Mobile Sensing of Spatial Fields

Challenges and Opportunities

Jayakrishnan Unnikrishnan

Audiovisual Communications Laboratory
Ecole Polytechnique Fédérale de Lausanne

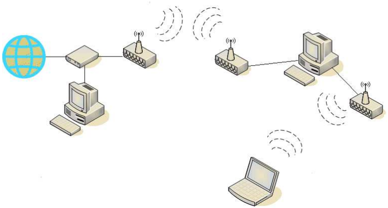
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Collaborators: Martin Vetterli, José Luis Romero, Karlheinz Gröchenig, Farid M. Naini

University of Michigan
Oct. 9, 2013

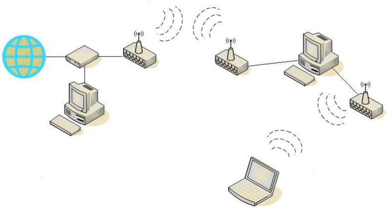
Life in a networked world

- Communication and computing devices are everywhere



Life in a networked world

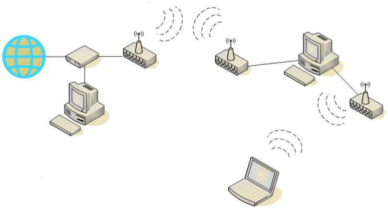
- Communication and computing devices are everywhere



- Sensors are everywhere

Life in a networked world

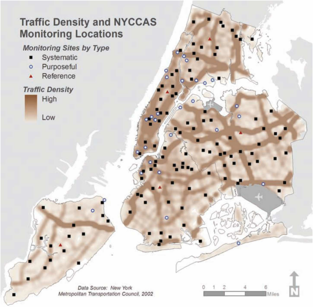
- Communication and computing devices are everywhere



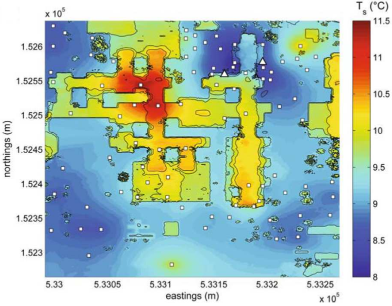
- Sensors are everywhere
- And getting smarter

Berkeley motes: "Smart Dust"

Environment monitoring like never before



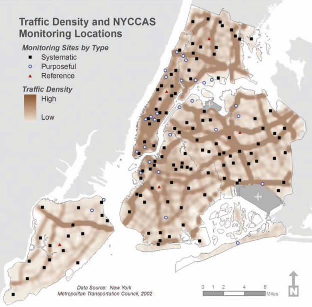
Traffic density, New York



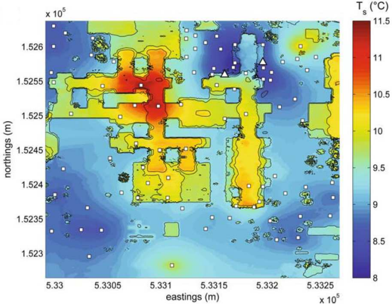
Surface temperature, EPFL,

[Nadeau et al., '09]

Environment monitoring like never before



Traffic density, New York

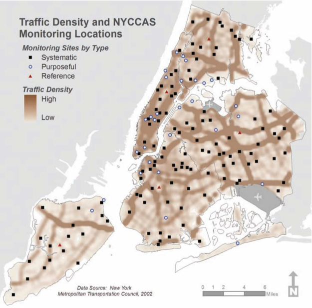


Surface temperature, EPFL,

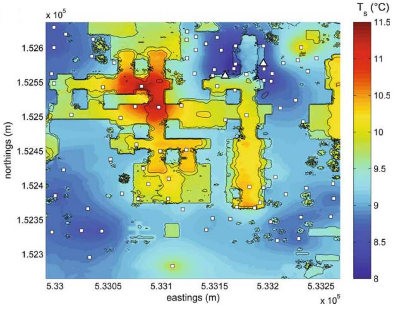
[Nadeau et al., '09]

- Emerging paradigm:
  - Sensing spatial fields with mobile sensors

Environment monitoring like never before



Traffic density, New York



Surface temperature, EPFL,

[Nadeau et al., '09]

- Emerging paradigm:
  - Sensing spatial fields with mobile sensors
  - Offers unique advantages over static sensing

Advantages of mobile sensing

- Mobile sensor can sample at arbitrarily high resolutions along path

Advantages of mobile sensing

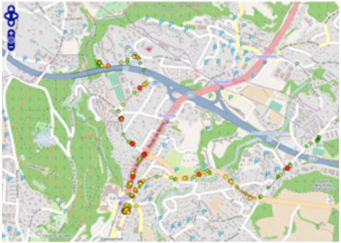
- Mobile sensor can sample at arbitrarily high resolutions along path

- Mobile sensors can implement spatial anti-aliasing via time-domain filtering (*See later*)

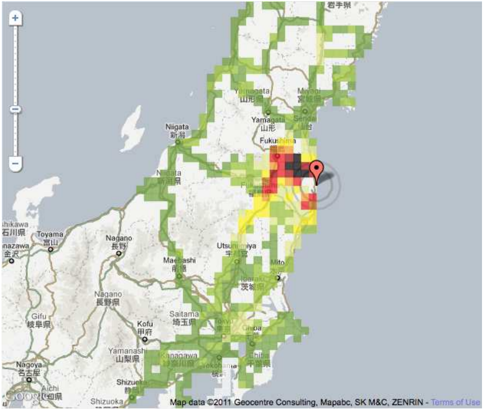
Advantages of mobile sensing

- Mobile sensor can sample at arbitrarily high resolutions along path
- Mobile sensors can implement spatial anti-aliasing via time-domain filtering (*See later*)
- Single mobile sensor can cover a wide area of interest
  - Potentially cost-effective and more practical - e.g. pollution monitoring in a city

Pollution monitoring in Lausanne



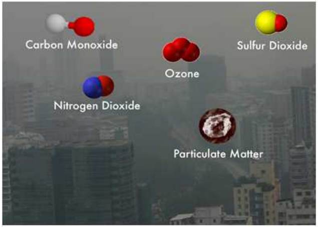
Mobile radiation sampling in Japan

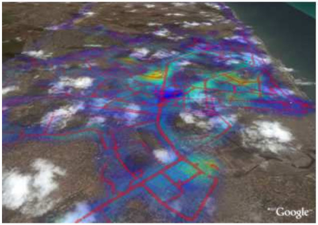


Mobile radiation sampling in Japan

Mobile sensing on roads

Citizen sensing



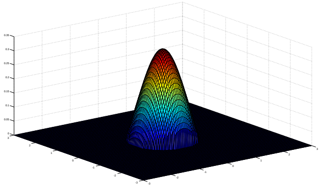


Images from http://www.urban-atmospheres.net/CitizenScience/

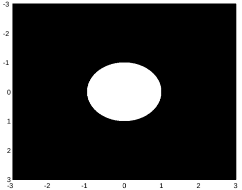
Outline of the talk

- Given: spatially bandlimited field $f : \mathbb{R}^d \mapsto \mathbb{R}$

$$\mathcal{F}(\omega) := \int f(r) e^{-j\langle \omega, r \rangle} dr = 0 \text{ for } \omega \notin \Omega$$



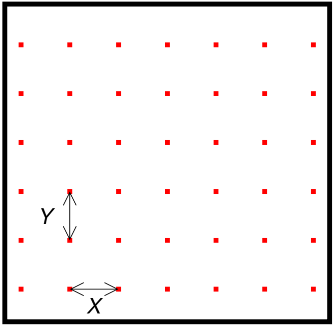
Spectrum $|\mathcal{F}(\omega_x, \omega_y)|$



