Learning Signals with Simple Fourier Transforms

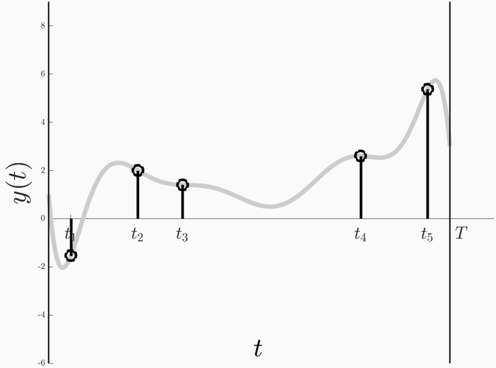
Christopher Musco, Princeton University

Solving an old problem in signal processing using tools from randomized algorithms.

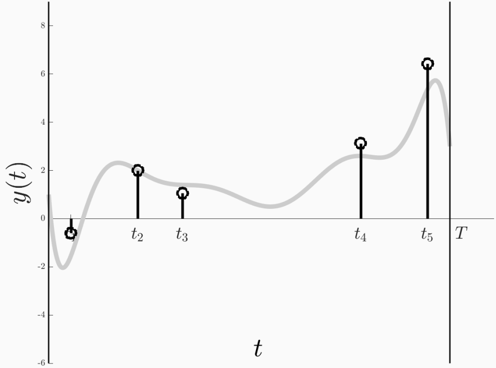
Solving an old problem in signal processing using tools from randomized algorithms.

(matrix sketching, randomized SVD, Laplacian linear systems.)

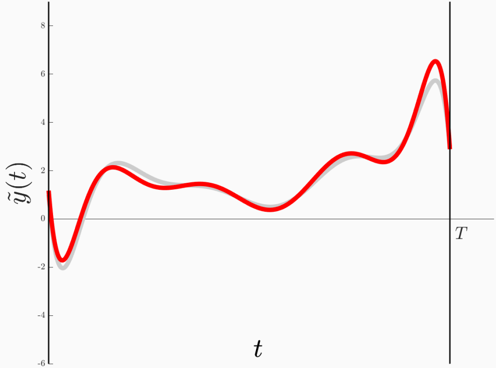
BASIC PROBLEM



Observe signal y at sample locations $t_1, \ldots, t_q \in [0, T]$.



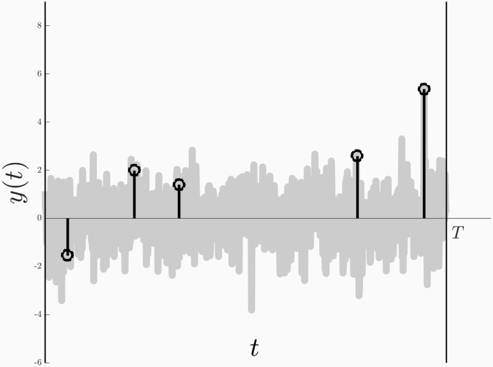
Observe signal y at sample locations $t_1, \ldots, t_q \in [0, T]$.
(possibly with noise)



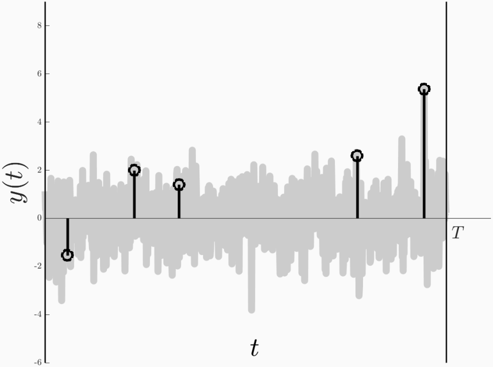
Goal: Recover signal $\tilde{y}$ which is close to y.

Central questions:

- How many samples do we need to approximately reconstruct y on $[0, T]$?
- How can we compute and represent $\tilde{y}$ in a computationally efficient way?



Naively, this problem is ill-posed.

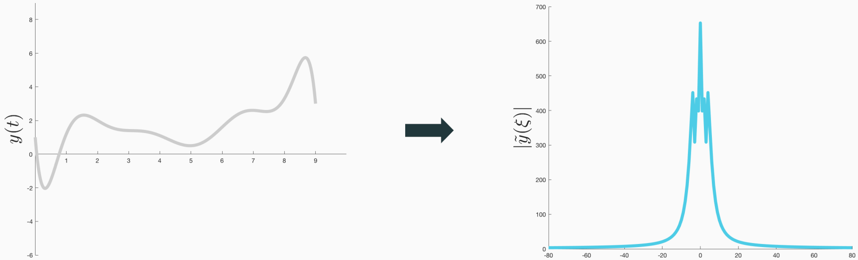


Naively, this problem is ill-posed.

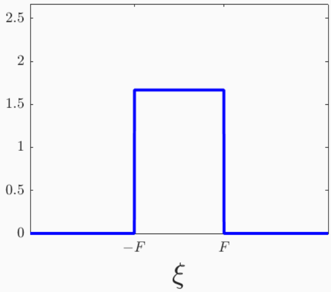
We need to assume y is smooth or structured in some way.

In science and engineering, we often impose structure by assuming y has a **“simple” Fourier transform.**

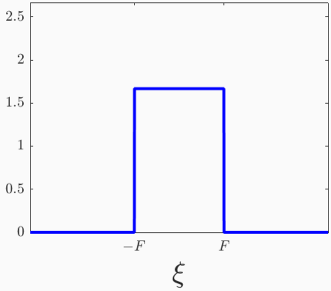
$$\hat{y}(\xi) = \int_{-\infty}^{\infty} y(t) e^{-2\pi i t \xi}\, dt.$$



Standard assumption: y is bandlimited, i.e. $\hat{y}(\xi) = 0$ for $|\xi| > F$.

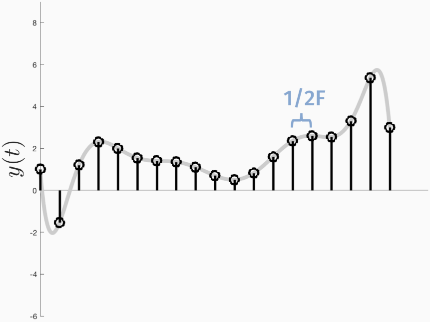


Standard assumption: y is **bandlimited**, i.e. $\hat{y}(\xi) = 0$ for $|\xi| > F$.



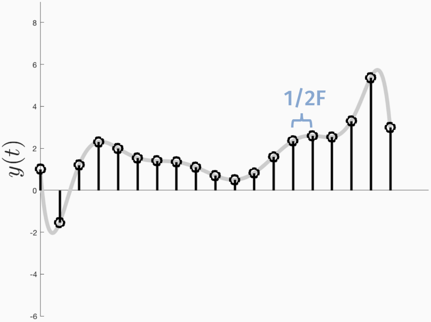
Shannon, Whitaker, Nyquist, Kotelnikov – foundations of modern signal processing and information theory.

Shannon-Nyquist Theorem: O(F) samples per second suffice.

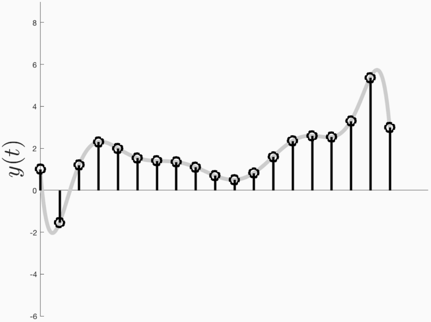


Uniform Nyquist sampling.

Shannon-Nyquist Theorem: O(F) samples per second suffice.

