From Fourier to Koopman

Spectral Methods for Long-term Time Series Prediction

arXiv:2004.00574

Henning Lange, Steven L. Brunton, J. Nathan Kutz

UNIVERSITY of WASHINGTON

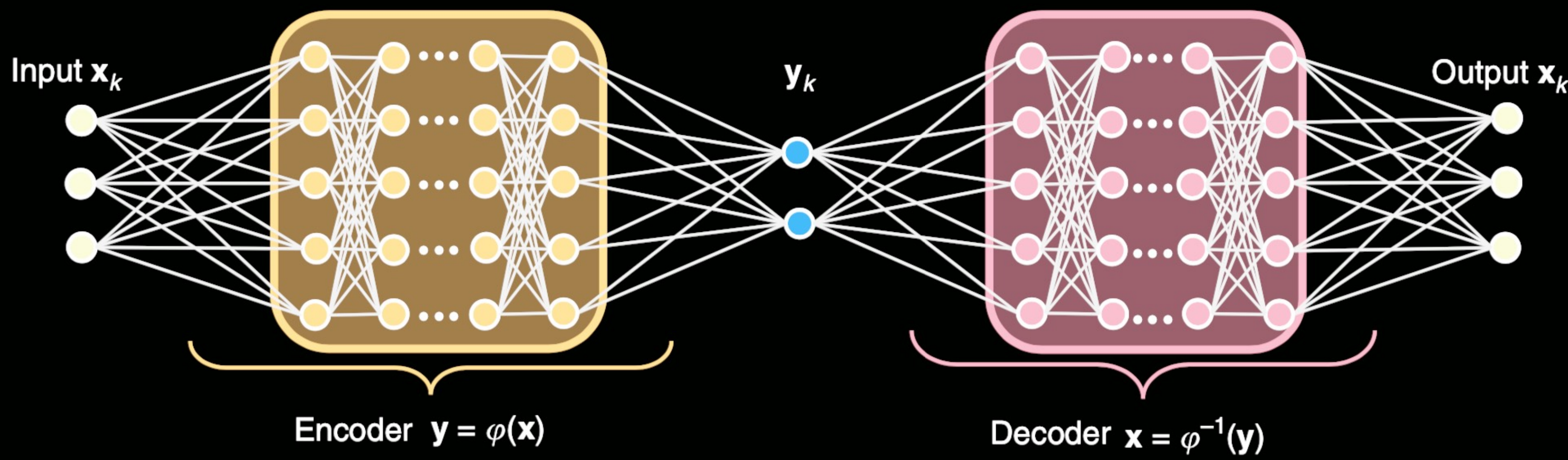
Objective

> Given data snapshots x_t from $t = 1$ to $t = T$

> Predict temporal snapshots x_{T+h}
>
> > h in the order of 10.000

> Assumption:
>
> > x_t is produced by quasi-periodic system

Spatio-Temporal Systems



UNIVERSITY *of* WASHINGTON

Outline

> Fourier Forecast
>> Similar to Fourier Transform
>> No implicit periodicity assumption

> Koopman Forecast
>> Based on Koopman theory
>> Fourier Transform in non-linear basis

Outline

- Fourier Forecast
  - Non-convex objective
- Koopman Forecast
  - Non-linear and non-convex objective
- FFT allows for obtaining global optima

Solution strategy

- Both learning objectives contain easy and hard to optimize parameters
- For both algorithms, the strategy for obtaining the global optimum of a single value of the hard to optimize parameters is introduced
  - Apply coordinate descent
  - Alternately optimize hard and easy quantities

Fourier Forecast

UNIVERSITY *of* WASHINGTON

Objective

- Goal: Fit linear dynamical y_t system to data x_t

$$\text{minimize} \quad E(\mathbf{A}, \mathbf{B}) = \sum_{t=1}^{T} (\mathbf{x}_t - \mathbf{A}\mathbf{y}_t)^2$$

$$\text{subject to} \quad \mathbf{y}_t = \mathbf{B}\mathbf{y}_{t-1}$$

$$Re[eig(B)] = 0$$

Objective

> Goal: Fit linear dynamical y_t system to data x_t

$$E(A,\omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - \mathbf{A} \begin{bmatrix} \sin(\omega_1 t) \\ \vdots \\ \sin(\omega_N t) \\ \cos(\omega_1 t) \\ \vdots \\ \cos(\omega_N t) \end{bmatrix} \right)^2$$

Objective

- Goal: Fit linear dynamical y_t system to data x_t

$$E(A, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - \mathbf{A}\Omega(\omega t)\right)^2$$