Support of spectrum Ω

Classical sampling in higher dimensions [PM 1962]

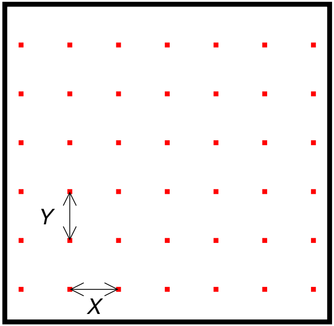
- Sampling on a lattice



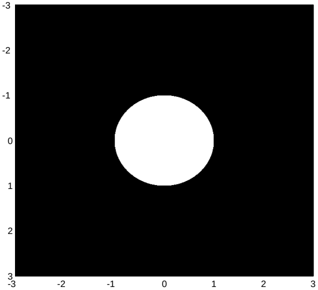
Sampling lattice

Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice



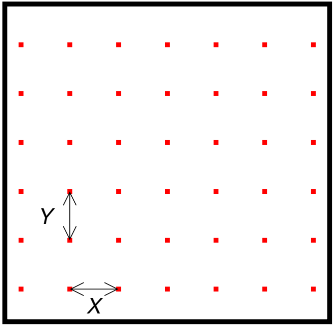
Sampling lattice



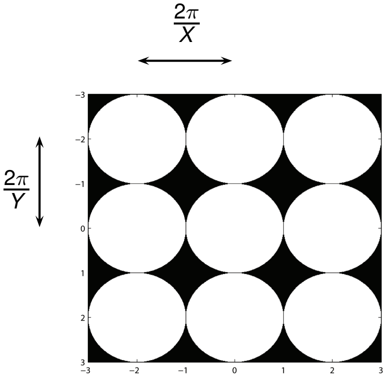
Original spectrum

Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice



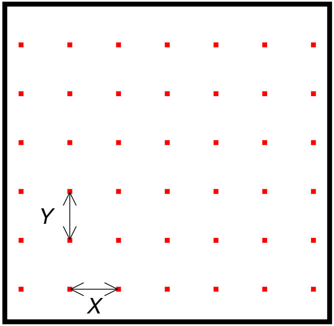
Sampling lattice



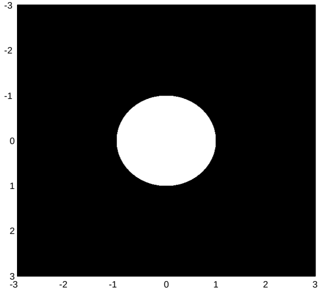
No aliasing in sampled spectrum for $X = Y \leq \frac{\pi}{R}$

Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice



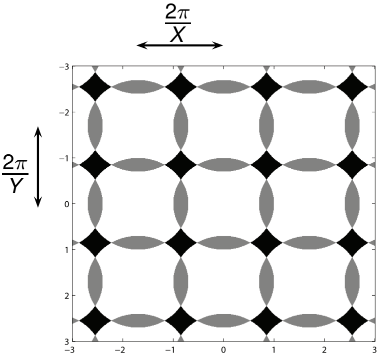
Sampling lattice



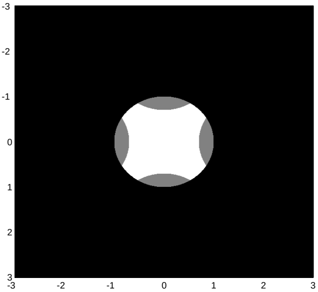
Perfect recovery of original spectrum

Classical sampling in higher dimensions [PM 1962]

- Sampling on a lattice



Aliased sampled spectrum



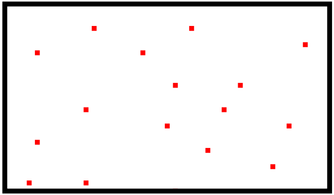
Perfect recovery impossible

- Lattice should be fine enough $\equiv$ Nyquist criterion in $\mathbb{R}^d$

Classical sampling vs Mobile sampling

Classical sampling

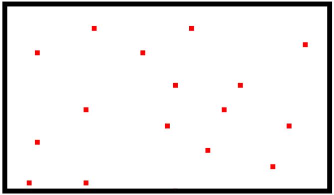
Static sensors record field values at points



Classical sampling vs Mobile sampling

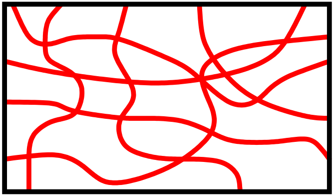
Classical sampling

Static sensors record field values at points



Mobile sampling

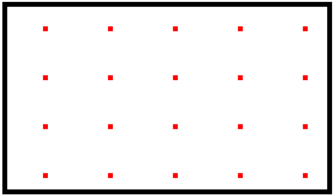
Mobile sensors record field values on trajectories



Classical sampling vs Mobile sampling

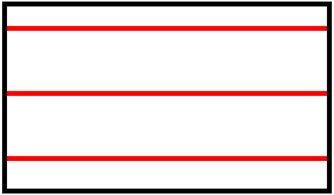
Classical sampling

Static sensors record field values at points (e.g. a lattice)



Mobile sampling

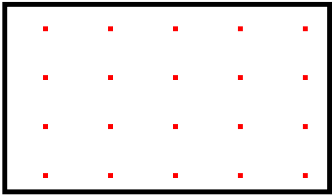
Mobile sensors record field values on trajectories



Classical sampling vs Mobile sampling

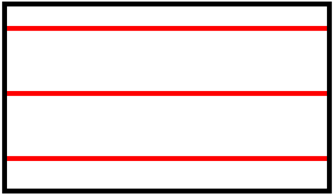
Classical sampling

Static sensors record field values at points (e.g. a lattice)



Mobile sampling

Mobile sensors record field values on trajectories



Focus on time-invariant fields

Sampling trajectories

- A trajectory set p is a countable collection of paths p_i:

$$p = \{p_i : i \in \mathbb{I}\}$$

Sampling trajectories

- A trajectory set p is a countable collection of paths p_i:

$$p = \{p_i : i \in \mathbb{I}\}$$

- Path density of p: Total path-length per unit spatial volume

$$\ell(p) := \lim_{a \to \infty} \frac{\mathcal{D}^p(a)}{\mathrm{Vol}_d(a)}$$

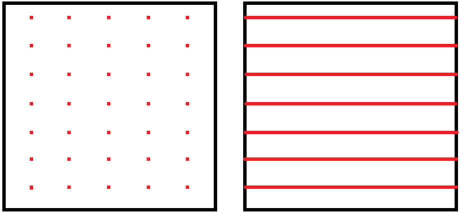
Sampling trajectories

- A trajectory set p is a countable collection of paths p_i:

$$p = \{p_i : i \in \mathbb{I}\}$$

- Path density of p: Total path-length per unit spatial volume

$$\ell(p) := \lim_{a \to \infty} \frac{\mathcal{D}^p(a)}{\mathrm{Vol}_d(a)}$$



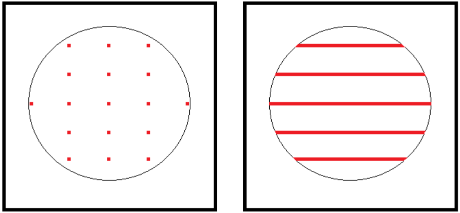
Sampling trajectories

- A trajectory set p is a countable collection of paths p_i:

$$p = \{p_i : i \in \mathbb{I}\}$$

- Path density of p: Total path-length per unit spatial volume

$$\ell(p) := \lim_{a \to \infty} \frac{\mathcal{D}^p(a)}{\mathrm{Vol}_d(a)}$$



Sampling trajectories for bandlimited fields

- $\mathcal{N}_\Omega$: collection of Nyquist trajectory sets p for fields f bandlimited to Ω