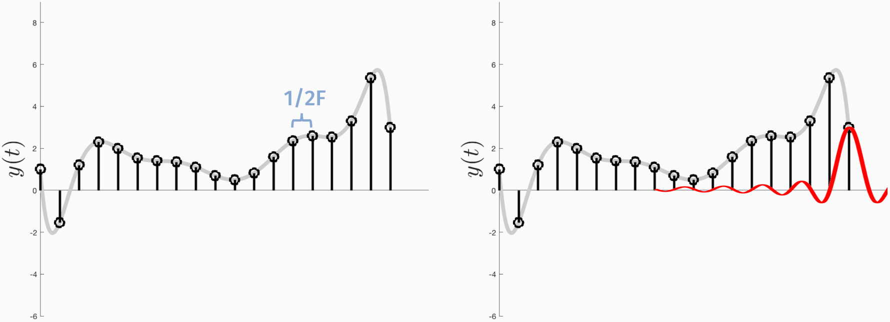


Uniform Nyquist sampling.



Sinc interpolation.

Shannon-Nyquist Theorem: O(F) samples per second suffice.

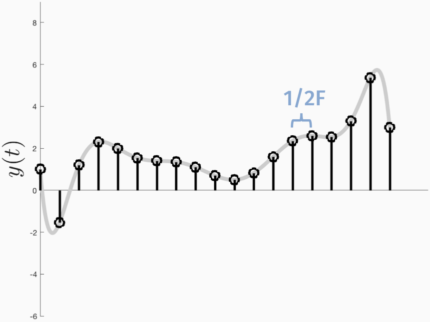


Uniform Nyquist sampling.

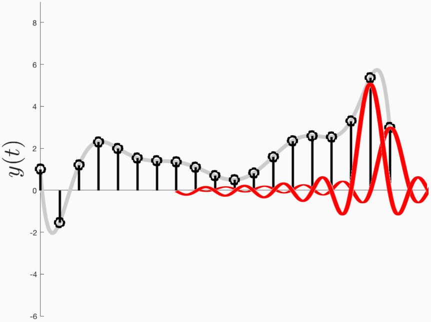
Sinc interpolation.

$$\tilde{y}(t) = \sum_{s=-\infty}^{\infty} \text{sinc}(y(t+s) \cdot F)$$

Shannon-Nyquist Theorem: O(F) samples per second suffice.



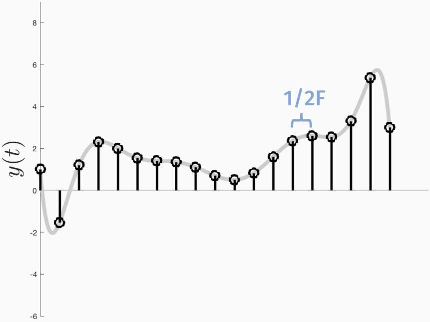
Uniform Nyquist sampling.



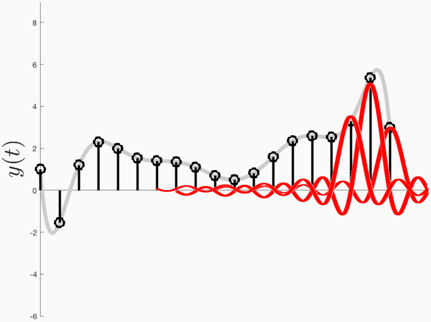
Sinc interpolation.

$$\tilde{y}(t) = \sum_{s=-\infty}^{\infty} \text{sinc}(y(t+s) \cdot F)$$

Shannon-Nyquist Theorem: O(F) samples per second suffice.



Uniform Nyquist sampling.



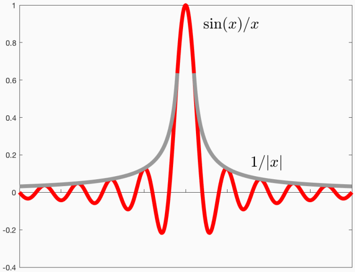
Sinc interpolation.

$$\tilde{y}(t) = \sum_{s=-\infty}^{\infty} \mathrm{sinc}(y(t+s) \cdot F)$$

Sinc interpolation doesn't actually work over $[0, T]$...

Sinc interpolation doesn't actually work over $[0, T]$...
(in theory or practice)

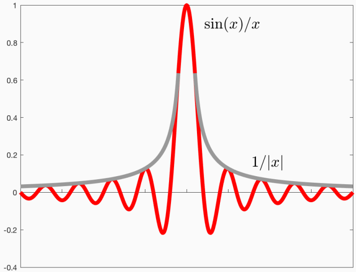
Sinc interpolation doesn't actually work over $[0, T]$...
(in theory or practice)



$$\tilde{y}(t) = \sum_{s=0}^{FT} \text{sinc}(y(t+s) \cdot F)$$

PROBLEMS WITH SINC INTERPOLATION

Sinc interpolation doesn't actually work over $[0, T]$...
(in theory or practice)



$$\tilde{y}(t) = \sum_{s=0}^{FT} \text{sinc}(y(t+s) \cdot F)$$

$O(FT/\epsilon)$ samples for ϵ error at best.

This problem was resolved in the 1960s (+ some):

Theorem (Slepian, Landau, Pollak and Rokhlin et al.)

Suppose $y = \mathcal{F}^* x$ *for bandlimited* x *and we observe* $y + n$*.*
With $O(FT + \log(1/\epsilon))$ *samples, it's possible to recover* $\tilde{y}$ *with:*

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_2^2 + c \cdot \|n\|_2^2$$

This problem was resolved in the 1960s (+ some):

Theorem (Slepian, Landau, Pollak and Rokhlin et al.)

Suppose $y = \mathcal{F}^* x$ for bandlimited x and we observe $y + n$. With $O(FT + \log(1/\epsilon))$ samples, it's possible to recover $\tilde{y}$ with:

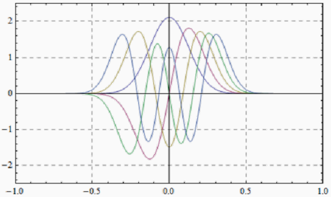
$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_2^2 + c \cdot \|n\|_2^2$$

This problem was resolved in the 1960s (+ some):

Theorem (Slepian, Landau, Pollak and Rokhlin et al.)

Suppose $y = \mathcal{F}^* x$ for bandlimited x and we observe $y + n$. With $O(FT + \log(1/\epsilon))$ samples, it's possible to recover $\tilde{y}$ with:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_2^2 + c \cdot \|n\|_2^2$$



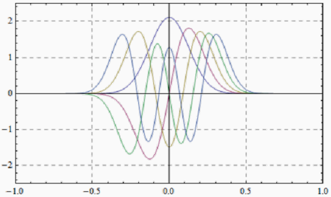
Prolate spheroid wave functions provide a smooth basis for bandlimited interpolation.

This problem was resolved in the 1960s (+ some):

Theorem (Slepian, Landau, Pollak and Rokhlin et al.)

Suppose $y = \mathcal{F}^* x$ *for bandlimited* x *and we observe* $y + n$*.*
With $O(FT + \log(1/\epsilon))$ *samples, it's possible to recover* $\tilde{y}$ *with:*

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_2^2 + c \cdot \|n\|_2^2$$

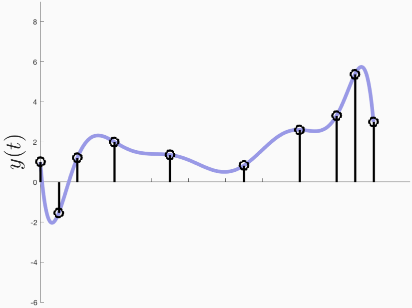


Prolate spheroid wave functions provide a smooth basis for bandlimited interpolation.

Can project to this basis with numerical quadrature.

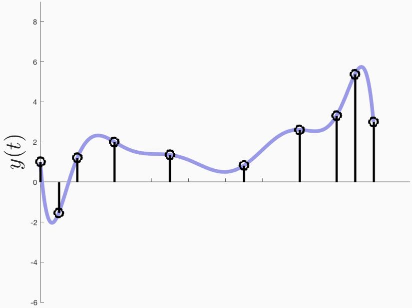
Important conclusion: For interpolation on a finite interval $[0, T]$, uniform sampling is suboptimal.

Important conclusion: For interpolation on a finite interval $[0, T]$, uniform sampling is suboptimal.



Non-uniform distribution required to get $\tilde{O}(FT)$ samples.

