Stochastic system consideration for EDMD

October 23, 2024

1 Notation introduction

A list of all the notation and relevant definition:

1. (Autonomous) SDE in the dynamical system (Ω, μ):

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t.$$

2. A deterministic/stochastic system with flow map: $y = \Phi^t(x)$.
3. m: Number of (i.i.d.) data points
4. N: Dictionary size
5. $n, t = \frac{1}{n}$: sampling frequency and time step between two data snapshots
6. Data(i.i.d.) snapshots:

$$X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_m \\ | & & | \end{bmatrix}, \quad Y = \begin{bmatrix} | & & | \\ y_1 & \dots & y_m \\ | & & | \end{bmatrix}$$

7. Dictionary: $\mathbf{\Psi} = [\psi_1, \dots, \psi_N]$ where each $\psi_i \in L^2(\Omega, \mu)$.
8. Matrices of dictionary $\{\psi_j\}_{j=1}^N$ evaluated on data $\{x_i, y_i\}_{i=1}^m$:

$$\Psi_X = \begin{bmatrix} \rule{1cm}{0.4pt} & \mathbf{\Psi}(x_1) & \rule{1cm}{0.4pt} \\ & \vdots & \\ \rule{1cm}{0.4pt} & \mathbf{\Psi}(x_m) & \rule{1cm}{0.4pt} \end{bmatrix} = \begin{bmatrix} \psi_1(x_1) & \cdots & \psi_N(x_1) \\ \vdots & \ddots & \vdots \\ \psi_1(x_m) & \cdots & \psi_N(x_m) \end{bmatrix}$$

$$\Psi_Y = \begin{bmatrix} \rule{1cm}{0.4pt} & \mathbf{\Psi}(y_1) & \rule{1cm}{0.4pt} \\ & \vdots & \\ \rule{1cm}{0.4pt} & \mathbf{\Psi}(y_m) & \rule{1cm}{0.4pt} \end{bmatrix} = \begin{bmatrix} \psi_1(y_1) & \cdots & \psi_N(y_1) \\ \vdots & \ddots & \vdots \\ \psi_1(y_m) & \cdots & \psi_N(y_m) \end{bmatrix}$$

9. Definition of Koopman operator in deterministic system:

$$\mathcal{K}f = f \circ \Phi$$

Definition of stochastic Koopman operator(**semigroup**):

$$\mathcal{K}^t f = \mathbb{E}[f \circ \Phi^t]$$

Remark: In the deterministic system, we can still use $\mathcal{K}^t, \Phi^t$ where t is the (fixed) time interval between two sampling snapshots.

10. Definition of Koopman generator:

$$\begin{aligned}\mathcal{L}f(x) &:= \lim_{t\to 0} \frac{[\mathcal{K}^t f(x)] - f(x)}{t} \\ &= \lim_{t\to 0} \frac{\mathbb{E}[f \circ \Phi^t(x)] - f(x)}{t}\end{aligned} \tag{1}$$

EDMD framework review

Let $f(x) = \mathbf{\Psi}(x)a$. Then

$$\mathcal{K}f(x) = f(y) = \mathbf{\Psi}(y)a = \mathbf{\Psi}(x)Ka + r(x)$$

where $r(x)$ is the residual. Notice that the span of $\{\psi_1, \dots, \psi_N\}$ is not necessarily invariant under $\mathcal{K}$.

In the EDMD framework, given a dataset $\{(x_i, y_i)\}_{i=1}^M$, we want to minimize the residual over the finite data set:

$$J := \sum_{i=1}^M |r(x_i)|^2 = \sum_{i=1}^M |(\mathbf{\Psi}(y_i) - \mathbf{\Psi}(x_i)K)\,\mathbf{a}|^2,$$

so, this is equivalent to $\min_K \|\Psi_Y - \Psi_X K\|_F$ and the minimal K is

$$K = \Psi_X^\dagger \Psi_Y = (\Psi_X^T \Psi_X)^\dagger (\Psi_X^T \Psi_Y) = G^\dagger A$$

where $\dagger$ is the pseudoinverse and $G = \frac{1}{m}\Psi_X^T \Psi_X, A = \frac{1}{m}\Psi_X^T \Psi_Y$.