Sampling trajectories for bandlimited fields

- $\mathcal{N}_\Omega$: collection of Nyquist trajectory sets p for fields f bandlimited to Ω
- Requirements:
  - Field $f(.)$ can be reconstructed stably from values on trajectories
    - There exists $\Lambda \subset \{p_i(t) : t \in \mathbb{R}, i \in \mathbb{I}\}$ and $A, B < \infty$ s.t.,

$$A\|f\|^2 \leq \sum_{x \in \Lambda} |f(x)|^2 \leq B\|f\|^2, \text{ for all } f \in \mathcal{B}_\Omega$$

Sampling trajectories for bandlimited fields

- $\mathcal{N}_\Omega$: collection of Nyquist trajectory sets p for fields f bandlimited to Ω
- Requirements:
  - Field $f(.)$ can be reconstructed stably from values on trajectories
    - There exists $\Lambda \subset \{p_i(t) : t \in \mathbb{R}, i \in \mathbb{I}\}$ and $A, B < \infty$ s.t.,

$$A\|f\|^2 \leq \sum_{x \in \Lambda} |f(x)|^2 \leq B\|f\|^2, \text{ for all } f \in \mathcal{B}_\Omega$$

  - Regularity conditions

Main contributions

- Examples of trajectory sets in $\mathcal{N}_\Omega$
  - Some new results

Main contributions

- Examples of trajectory sets in $\mathcal{N}_\Omega$
  - Some new results
- Designing trajectory-sets that are minimal in path density
  - New formulation
  - Optimality results from restricted classes

Related work

- Scanning trajectories for MRI
  - Trajectories in Fourier space indicate how to vary magnetic field in time

Related work

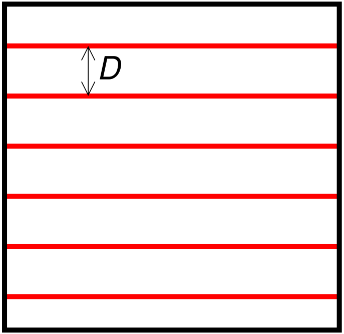
- Scanning trajectories for MRI
  - Trajectories in Fourier space indicate how to vary magnetic field in time

- Reconstructing bandlimited fields from readings on circles [Tewfik, Levy, Willsky '88], [Myridis, Chamzas '98] and spirals [Benedetto, Wu '00]

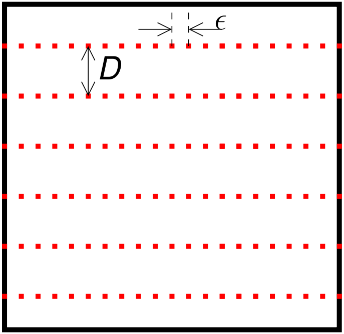
Related work

- Scanning trajectories for MRI
  - Trajectories in Fourier space indicate how to vary magnetic field in time
- Reconstructing bandlimited fields from readings on circles [Tewfik, Levy, Willsky '88], [Myridis, Chamzas '98] and spirals [Benedetto, Wu '00]
- Adaptive path-planning in mobile sensor networks etc.

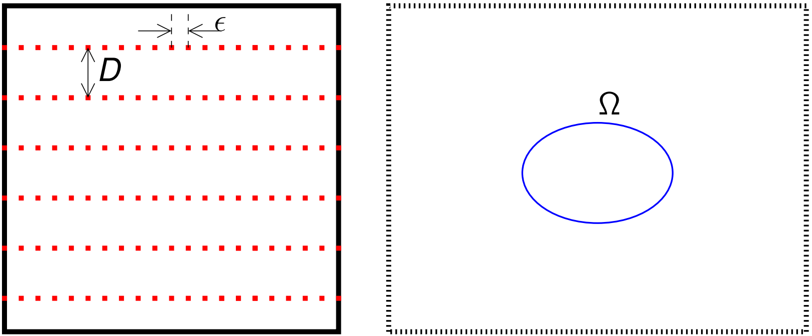
A Uniform trajectory set in $\mathbb{R}^2$



Proof - Uniform set

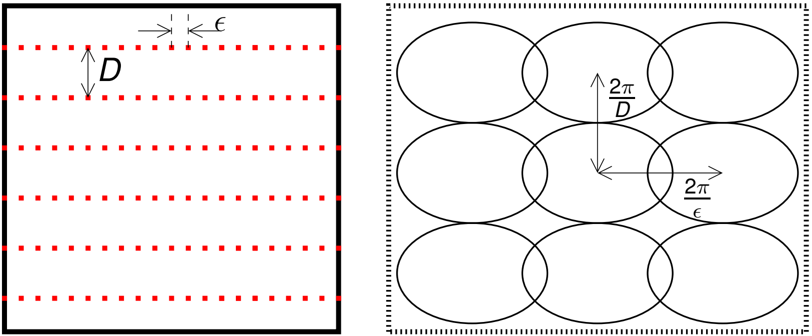


Proof - Uniform set



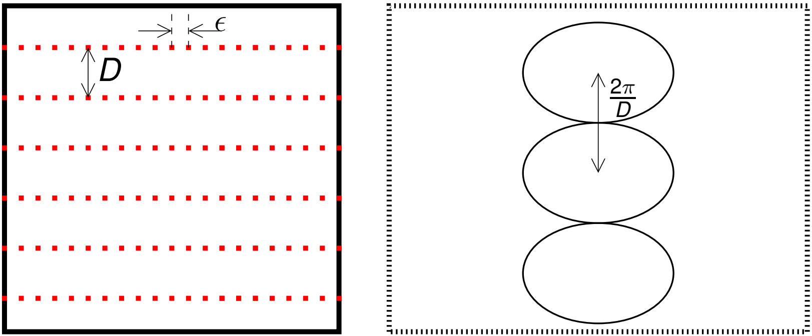
Original field spectrum

Proof - Uniform set



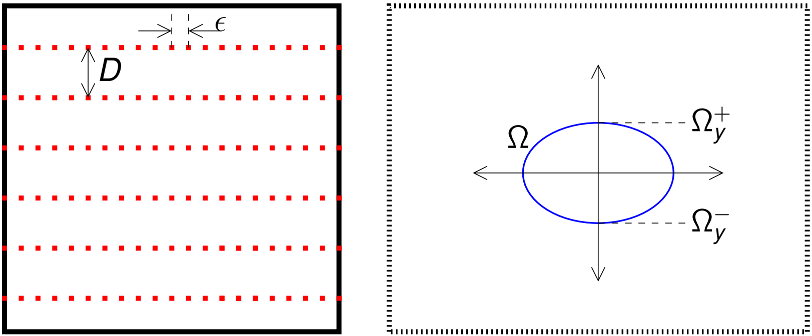
Sampled field spectrum

Proof - Uniform set



Sampled field spectrum as $\epsilon \to 0$

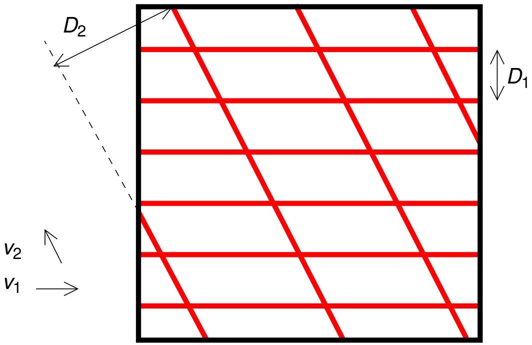
Proof - Uniform set



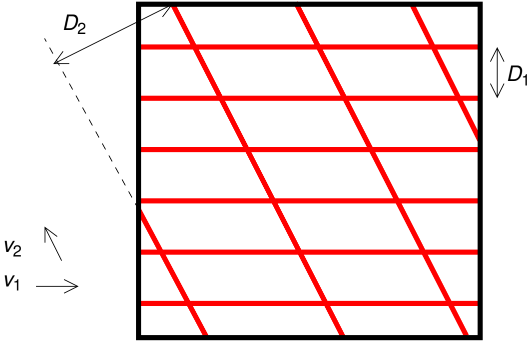
Perfect recovery provided $D \leq \frac{2\pi}{\Omega_y^+ - \Omega_y^-}$

