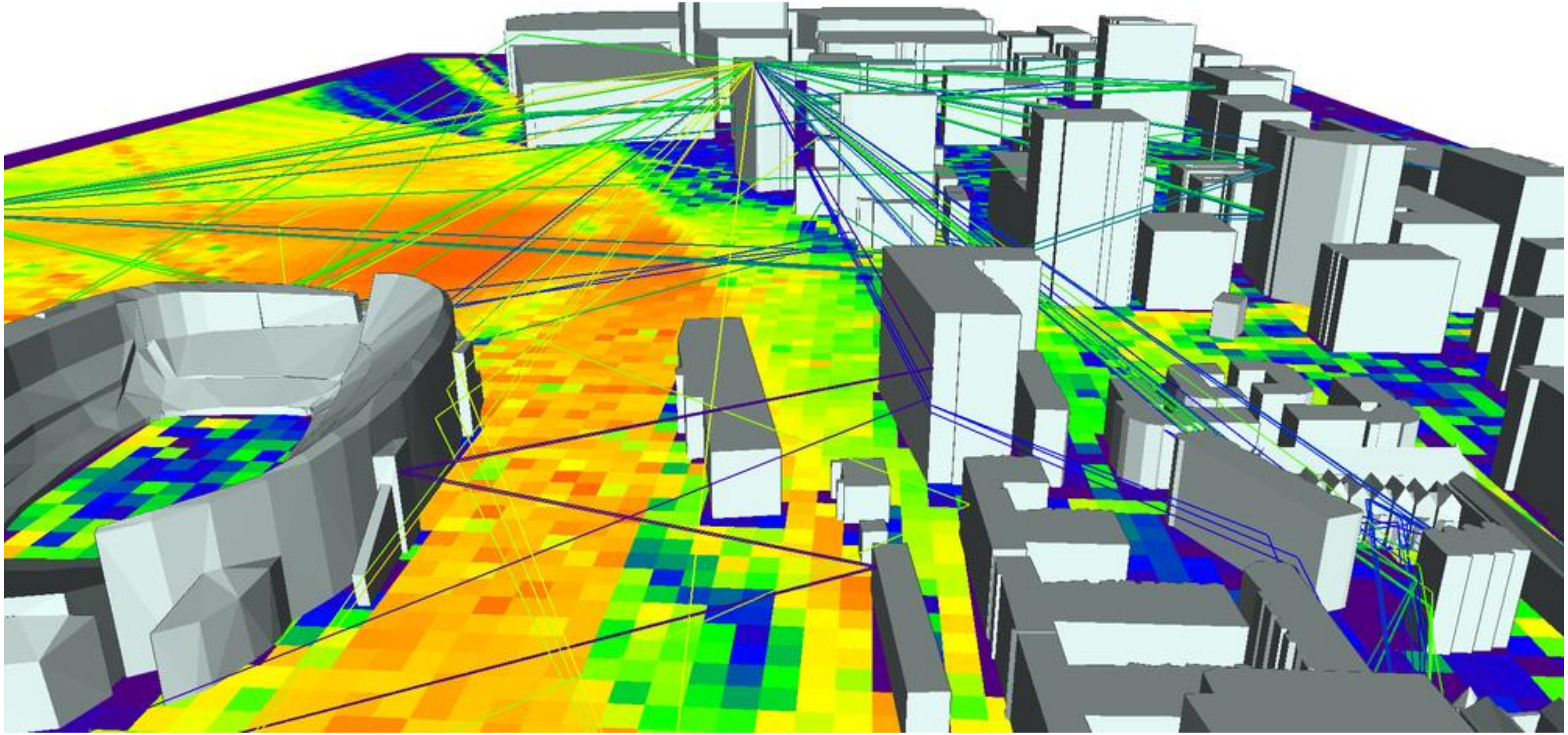
Dual-Kernel Online Reconstruction of Power Maps

Rasoul Nikbakht, Anders Jonsson, Angel Lozano

Universitat Pompeu Fabra Barcelona

Power map

An indication of the average received power from one or several base station(BSs)