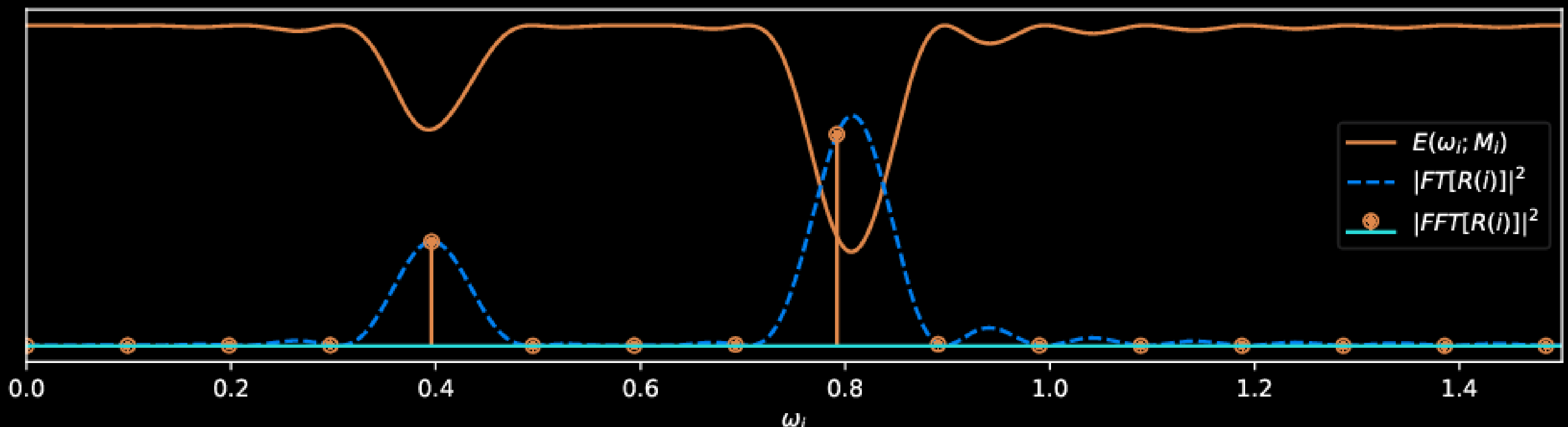
Objective

- Goal: Fit linear dynamical y_t system to data x_t
- Because of linearity of A and Ω
  - Analytic solution for ω_i
  - Symmetry relationship to Fourier Transform

$$E(A, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - \mathbf{A}\Omega(\omega t)\right)^2$$

Symmetry

$$E(A, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - \mathbf{A}\Omega(\omega t) \right)^2$$

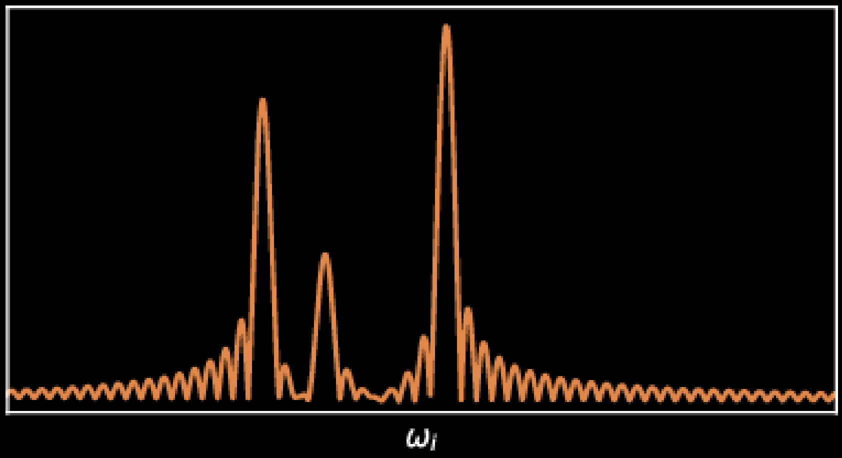


Jaynes, E. T. "Bayesian spectrum and chirp analysis." Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems. Springer, Dordrecht, 1987. 1-37.

UNIVERSITY *of* WASHINGTON

Spectral leakage

> For quasi-periodic systems, FT/error surface is superposition of sinc-functions



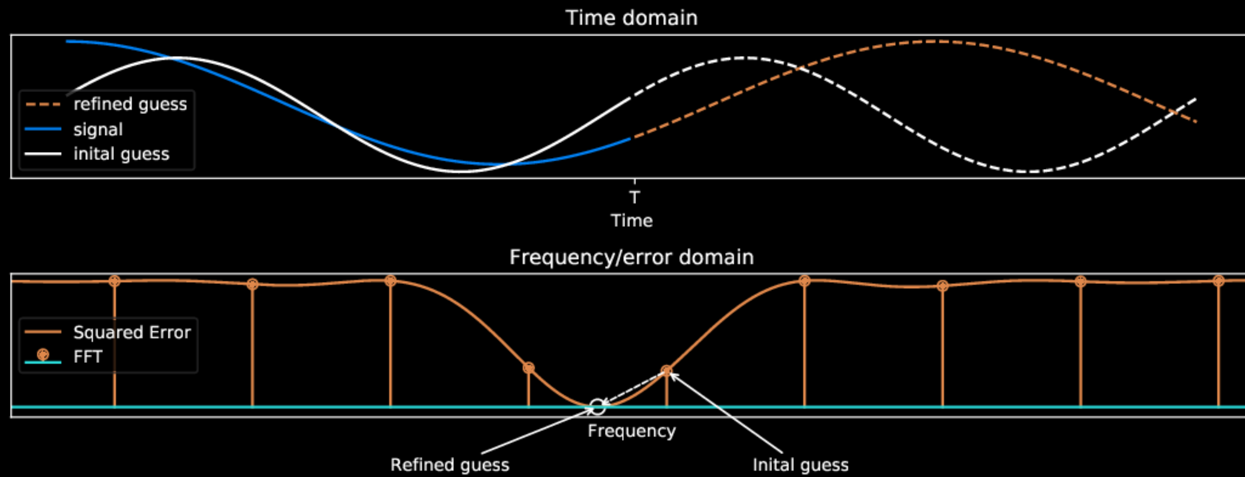
Combining FFT and GD

- Fast Fourier Transform
  - evaluates the Fourier Transform at frequencies with period T
    - harmful for forecasting
- Gradient Descent
  - because of non-convexity, will get stuck in bad local minimum