i.e. $p \in \mathcal{N}_\Omega$

Union of Uniform sets in $\mathbb{R}^2$

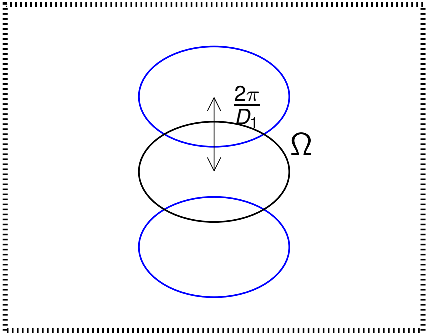


Union of Uniform sets in $\mathbb{R}^2$



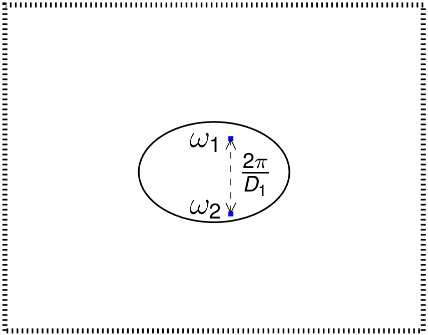
- Can identify exact conditions on D_i, v_i to ensure $p \in \mathcal{N}_\Omega$

Proof intuition - union of uniform sets



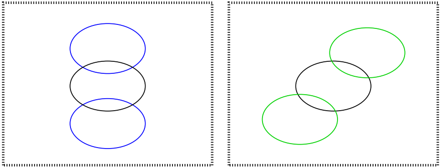
Aliased spectrum from first uniform set

Proof intuition - union of uniform sets



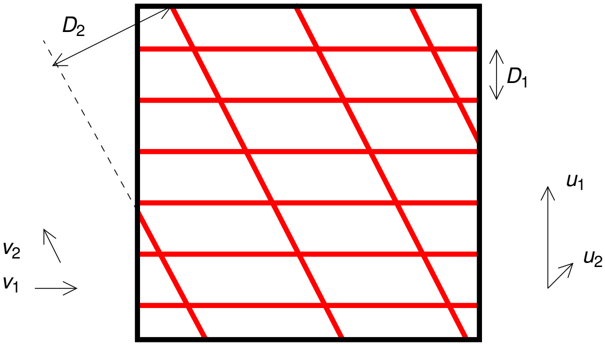
Aliased reconstruction: $\hat{\mathcal{F}}(\omega_1) = \mathcal{F}(\omega_1) + \mathcal{F}(\omega_2)$

Proof intuition - union of uniform sets



Linear equations solvable under condition

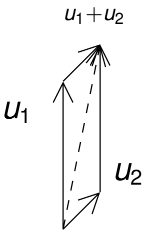
Union of regular parallel trajectories for $\mathbb{R}^2$



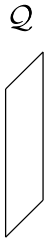
- For each i define

$$u_i = \frac{2\pi v_i^{\perp}}{D_i}$$

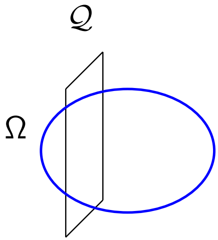
Union of regular parallel trajectories for $\mathbb{R}^2$



Union of regular parallel trajectories for $\mathbb{R}^2$



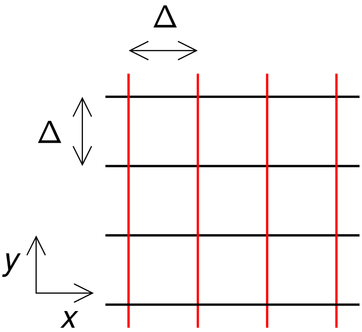
Union of regular parallel trajectories for $\mathbb{R}^2$



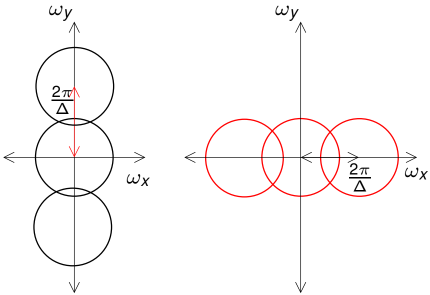
Theorem

For convex, compact Ω we have $p \in \mathcal{N}_\Omega$ if and only if $\mathcal{Q}$ is not contained within Ω or its translates