Motivation

- Site-specific modeling rather than a statistical modeling (Hata, COST-231)
- More accurate pathloss values
- Network coverage analysis
- Proactive resource allocation

Problem statement

Scenario:

- Users are randomly distributed
- Each BS has a coverage area
- The coverage area is obtained by setting a threshold on the pathloss value
- One user can be in the coverage area of several BSs

Measurements:

- Average received signal power at each user location from assigned BSs

Objective:

- Build a power map for each BS: for some input coordinates, produce power values

Problem formulation

Naturally falls into the regression task

- Averaged received power is given for some locations (training set)
- Learn a function that connects the locations to the power levels
- Evaluate the performance of the learned function in the test set

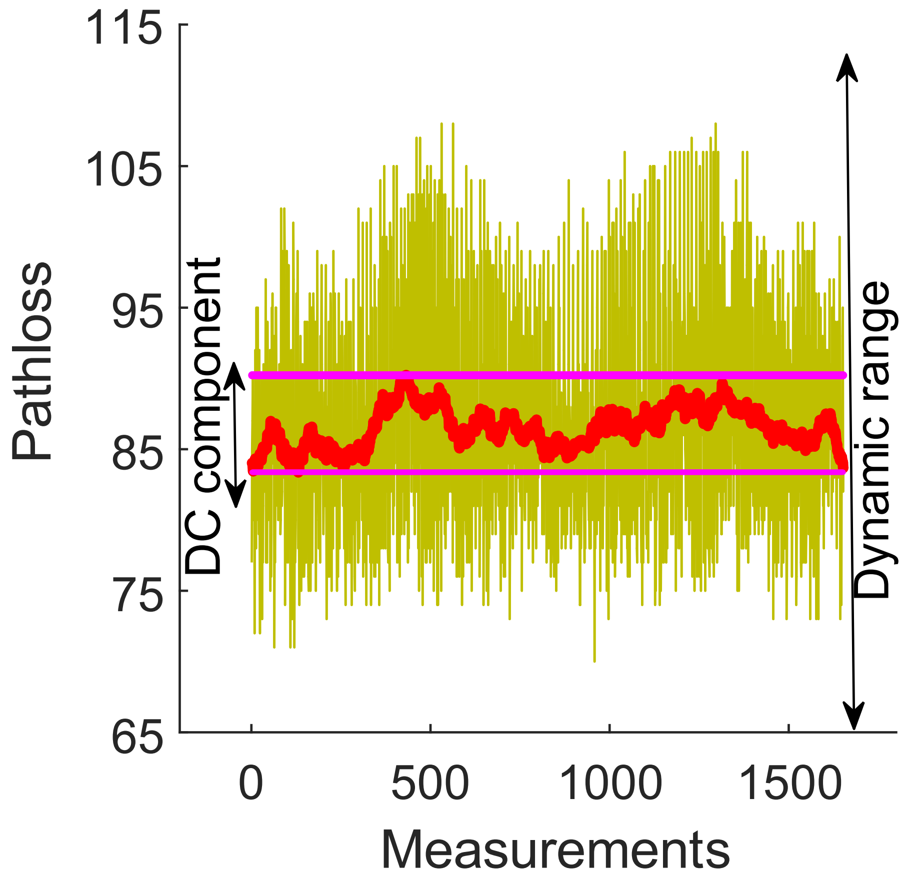
Regression task can be categorized

- Online regression
  Process measurement one by one (simpler and faster)
- Batch regression
  Process measurements altogether (slower but more accurate)

Contributions

- Pathloss or average received power is well bounded, let's say between 50 and 160 dB
- Majority of users have pathloss in order of 60-120 dB

  The center of this interval is called the DC component
- Extract the DC component from the pathloss values and find the regression function for the rest
- Advantages:
  - Faster convergence
  - Better accuracy in terms of std and MSE



How can we choose the regression function?