Combining FFT and GD

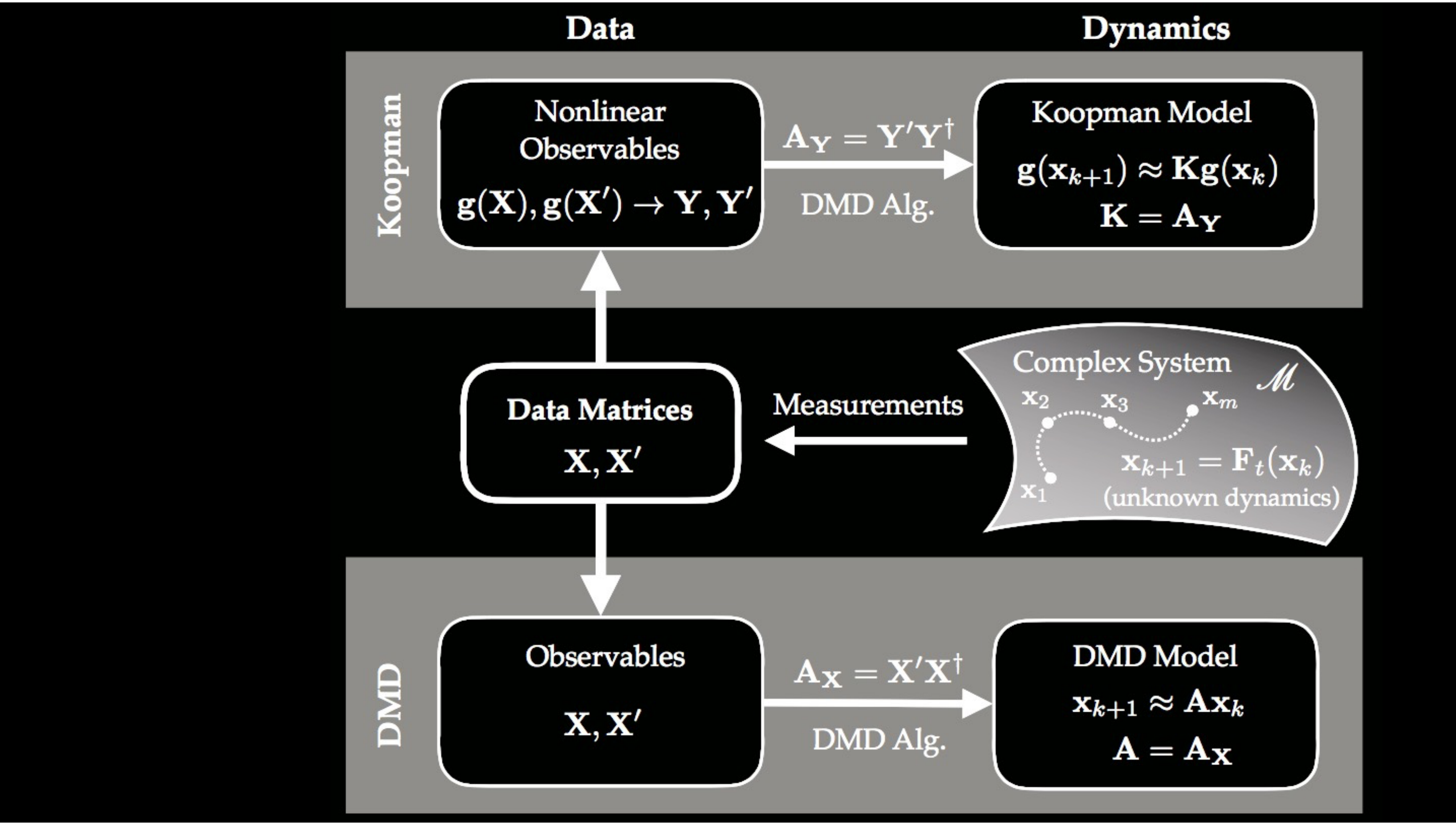
- Use Fast Fourier Transform
  - to locate global valley of error surface
- Use Gradient Descent
  - to improve initial guess of FFT to break implicit periodicity assumptions

Combining FFT and GD



Koopman Forecast

UNIVERSITY *of* WASHINGTON



Data
Dynamics
Koopman
Nonlinear Observables
$\mathbf{g}(\mathbf{X}), \mathbf{g}(\mathbf{X}') \rightarrow \mathbf{Y}, \mathbf{Y}'$
$\mathbf{A}_{\mathbf{Y}} = \mathbf{Y}'\mathbf{Y}^{\dagger}$
DMD Alg.
Koopman Model
$\mathbf{g}(\mathbf{x}_{k+1}) \approx \mathbf{K}\mathbf{g}(\mathbf{x}_k)$
$\mathbf{K} = \mathbf{A}_{\mathbf{Y}}$
Data Matrices
$\mathbf{X}, \mathbf{X}'$
Measurements
Complex System
$\mathscr{M}$
$\mathbf{x}_2$ $\mathbf{x}_3$ $\mathbf{x}_m$ $\mathbf{x}_1$
$\mathbf{x}_{k+1} = \mathbf{F}_t(\mathbf{x}_k)$
(unknown dynamics)
DMD
Observables
$\mathbf{X}, \mathbf{X}'$
$\mathbf{A}_{\mathbf{X}} = \mathbf{X}'\mathbf{X}^{\dagger}$
DMD Alg.
DMD Model
$\mathbf{x}_{k+1} \approx \mathbf{A}\mathbf{x}_k$
$\mathbf{A} = \mathbf{A}_{\mathbf{X}}$

Koopman Theory

> Koopman showed in 1931:
>
> > any non-linear dynamical system can be lifted by *non-linear* but *time-invariant* function into space where time evolution is linear

Koopman, Bernard O. "Hamiltonian systems and transformation in Hilbert space." *Proceedings of the National Academy of Sciences of the United States of America* 17.5 (1931): 315

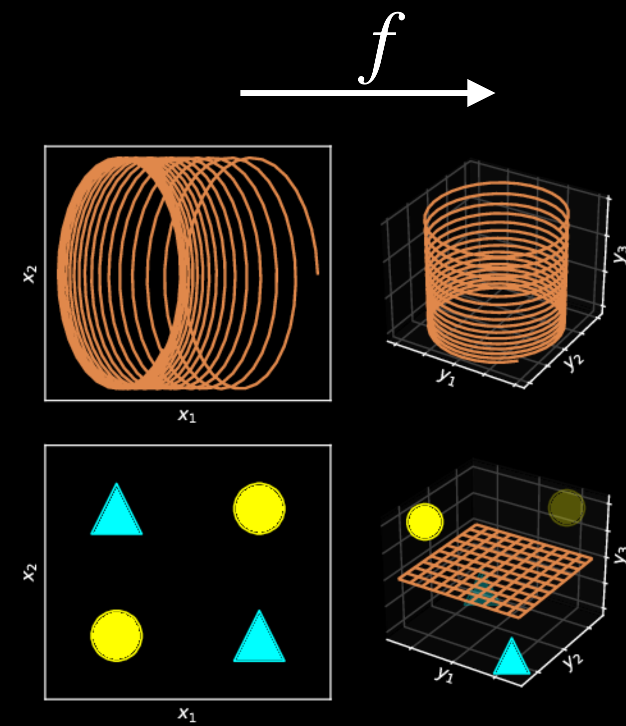
> Analogous to Cover's theorem (1965)
>
> > Theoretical underpinning of Kernel methods and Deep Learning

