New views of the Kalman Filter connect machine learning, signal processing, and system identification

Dr. Richard E. Turner (ret26@cam.ac.uk)

Computational and Biological Learning Lab
Department of Engineering
University of Cambridge
signal

audio
video
brain recordings

applied

speech recognition
source separation
hearing devices
voice manipulation
object recognition
EEG analysis
signal
audio
video
brain recordings

speech recognition
source separation
hearing devices
voice manipulation
object recognition
EEG analysis
signal
representation
application

audio
video
brain recordings

filter 1
filter 2
filter 3

speech recognition
source separation
hearing devices

voice manipulation
object recognition
EEG analysis
signal application representation

filter 1
filter 2
filter 3

speech recognition
source separation
hearing devices
voice manipulation
object recognition
EEG analysis

audio
video
brain recordings

current methods:
not robust
not adaptive
signal representation application

audio video brain recordings

filter 1

filter 2

filter 3

speech recognition
source separation
hearing devices

voice manipulation
object recognition
EEG analysis

current methods:
not robust
state estimation
not adaptive
system identification
signal application
representation
application

speech recognition
source separation
hearing devices

voice manipulation
object recognition
EEG analysis

current methods:
not robust
state estimation
not adaptive
system identification

audio
video
brain recordings

filter 3
filter 2
filter 1
Core problem

\[ y(t) \]
Core problem

\[ y(t) = c(t) \times a(t) \]

\[ c(t) = \cos(\phi(t)) \]
III posed

\[ y(t) = \cos(\phi(t)) \]
Constraints

\[ y(t) = c(t) = \cos(\phi(t)) \]

\[ a(t) = \text{frequency} \frac{d}{dt} \phi(t) \]
Problems with current approaches
Problems with current approaches

Problems with current approaches

Problems with current approaches

Problems with current approaches

**Principled approach**

- **Ill posed** problem - explicitly state assumptions
- **Soft constraints** - interpolate through noise
- **Adaptability** - learn time-scales
- Solution: **probabilistic approach**
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i \phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i \phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \mathcal{R} \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i \phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \mathcal{R} [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re[a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i \phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re [a(t) \exp(i\phi(t))] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach

\[ y(t) = \Re \left[ a(t) \exp(i\phi(t)) \right] \]
New twist on standard approach
New twist on standard approach

soft constraints

match signal

\[ y_t = [1, 0]x_t + \sigma_y \eta_t \]

rotate slowly

\[ x_t = Rx_{t-1} + \sigma \epsilon_t \]
New twist on standard approach

soft constraints
match signal
\[ y_t = [1, 0]x_t + \sigma_y \eta_t \]
rotate slowly
\[ x_t = R x_{t-1} + \sigma \epsilon_t \]

Estimation: Kalman Smoother
\[ p(x_t | y_{1:T}, \theta) \]

Turner, 2010
Running example
Running example
new approach $\equiv$ Hilbert + prefiltering
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude

![Graph of phasor with large initial amplitude]
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude

![Graph showing phasors with large initial amplitude.](image)
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude

![Phasor diagram](image-url)
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude
Phasor - large initial amplitude

Large amplitude: fixed frequency
Phasor dynamics - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude

Diagram showing a phasor with a small initial amplitude.
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude

![Graph showing phasor with small initial amplitude.](image-url)
Phasor - small initial amplitude
Phasor - small initial amplitude

![Phasor Diagram]
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude
Phasor - small initial amplitude

Small amplitude:
unconstrained frequency
Summary

Observed

• Large amplitude $\implies$ strong constraint on frequency
• Small amplitude $\implies$ no constraint on frequency

Desired

• Large amplitude $\implies$ no constraint on frequency
• Small amplitude $\implies$ strong constraint on frequency