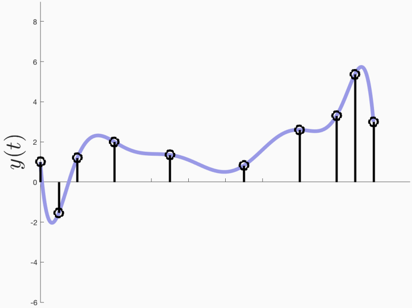
Important conclusion: For interpolation on a finite interval $[0, T]$, uniform sampling is suboptimal.



Non-uniform distribution required to get $\tilde{O}(FT)$ samples.

Uniform would give $\tilde{O}(F^2T^2)$ or $O(FT/\epsilon)$.

Important conclusion: For interpolation on a finite interval $[0, T]$, uniform sampling is suboptimal.



Non-uniform distribution required to get $\tilde{O}(FT)$ samples.

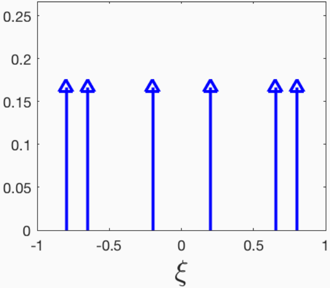
Uniform would give $\tilde{O}(F^2T^2)$ or $O(FT/\epsilon)$.

Not surprising if you think about polynomial interpolation.

WHAT ABOUT OTHER STRUCTURE?

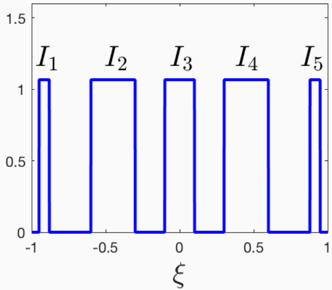
What about Fourier structure beyond a bandlimit?

E.g. y is **Fourier sparse**. $\hat{y}(\xi)$ is supported on k frequencies.

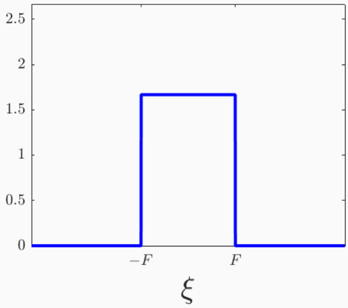


Compressed sensing, applications in medical imaging, microscopy, RADAR, etc.

E.g. y is multiband, i.e. $\hat{y}(\xi)$ supported on k intervals.

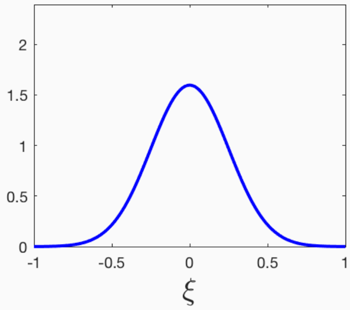


Bayesian perspective: instead of “allowing” or “disallowing” certain frequencies, we can consider any prior distribution on y’s power spectral density.



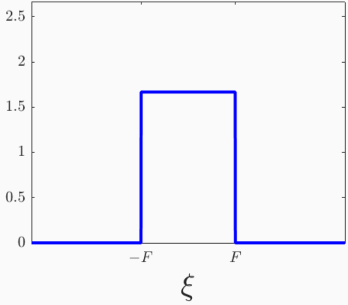
Bandlimited.

vs.



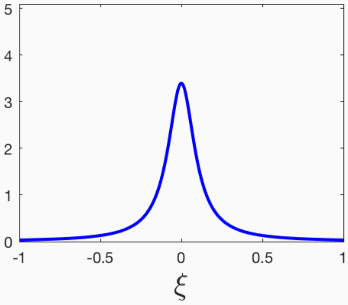
Gaussian prior.

Bayesian perspective: instead of "allowing" or "disallowing" certain frequencies, we can consider any prior distribution on y's power spectral density.



Bandlimited.

vs.



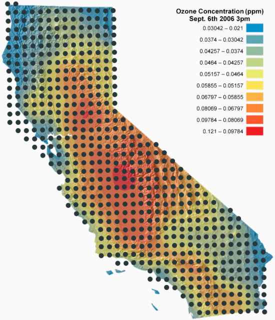
Cauchy-Lorentz Prior.

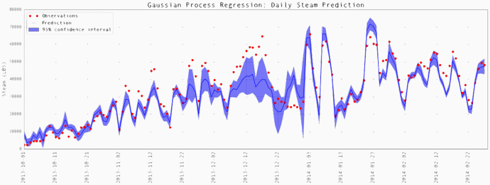
OTHER PRIORS

Smooth penalties underly Gaussian process regression, kriging, kernel ridge regression, etc.

OTHER PRIORS

Smooth penalties underly Gaussian process regression, kriging, kernel ridge regression, etc.





Countless applications in environmental science, geostatistics, image processing, economics, time series analysis, etc.

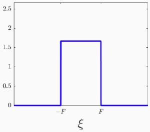
Knowledge gap: 50 years after PSWFs, no finite sample guarantees or efficient recovery algorithms for these popular problems on a finite interval $[0, T]$.

Knowledge gap: 50 years after PSWFs, no finite sample guarantees or efficient recovery algorithms for these popular problems on a finite interval $[0, T]$.

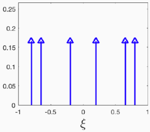
With the exception of Fourier sparse functions.
(Chen, Kane, Price, Song FOCS 2016, Chen, Kane 2018).

Our results:

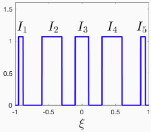
1. Characterize optimal sample complexity for any prior distribution μ.



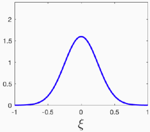
Bandlimited.



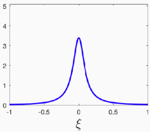
Sparse.



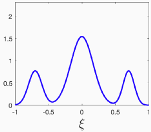
Multiband.



Gaussian.



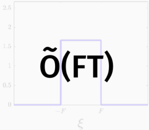
Cauchy-Lorentz.



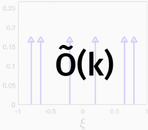
Gaussian mixture.

Our results:

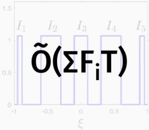
1. Characterize optimal sample complexity for any prior distribution μ.



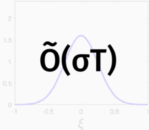
Bandlimited.



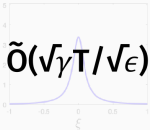
Sparse.



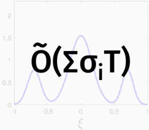
Multiband.



Gaussian.



Cauchy-Lorentz.



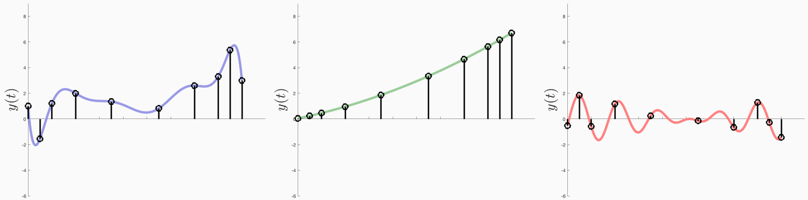
Gaussian mixture.

Our results:

1. Characterize optimal sample complexity for any prior distribution μ.
2. Universal non-uniform sampling scheme that matches this complexity up to logarithmic factors.

Our results:

1. Characterize optimal sample complexity for any prior distribution μ.
2. Universal non-uniform sampling scheme that matches this complexity up to logarithmic factors.



Typically a quadratic improvement on uniform sampling.

Our results:

1. Characterize optimal sample complexity for any prior distribution μ.
2. **Universal non-uniform sampling scheme** that matches this complexity up to logarithmic factors.
3. Efficient algorithm to pair with this sampling scheme that works for essentially all priors used in practice.

All using tools from discrete randomized algorithms!