Example 1: Orthogonal trajectories and Isotropic field

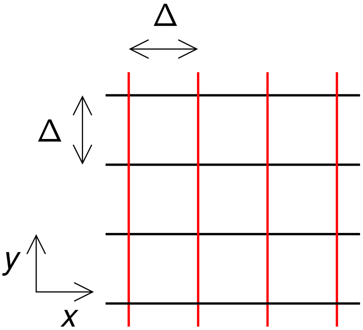


Orthogonal sets of trajectories

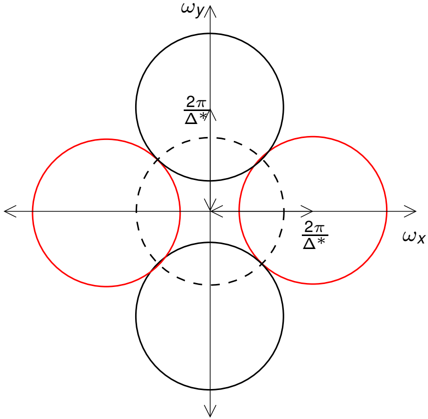


Sampled spectra from the two sets

Example 1: Orthogonal trajectories and Isotropic field

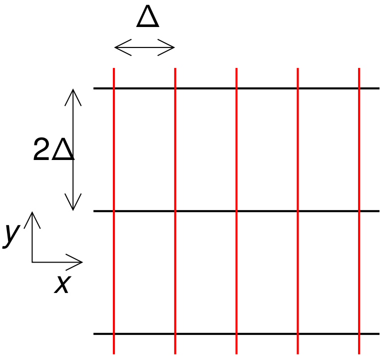


Orthogonal sets of trajectories



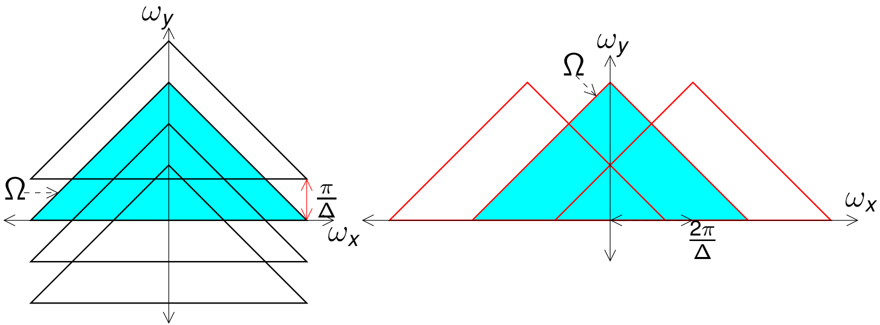
Critical sampling

Example 2: Non-isotropic field



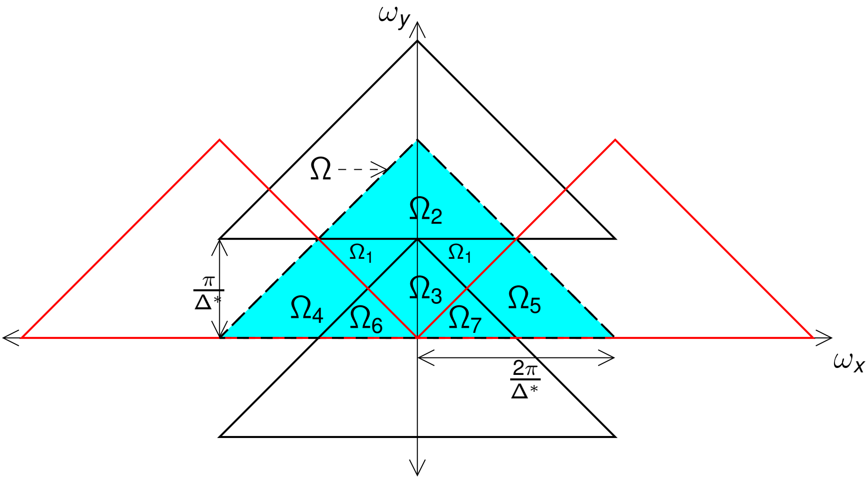
Orthogonal sets of trajectories

Example 2: Non-isotropic field



Sampled spectra from the two sets

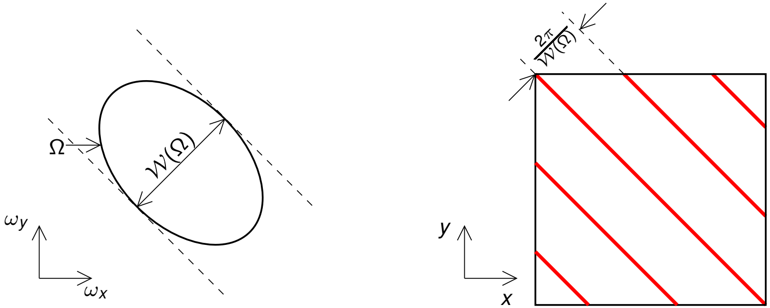
Example 2: Non-isotropic field



Critical sampling

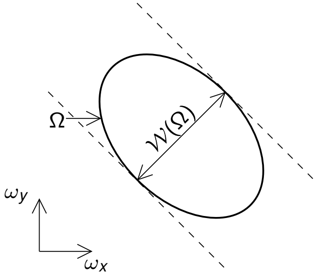
Optimality of a Uniform set in $\mathbb{R}^2$

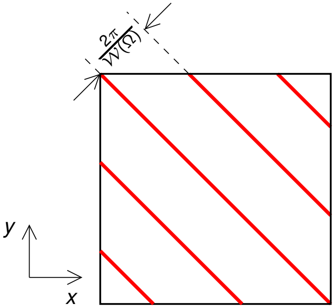
- Assume Ω is a convex set



Optimality of a Uniform set in $\mathbb{R}^2$

- Assume Ω is a convex set

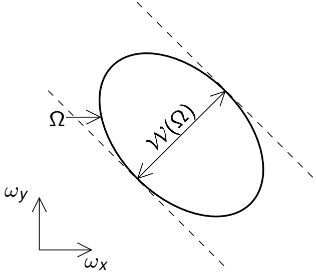


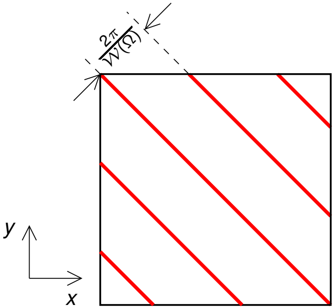


- Optimal Uniform set p is $\perp$ direction in which Ω is narrowest

Optimality of a Uniform set in $\mathbb{R}^2$

- Assume Ω is a convex set





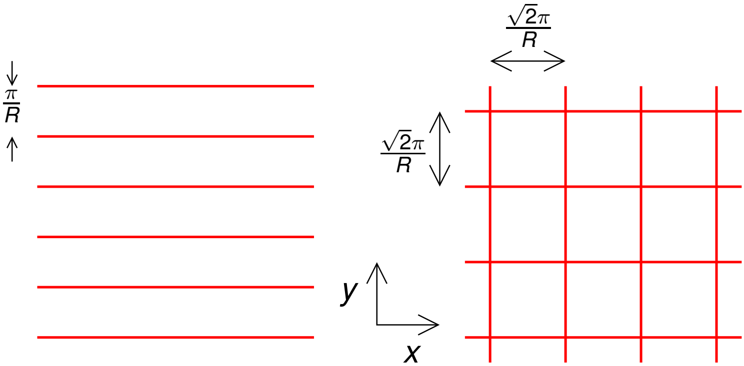
- Optimal Uniform set p is $\perp$ direction in which Ω is narrowest

Theorem

For any union of uniform sets $q \in \mathcal{N}_\Omega$, we have

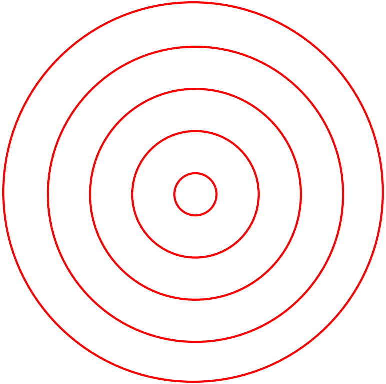
$$\ell(q) \geq \ell(p)$$

Implications for isotropic fields



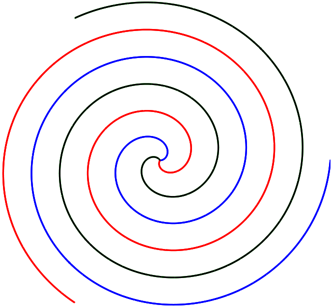
Optimal uniform set has lower path density than any union of orthogonal uniform sets

Other trajectory configurations



Equispaced concentric circles

Other trajectory configurations



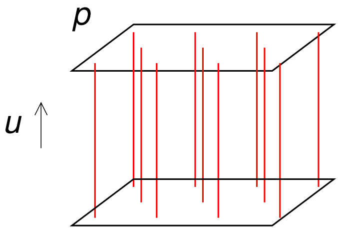
Interleaved spirals

Uniform sets for $\mathbb{R}^3$

- Assume $\Omega \subset \mathbb{R}^3$ is convex and symmetric

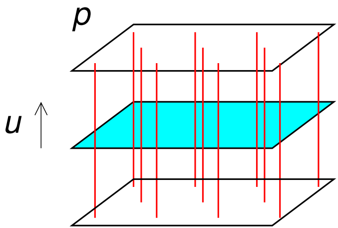
Uniform sets for $\mathbb{R}^3$

- Assume $\Omega \subset \mathbb{R}^3$ is convex and symmetric
- Consider Uniform sets in $\mathbb{R}^3$



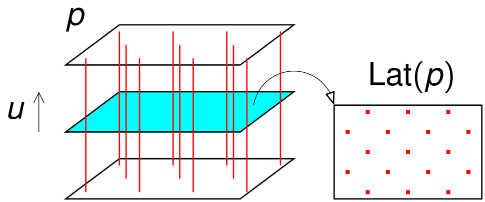
Uniform sets for $\mathbb{R}^3$

- Assume $\Omega \subset \mathbb{R}^3$ is convex and symmetric
- Consider Uniform sets in $\mathbb{R}^3$



