Lcos\theta\boldsymbol{j} \\
\boldsymbol{r}_{B_1} &= (x + Lcos\theta)\boldsymbol{i} + L\dot{\theta}sin\theta\boldsymbol{j} \\
\dot{\boldsymbol{r}}_{B_1} &= (\dot{x} + L\dot{\theta}cos\theta)\boldsymbol{i} - L\dot{\theta}sin\theta\boldsymbol{j}
\end{aligned}$$

Linear Momentum, $\boldsymbol{\rho} = \sum_i m_i\dot{\boldsymbol{r}}_{i/o} = m\dot{\boldsymbol{r}}_{B_1} + M\dot{\boldsymbol{r}}_P$:

$$\boldsymbol{\rho} = \begin{bmatrix} (M+m)\dot{x} + mL\dot{\theta}cos\theta \\ -mL\dot{\theta}sin\theta \\ 0 \end{bmatrix}$$

Moment of momentum about P, $\boldsymbol{h}_P = \boldsymbol{r}_{B_1/P} \times m\dot{\boldsymbol{r}}_{B_1}$:

$$\begin{aligned}
\boldsymbol{h}_P &= -mL(L\dot{\theta} + \dot{x}cos\theta)\boldsymbol{k} \\
\therefore \dot{\boldsymbol{h}}_P &= -mL(L\ddot{\theta} + \ddot{x}cos\theta - \dot{x}\dot{\theta}sin\theta)\boldsymbol{k}
\end{aligned}$$

We can balance moments using $\dot{\boldsymbol{h}}_P + \dot{\boldsymbol{r}}_P \times \boldsymbol{\rho} = \boldsymbol{Q}_e$ and $\boldsymbol{Q}_e = \boldsymbol{r}_{B_1/P} \times -mg\boldsymbol{j} + \boldsymbol{r}_{B_2/P} \times F_2\boldsymbol{i}$:

$$\dot{\boldsymbol{h}}_P + \dot{\boldsymbol{r}}_P \times \boldsymbol{\rho} = \begin{bmatrix} 0 \\ 0 \\ -mL(\ddot{x}cos\theta + L\ddot{\theta}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -L(mgsin\theta + 2F_2cos\theta) \end{bmatrix} = \boldsymbol{Q}_e$$

And also balance linear momentum using $\boldsymbol{F}_e = \dot{\boldsymbol{\rho}}$:

$$\dot{\boldsymbol{\rho}} = \begin{bmatrix} (m+M)\ddot{x} + mL(\ddot{\theta}cos\theta - \dot{\theta}^2sin\theta) \\ -mL(\ddot{\theta}sin\theta + \dot{\theta}^2cos\theta) \\ 0 \end{bmatrix} = \begin{bmatrix} F_1 + F_2 \\ R - (M+m)g \\ 0 \end{bmatrix} = \boldsymbol{F}_e$$

Finally we can write the system dynamics in terms of $\ddot{\theta}$ and $\ddot{x}$:

$$\ddot{\theta}\big(M + msin^2\theta\big)L = \left(\frac{2M+m}{m}F_2 - F_1\right)cos\theta + g(M+m)sin\theta - mL\dot{\theta}^2 sin\theta cos\theta \quad (1.1)$$

$$\ddot{x}(M + msin^2\theta) = F_1 - F_2cos(2\theta) + msin\theta(L\dot{\theta}^2 - gcos\theta) \quad (1.2)$$

Simplifying this for our problem by substituting in constants, we can write the full state space equation:

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \\ \dot{\theta} \\ \ddot{\theta} \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \frac{\left(\frac{2M+m}{m}F_2 - F_1\right)cos\theta + g(M+m)sin\theta - mL\dot{\theta}^2 sin\theta cos\theta}{(M+msin^2\theta)} \\ \dot{\theta} \\ \frac{F_1 - F_2cos(2\theta) + msin\theta(L\dot{\theta}^2 - gcos\theta)}{L(M+msin^2\theta)} \end{bmatrix} = \begin{bmatrix} f_1(\boldsymbol{x}, F_1, F_2) \\ f_2(\boldsymbol{x}, F_1, F_2) \\ f_3(\boldsymbol{x}, F_1, F_2) \\ f_4(\boldsymbol{x}, F_1, F_2) \end{bmatrix} \quad (1.3)$$