On arXiv soon:

"Universal Sampling Strategies for Learning Signals with Simple Fourier Transforms"

Joint work with:

Haim Avron (TAU)
Michael Kapralov (EPFL)
Cameron Musco (MSR)
Ameya Velingker (EPFL)
Amir Zandieh (EPFL)

Definition (Weighted Inverse Fourier Transform)

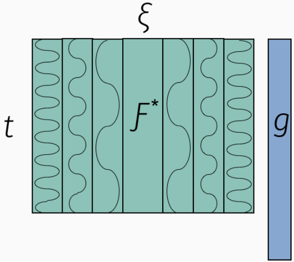
For any probability distribution μ over $\mathbb{R}$ and frequency domain function g, let:

$$\left[\mathcal{F}_\mu^* g\right](t) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i \xi t} \mu(\xi) d\xi.$$

Definition (Weighted Inverse Fourier Transform)

For any probability distribution μ over $\mathbb{R}$ and frequency domain function g, let:

$$\left[\mathcal{F}_\mu^* g\right](t) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i \xi t} \mu(\xi) d\xi.$$



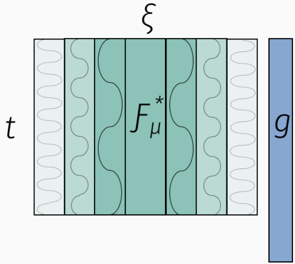
$$[\mathcal{F}^* g](t) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i \xi t} d\xi.$$

Standard inverse Fourier transform.

Definition (Weighted Inverse Fourier Transform)

For any probability distribution μ over $\mathbb{R}$ and frequency domain function g, let:

$$\left[\mathcal{F}_\mu^* g\right](t) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i \xi t} \mu(\xi) d\xi.$$



$$\left[\mathcal{F}_\mu^* g\right](t) = \int_{-\infty}^{\infty} g(\xi) e^{2\pi i \xi t} \mu(\xi) d\xi.$$

μ weighted inverse Fourier transform.

One possible formulation:

Suppose that y can be written as:

$$y = \mathcal{F}_\mu^* x.$$

We observe $y + n$ for some fixed noise function n.

One possible formulation:

Suppose that y can be written as:

$$y = \mathcal{F}_\mu^* x.$$

We observe $y + n$ for some fixed noise function n.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

One possible formulation:

Suppose that y can be written as:

$$y = \mathcal{F}_\mu^* x.$$

We observe $y + n$ for some fixed noise function n.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

- $\|z\|_T^2 = \frac{1}{T}\int_0^T |z(t)|^2 dt =$ average squared error.

One possible formulation:

Suppose that y can be written as:

$$y = \mathcal{F}_\mu^* x.$$

We observe $y + n$ for some fixed noise function n.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

- $\|z\|_T^2 = \frac{1}{T}\int_0^T |z(t)|^2 dt =$ average squared error.
- $\|x\|_\mu^2 = \int_{\mathbb{R}} |x(\xi)|^2 \mu(\xi) d\mu =$ signal energy under μ.

One possible formulation:

Suppose that y can be written as:

$$y = \mathcal{F}_\mu^* x.$$

We observe $y + n$ for some fixed noise function n.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

- $\|z\|_T^2 = \frac{1}{T}\int_0^T |z(t)|^2 dt =$ average squared error.
- $\|x\|_\mu^2 = \int_{\mathbb{R}} |x(\xi)|^2 \mu(\xi) d\mu =$ signal energy under μ.
- ϵ is a tunable parameter. Smaller ϵ requires more samples.
- Any fixed constant (e.g. 2) will do for c.

One possible formulation:

Suppose that y can be written as:

$$y = \mathcal{F}_\mu^* x.$$

We observe $y + n$ for some fixed noise function n.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

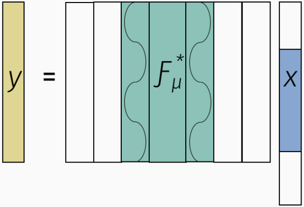
- $\|z\|_T^2 = \frac{1}{T}\int_0^T |z(t)|^2 dt =$ average squared error.
- $\|x\|_\mu^2 = \int_{\mathbb{R}} |x(\xi)|^2 \mu(\xi) d\mu =$ signal energy under μ.
- ϵ is a tunable parameter. Smaller ϵ requires more samples.
- Any fixed constant (e.g. 2) will do for c.

Given: $y = \mathcal{F}_\mu^* x + n$.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

Easiest to understand for bandlimited, sparse, or multiband.

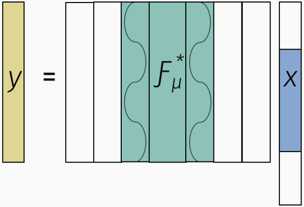


Given: $y = \mathcal{F}_\mu^* x + n$.

Goal: Find an efficient representation of a function $\tilde{y}$ such that:

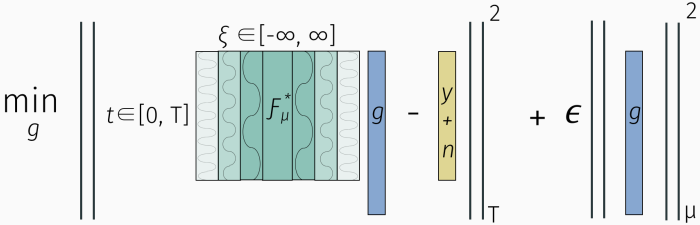
$$\|y - \tilde{y}\|_T^2 \leq \epsilon \|x\|_\mu^2 + c \cdot \|n\|_T^2.$$

Easiest to understand for bandlimited, sparse, or multiband.

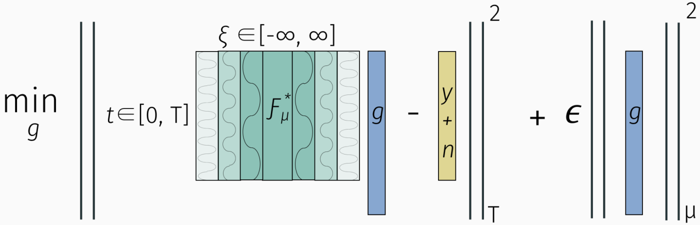


There's a natural Bayesian formulation for non-uniform priors.

Suffices to return $\tilde{y} = F_\mu^* \tilde{g}$ for any constant factor approximation $\tilde{g}$ to the regression problem:



Suffices to return $\tilde{y} = F_\mu^* \tilde{g}$ for any constant factor approximation $\tilde{g}$ to the regression problem:

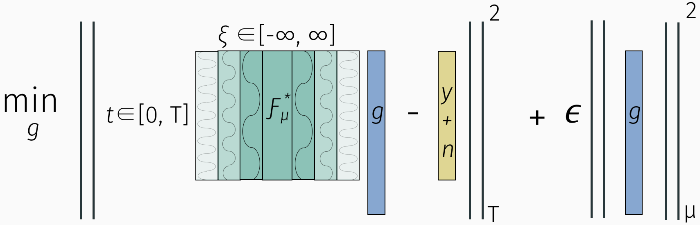


If $\tilde{g}$ satisfies:

$$\|y + n - \mathcal{F}_\mu^* \tilde{g}\|_T^2 + \epsilon \|\tilde{g}\|_\mu^2 \leq C \cdot \left[\min_g \|y + n - \mathcal{F}_\mu^* g\|_T^2 + \epsilon \|g\|_\mu^2\right],$$

$$\text{then: } \|y - \mathcal{F}_\mu^* \tilde{g}\|_T^2 \leq O(C) \cdot \left[\|n\|_T^2 + \epsilon \|x\|_\mu^2\right].$$

Suffices to return $\tilde{y} = F_\mu^* \tilde{g}$ for any constant factor approximation $\tilde{g}$ to the regression problem:



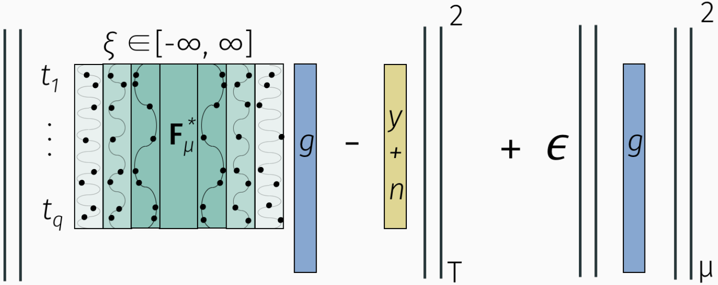
If $\tilde{g}$ satisfies:

$$\|y + n - \mathcal{F}_\mu^* \tilde{g}\|_T^2 + \epsilon \|\tilde{g}\|_\mu^2 \leq C \cdot \left[\min_g \|y + n - \mathcal{F}_\mu^* g\|_T^2 + \epsilon \|g\|_\mu^2\right],$$