- Define a similarity measure or kernel function between measurements
- Build the regression function based on the combination of kernel basis

Example: kernel as the correlation between average received power in the location of users

$$\kappa(x_1, x_2) = e^{-\frac{|x_1 - x_2|}{2\sigma^2}}$$

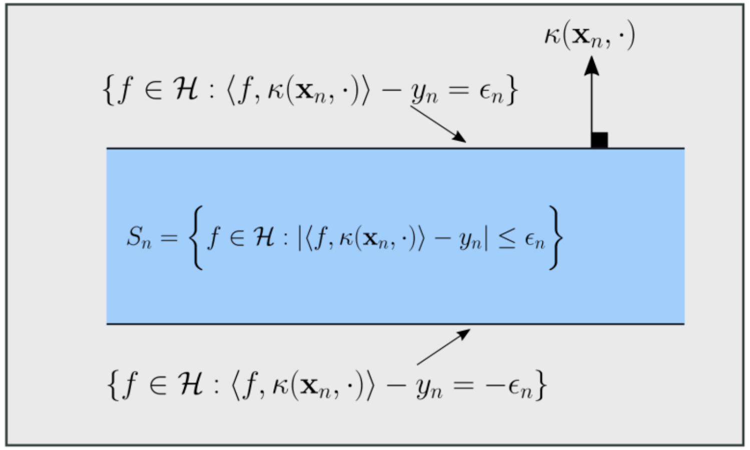
Then the optimum regressor function is

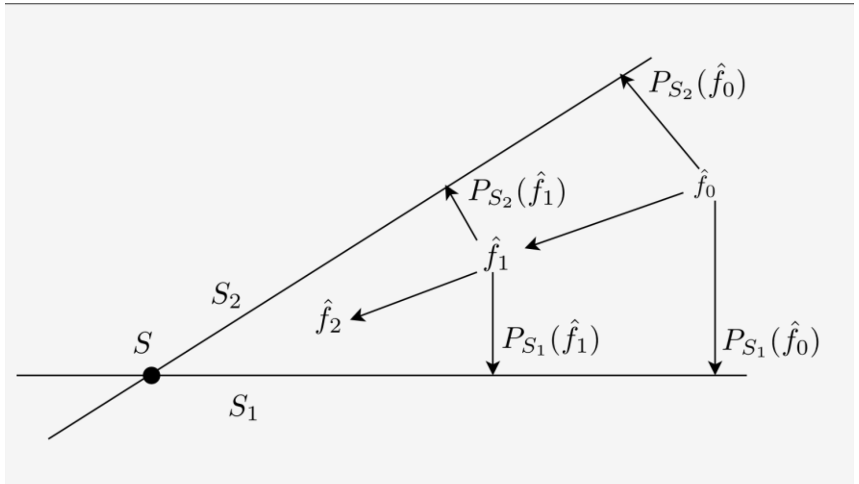
$$f^\star(x) = \sum_n \alpha_n e^{-\frac{|x - x_n|}{2\sigma^2}}$$

Each term in summation models the effect of one measurement and called a dictionary element

Adaptive projected sub-gradient method (APSM)
An online regression tool

- Create a set of functions that can explain a specific measurement
- Find the intersection of these sets, or functions that can explain all or most of the measurements data
  - Use projection to a convex set to find the intersection