Remark. *Regularization through truncated SVD or by adding a small perturbation is typically applied.*

S-EDMD framework

First, we can expand the Koopman operator(semigroup) $\mathcal{K}(t)$ similar as in EDMD:

$$\begin{aligned}\mathcal{K}(t)f(x) &= \mathbb{E}[f(y)] \\ &= \mathbf{\Psi}(x)K(t)\mathbf{a} + r(x)\end{aligned} \tag{2}$$

Alternatively, we approximate the Koopman operator(semigroup) $\mathcal{K}(t)$ analogous to Taylor expansion using the definition of generator(1):

$$\begin{aligned}\mathcal{K}(t)f(x) &= \mathbb{E}[f(y)]\\ &= \mathbb{E}[\mathbf{\Psi}(y)]\mathbf{a}\\ &= \left[\mathbf{\Psi}(x) + t\cdot\boldsymbol{\mathcal{L}}\mathbf{\Psi}(x) + O(t^2)\right]\mathbf{a}\end{aligned} \tag{3}$$

where $\boldsymbol{\mathcal{L}}\mathbf{\Psi}(x) = [\mathcal{L}\psi_1(x), \dots, \mathcal{L}\psi_N(x)]$ and

$$\mathcal{L}\psi_j(x) = b(x)\cdot\nabla\psi_j(x) + \frac{1}{2}\sigma^2(x) : \nabla^2\psi_j(x) \tag{4}$$

Remark. *We can get explicit higher order term $O(t^2)$ in (3) by applying Itô's formula recursively. See A for more details.*

Remark. *In order to obtain (4), we can use Itô's formula. From Itô's formula, we know that given $f \in C_b^2(\Omega)$ and $X_0 = x$,*

$$df(X_t) = \left(b(X_t)\cdot\nabla f(X_t) + \frac{1}{2}\sigma^2(x) : \nabla^2 f(X_t)\right)dt + \sigma(X_t)\cdot\nabla f(X_t)dW_t$$

Taking expectation on its integral form, we have:

$$\mathbb{E}[f(X_t)|X_0 = x] - f(x) = \int_0^t \left(b(X_s)\cdot\nabla f(X_s) + \frac{1}{2}\sigma^2(X_s) : \nabla^2 f(X_s)\right)ds$$

Let $t \to 0$, we have

$$\begin{aligned}\frac{d}{dt}f = \mathcal{L}f(x) &\coloneqq \lim_{t\to 0}\frac{\mathbb{E}[f(X_t) - f(x)|X_0 = x]}{t}\\ &= b(X_t)\cdot\nabla f(X_t) + \frac{1}{2}\sigma^2(x) : \nabla^2 f(X_t)\end{aligned}$$

In our case, combning (3) and (2), we can minimize the following:

$$J \coloneqq \sum_{i=1}^m |r(x_i)|^2 = \sum_{i=1}^m \left|\left[\mathbf{\Psi}(x_i) + t\cdot\boldsymbol{\mathcal{L}}\mathbf{\Psi}(x_i) + O(t^2) - \mathbf{\Psi}(x_i)K\right]\mathbf{a}\right|^2$$

which is equivalent to minimize

$$\boxed{J = \min_K \|\Psi_X + t\cdot\mathcal{L}\Psi_X + O(t^2) - \Psi_X K\|_F^2} \tag{5}$$

and the minimal $K(t)$ is

$$\boxed{K(t) = \Psi_X^\dagger(\Psi_X + t\cdot\mathcal{L}\Psi_X + O(t^2))} \tag{6}$$

where the generator(time differential) matrix approximation $[\mathcal{L}\Psi_X]_{ij} = \mathcal{L}\psi_j(x_i)$.

Remark. *$O(t^2)$ in (7), (3) and (5) are different. They are scalar, vector and matrix respectively. For example, each element of $O(t^2)$ in (3) and (5) is a scalar $O(t^2)$ in (7).*

Empirically, if t is very small, we update the Koopman matrix approximation in the following way with $O(t^2)$ omitted:

$$\boxed{\widehat{K}(t) = \widehat{\Psi}_X^{\dagger}(\widehat{\Psi}_X + t \cdot \widehat{\mathcal{L}\Psi}_X)}$$

Remark. *For the case of trained basis by NN, we replace $\Psi_X + t \cdot \mathcal{L}\Psi_X + O(t^2)$ by Ψ_Y in (5),*

Now, we will show that the computed $K_{N,n,m}$ by our S-EDMD method converges not only in large data m and large dictionary size N, but also in zero-limit of sampling time $t = 1/n$.