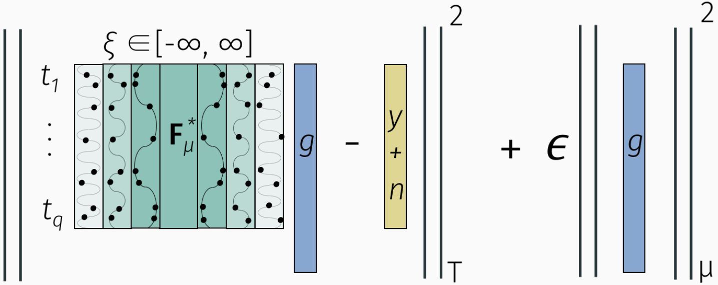
$$\text{then: } \|y - \mathcal{F}_\mu^* \tilde{g}\|_T^2 \leq O(C) \cdot \left[\|n\|_T^2 + \epsilon \|x\|_\mu^2\right].$$

Solution by discretization.

Selecting time samples discretizes time domain.



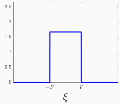
Selecting time samples discretizes time domain.



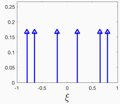
What about Fourier domain?

We can avoid discretization entirely as long as we have a closed form representation of $\hat{\mu}(t)$.

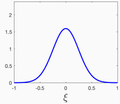
We can avoid discretization entirely as long as we have a closed form representation of $\hat{\mu}(t)$.



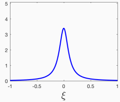
$$\hat{\mu}(t) = \text{sinc}(t)$$



$$\hat{\mu}(t) = \sum_{j=1}^{k} e^{-2\pi i(t)}$$

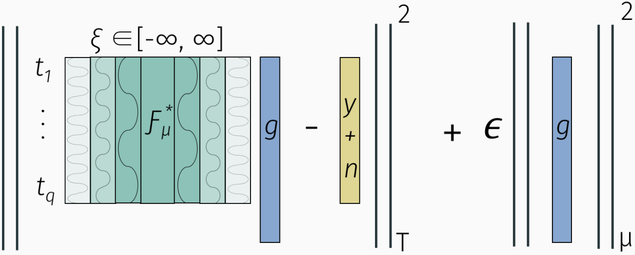


$$\hat{\mu}(t) = e^{-|t|^2}$$

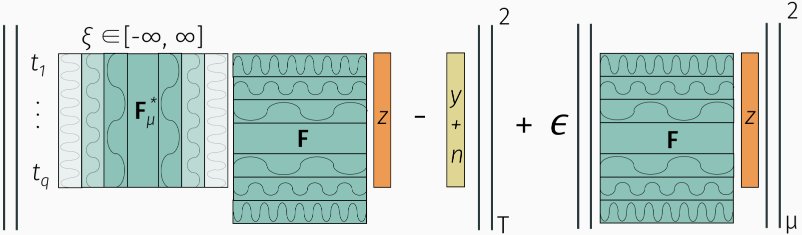


$$\hat{\mu}(t) = e^{-|t|}$$

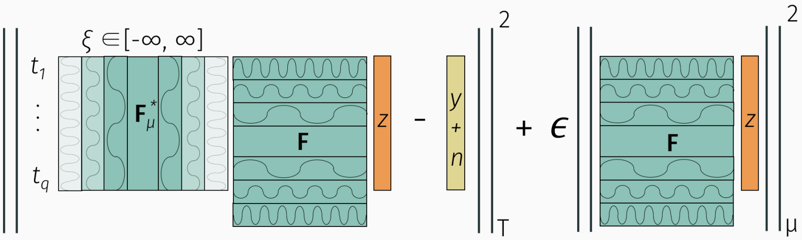
$\hat{\mu}$ is referred to as the <u>autocorrelation function</u>, the <u>semivariogram</u>, the <u>kernel function</u>, etc.



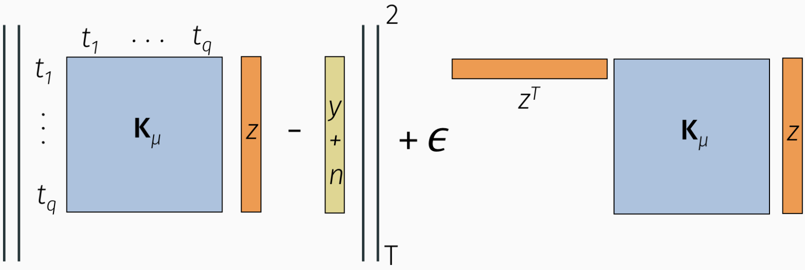
equivalent to

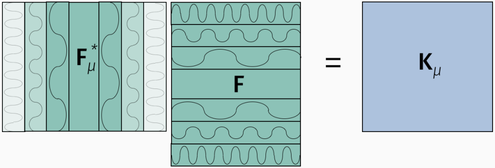


HANDLING FOURIER DOMAIN



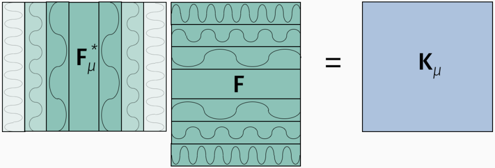
equivalent to



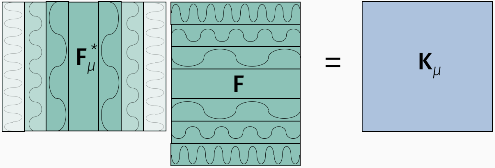


$$\mathrm{K}_\mu(i,j) = \int_{-\infty}^{\infty} e^{2\pi i t_i \xi} e^{-2\pi i t_j \xi} \mu(\xi) d\xi$$

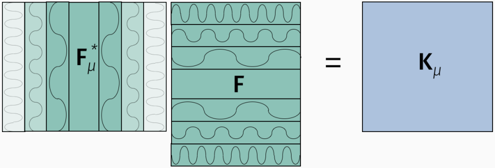
EQUIVALENCE TO KERNEL REGRESSION



$$\mathrm{K}_\mu(i,j) = \int_{-\infty}^{\infty} e^{2\pi i t_i \xi} e^{-2\pi i t_j \xi} \mu(\xi) d\xi$$
$$= \int_{-\infty}^{\infty} \mu(\xi) e^{-2\pi i (t_j - t_i)\xi} d\xi$$



$$\begin{aligned}
\mathrm{K}_\mu(i,j) &= \int_{-\infty}^{\infty} e^{2\pi i t_i \xi} e^{-2\pi i t_j \xi} \mu(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \mu(\xi) e^{-2\pi i (t_j - t_i)\xi} d\xi \\
&= \hat{\mu}(t_j - t_i).
\end{aligned}$$

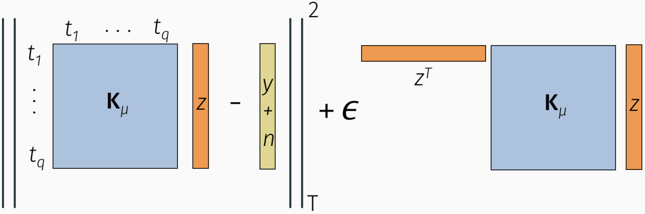


$$
\begin{aligned}
\mathbf{K}_\mu(i,j) &= \int_{-\infty}^{\infty} e^{2\pi i t_i \xi} e^{-2\pi i t_j \xi} \mu(\xi) d\xi \\
&= \int_{-\infty}^{\infty} \mu(\xi) e^{-2\pi i (t_j - t_i)\xi} d\xi \\
&= \hat{\mu}(t_j - t_i).
\end{aligned}
$$

We can construct $\mathbf{K}$ in $O(q^2)$ time.