Cover, T.M. (1965). "Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition" (PDF). IEEE Transactions on Electronic Computers. EC-14 (3): 326–334

Koopman Theory

Koopman:

Cover:



Objective: Koopman

> Recap: Stable Linear Dynamical System

$$\Omega(\omega t) = \begin{bmatrix} \sin(\omega_1 t) \\ \vdots \\ \sin(\omega_N t) \\ \cos(\omega_1 t) \\ \vdots \\ \cos(\omega_N t) \end{bmatrix}$$

Objectives

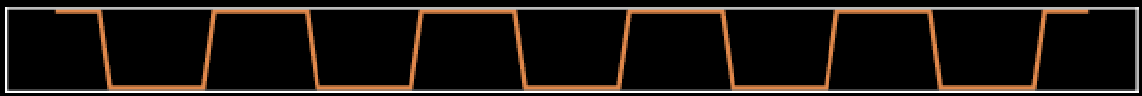
Koopman: $$E(\Theta, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - f_\Theta(\Omega(\omega t))\right)^2$$

Fourier: $$E(A, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - \mathbf{A}\Omega(\omega t)\right)^2$$

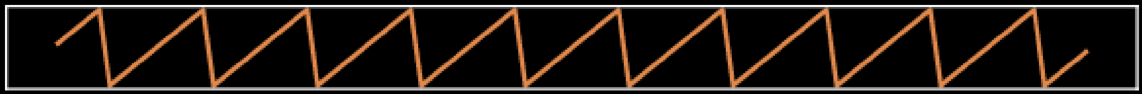
Objectives

Koopman: $$E(\Theta, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - f_{\Theta}(\Omega(\omega t))\right)^2$$

a)



b)



Objective: Koopman

Koopman: $$E(\Theta, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - f_{\Theta}(\Omega(\omega t)) \right)^2$$

Neural Network parameterized by Θ

Objective: Koopman

Koopman: $$E(\Theta, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - f_\Theta(\Omega(\omega t))\right)^2$$

Because of non-linearity, no analytical solution for ω_i

Objective: Koopman

Koopman: $$E(\Theta, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - f_{\Theta}(\Omega(\omega t))\right)^2$$

However, in spite of *non-linearity* and *non-convexity*, computing global optima in direction of ω_i possible!

Objective: Koopman

Koopman:
$$E(\Theta, \omega) = \sum_{t=1}^{T} \left(\mathbf{x}_t - f_\Theta(\Omega(\omega t))\right)^2$$
$$= \sum_{t=1}^{T} L(\Theta, \omega, t)$$

$$L(\Theta, \omega, t) = \left(\mathbf{x}_t - f_\Theta(\Omega(\omega t))\right)^2$$

Periodicity in loss

$$L(\Theta, \omega + \frac{2\pi}{t}, t) = \left(\mathbf{x}_t - f_\Theta(\Omega((\omega + \frac{2\pi}{t})t))\right)^2$$