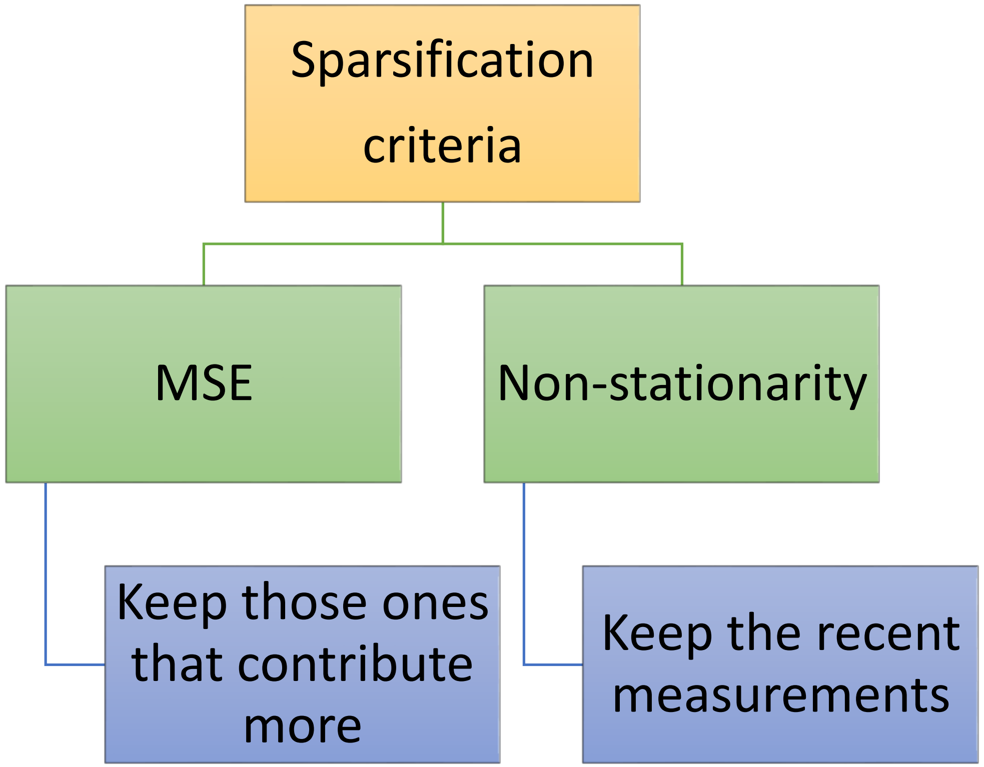
Regularization

Regression is a convex optimization problem

- In a convex optimization problem, regularization is
  Including the function norm in the objective
- In APSM algorithm regularization is
  Projection over norm ball

Projection over norm ball can also be seen as the forgetting factor

Sparsification