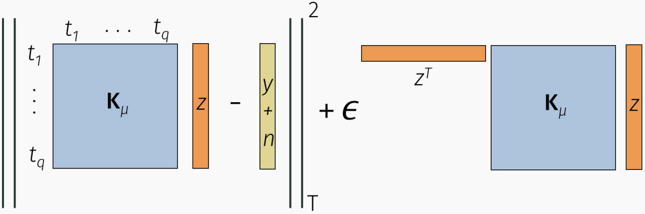
EQUIVALENCE TO KERNEL REGRESSION

- Sample $t_1, \ldots, t_q$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.
- Solve $\mathbf{z} = (K + \epsilon \mathbf{I})^{-1} [y_n(t_1), \ldots, y_n(t_q)]$.
- Evaluate $\tilde{y}(t) = \sum_{i=1}^{1} \mathbf{z}_i \hat{\mu}(t_i - t) \cdot f(t_i)$.

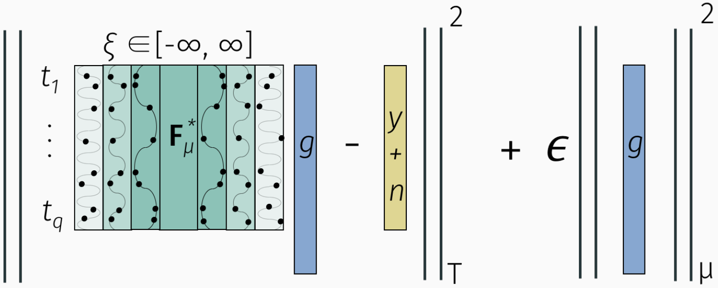


EQUIVALENCE TO KERNEL REGRESSION

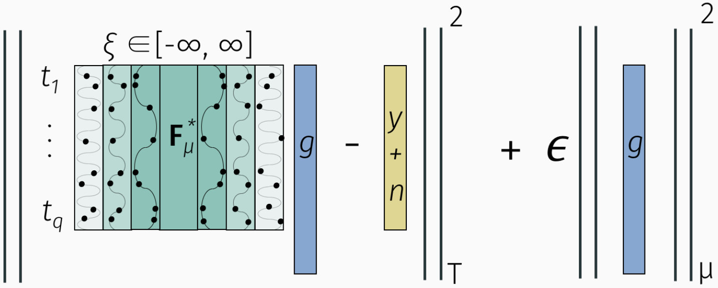
- Sample $t_1, \ldots, t_q$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.
- Solve $\mathbf{z} = (K + \epsilon \mathbf{I})^{-1} [y_n(t_1), \ldots, y_n(t_q)]$.
- Evaluate $\tilde{y}(t) = \sum_{i=1}^{1} \mathbf{z}_i \hat{\mu}(t_i - t) \cdot f(t_i)$.



Key Challenge: How to select samples in time domain.



Key Challenge: How to select samples in time domain.



Approach: Lean on well developed theory for randomly sampling discrete regression problems.

For an approximate solution, suffices to sample rows (i.e. time points) by their statistical leverage score:

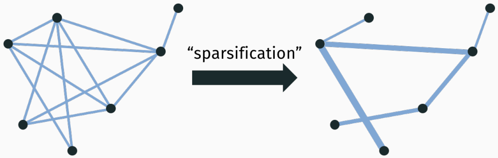
$$\tau_{\mu,\epsilon}(t) = \max_{g} \frac{\frac{1}{T}\left|\mathcal{F}_{\mu}^{*}g(t)\right|^{2}}{\|\mathcal{F}_{\mu}^{*}g\|_{T}^{2} + \epsilon\|g\|_{\mu}^{2}}$$

For an approximate solution, suffices to sample rows (i.e. time points) by their statistical leverage score:

$$\tau_{\mu,\epsilon}(t) = \max_{g} \frac{\frac{1}{T}\left|\mathcal{F}_\mu^* g(t)\right|^2}{\|\mathcal{F}_\mu^* g\|_T^2 + \epsilon\|g\|_\mu^2}$$

$$0 \leq \tau_{\mu,\epsilon}(t) \leq 1$$

$\tau_{\mu,\epsilon}$ is a regularized version of effective resistance, a central quantity in recent work on randomized algorithms for graph problems and matrix sketching.



[Drineas, Mahoney, Muthukrishnan 2006]

[Spielman, Srivastava 2008]

How many samples are required?

How many samples are required?

$$\int_0^T \tau_{\mu,\epsilon}(t)dt = S_{\mu,\epsilon} = \text{“statistical dimension”}.$$

We need to take $S_{\mu,\epsilon}$ total samples to approximate the regression problem.

How many samples are required?

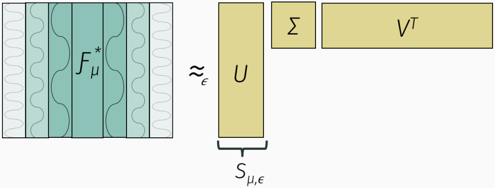
$$\int_0^T \tau_{\mu,\epsilon}(t)dt = S_{\mu,\epsilon} = \text{“statistical dimension”}.$$

We need to take $S_{\mu,\epsilon}$ total samples to approximate the regression problem.

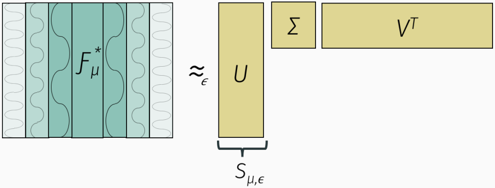
For finite dimension problems, $S_{\mu,\epsilon}$ is bounded by d.

$$S_{\mu,\epsilon} = \operatorname{tr}\left(\mathcal{K}_\mu + \epsilon\mathcal{I}\right)^{-1}\mathcal{K}_\mu = \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + \epsilon}$$

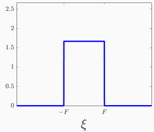
$$S_{\mu,\epsilon} = \operatorname{tr}\left(\mathcal{K}_\mu + \epsilon\mathcal{I}\right)^{-1}\mathcal{K}_\mu$$
$$= \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + \epsilon} \approx \text{number of eigenvalues larger than } \epsilon.$$



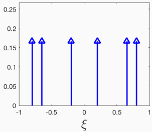
$$
\begin{aligned}
S_{\mu,\epsilon} &= \operatorname{tr}\left(\mathcal{K}_\mu + \epsilon\mathcal{I}\right)^{-1}\mathcal{K}_\mu \\
&= \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + \epsilon} \approx \text{number of eigenvalues larger than } \epsilon.
\end{aligned}
$$



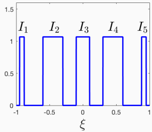
$$S_{\mu,\epsilon} = \operatorname{tr}\left(\mathcal{K}_\mu + \epsilon\mathcal{I}\right)^{-1}\mathcal{K}_\mu$$
$$= \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + \epsilon} \approx \text{number of eigenvalues larger than } \epsilon.$$



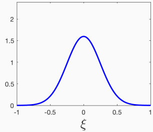
$O\left(FT + \log\frac{1}{\epsilon}\right)$



k

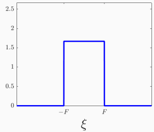


$O\left(\sum F_i T + \log\frac{1}{\epsilon}\right)$

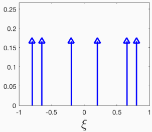


$\tilde{O}\left(\sigma T + \log\frac{1}{\epsilon}\right)$

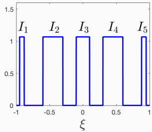
$$S_{\mu,\epsilon} = \operatorname{tr}\left(\mathcal{K}_\mu + \epsilon\mathcal{I}\right)^{-1}\mathcal{K}_\mu$$
$$= \sum_{i=1}^{\infty} \frac{\lambda_i}{\lambda_i + \epsilon} \approx \text{number of eigenvalues larger than } \epsilon.$$



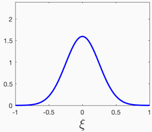
$O\left(FT + \log\frac{1}{\epsilon}\right)$



k



$O\left(\sum F_i T + \log\frac{1}{\epsilon}\right)$

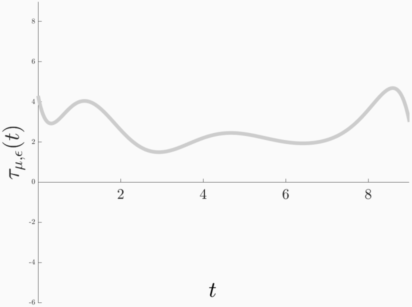


$\tilde{O}\left(\sigma T + \log\frac{1}{\epsilon}\right)$

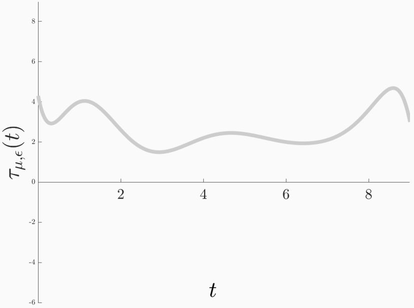
Bound of $S_{\mu,\epsilon}$ samples is tight.