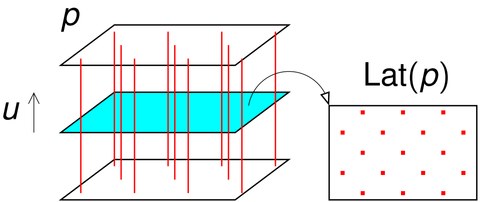
Uniform sets for $\mathbb{R}^3$

- Assume $\Omega \subset \mathbb{R}^3$ is convex and symmetric
- Consider Uniform sets in $\mathbb{R}^3$



Uniform sets for $\mathbb{R}^3$

- Assume $\Omega \subset \mathbb{R}^3$ is convex and symmetric
- Consider Uniform sets in $\mathbb{R}^3$

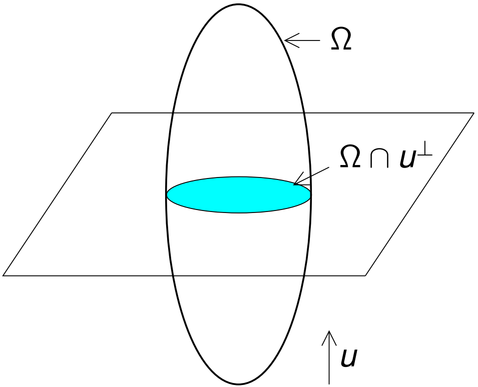


Theorem

$p \in \mathcal{N}_\Omega$ iff Lat(p) forms sampling lattice for $\Omega \cap u^\perp$. Furthermore

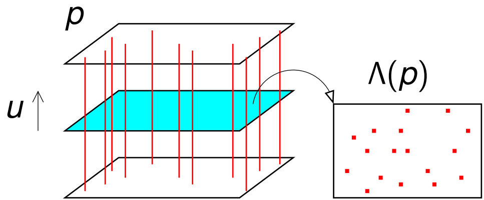
$$\ell(p) = \text{Sampling density}(\text{Lat}(p))$$

Section of a set



Non-uniform parallel trajectory sets for $\mathbb{R}^d$

- Consider non-uniform parallel trajectory sets



- If p is homogenous (i.e., $\Lambda(p)$ has equal density about every point), then

$$\ell(p) = \text{Sampling density}(\Lambda(p))$$

Optimal parallel trajectory sets for $\mathbb{R}^d$

Theorem

For Ω compact, convex, and symmetric let

$$u^* = \underset{u \in \mathbb{R}^d : \|u\| = 1}{\arg\min} \; |\Omega \cap u^\perp|$$

Then the optimal parallel trajectory set p^* is parallel to u^* and

$$\ell(p^*) = \frac{|\Omega \cap u^{*\perp}|}{(2\pi)^{d-1}}.$$

Optimal parallel trajectory sets for $\mathbb{R}^d$

Theorem

For Ω compact, convex, and symmetric let

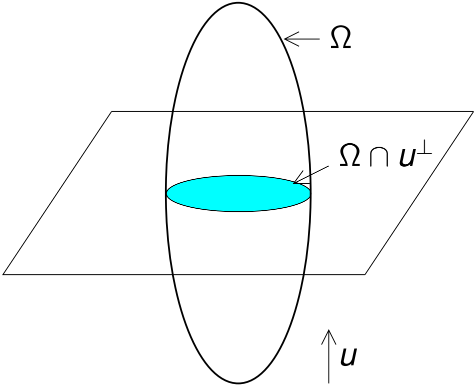
$$u^* = \operatorname*{arg\,min}_{u \in \mathbb{R}^d : \|u\| = 1} |\Omega \cap u^\perp|$$

Then the optimal parallel trajectory set p^* is parallel to u^* and

$$\ell(p^*) = \frac{|\Omega \cap u^{*\perp}|}{(2\pi)^{d-1}}.$$

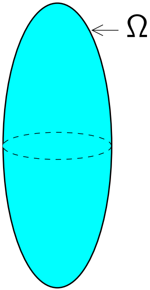
- I.e., minimum path density $\propto$ minimum section through the origin

Optimal parallel trajectory sets for $\mathbb{R}^d$



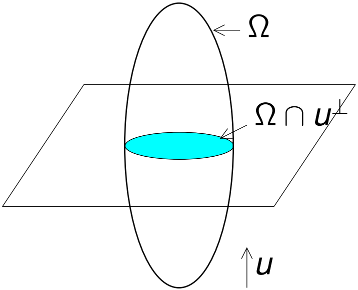
Classical sampling vs Mobile sampling

Classical sampling



Minimum sampling density
$\propto \mathrm{Vol}(\Omega)$ [Landau '67s]

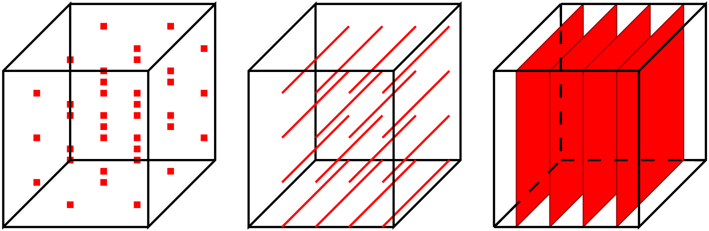
Sampling on parallel lines



Minimum path density
$\propto$ Min. section(Ω)

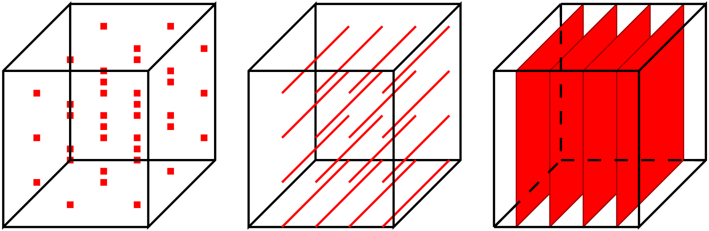
Extensions

- Generalize trajectories further to manifolds



Extensions

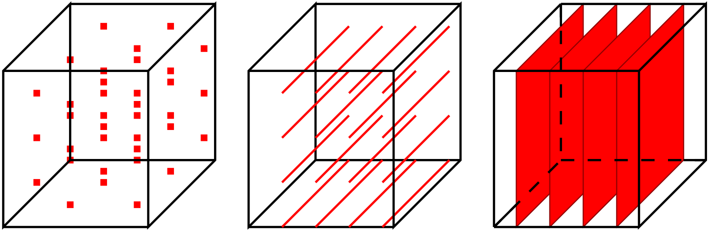
- Generalize trajectories further to manifolds



- Sampling on arbitrary curves: Ill-posed but can be fixed

Extensions

- Generalize trajectories further to manifolds

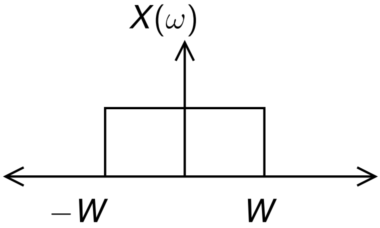


- Sampling on arbitrary curves: Ill-posed but can be fixed
- Time-varying bandlimited fields e.g. spatial audio fields
  - Reduce sensor density by increasing temporal sampling rate

Outline of the talk

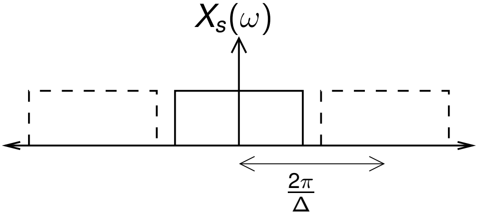
1. Sampling Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories
2. Spatial Anti-aliasing via Mobile Sensing
3. Privacy of Mobility traces
4. Recap