Using Lyapunov's indirect method, we can write the linearised equations about the equilibrium, $\boldsymbol{x}_e = [x_e, \dot{x}_e, \theta_e, \dot{\theta}_e]^T = [0,0,0,0]^T$, as:

$$\begin{bmatrix} \delta\dot{x} \\ \delta\ddot{x} \\ \delta\dot{\theta} \\ \delta\ddot{\theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_1}{\partial \dot{x}}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_1}{\partial \theta}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_1}{\partial \dot{\theta}}\Big|_{\boldsymbol{x}_e} \\ \frac{\partial f_2}{\partial x}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_2}{\partial \dot{x}}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_2}{\partial \theta}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_2}{\partial \dot{\theta}}\Big|_{\boldsymbol{x}_e} \\ \frac{\partial f_3}{\partial x}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_3}{\partial \dot{x}}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_3}{\partial \theta}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_3}{\partial \dot{\theta}}\Big|_{\boldsymbol{x}_e} \\ \frac{\partial f_4}{\partial x}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_4}{\partial \dot{x}}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_4}{\partial \theta}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_4}{\partial \dot{\theta}}\Big|_{\boldsymbol{x}_e} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta\dot{x} \\ \delta\theta \\ \delta\dot{\theta} \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial F_1}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_1}{\partial F_2}\Big|_{\boldsymbol{x}_e} \\ \frac{\partial f_2}{\partial F_1}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_2}{\partial F_2}\Big|_{\boldsymbol{x}_e} \\ \frac{\partial f_3}{\partial F_1}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_3}{\partial F_2}\Big|_{\boldsymbol{x}_e} \\ \frac{\partial f_4}{\partial F_1}\Big|_{\boldsymbol{x}_e} & \frac{\partial f_4}{\partial F_2}\Big|_{\boldsymbol{x}_e} \end{bmatrix} \begin{bmatrix} \delta F_1 \\ \delta F_2 \end{bmatrix} \quad (1.4)$$

$$\begin{bmatrix} \delta\dot{x} \\ \delta\ddot{x} \\ \delta\dot{\theta} \\ \delta\ddot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{mg}{M} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{(m+M)}{ML}g & 0 \end{bmatrix} \begin{bmatrix} \delta x \\ \delta\dot{x} \\ \delta\theta \\ \delta\dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{1}{M} & -\frac{1}{M} \\ 0 & 0 \\ -\frac{1}{ML} & \frac{2M+m}{mML} \end{bmatrix} \begin{bmatrix} \delta F_1 \\ \delta F_2 \end{bmatrix} \quad (1.5)$$

The eigenvalues are given by $det(\lambda I - A) = \lambda^2(\lambda^2 - \frac{(m+M)}{ML}g) = 0$. Therefore, the system is unstable about $\boldsymbol{x}_e$ due to the right half plane pole, $\lambda = \sqrt{\frac{(m+M)}{ML}g}$. Additionally, the time constant of this unstable system is $\tau = \sqrt{\frac{ML}{g(m+M)}}$. Note, if $M >> m, \tau \rightarrow \sqrt{\frac{L}{g}}$, which is the time constant for a simple pendulum.

It can be proved that the inverted pendulum system is controllable by showing:

$$rank[\boldsymbol{B}\ \boldsymbol{AB}\ \boldsymbol{A^2B}\ \boldsymbol{A^3B}] = 4 \quad (1.6)$$

Therefore for any initial condition we can reach $\boldsymbol{x}_e$ in finite time under these linear assumptions.

1.2 Propagation of Quantisation Error

The state space model for the quantisation of the linearised inverted pendulum can be written as:

$$\boldsymbol{x}_t^{(2D)} = C\boldsymbol{x}_t + \boldsymbol{V}_t \qquad \boldsymbol{V}_t \sim \mathcal{U}\left(\begin{bmatrix}\frac{1}{\delta x}\\ \frac{1}{\delta\theta}\end{bmatrix}\right) \tag{1.7}$$

$$\boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + B\boldsymbol{u}_t \tag{1.8}$$

Where A and B are the linearised system dynamics (valid for small time steps), and C is the linear transformation to a 2D state space, with quantisation noise **V**.