Leverage scores are hard to compute, even for discrete regression problems.

Leverage scores are hard to compute, even for discrete regression problems. For our problem the challenge is even more daunting... we need scores for a continuum of values.



Leverage scores are hard to compute, even for discrete regression problems. For our problem the challenge is even more daunting... we need scores for a continuum of values.



But... we have structure on our side.

What is the leverage score?

$$\tau_{\mu,\epsilon}(t) = \frac{1}{T} \max_{g} \frac{\left|\mathcal{F}_{\mu}^{*} g(t)\right|^{2}}{\|\mathcal{F}_{\mu}^{*} g\|_{T}^{2} + \epsilon \|g\|_{\mu}^{2}}$$

What is the leverage score?

$$\tau_{\mu,\epsilon}(t) = \frac{1}{T} \max_{g} \frac{\left|\mathcal{F}_\mu^* g(t)\right|^2}{\|\mathcal{F}_\mu^* g\|_T^2 + \epsilon\|g\|_\mu^2}$$

Squared value of a function at t over the average squared value. I.e. how much can the function "spike" at time t.

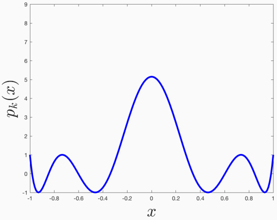
What is the leverage score?

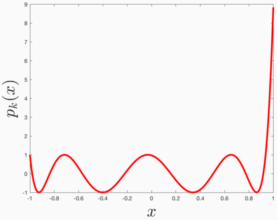
$$\tau_{\mu,\epsilon}(t) = \frac{1}{T} \max_{g} \frac{\left|\mathcal{F}_{\mu}^{*} g(t)\right|^2}{\|\mathcal{F}_{\mu}^{*} g\|_T^2 + \epsilon \|g\|_{\mu}^2}$$

Squared value of a function at t over the average squared value. I.e. how much can the function "spike" at time t.

Worst case, but over a restricted class of functions – need to have small norm under μ.

Leverage for degree k polynomials:



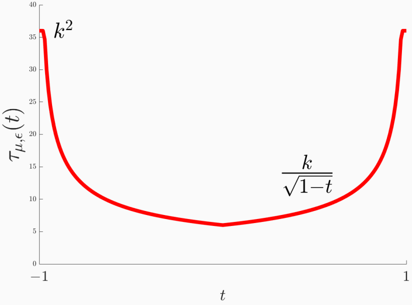


Bernstein Inequality.
$\tau(t) \leq k/\sqrt{\min(t, T-t)}$

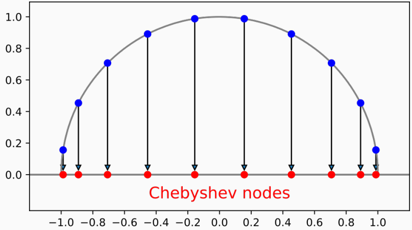
Markov Brother's Inequality.
$\tau(t) \leq k^2$

In general, a polynomial can "spike" more near the edge of an interval.

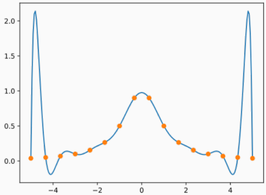
Leverage for degree k polynomials:



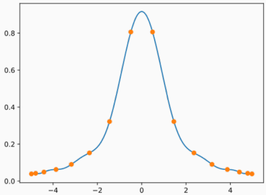
Leverage for degree *k* polynomials:



Leverage for degree k polynomials:

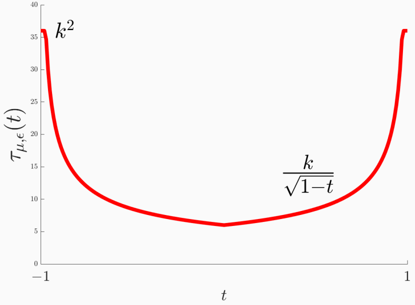


Uniform samples.



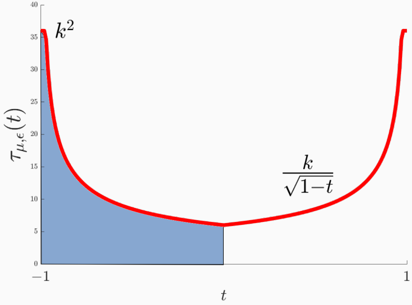
Chebyshev samples.

Leverage for degree k polynomials:



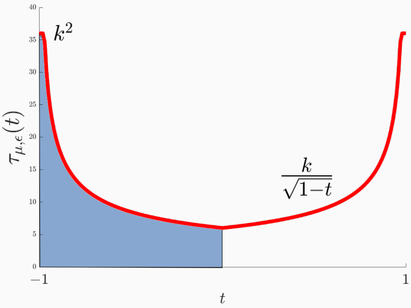
Total leverage:

Leverage for degree k polynomials:



Total leverage: $O(k)$

Leverage for degree k polynomials:



Total leverage: $O(k)$

Extends to bandlimited functions, which can be approximated by degree $k = O(FT + \log(1/\epsilon))$ degree polynomials.

[Chen, Kane, Price, Song, FOCS 2016], [Chen, Price 2018]

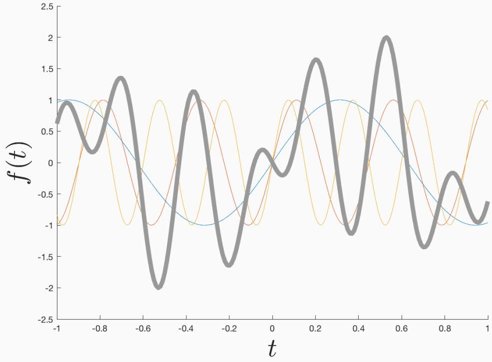
Nearly the same bounds holds for k-sparse Fourier functions.

[Chen, Kane, Price, Song, FOCS 2016], [Chen, Price 2018]

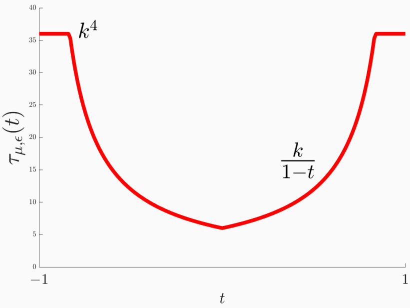
Nearly the same bounds holds for k-sparse Fourier functions.

$$\frac{|f_k(t)|^2}{\|f_k\|_T^2} = \tilde{O}\left(\min\left[k^4, k/\min(t, T-t)\right]\right)$$

Intuition: Sums of close frequencies look like modulated polynomials. Far frequencies are nearly orthogonal.