The act of selecting some measurements and throwing away the rest



Sparsification threshold

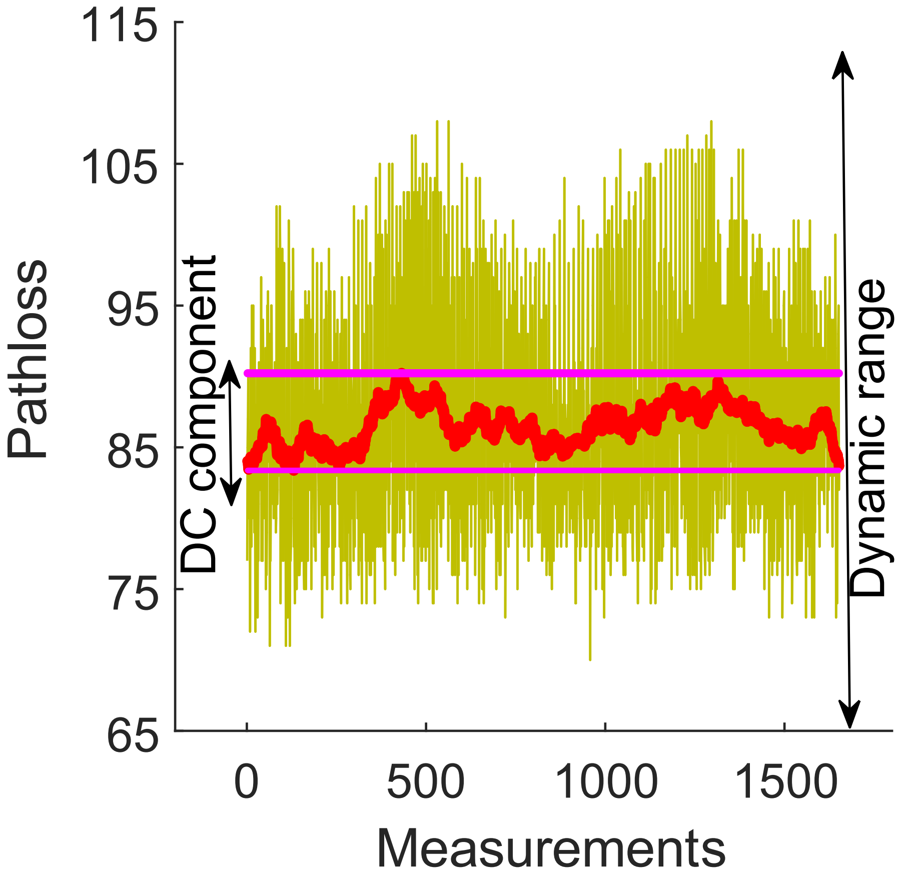
- Projection on the norm ball takes care of non-stationarity
- A threshold is defined:
  - To keep the most effective dictionary elements (in MSE sense)
  - The larger the dictionary size, the better the MSE error
  - The bigger the threshold, the more sparse the dictionary