$$= \left(\mathbf{x}_t - f_\Theta(\Omega(\omega t))\right)^2$$

$$= L(\Theta, \omega, t)$$

UNIVERSITY *of* WASHINGTON

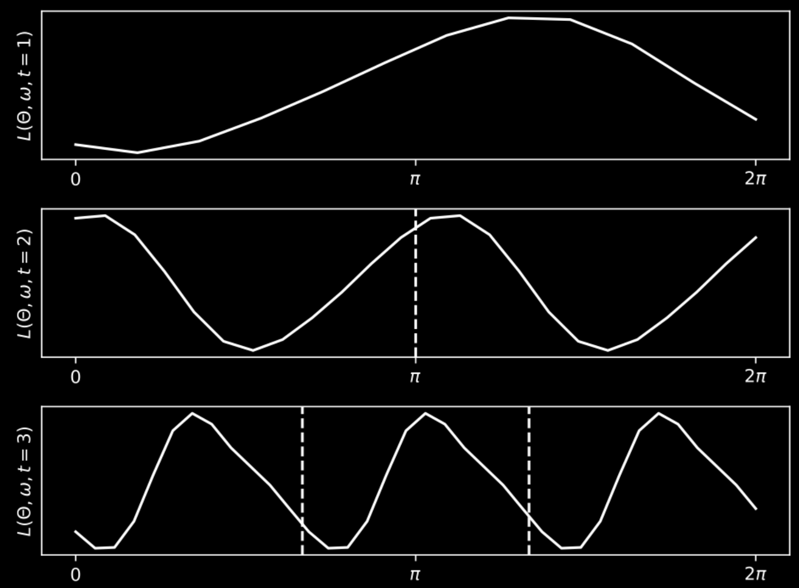
Periodicity in loss

$$L(\Theta, \omega, t) = L(\Theta, \omega + \frac{2\pi}{t}, t)$$

$$\sin((\omega + \frac{2\pi}{t})t) = sin(\omega t + 2\pi) = \sin(\omega t)$$

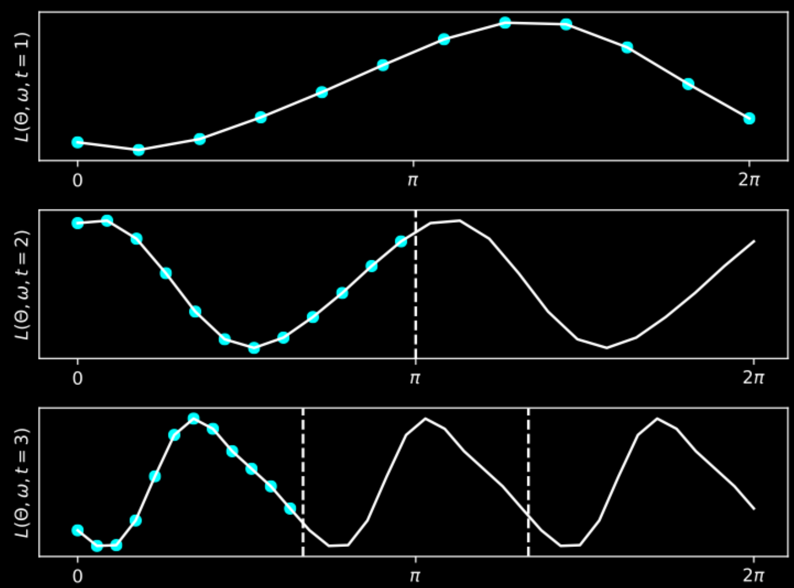
Periodicity in loss

$$L(\Theta, \omega, t) = L(\Theta, \omega + \frac{2\pi}{t}, t)$$



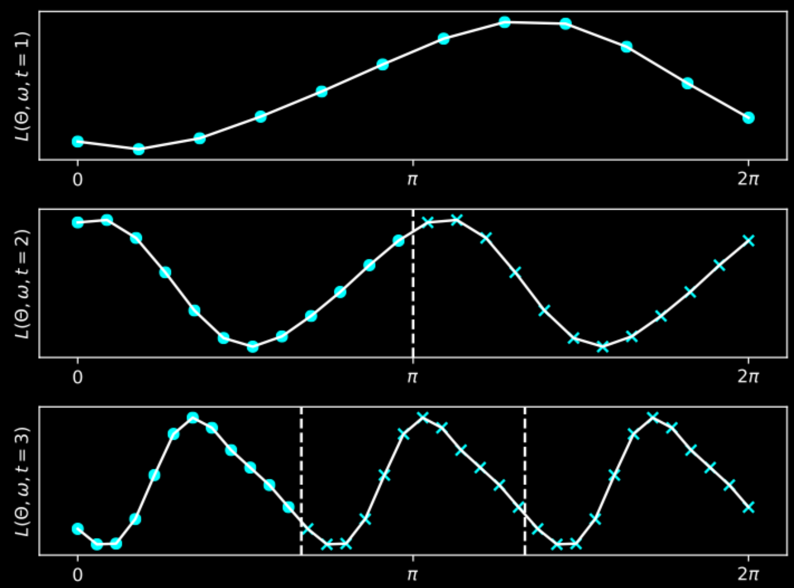
Computing the loss

For all t, compute loss within $\frac{2\pi}{t}$



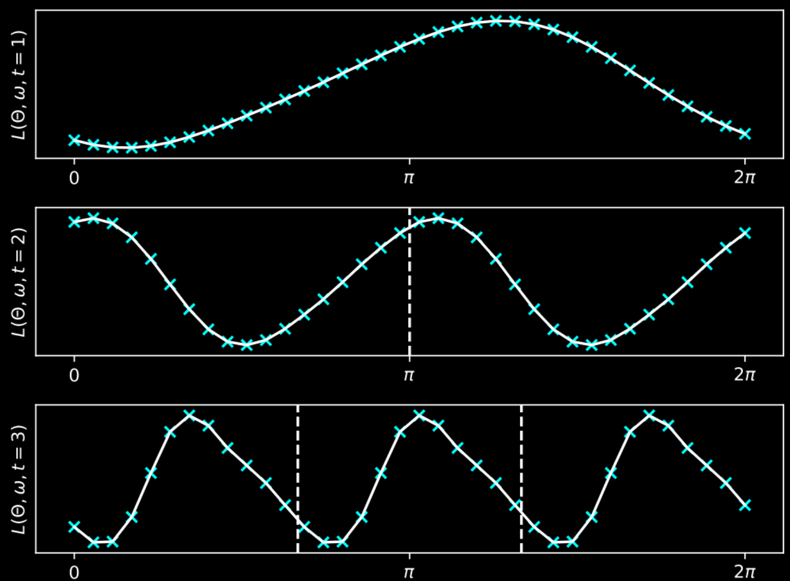
Computing the loss

For all t, repeat computed loss t times



Computing the loss

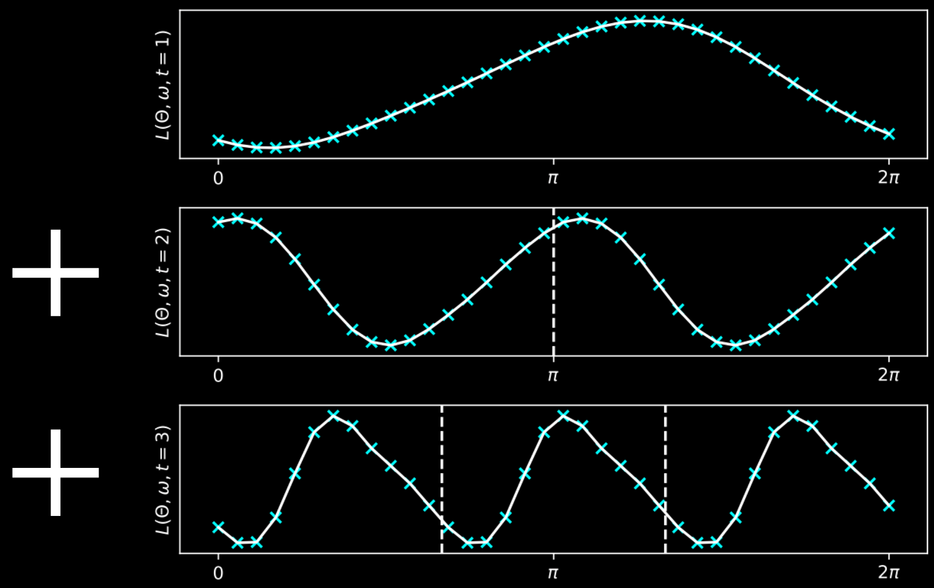
For all t, resample loss



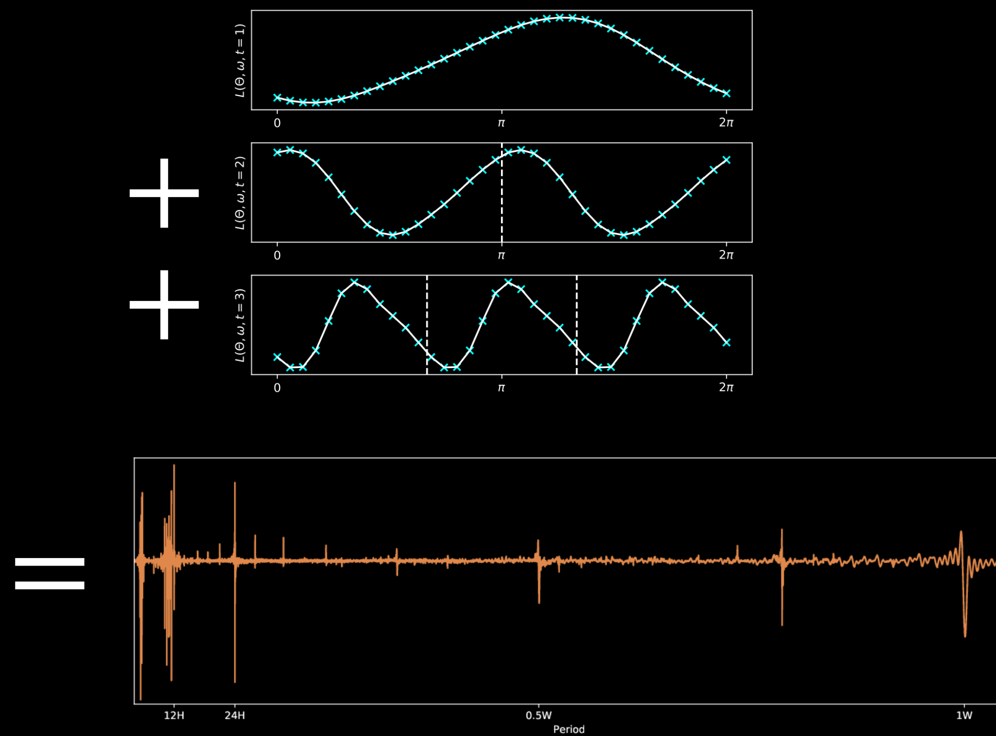
UNIVERSITY *of* WASHINGTON

Computing the loss

Sum all ‘temporally local’ losses



Computing the loss