Solution

• Place independent constraints on amplitude and frequency
Bivariate Isotropic Gaussian: $p(x_1, x_2) = \text{Norm}(x; \mu, \sigma^2 I)$
Restrict to unit circle: $p(\theta | x_1^2 + x_2^2 = 1)$
von Mises: \( p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \)
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \( p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \)
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: $p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu))$
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \( p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \)
von Mises: \( p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \)
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \( p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \)
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
**von Mises:** \( p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \)
von Mises:  

\[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
von Mises: $p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu))$
von Mises: \[ p(\theta) = \frac{1}{Z(k)} \exp(k \cos(\theta - \mu)) \]
Bivariate Anisotropic Gaussian: \( p(x_1, x_2) = \text{Norm}(x; \mu, \Sigma) \)
\[
p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2)))
\]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
$$p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2)))$$
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
\[ p(\theta) = \frac{1}{Z(k)} \exp(k_1 \cos(\theta - \mu_1) + k_2 \cos(2(\theta - \mu_2))) \]
Time-series version: Dynamics (Prior)
Time-series version: Estimation (Posterior)
Estimation: Soften constraints
Estimation: Soften constraints

$\mathcal{G}(x; \mu, \Sigma) I(x_1^2 + x_2^2 = 1)$
Estimation: Soften constraints

\( g(x; \mu, \Sigma) 1(x_1^2 + x_2^2 = 1) \approx g(x; \mu, \Sigma) g(x; \mu^*, \Sigma^*) \)
Estimation: Soften constraints

\[ G(x; \mu, \Sigma) 1(x_1^2 + x_2^2 = 1) \]

\[ \overset{\text{MOM}}{=} G(x; \mu, \Sigma) G(x; \mu^*, \Sigma^*) \]
Expectation Propagation

\[ p(x_1:T) = \prod_{t} \phi(x_t, x_{t-1}) \]

Minka, 2001
Expectation Propagation

\[ p(x_{1:T}) = \prod_{t} \phi(x_t, x_{t-1}) \approx \prod_{t} \hat{\phi}(x_t, x_{t-1}) \]

Minka, 2001
Expectation Propagation

$\mathbf{1)}$ delete approximate potential  

$\mathbf{2)}$ add true potential  

$\mathbf{3)}$ update approximate potential (moment matching)

\[ p(x_{1:T}) = \prod_t \phi(x_t, x_{t-1}) \approx \prod_t \hat{\phi}(x_t, x_{t-1}) \]

\[ \frac{1}{\hat{\phi}(x_a, x_{a-1})} \prod_t \hat{\phi}(x_t, x_{t-1}) \]

1) delete approximate potential

\[ \text{Minka, 2001} \]
Expectation Propagation

\[ p(x_1:T) = \prod_t \phi(x_t, x_{t-1}) \approx \prod_t \hat{\phi}(x_t, x_{t-1}) \]

\[ \frac{\phi(x_a, x_{a-1})}{\hat{\phi}(x_a, x_{a-1})} \prod_t \hat{\phi}(x_t, x_{t-1}) \]

1) delete approximate potential

2) add true potential

Minka, 2001
Expectation Propagation

\[ p(x_{1:T}) = \prod_{t} \phi(x_t, x_{t-1}) \approx \prod_{t} \hat{\phi}(x_t, x_{t-1}) \]

\[ \frac{\phi(x_a, x_{a-1})}{\hat{\phi}(x_a, x_{a-1})} \prod_{t} \hat{\phi}(x_t, x_{t-1}) \overset{\text{MOM}}{=} \prod_{t} \hat{\phi}(x_t, x_{t-1}) \]

1) delete approximate potential
2) add true potential
3) update approximate potential (moment matching)

Minka, 2001
Estimation via EP
Estimation via EP
Estimation via EP

\[ (t-2) \quad (t-1) \quad t \quad (t+1) \quad (t+2) \]
Estimation via EP
Estimation via EP

t−2  t−1  t  t+1  t+2
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP

\[ t-2 \quad t-1 \quad t \quad t+1 \quad t+2 \]
Estimation via EP
Estimation via EP

t−2  t−1  t  t+1  t+2
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Estimation via EP
Summary

- **Powerful method** for time-series of circular variables
- **Connects** simple Gaussians to difficult circular variables
- **Estimation**: Kalman smoothing with a novel moment-matching step
- **System identification/learning**: EP used to adapt parameters too
RESULTS
Synthetic signals
Synthetic signals: Hilbert
Synthetic signals: New approach
Synthetic signals - Missing Data
Synthetic signals - Missing Data