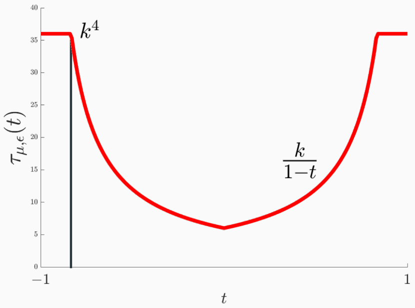


Leverage for *k* sparse Fourier functions:



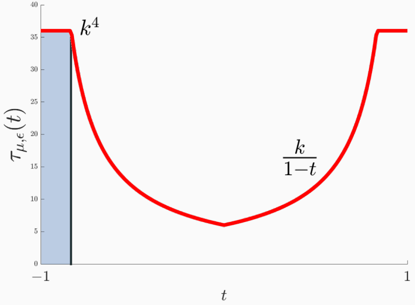
Total leverage:

Leverage for *k* sparse Fourier functions:



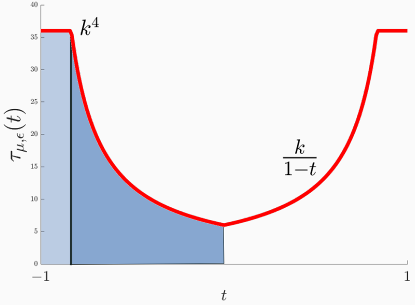
Total leverage:

Leverage for k sparse Fourier functions:



Total leverage: k

Leverage for k sparse Fourier functions:



Total leverage: $k + O(k \log k)$

How do we extend these bounds to more general constraint distributions μ? Want $\tilde{O}(S_{\mu,\epsilon})$ samples.

How do we extend these bounds to more general constraint distributions μ? Want $\tilde{O}(S_{\mu,\epsilon})$ samples.

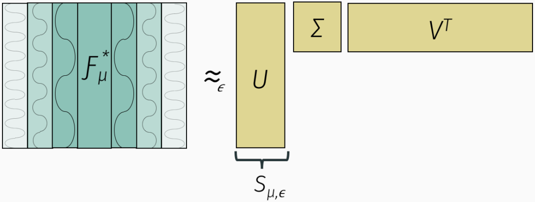
More tools from randomized matrix algorithms!

Lesson from past decade: Top q singular vectors of a matrix are approximately spanned by $O(q)$ columns from that matrix.

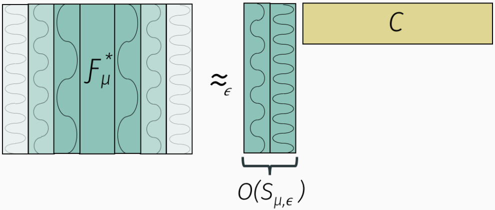
Lesson from past decade: Top q singular vectors of a matrix are approximately spanned by $O(q)$ columns from that matrix.

(rank-revealing QR, randomized SVD, columns subset selection, CUR decomposition, Nyström approximation, graph sparsification, random Fourier features, etc.)

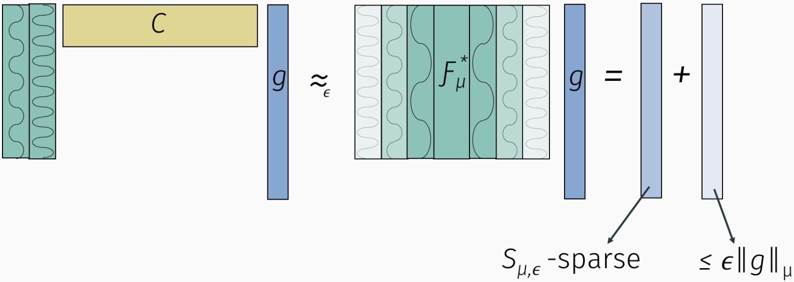
SPARSE APPROXIMATION OF WEIGHTED FOURIER TRANSFORM



SPARSE APPROXIMATION OF WEIGHTED FOURIER TRANSFORM



SPARSE APPROXIMATION OF WEIGHTED FOURIER TRANSFORM

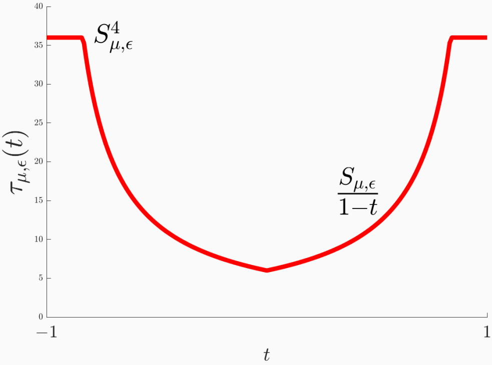


FINAL BOUND

$$\tau_{\mu,\epsilon}(t) = \max_g \frac{\frac{1}{T}\left|\mathcal{F}_\mu^* g(t)\right|^2}{\|\mathcal{F}_\mu^* g\|_T^2 + \epsilon\|g\|_\mu^2} \leq \tilde{O}\left(\min\left[S_{\mu,\epsilon}^4, S_{\mu,\epsilon}/\min(t, T-t)\right]\right)$$

FINAL BOUND

$$\tau_{\mu,\epsilon}(t) = \max_g \frac{\frac{1}{T}\left|\mathcal{F}_\mu^* g(t)\right|^2}{\|\mathcal{F}_\mu^* g\|_T^2 + \epsilon\|g\|_\mu^2} \leq \tilde{O}\left(\min\left[S_{\mu,\epsilon}^4, S_{\mu,\epsilon}/\min(t, T-t)\right]\right)$$



Total number of samples: $\tilde{O}(S_{\mu,\epsilon})$.

Simple Fourier function fitting:

- Sample $t_1, \ldots, t_q$ according to $\mathcal{D}$.

Simple Fourier function fitting:

- Sample $t_1, \ldots, t_q$ according to $\mathcal{D}$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.

Simple Fourier function fitting:

- Sample $t_1, \ldots, t_q$ according to $\mathcal{D}$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.
- Solve $\mathbf{z} = (K + \epsilon \mathbf{I})^{-1} [y_n(t_1), \ldots, y_n(t_q)]$.

Simple Fourier function fitting:

- Sample $t_1, \ldots, t_q$ according to $\mathcal{D}$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.
- Solve $\mathbf{z} = (K + \epsilon \mathbf{I})^{-1} [y_n(t_1), \ldots, y_n(t_q)]$.
- Evaluate $\tilde{y}(t) = \sum_{i=1}^{1} \mathbf{z}_i \hat{\mu}(t_i - t) \cdot f(t_i)$.

Simple Fourier function fitting:

- Sample $t_1, \ldots, t_q$ according to $\mathcal{D}$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.
- Solve $\mathbf{z} = (K + \epsilon \mathbf{I})^{-1} [y_n(t_1), \ldots, y_n(t_q)]$.
- Evaluate $\tilde{y}(t) = \sum_{i=1}^{1} \mathbf{z}_i \hat{\mu}(t_i - t) \cdot f(t_i)$.

That's it!

Simple Fourier function fitting:

- Sample $t_1, \ldots, t_q$ according to $\mathcal{D}$.
- Compute $\hat{\mu}(t_i - t_j)$ to build $q \times q$ kernel matrix $\mathbf{K}$.
- Solve $\mathbf{z} = (K + \epsilon \mathbf{I})^{-1} [y_n(t_1), \ldots, y_n(t_q)]$.
- Evaluate $\tilde{y}(t) = \sum_{i=1}^{1} \mathbf{z}_i \hat{\mu}(t_i - t) \cdot f(t_i)$.

That's it!

Matches known results for sparse and bandlimited function up to log factors, while achieving nearly optimal sample complexity for any other Fourier constraints.

Lots of open questions:

Lots of open questions:

- Streaming reconstruction? Anything more local?

Lots of open questions:

- Streaming reconstruction? Anything more local?
- Does uniform sampling work if we look outside $[0, T]$?

Lots of open questions:

- Streaming reconstruction? Anything more local?
- Does uniform sampling work if we look outside $[0, T]$?
- Deterministic sampling patterns in noiseless case, or under reasonable noise model?

Lots of open questions:

- Streaming reconstruction? Anything more local?
- Does uniform sampling work if we look outside $[0, T]$?
- Deterministic sampling patterns in noiseless case, or under reasonable noise model?

Other connections between random graph/matrix sampling and classic function interpolation?

THANK YOU!