Anti-aliasing in one dimension



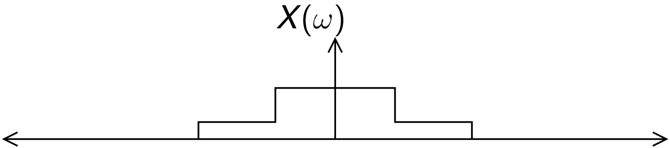
Bandlimited signal spectrum

Anti-aliasing in one dimension



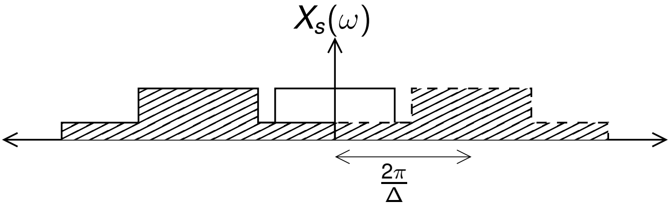
Sampled spectrum

Anti-aliasing in one dimension



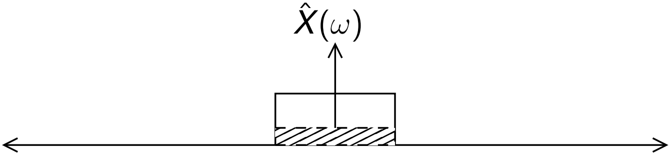
Imperfectly bandlimited signal

Anti-aliasing in one dimension



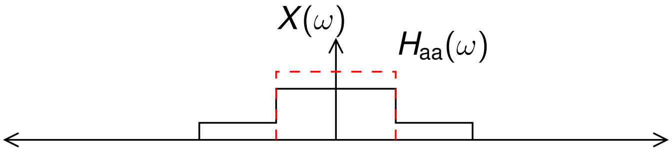
Sampled spectrum

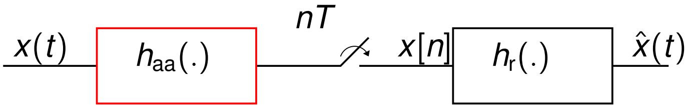
Anti-aliasing in one dimension



Aliasing in reconstructed spectrum

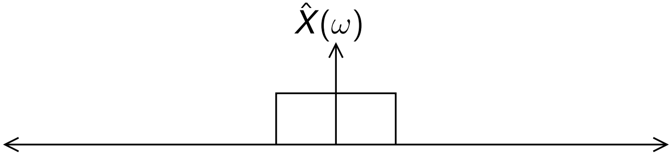
Anti-aliasing in one dimension

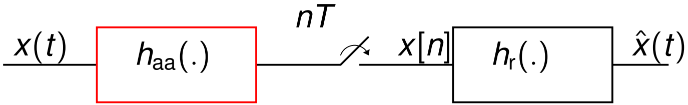




Anti-aliasing filtering

Anti-aliasing in one dimension





No aliasing in reconstruction

Spatial anti-aliasing

- Impossible with static sensors
  - Cannot integrate over continuous space

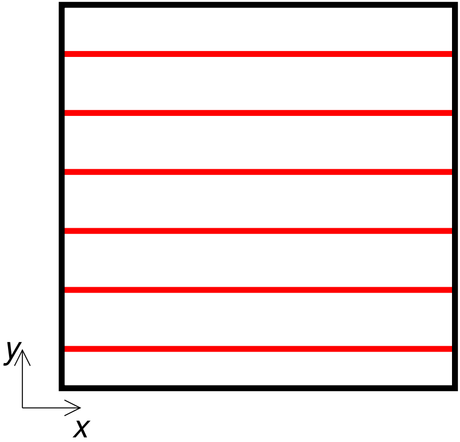
Spatial anti-aliasing

- Impossible with static sensors
  - Cannot integrate over continuous space
- Mobile sensor sees field as function of time
  - $s(t) = f(r(t))$
  - If constant velocity then $s(t)$ is bandlimited

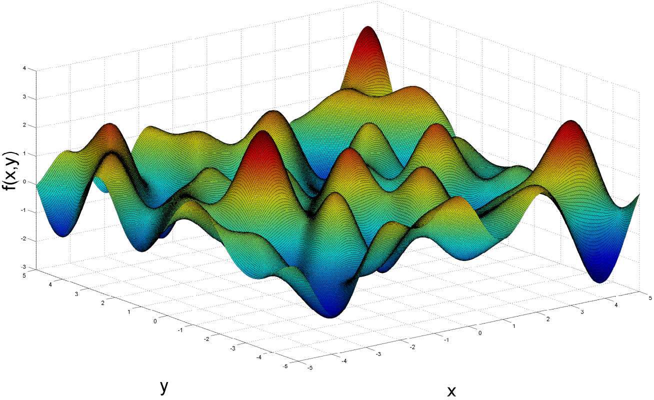
Spatial anti-aliasing

- Impossible with static sensors
  - Cannot integrate over continuous space
- Mobile sensor sees field as function of time
  - $s(t) = f(r(t))$
  - If constant velocity then $s(t)$ is bandlimited
- In presence of noise time-domain filtering can suppress spatial aliasing along direction of motion