2 Convergence in the limit of large data

(1) i.i.d. data:

$$
\begin{aligned}
&P\left(\|\widehat{K}_{N,n,m} - K_{N,n,m}\|_F > \epsilon\right) \\
&= P\left(\|\widehat{\Psi}_X^{\dagger}\big(\widehat{\Psi}_X + t\widehat{\mathcal{L}\Psi_X}\big) - \Psi_X^{\dagger}\mathbb{E}[\Psi_Y]\|_F > \epsilon\right) \\
&= P\left(\|t\widehat{\Psi}_X^{\dagger}\widehat{\mathcal{L}\Psi_X} - \Psi_X^{\dagger}\big(\Psi_X + t\mathcal{L}\Psi_X + O(t^2)\big)\|_F > \epsilon\right) \\
&= P\left(\|t\widehat{\Psi}_X^{\dagger}\widehat{\mathcal{L}\Psi_X} - t\Psi_X^{\dagger}\mathcal{L}\Psi_X + O(t^2)\|_F > \epsilon\right) \\
&\leq P\left(t\|\widehat{\Psi}_X^{\dagger}\widehat{\mathcal{L}\Psi_X} - \Psi_X^{\dagger}\mathcal{L}\Psi_X\|_F + O(t^2) > \epsilon\right) \\
&= P\left(t\|\widehat{G}^{-1}\widehat{A} - G^{-1}A\|_F + O(t^2) > \epsilon\right)
\end{aligned}
$$

where $G = \frac{1}{m}\Psi_X^*\Psi_X$ and $A = \frac{1}{m}\Psi_X^*\mathcal{L}\Psi_X$.

(From Wiki)

A function $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfies the *bounded differences property* if substituting the value of the ith coordinate x_i changes the value of f by at most c_i. More formally, if there are constants $c_1, c_2, \ldots, c_n$ such that for all $i \in [n]$, and all $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \ldots, x_n \in \mathcal{X}_n$,

$$\sup_{x_i' \in \mathcal{X}_i} |f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i.$$

Lemma 1 (McDiarmid's Inequality)**.** *Let $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ satisfy the bounded differences property with bounds $c_1, c_2, \ldots, c_n$. Consider independent random variables $X_1, X_2, \ldots, X_n$ where $X_i \in \mathcal{X}_i$ for all i. Then, for any $\epsilon > 0$,*

$$\mathbb{P}(|f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)]| \geq \epsilon) \leq 2\exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^n c_i^2}\right).$$

Proof. Omitted. □

Lemma 2 (Hoeffding's Inequality)**.** *Assume $X_i \in [a_i, b_i]$. Then, for any $\epsilon > 0$,*

$$\mathbb{P}\left(\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right] \geq \epsilon\right) \leq e^{-\frac{2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

and

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right]\right| \geq \epsilon\right) \leq 2e^{-\frac{2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

Proof. The McDiarmid's Inequality(1) directly implies Hoeffding's inequality by taking $f(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$. □

Theorem 3. *Let $|\psi_i(X_k)| \le C$ $\mathbb{P}-a.e.$ for all $1 \le i \le N$ and $1 \le k \le m$. Suppose $\|\mathcal{L}_{N,n,m}\| \le L$. Then, for any $\epsilon > 0$,*

$$\mathbb{P}\left(\left\|\widehat{G} - G\right\|_F \ge \epsilon\right) \le 2N^2 \exp\left(-\frac{m\epsilon^2}{8N^2C^4}\right),$$

$$\mathbb{P}\left(\left\|\widehat{A} - A\right\|_F \ge \epsilon\right) \le 2N^2 \exp\left(-\frac{m\epsilon^2}{8N^2C^4L^2}\right).$$

Proof. Define the random matrix $\eta(X) := \Psi(X)^T\Psi(X) \in \mathbb{R}^{N\times N}$, i.e., $\eta_{ij}(X) = \psi_i(X)\psi_j(X)$. Let $G = \mathbb{E}[\eta(X_1)], \widehat{G} = \frac{1}{m}\sum_{k=1}^m \eta(X_k)$. Then,

$$\begin{aligned}
\|\widehat{G} - G\|_F^2 &= \sum_{i=1}^N\sum_{j=1}^N \left|\widehat{G}_{ij} - G_{ij}\right|^2 \\
&= \sum_{i=1}^N\sum_{j=1}^N \left|\frac{1}{m}\sum_{k=1}^m \eta_{ij}(X_k) - \mathbb{E}\left[\eta_{ij}(X_k)\right]\right|^2
\end{aligned}$$