Computing the loss

Easy and efficient to implement in freq. domain!

```
for t in range(T):
      E_ft[range(K)*t] += fft(L[t])
E = ifft(E_ft)
```

Results

UNIVERSITY *of* WASHINGTON

Results: Theoretical

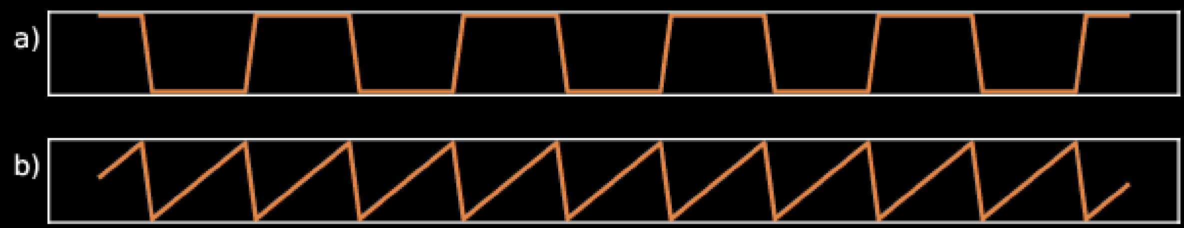
> **Fourier algorithm has universal approximation properties on finite datasets**

> Sines and cosine form an orthogonal basis
> > which is periodic in T

> Analogous to Cover's theorem, requires N dimensional space

Results: Theoretical

> For infinite data, Koopman algorithm is more expressive than Fourier counterpart



Results: Theoretical

- Close relationship to Bayesian Spectral analysis
- Error grows *linear* in time and with noise variance
- But shrinks superlinearly with amount of data

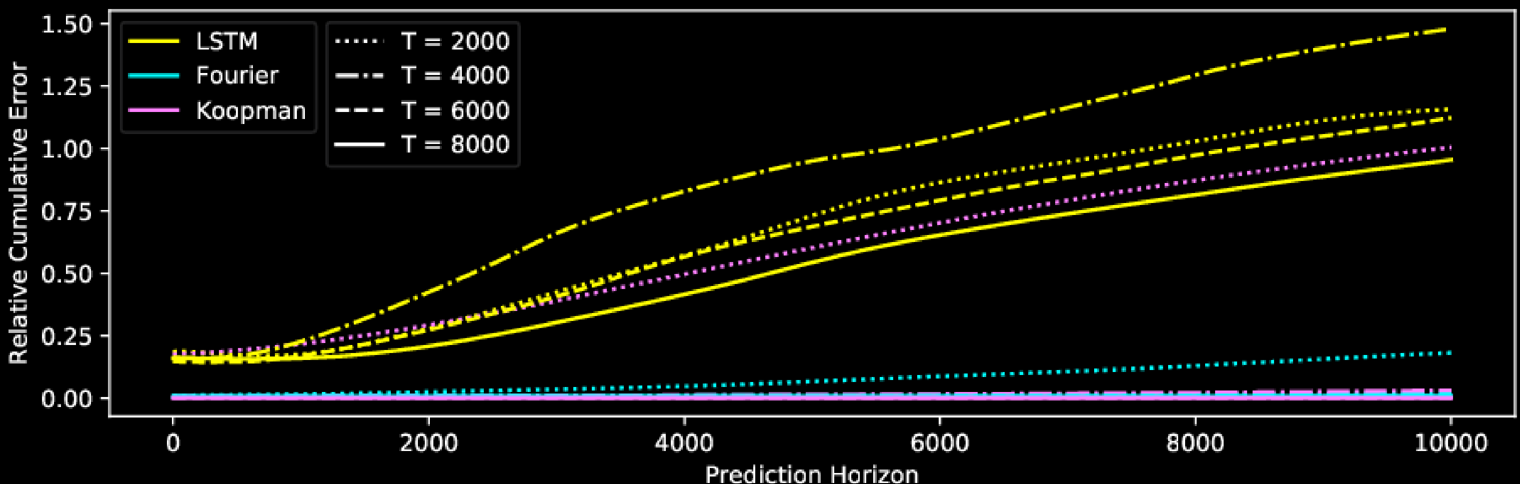
$$|\hat{x}_t(\omega) - \hat{x}_t(\omega^*)| \in \mathcal{O}\left(\frac{t}{\sqrt{T^3}} \sum_i \frac{\sigma^2}{A_i}\right)$$