The diagram illustrates synthetic signals measured over time, with time plotted along the x-axis and signal strength along the y-axis. The plot also shows frequency in Hertz, with shaded areas indicating periods of missing data.
Synthetic signals - Missing Data

The graph shows a signal with missing data indicated by shaded regions. The signal is plotted against time (x-axis) and frequency (y-axis). The shaded areas represent the intervals where data is missing. The graph illustrates how the signal behaves over time and its corresponding frequency spectrum.
Synthetic signals - Missing Data
Synthetic signals - Missing Data

![Graph showing synthetic signals with missing data. The graph plots signal strength over time and frequency. The x-axis represents time in seconds, ranging from 0 to 0.4 seconds. The y-axis on the left represents signal strength ranging from -2 to 2. The y-axis on the right represents frequency in Hz, ranging from 0 to 100. Grey shaded areas indicate missing data points.](image-url)
Synthetic signals - Missing Data
Natural signals: running example
Natural signals: running example
Natural signals: running example
Natural signals: Speech
Natural signals: Speech
EEG Data
Summary

- state of the art demodulation
- higher quality
- higher computational cost

applications: voice manipulation
brain recording (EEG analysis)
cochlear implants
auditory science
Multiple channels
Audio modelling

Turner, 2010
Audio modelling

- fire
- stream
- wind
- rain
- tent-zip
- foot step

Turner, 2010
Audio modelling

Turner, 2010
Statistical texture synthesis

• Old approach: build detailed physical models (e.g. rain drops)

• New approach
  – train model on your favourite texture
  – sample from the prior, and then from the likelihood.

• Waveform unique, but statistically matched to original

• Often perceptually indistinguishable
Audio denoising

Turner, 2010
Probabilistic signal processing

Classical signal processing
Probabilistic signal processing

Classical signal processing

fixed

Gaussian

STFT
filter bank & Hilbert spectrogram

estimation

Turner et al. in prep.
Probabilistic signal processing

Classical signal processing

robustness adaptation

fast methods

important variables

STFT

filter bank & Hilbert spectrogram

Turner et al. in prep.

fixed

Gaussian

estimation
Probabilistic approach is powerful and general

- Probabilistic language **unifying** frame-work
- **Robust to noise**
- **Adapt to the signal**
video signal processing

learning

Slow Feature Analysis
(classical method: object recognition gesture recognition)

Turner et al. 2007
video signal processing $\rightarrow$ audio signal processing

state of the art video denoising

Turner, submitted.
Synthetic signals: Hilbert
Synthetic signals: Filter+Hilbert
Synthetic signals: Probabilistic

![Graph showing synthetic signals with probabilistic signals over time and frequency.](image-url)
Probabilistic signal processing

Cemgil & Godsill
X

Classical signal processing

Filter Bank & Hilbert
X

____
estimation

Turner et al. in prep.
Probabilistic signal processing

Cemgil & Godsill

freq shift

X \leftrightarrow Z

Qi & Minka

Filter Bank & Hilbert

freq shift

X \leftrightarrow Z

STFT

estimation

Classical signal processing

Turner et al. in prep.
Probabilistic signal processing

Cemgil & Godsill

freq shift

X ←→ Z

abs

a

Amplitudes

Qi & Minka

Filter Bank & Hilbert

freq shift

STFT

X ←→ Z

estimation

abs

a

Spectrogram

Turner et al. in prep.

Classical signal processing
Probabilistic signal processing \(\xrightarrow{\text{freq shift}}\) Classical signal processing

Cemgil & Godsill

\(X\) \(\xrightarrow{\text{freq shift}}\) \(Z\)

\(\text{abs}\) \(\xrightarrow{\text{freq shift}}\) \(\text{abs}\)

Amplitudes

Qi & Minka

Filter Bank & Hilbert

\(X\) \(\xrightarrow{\text{freq shift}}\) \(Z\)

\(\text{abs}\) \(\xrightarrow{\text{freq shift}}\) \(\text{abs}\)

Spectrogram

\(\text{estimation}\)

Turner et al. in prep.