where $\tilde{\eta}(X) \coloneqq \eta(X) - \mathbb{E}\left[\eta(X_1)\right]$ and thus $|\tilde{\eta}_{ij}(X)| \le 2C^2$ for all $1 \le i, j \le N$. Next, applying Hoeffding's inequality (2) and union bound, we have

$$\begin{aligned}
\mathbb{P}\left(\left\|\widehat{G} - G\right\|_F \ge \epsilon\right) &= \mathbb{P}\left(\left\|\widehat{G} - G\right\|_F^2 \ge \epsilon^2\right) \\
&\le \mathbb{P}\left(\sum_{i=1}^N\sum_{j=1}^N \left|\frac{1}{m}\sum_{k=1}^m \tilde{\eta}_{ij}(X_k)\right|^2 \ge \epsilon^2\right) \\
&\le N^2\mathbb{P}\left(\left|\frac{1}{m}\sum_{k=1}^m \tilde{\eta}_{ij}(X_k)\right|^2 \ge \epsilon^2/N^2\right) \\
&= N^2\mathbb{P}\left(\left|\frac{1}{m}\sum_{k=1}^m \tilde{\eta}_{ij}(X_k)\right| \ge \epsilon/N\right) \\
&\le 2N^2 \exp\left(-\frac{2(m\epsilon/N)^2}{m(4C^2)^2}\right) \\
&= 2N^2 \exp\left(-\frac{m\epsilon^2}{8N^2C^4}\right)
\end{aligned}$$

Similarly, define $\xi(X) \coloneqq \Psi(X)^T\mathcal{L}_{N,n,m}\Psi(X)$. Since

$$|\psi_i(X)\mathcal{L}\psi_j(X)| \le C^2\|\mathcal{L}_{N,n,m}\| \le C^2L,$$

we have

$$\mathbb{P}\left(\|\widehat{A} - A\|_F \ge \epsilon\right) \le 2N^2 \exp\left(-\frac{m\epsilon^2}{8N^2C^4L^2}\right),$$

□

Lemma 4 (Philipp, Lemma C.5)**.** *Let $G, A \in \mathbb{R}^{N\times N}$ be such that G is invertible and $A \neq 0$. Let $\widehat{G}, \widehat{A} \in \mathbb{R}^{N\times N}$ be random matrices such that $\widehat{G}$ is invertible a.s. Then for any sub-multiplicative matrix norm $\|\cdot\|$ on $\mathbb{R}^{N\times N}$ and any $\epsilon > 0$ we have*

$$\mathbb{P}\left(\|G^{-1}A - \widehat{G}^{-1}\widehat{A}\| > \epsilon\right) \leq \mathbb{P}\left(\|A - \widehat{A}\| > \frac{\epsilon}{\tau}\|A\|\right) + \mathbb{P}\left(\|G - \widehat{G}\| > \frac{\epsilon}{\tau}\|G^{-1}\|^{-1}\right),$$

where $\tau = 2\|G^{-1}\|\|A\| + \epsilon$.

Theorem 5. *Let $\epsilon > 0$. Define $\epsilon_t = \left(\epsilon - O(t^2)\right)/t$. Then, with same conditions defined in Theorem 3 and Lemma 4, we have*

$$P\left(\|\widehat{K}_{N,n,m} - K_{N,n,m}\|_F > \epsilon\right) \leq 2N^2\left[\exp\left(-\frac{m}{8}\left(\frac{\epsilon_t}{NC^2\tau\|G^{-1}\|}\right)^2\right) + \exp\left(-\frac{m}{8}\left(\frac{\epsilon_t\|A\|}{NC^2\tau L}\right)^2\right)\right]$$

Proof.

$$\begin{aligned}
P\left(\|\widehat{K}_{N,n,m} - K_{N,n,m}\|_F > \epsilon\right) &= P\left(t\|\widehat{G}^{-1}\widehat{A} - G^{-1}A\|_F + O(t^2) > \epsilon\right) \\
&= P\left(\|\widehat{G}^{-1}\widehat{A} - G^{-1}A\|_F > \epsilon_t\right) \\
&\leq \mathbb{P}\left(\|G - \widehat{G}\| > \frac{\epsilon_t}{\tau}\|G^{-1}\|^{-1}\right) + \mathbb{P}\left(\|A - \widehat{A}\| > \frac{\epsilon_t}{\tau}\|A\|\right) \\
&\leq 2N^2\exp\left(-\frac{m(\frac{\epsilon_t}{\tau}\|G^{-1}\|^{-1})^2}{8N^2C^4}\right) + 2N^2\exp\left(-\frac{m(\frac{\epsilon_t}{\tau}\|A\|)^2}{8N^2C^4L^2}\right) \\
&= 2N^2\left[\exp\left(-\frac{m}{8}\left(\frac{\epsilon_t\|G^{-1}\|^{-1}}{NC^2\tau}\right)^2\right) + \exp\left(-\frac{m}{8}\left(\frac{\epsilon_t\|A\|}{NC^2\tau L}\right)^2\right)\right]
\end{aligned}$$