$$P_{B[0,\delta]}(f) = \begin{cases} f & \|f\| \le \delta \\ f/\|f\| & \|f\| > \delta. \end{cases}$$

$$\mathcal{J}_n = \left\{ j \mid \frac{(\alpha_{n,j})^2}{\sum_{\ell \in \mathcal{L}_n} (\alpha_{n,\ell})^2} > \tau \right\}$$

$$f^\star(x) = \sum_n \alpha_n \, e^{-\frac{|x - x_n|}{2\sigma^2}}$$

Dual-kernel approach:
One kernel for DC part and another for varying part



$$\kappa(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|}{D_{\text{DC}}}\right) + \exp\left(-\frac{\|x_1 - x_2\|}{D_{\text{Var}}}\right)$$

Numerical evaluation

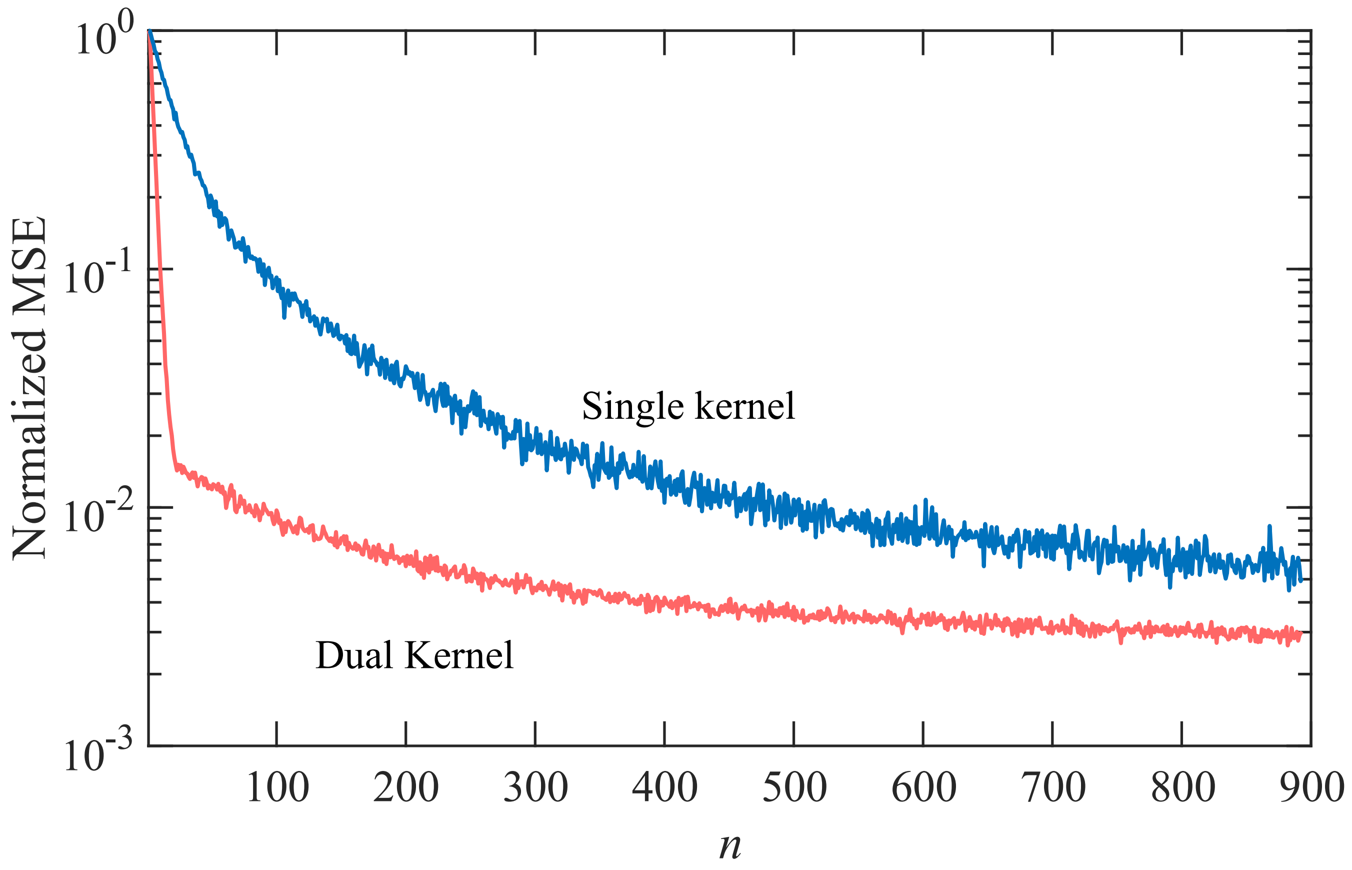
- Campus scenario:
  - High-resolution data (3m)
  - Measurement data in an 802.11 network
- Urban scenario:
  - Low-resolution data (50 m)
  - Ray-tracing data
- Available data samples are divided into:
  - 70% of learning data
  - 30% of test data

The dual-kernel regression is applied to 1000 random perturbations of available data samples, then averaged MSE and std are computed

Campus scenario

Dual kernel advantages:

- Better convergence speed
- Better MSE error
- Performs close to the batch SVM
- Performs better than Multi-kernel approach