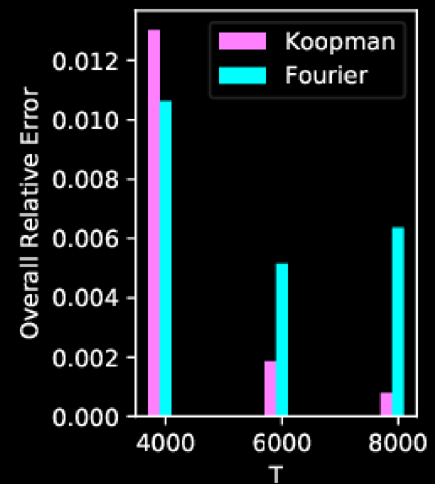
Bretthorst, G. Larry. Bayesian spectrum analysis and parameter estimation. Vol. 48. Springer Science & Business Media, 2013.

Jaynes, E. T. "Bayesian spectrum and chirp analysis." Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems. Springer, Dordrecht, 1987. 1-37.

Results: Practical

$$x_t = \sin\left(\frac{2\pi}{24}t\right)^{17} + \epsilon_t$$

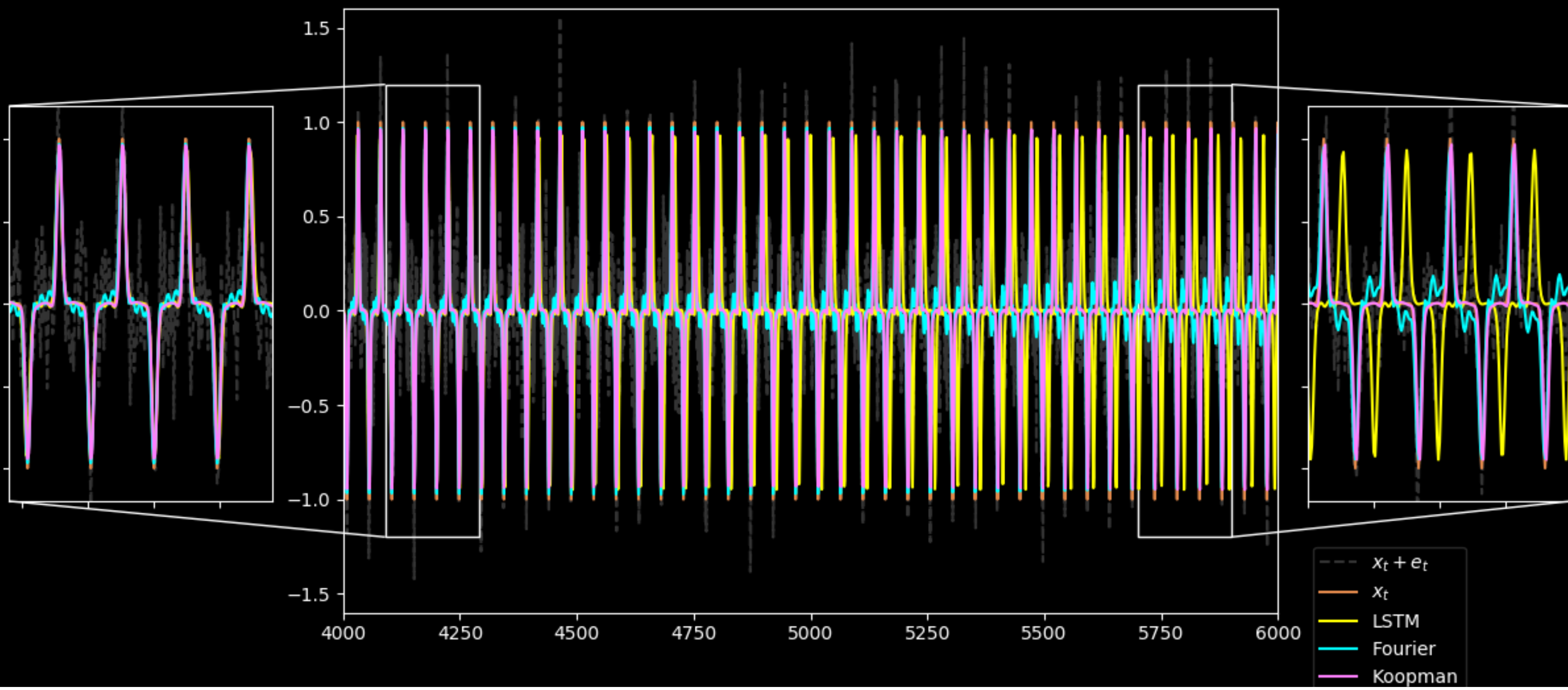




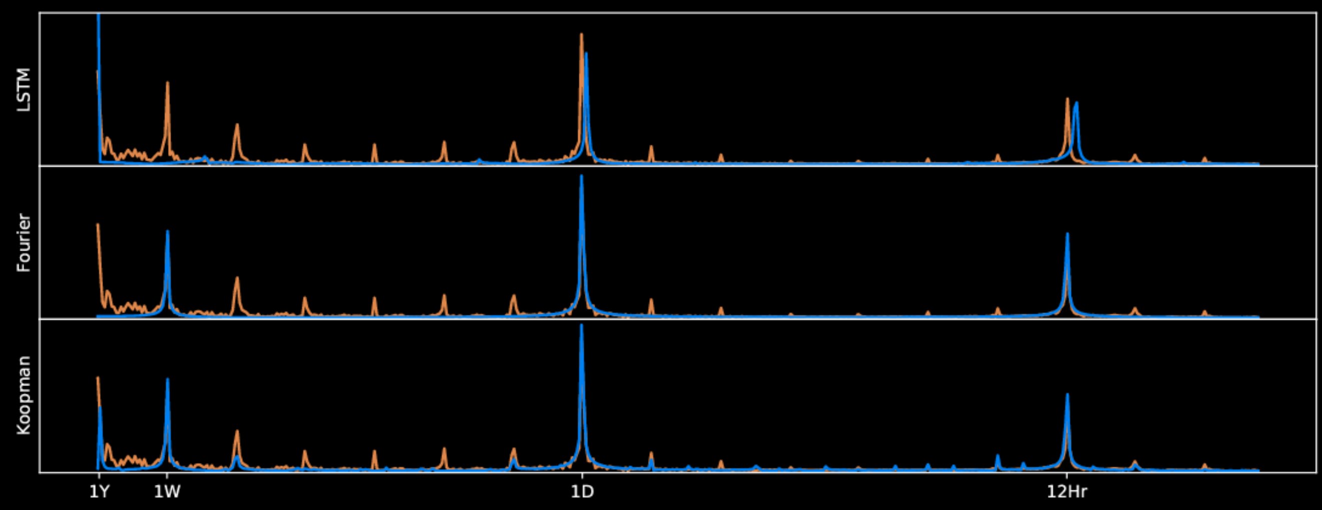
UNIVERSITY *of* WASHINGTON

Algorithm	Forecast Horizon				Patterns		
	25%	50%	75%	100%	D	W	Y
Koopman Forecast	**0.19**	**0.21**	**0.19**	**0.19**	✓	✓	✓
Fourier Forecast	0.31	0.39	0.33	0.3	✓	✓	✓
LSTM	0.37	0.4	0.42	0.45	✓	×	×
GRU	0.53	0.55	0.52	0.5	✓	×	×