□

3 Convergence in zero-limit of sampling time

Now we have Koopman matrix approximation at discrete time $K_{N,n}(1/n)$. We will use it to construct a sequence of generator approximant $\{A_{N,n}\}_{n\geq 1}$ and find its limit as $n \to \infty$, if the limit exists in some sense. Once we find the limit, we can create the continuous time N-dimensional Koopman semigroup.

First, for each $N > 0$, construct the following sequence of matrices:

$$
\begin{aligned}
A_{N,n} &\coloneqq \frac{K_{N,n}\left(\frac{1}{n}\right) - I}{\frac{1}{n}} \\
&= \frac{\left[\Psi_X^\dagger \left(\Psi_X + \frac{1}{n}\cdot \mathcal{L}\Psi_X + O(\frac{1}{n^2})\right)\right] - I}{\frac{1}{n}} \\
&= \frac{\left[I + \frac{1}{n}\Psi_X^\dagger \mathcal{L}\Psi_X + O(\frac{1}{n^2})\right] - I}{\frac{1}{n}} \\
&= \Psi_X^\dagger \mathcal{L}\Psi_X + O(\tfrac{1}{n}) \\
&\quad \text{(Here the first term is not dependent on } n\text{)}
\end{aligned}
$$

Next, choose $f = \mathbf{\Psi a} \in C_b^2(\Omega)$, we have

$$
\begin{aligned}
&\|\lim_{n\to\infty} \mathbf{\Psi} A_{N,n}\mathbf{a}\| \\
&= \|\lim_{n\to\infty} \mathbf{\Psi}\left(\Psi_X^\dagger \mathcal{L}\Psi_X + O(\tfrac{1}{n})\right)\mathbf{a}\| \\
&= \|\mathbf{\Psi}\left(\Psi_X^\dagger \mathcal{L}\Psi_X\right)\mathbf{a} + \lim_{n\to\infty}\mathbf{\Psi}O(\tfrac{1}{n})\mathbf{a}\| \quad \text{(Here the } O(\tfrac{1}{n}) \text{ is a matrix)}
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\|\lim_{n\to\infty} \mathbf{\Psi}O(\tfrac{1}{n})\mathbf{a}\| \\
&= \|\lim_{n\to\infty} \begin{bmatrix}\psi_1 & \cdots & \psi_N\end{bmatrix} \begin{bmatrix} O(\frac{1}{n}) & \cdots & O(\frac{1}{n}) \\ \vdots & \ddots & \vdots \\ O(\frac{1}{n}) & \cdots & O(\frac{1}{n})\end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_N\end{bmatrix}\| \\
&= \|\lim_{n\to\infty} \begin{bmatrix}\psi_1 & \cdots & \psi_N\end{bmatrix} \begin{bmatrix} a_1' \\ \vdots \\ a_N'\end{bmatrix}\| \\
&\text{(where } a_i' = \sum_{j=1}^{N}[O(\tfrac{1}{n})]_{ij}a_j\text{)} \\
&= \|\lim_{n\to\infty} \sum_{i=1}^{N}\psi_i a_i'\| \\
&= \|\lim_{n\to\infty} \sum_{i=1}^{N}\psi_i \left(\sum_{j=1}^{N}[O(\tfrac{1}{n})]_{ij}a_j\right)\|
\end{aligned}
$$

$$\leq \| \lim_{n\to\infty} \sum_{i=1}^{N} \psi_i \left(\left(N \cdot O(\tfrac{1}{n})\right) \sum_{j=1}^{N} a_j \right) \|$$

$$= \| \lim_{n\to\infty} \left(N \cdot O(\tfrac{1}{n})\right) \sum_{i,j=1}^{N} \psi_i \left(\sum_{i,j=1}^{N} a_j \right) \| = 0$$

Therefore, for $f = \mathbf{\Psi a}$, we have

$$\lim_{n\to\infty} \mathbf{\Psi} A_{N,n}\mathbf{a} = \mathbf{\Psi} A_N \mathbf{a}$$

where $A_N = \Psi_X^\dagger \mathcal{L} \Psi_X$. In other words,

$$\lim_{n\to\infty} \mathcal{A}_{N,n} f = \mathcal{A}_N f$$

where $\mathcal{A}_{N,n}, \mathcal{A}_N$ are the operators whose matrix representation are $A_{N,n}, A_N$ respectively.(Not sure if it is appropriate and necessary to mention $\mathcal{A}_{N,n}, \mathcal{A}_N$ here. If yes, do we need to define the domain for these operators?)