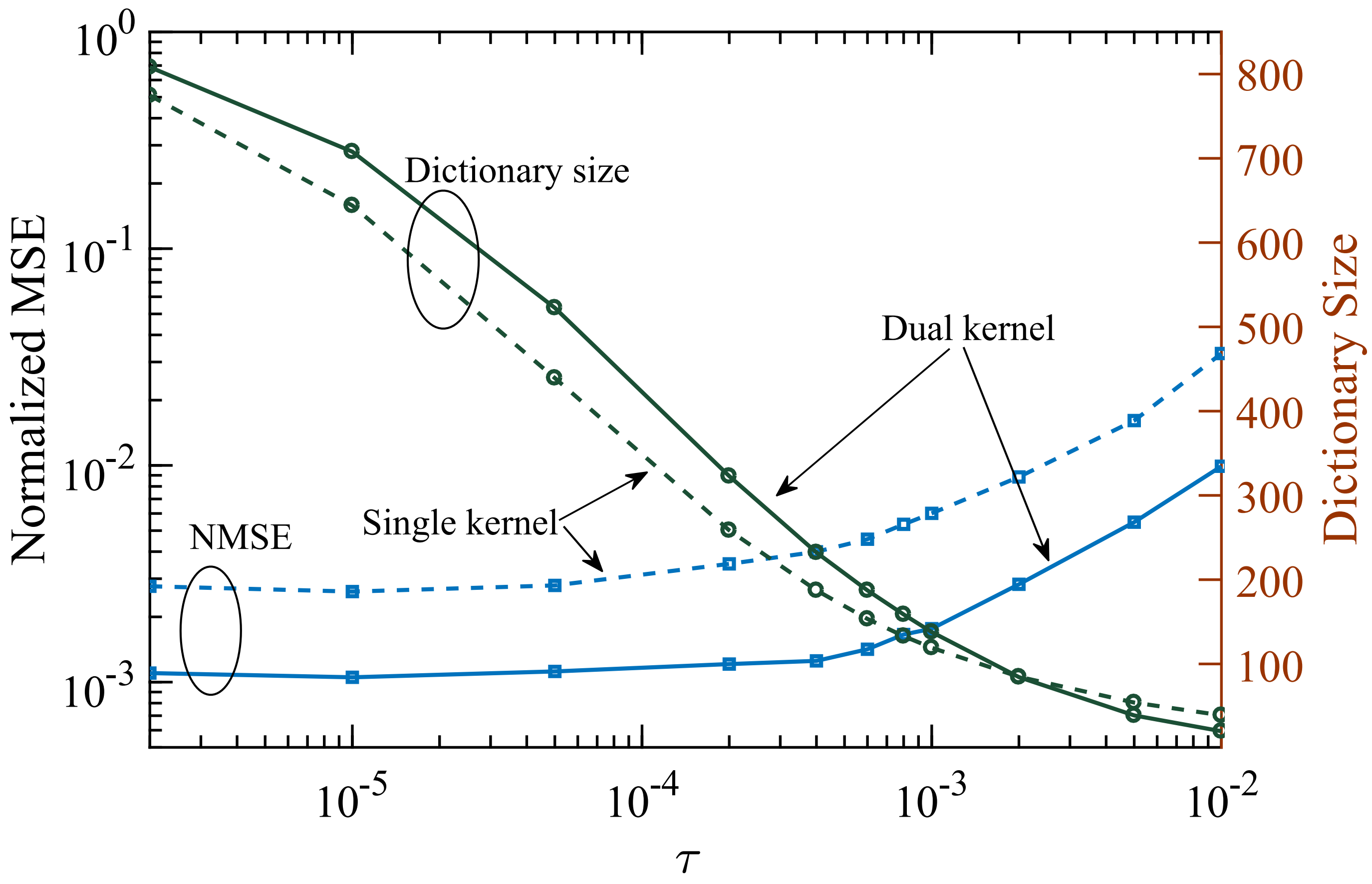


	Batch SVM	Online Dual Kernel	Online Single Kernel	Online Multikernel
NMSE ×10e3	1.7	**2.7**	4.4	6
std ×103	2.5	**3.9**	11.2	--

Urban scenario

Dual kernel approach:

- Has a smaller dictionary size for a given NMSE error
- Performs better than the single kernel and dual kernel approaches



	Batch SVM	Online Dual Kernel	Online Single Kernel	Online Multikernel
NMSE 10e3	.4	**1.1**	2.8	6
std 103	3.5	**5**	20	--

Conclusion:

- Power maps are data-driven maps for received power level
- Power maps model site-specific behaviors rather than the statistical ones
- Power maps have a fairly stable DC component, which can be captured using DC kernel
- The online Dual kernel is an online regression tool that uses one kernel for DC component and another kernel for the varying part of power map
- The online Dual kernel outperforms other states of the art methods in terms of:
  - Convergence speed
  - Normalized MSE error

Questions

Define the solution set

1. For each measurement, define a solution set

$$S_n = \left\{ \mathsf{f} \in \mathcal{H} : |y_n - \langle \mathsf{f}, k(x_n, \cdot) \rangle| \leq \epsilon \right\}$$

- Every point in Hilbert space is a function
- We assign a function (or a point in Hilbert space) to each user location (concept of feature projection)
- Computing the value of pathloss estimator function in a user location is equivalent to performing an inner product between their representatives in Hilbert space

2. Find the intersection or
    - Solution set for all but limited number of measurements
    - Solution set for pathloss estimator function