Assuming the quantisation bin sizes, δx and $\delta\theta$, are small and that x and θ are independent within the bin, the quantisation noise can be modelled as uniform random variables with covariance, $cov(\boldsymbol{V}, \boldsymbol{V}) = \mathbb{E}[\boldsymbol{V}\boldsymbol{V}^T]$:

$$= \mathbb{E}\begin{bmatrix} x^2 & x\theta \\ \theta x & \theta^2 \end{bmatrix} = \begin{bmatrix} \int_{-\delta x/2}^{\delta x/2} x^2 \cdot \frac{1}{\delta x} dx & 0 \\ 0 & \int_{-\delta\theta/2}^{\delta\theta/2} \theta^2 \cdot \frac{1}{\delta\theta} d\theta \end{bmatrix} = \begin{bmatrix} \frac{\delta x^2}{12} & 0 \\ 0 & \frac{\delta\theta^2}{12} \end{bmatrix} \tag{1.9}$$

For simplicity, let $\delta x = \delta\theta$, and therefore, $cov(\boldsymbol{V}, \boldsymbol{V}) = \sigma_v^2 I$.

Kalman filtering can be used to find an optimal estimate $\hat{\boldsymbol{x}}_n = K[\boldsymbol{x}_n|\boldsymbol{y}_{n-k:n}]$ using the algorithm (derived in [3]).

Algorithm 1 Multivariate Kalman Filtering

1: **Given:** $\hat{\boldsymbol{x}}_n = K[\boldsymbol{x}_n|\boldsymbol{y}_{1:n}]$ and $\Sigma_n = \mathbb{E}[(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n)(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n)^T]$

Prediction:

2: $\bar{\boldsymbol{x}}_{n+1} = K[\boldsymbol{x}_n|\boldsymbol{y}_{1:n}] = A\hat{\boldsymbol{x}}_n + B\boldsymbol{u}_{n+1}$

3: $\bar{\Sigma}_{n+1} = \mathbb{E}[(\boldsymbol{x}_{n+1} - \bar{\boldsymbol{x}}_{n+1})(\boldsymbol{x}_{n+1} - \bar{\boldsymbol{x}}_{n+1})^T] = A\Sigma_n A^T + \Sigma_w$ ▷ $\Sigma_w = \mathbf{0}$

Update:

4: $\tilde{\boldsymbol{y}}_{n+1} = \boldsymbol{y}_{n+1} - C\bar{\boldsymbol{x}}_{n+1}$ ▷ Calculate Innovation residual

5: $S_{n+1} = \Sigma_v + C\bar{\Sigma}_{n+1}C^T$ ▷ Calculate Innovation Covariance

6: $\hat{\boldsymbol{x}}_{n+1} = \bar{\boldsymbol{x}}_{n+1} + \bar{\Sigma}_{n+1}C^T S_{n+1}^{-1}\tilde{\boldsymbol{y}}_{n+1}$ ▷ $\Sigma_v = \sigma_v^2 I$

7: $\Sigma_{n+1} = (I - \bar{\Sigma}_{n+1}C^T S_{n+1}^{-1} C)\bar{\Sigma}_{n+1}$

Thus we can find the optimal linear estimate of the covariance of Equation (1.7) from algorithm 1 line 7:

$$\Sigma_{n+1} = (I - \bar{\Sigma}_{n+1}C^T S_{n+1}^{-1} C)\bar{\Sigma}_{n+1} \tag{1.10}$$

$$= (I - \bar{\Sigma}_{n+1}C^T(\sigma_v^2 I + C\bar{\Sigma}_{n+1}C^T)^{-1}C)\bar{\Sigma}_{n+1}, \qquad where\ \ \bar{\Sigma}_{n+1} = A\Sigma_n A^T \tag{1.11}$$

For the univariate case this reduces to:

$$\sigma_{n+1}^2 = \bar{\sigma}_{n+1}^2(1 - \frac{C^2\bar{\sigma}_{n+1}^2}{\sigma_v^2 + C^2\bar{\sigma}_{n+1}^2}) \tag{1.12}$$

$$= A^2\sigma_n^2(1 - \frac{C^2A^2\sigma_n^2}{\sigma_v^2 + C^2A^2\sigma_n^2}) \tag{1.13}$$

It is obvious from this that as $\sigma_v^2 \to 0$, $\sigma_{n+1}^2 \to 0$. Furthermore, if the spectral radius, $\rho\left(A^2(1 - \frac{C^2A^2\sigma_n^2}{\sigma_v^2+C^2A^2\sigma_n^2})\right) < 1$, then the mean squared error from quantisation will reduce to zero. Since $0 < ||\frac{C^2A^2\sigma_n^2}{\sigma_v^2+C^2A^2\sigma_n^2}||_2^2 < 1$ if $\sigma_v^2 > 0$, a sufficient condition is that $A^2 \leq 1 \implies \rho(A) \leq 1$.