Next, by Trotter-Kato Approximation theorem, we have

$$\lim_{n\to\infty} \mathbf{\Psi} K_{N,n}\mathbf{a} = \mathbf{\Psi} K_N \mathbf{a}$$

where $K_{N,n}(t) = e^{tA_{N,n}}, K_N(t) = e^{tA_N}$. In other words, we have

$$\lim_{n\to\infty} \mathcal{K}_{N,n}(t) f = \mathcal{K}_N(t) f$$

where $\mathcal{K}_{N,n}(t) \coloneqq e^{t\mathcal{A}_{N,n}}, \mathcal{K}_N(t) \coloneqq e^{t\mathcal{A}_N}$.

First Trotter–Kato Approximation Theorem.

(Trotter 1958, Kato 1959). Let $(T(t))_{t\geq 0}$ and $(T_n(t))_{t\geq 0}$, $n \in \mathbb{N}$, be strongly continuous semigroups on X with generators A and A_n, respectively, and assume that they satisfy the estimate

$$\|T(t)\|, \|T_n(t)\| \leq Me^{wt} \quad \text{for all } t \geq 0, n \in \mathbb{N},$$

and some constants $M \geq 1$, $w \in \mathbb{R}$. Take D to be a core for A and consider the following assertions.

(a) $D \subset D(A_n)$ for all $n \in \mathbb{N}$ and $A_n x \to Ax$ for all $x \in D$.

(b) For each $x \in D$, there exists $x_n \in D(A_n)$ such that $x_n \to x$ and $A_n x_n \to Ax$.

(c) $R(\lambda, A_n)x \to R(\lambda, A)x$ for all $x \in X$ and some/all $\lambda > w$.

(d) $T_n(t)x \to T(t)x$ for all $x \in X$, uniformly for t in compact intervals.

Then the implications

$$\text{(a)} \implies \text{(b)} \iff \text{(c)} \iff \text{(d)}$$

hold, while (b) does not imply (a).

4 Convergence in large dictionary size N

In this section, we want to show $\mathcal{P}_N\mathcal{K}(t)\mathcal{P}_N = \mathcal{K}_N(t)\mathcal{P}_N \to \mathcal{K}(t)$ or $\mathcal{A}_N \to \mathcal{A}$ as $N \to \infty$ in strong operator topology. If we can assume the $\mathcal{K}(t)$ in (3) is bounded, then we can use the result from Igor's paper, as stated in Theorem 6.
Here are some questions:

1. Do we show whether $\lim_{N\to\infty} \mathcal{A}_N$ or $\lim_{N\to\infty} \mathcal{K}_N(t)$ exists? Which way is better?

2. Can we assume the true Koopman operator $\mathcal{K}(t)$ is bounded in order to show $\lim_{N\to\infty} e^{t\mathcal{A}_N}$ exists? For example, if the dynamical system (Ω, μ) has an invariant measure μ, the Koopman operator $\mathcal{K}(t)$ is not only bounded but also unitary.

3. Can we also say that $K_N(t) = e^{tA_N}$ is a Galerkin approximation of true Koopman operator $\mathcal{K}(t)$?

4.1 Convergence of $\mathcal{K}_N$ to $\mathcal{K}(t)$ (in the L^2 norm)

Review on convergence analysis of EDMD: The following theorem and the assumption is from Milan and Igor's paper: *On Convergence of Extended Dynamic Mode Decomposition to the Koopman Operator.*
Assumption 2 The following conditions hold:

1. The Koopman operator $\mathcal{K} : \mathcal{F} \to \mathcal{F}$ is bounded.

2. The observables $\psi_1, \dots, \psi_N$ defining $\mathcal{F}_N$ are selected from a given orthonormal basis of $\mathcal{F}$, i.e., $(\psi_i)_{i=1}^{\infty}$ is an orthonormal basis of $\mathcal{F}$.