Echo State Network	0.67	0.73	0.76	0.73	✓	×	×
AR(1,12,24,168,4380,8760)	0.75	0.95	1.07	1.13	✓	✓	✓
CW-RNN (data clocks)	1.1	1.14	1.14	1.15	(✓)	×	×
CW-RNN	1.05	1.08	1.08	1.09	(✓)	×	×
AutoARIMA	0.83	1.11	1.18	1.26	×	×	×
Fourier Neural Networks	1.1	1.15	1.21	1.21	✓	×	×

Results: Practical



Results: Practical



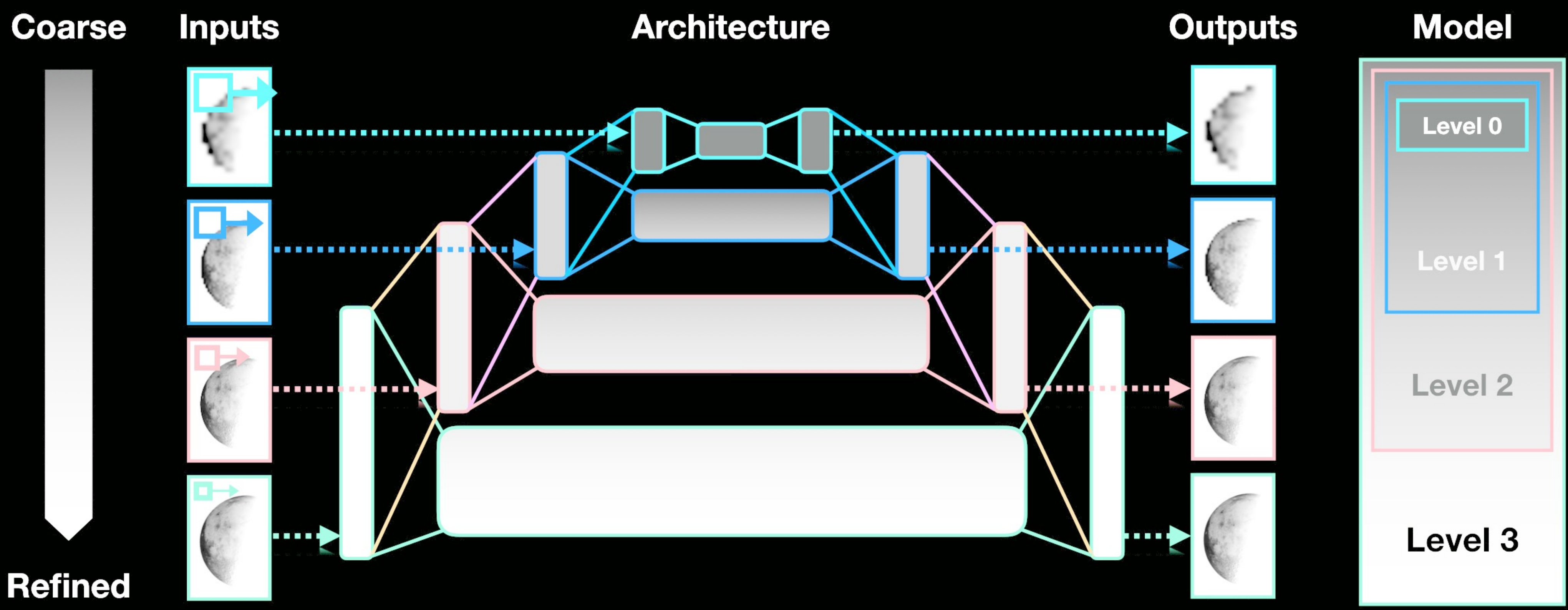
UNIVERSITY *of* WASHINGTON

Results: Practical

Ground Truth

Prediction

Spatio-Temporal Systems



Summary

- Fit linear and non-linear oscillators to data
  - non-convex and non-linear objective
- Many real world phenomena are quasi-periodic
  - gait, (space) weather, fluid flows, epidemiological data, power systems, sales, room occupancy, …
- Code is available:
  - https://github.com/helange23/from_fourier_to_koopman