For the multivariate case, it can similarly be shown (by Gelfand's Theorem, $\rho(M_1...M_n) \leq \rho(M_1)...\rho(M_n)$) that $\rho\left(\bar{\Sigma}_{n+1}C^T(\sigma_v^2 I + C\bar{\Sigma}_{n+1}C^T)^{-1}C)\right) < 1$, therefore a sufficient condition for the decay of the covariance is $\rho(A\Sigma_n A^T) \leq \rho(\Sigma_n) \implies \rho(AA^T) \leq 1$. For the linearised dynamics derived in eq. (1.5) the eigenvalues, λ are given by $\lambda^2(1-\lambda)^2 = 0 \implies \rho(AA^T) \leq 1$. Therefore, the covariance decays to zero in the linearised multivariate case and the 2D state becomes a lossless estimate of the state.

The rate of decay for the univariate case can be determined with a first order approximation about $\sigma_v^2 = 0$:

$$\sigma_{n+1}^2(\sigma_v^2) = A^2\sigma_n^2(1 - \frac{C^2A^2\sigma_n^2}{\sigma_v^2 + C^2A^2\sigma_n^2}) \tag{1.14}$$

$$\approx \sigma_{n+1}^2(0) + \delta\sigma_v^2\frac{\partial\sigma_{n+1}^2(\sigma_v^2)}{\partial\sigma_v^2}|_{\sigma_v^2=0} \tag{1.15}$$

$$= \frac{\delta\sigma_v^2}{(AC\sigma_n)^2} \tag{1.16}$$

As shown above, the mean square error decays with the square of the bin-size. Similarly, linearising about $\sigma_n^2 \approx 0$ gives $\sigma_{n+1}^2 = A^2$ for the linear case, which is constant.... Need to go to second order...

2 References

[1] M.C. Smith, I Lestas, *4F2: Robust and Non-Linear Control* Cambridge University Engineering Department, 2019

[2] G. Vinnicombe, K. Glover, F. Forni, *4F3: Optimal and Predictive Control* Cambridge University Engineering Department, 2019

[3] S. Singh, *4F7: Statistical Signal Analysis* Cambridge University Engineering Department, 2019

[4] Arthur E. Bryson Jr, *Optimal Control - 1950 to 1985.* IEEE Control Systems, 0272-1708/95 pg.26-33, 1996.

[5] I. Michael Ross, Ronald J. Proulx, and Mark Karpenko, *Unscented Optimal Control for Space Flight.* ISSFD S12-5, 2014.

[6] Zheng Jie Wang, Shijun Guo, Wei Li, *Modeling,Simulation and Optimal Control for an Aircraft of Aileron-less Folding Wing* WSEAS TRANSACTIONS on SYSTEMS and CONTROL, ISSN: 1991-8763, 10:3, 2008

[7] Giovanni Binet, Rainer Krenn and Alberto Bemporad, *Model Predictive Control Applications for Planetary Rovers.* imtlucca, 2012.

[8] Russ Tedrake, *Underactuated Robotics.* MIT OpenCourseWare, Ch.3, Spring 2009.

[9] Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction (2nd Edition).* The MIT Press, Cambridge, Massachusetts, London, England. 2018.

[10] I. Carlucho, M. De Paula, S. Villar, G. Acosta . *Incremental Q-learning strategy for adaptive PID control of mobile robots.* Expert Systems with Applications. 80. 10.1016, 2017

[11] Yuxi Li, *Deep Reinforcement Learning: An Overview.* CoRR, abs/1810.06339, 2018.

[12] Sandy H. Huang, Martina Zambelli, Jackie Kay, Murilo F. Martins, Yuval Tassa, Patrick M. Pilarski, Raia Hadsell, *Learning Gentle Object Manipulation with Curiosity-Driven Deep Reinforcement Learning.* arXiv 2019.

[13] David Silver, Julian Schrittwieser, Karen Simonyan et al, *Mastering the game of Go without human knowledge.* Nature, vol. 550, pg.354–359, 2017.

[14] David Silver, Thomas Hubert, Julian Schrittwieser et al, *A general reinforcement learning algorithm that masters chess, shogi and Go through self-pla.* Science 362:6419, pg.1140-1144, 2018.

[15] S. Thakoor, S. Nair and M. Jhunjhunwala, *Learning to Play Othello Without Human Knowledge* Stanford University Press, 2018 `https://github.com/suragnair/alpha-zero-general`

[16] HowlingPixel.com, *Elo Rating System.* `https://howlingpixel.com/i-en/Elo_rating_system` acc: 11/03/2019. published 2019