Lemma 2 *If $(\psi_i)_{i=1}^{\infty}$ form an orthonormal basis of $\mathcal{F} = L_2(\mu)$, then P_N^{μ} converge strongly to the identity operator I and in addition $\|I - P_N\| \leq 1$ for all N.*

Proof. Let $\phi = \sum_{i=1}^{\infty} c_i\psi_i$ *with* $\|\phi\| = 1$. *Then, by Parseval's identity* $\sum_{i=1}^{\infty} |c_i|^2 = 1$ *and*

$$\|P_N^{\mu}\phi - \phi\| = \left\| \sum_{i=N+1}^{\infty} c_i\psi_i \right\| = \sum_{i=N+1}^{\infty} |c_i|^2 \to 0$$

with $\sum_{i=N+1}^{\infty} |c_i|^2 \leq 1$ *for all* N. □

Theorem 6. *If Assumption 2 holds, then the sequence of operators $\mathcal{K}_N\mathcal{P}_N^{\mu} = \mathcal{P}_N^{\mu}\mathcal{K}\mathcal{P}_N^{\mu}$ converges strongly to $\mathcal{K}$ as $N \to \infty$, i.e.,*

$$\|\mathcal{P}_N^{\mu}\mathcal{K}\mathcal{P}_N^{\mu}\phi - \mathcal{K}\phi\| \to 0 \text{ as } N \to \infty$$

for all $\phi \in \mathcal{F}$.

Proof. Let $\phi \in \mathcal{F}$ be given. Then, writing $\phi = \mathcal{P}_N^\mu \phi + (I - \mathcal{P}_N^\mu)\phi$ we have

$$\begin{aligned}\|\mathcal{P}_N^\mu \mathcal{K} \mathcal{P}_N^\mu \phi - \mathcal{K}\phi\| &= \|(\mathcal{P}_N^\mu - I)\mathcal{K}\mathcal{P}_N^\mu \phi + \mathcal{K}(\mathcal{P}_N^\mu - I)\phi\| \\ &\leq \|(\mathcal{P}_N^\mu - I)\mathcal{K}\mathcal{P}_N^\mu \phi\| + \|\mathcal{K}\|\|(I - \mathcal{P}_N^\mu)\phi\| \\ &\leq \|(\mathcal{P}_N^\mu - I)\mathcal{K}\phi\| + \|(\mathcal{P}_N^\mu - I)\|\|\mathcal{K}\mathcal{P}_N^\mu \phi - \mathcal{K}\phi\| \\ &\quad + \|\mathcal{K}\|\|(I - \mathcal{P}_N^\mu)\phi\| \to 0\end{aligned}$$

by Lemma 2 and by the fact that $\mathcal{K}\mathcal{P}_N^\mu \phi \to \mathcal{K}\phi$ since $\mathcal{K}$ is continuous by Assumption 2. $\square$

So, if $\mathcal{K}(t)$ is bounded, we can conclude that $e^{t\mathcal{A}_N} = \mathcal{K}_N(t) \to \mathcal{K}(t)$ in strong operator topology as $N \to \infty$, using the results from Igor's paper.

4.2 Convergence of $\mathcal{A}_N$ to $\mathcal{A}$ (in the L^2 norm)

Assumption 1. *We make the following assumptions:*

1. *$\mathcal{A} : \mathcal{D}(\mathcal{A}) \to L^2(\Omega)$ is a closed operator (Koopman generator) with domain $\mathcal{D}(\mathcal{A}) = H^2(\Omega)$.*
2. *$\{\psi_i\}_{i=1}^{\infty}$ is an orthonormal basis of $H^2(\Omega)$.*
3. *$\mathcal{P}_N \mathcal{A} \mathcal{P}_N : A_N = \Psi_X^{\dagger} \mathcal{L} \Psi_X$ is the matrix representation of the finite-dimensional approximation of $\mathcal{A}$.*
4. *For any $f \in H^2(\Omega)$, its projection onto the first N basis functions is:*

$$\mathcal{P}_N f = f_N = \sum_{i=1}^{N} \langle f, \psi_i \rangle_{H^2} \psi_i$$

Definition 7. *For $f \in \mathcal{D}(\mathcal{A})$, the graph norm is defined as:*

$$\|f\|_0 = \left(\|f\|_{H^2}^2 + \|\mathcal{A} f\|_{L^2}^2 \right)^{1/2}$$

Theorem 8. *For any $f \in H^2(\Omega)$, $\|\mathcal{A} f - \mathcal{P}_N \mathcal{A} \mathcal{P}_N f\|_{L^2} \to 0$ as $N \to \infty$.*

Proof. For any $f \in H^2(\Omega)$, the error can be decomposed as:

$$\|\mathcal{A} f - \mathcal{P}_N \mathcal{A} \mathcal{P}_N f\|_{L^2} \leq \|\mathcal{A} f - \mathcal{A} \mathcal{P}_N f\|_{L^2} + \|\mathcal{A} \mathcal{P}_N f - \mathcal{P}_N \mathcal{A} \mathcal{P}_N f\|_{L^2}$$

1. (Show convergence of 1st term.) Using the graph norm, we have:

$$\|\mathcal{A} f - \mathcal{A} \mathcal{P}_N f\|_{L^2} \leq \|\mathcal{A}\| \|f - \mathcal{P}_N f\|_{H^2} \to 0$$

   Since $\mathcal{A}$ is bounded and we know that:

$$\|f - \mathcal{P}_N f\|_{H^2} \to 0 \quad \text{as } N \to \infty$$

2. (Show convergence of 2nd term.) Since $\mathcal{P}_N \to I$ strongly as $N \to \infty$:

$$\|\mathcal{A} \mathcal{P}_N f - \mathcal{P}_N \mathcal{A} \mathcal{P}_N f\|_{L^2} \to 0 \quad \text{as } N \to \infty$$

□

A "Taylor" expansion of stochastic Koopman operator

By applying Itô's formula to both $f(X_t)$ and $\mathcal{A}f(X_t)$, we can derive a "Taylor expansion" for $\mathbb{E}[f(X_t)]$ as in (3).

First, we apply Itô's formula to $f(X_t)$:

$$f(X_t) = f(x) + \int_0^t (\mathcal{A}f)(X_s)\, ds + \int_0^t f'(X_s)\, \sigma(X_s)\, dW_s.$$

Next, we treat $\mathcal{A}f$ as a function and apply Itô's formula to $\mathcal{A}f(X_s)$:

$$(\mathcal{A}f)(X_t) = (\mathcal{A}f)(x) + \int_0^t [\mathcal{A}(\mathcal{A}f)](X_s)\, ds + \int_0^t (\mathcal{A}f')(X_s)\, \sigma(X_s)\, dW_s.$$

Then, we substitute this expression for $\mathcal{A}f(X_t)$ back into the formula for $f(X_t)$:

$$\begin{aligned} f(X_t) &= f(x) + \int_0^t (\mathcal{A}f)(X_s)\, ds + \int_0^t f'(X_s)\, \sigma(X_s)\, dW_s \\ &= f(x) + \int_0^t \left[(\mathcal{A}f)(x) + \int_0^s [\mathcal{A}(\mathcal{A}f)](X_u)\, du + \int_0^s (\mathcal{A}f')(X_u)\, \sigma(X_u)\, dW_u \right] ds \\ &\quad + \int_0^t f'(X_s)\, \sigma(X_s)\, dW_s. \end{aligned}$$

After rearranging terms, we have:

$$\begin{aligned} f(X_t) = f(x) + (\mathcal{A}f)(x)\, t &+ \int_0^t \int_0^s [\mathcal{A}(\mathcal{A}f)](X_u)\, du\, ds \\ &+ \int_0^t \int_0^s (\mathcal{A}f')(X_u)\, \sigma(X_u)\, dW_u\, ds + \int_0^t f'(X_s)\, \sigma(X_s)\, dW_s. \end{aligned}$$

Taking expectations on both sides and noting that the stochastic integrals have zero mean (assuming appropriate integrability conditions), we get:

$$\mathbb{E}[f(X_t)] = f(x) + (\mathcal{A}f)(x)\, t + \int_0^t \int_0^s \mathbb{E}\left[\mathcal{A}(\mathcal{A}f)(X_u)\right] du\, ds.$$

The double integral term represents the accumulated effect of the higher-order derivatives of f over time. To understand its order, consider that if $\mathbb{E}\left[\mathcal{A}(\mathcal{A}f)(X_u)\right]$ is bounded by some constant M, then:

$$\left| \int_0^t \int_0^s \mathbb{E}\left[\mathcal{A}(\mathcal{A}f)(X_u)\right] du\, ds \right| \le M \int_0^t \int_0^s du\, ds = M \int_0^t s\, ds = M \frac{t^2}{2}.$$

This shows that the double integral is of order $O(t^2)$ when t is small.

Therefore, for small t, the expected value simplifies to:

$$\mathbb{E}[f(X_t)] = f(x) + (\mathcal{A}f)(x)\, t + O(t^2). \tag{7}$$