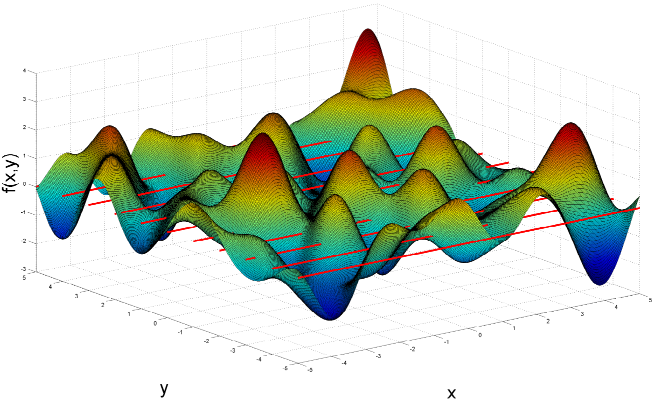
Spatial anti-aliasing filter



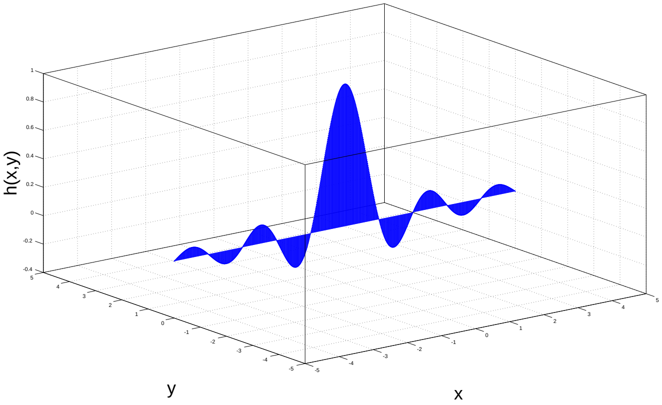
Spatial anti-aliasing filter



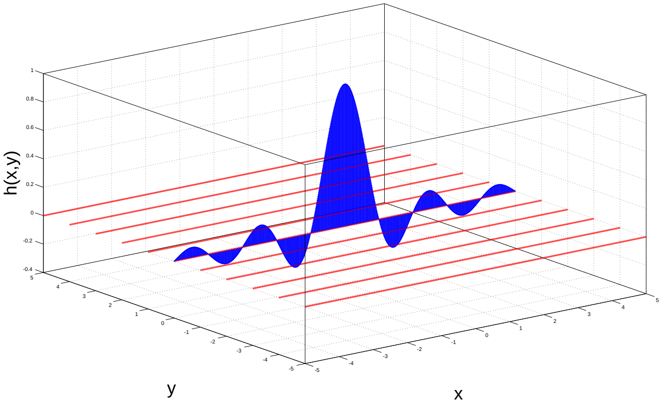
Spatial anti-aliasing filter



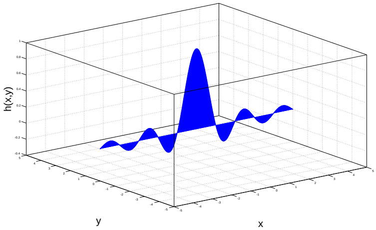
Spatial anti-aliasing filter



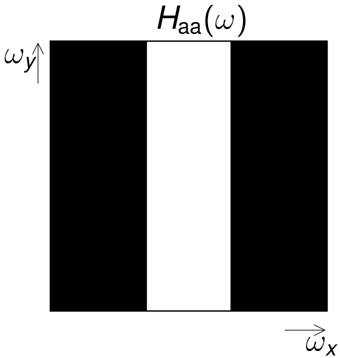
Spatial anti-aliasing filter



Spatial anti-aliasing filter

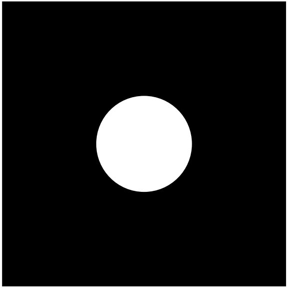


$\Longrightarrow$

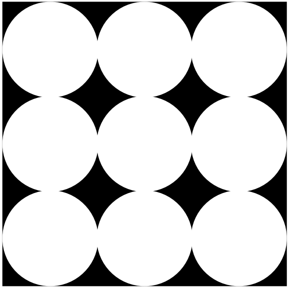


Spatial anti-aliasing illustrated

Static sampling



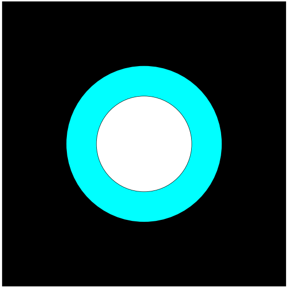
Original spectrum



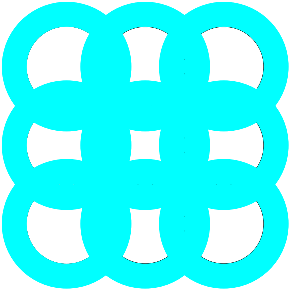
Sampled spectrum

Spatial anti-aliasing illustrated

Static sampling



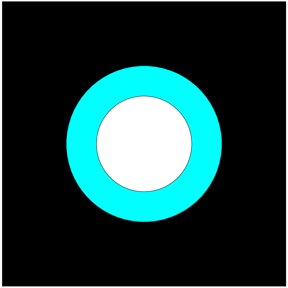
Original spectrum with noise



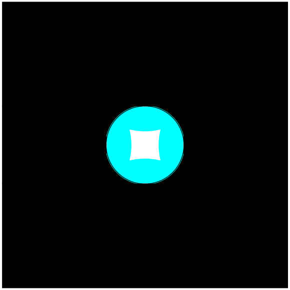
Sampled spectrum aliased in all directions

Spatial anti-aliasing illustrated

Static sampling



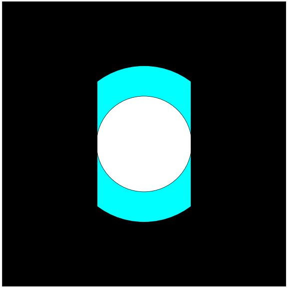
Original spectrum with noise



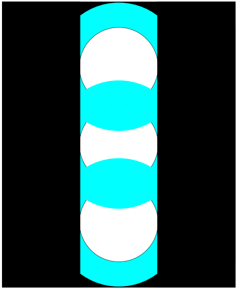
Reconstructed spectrum aliased in all directions

Spatial anti-aliasing illustrated

Mobile sampling



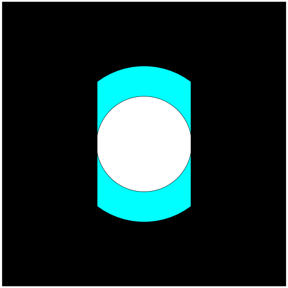
Spatially filtered field with noise



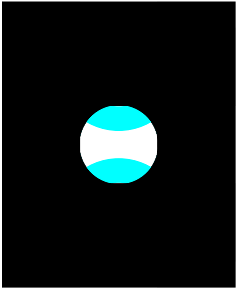
Sampled spectrum aliased only in two directions

Spatial anti-aliasing illustrated

Mobile sampling

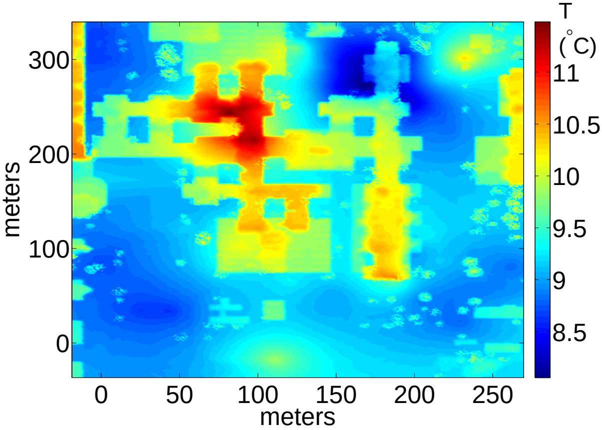


Spatially filtered field with noise



Reconstructed spectrum aliased in two directions

An example: Campus temperature



Performance comparison

Table : Reconstruction errors

Data type	Static sensing	Mobile sensing
Temperature	0.53%	0.45%
BL in noise	9.9%	1.5%

Performance comparison

Table : Reconstruction errors

Data type	Static sensing	Mobile sensing
Temperature	0.53%	0.45%
BL in noise	9.9%	1.5%

- Improvement in SNR can be quantified analytically

Performance comparison

Table : Reconstruction errors

Data type	Static sensing	Mobile sensing
Temperature	0.53%	0.45%
BL in noise	9.9%	1.5%

- Improvement in SNR can be quantified analytically
- Conclusion: Mobile sensing enables spatial anti-aliasing filtering

Outline of the talk

1. Sampling Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories
2. Spatial Anti-aliasing via Mobile Sensing
3. Privacy of Mobility traces
4. Recap

- Personal data being collected at unprecedented levels

Unique in the crowd

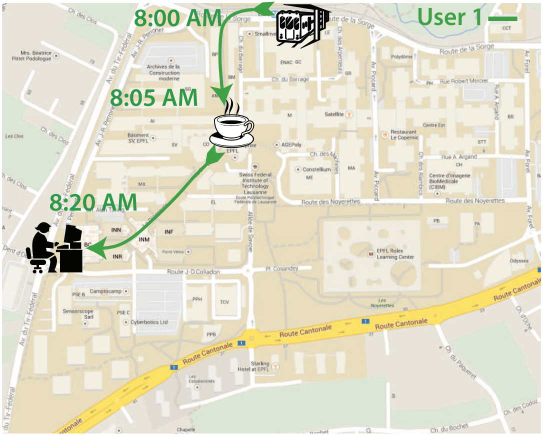
- Anonymized user data is not anonymous given auxiliary information, e.g., Netflix prize dataset [Narayanan, Shmatikov 2008]

Unique in the crowd

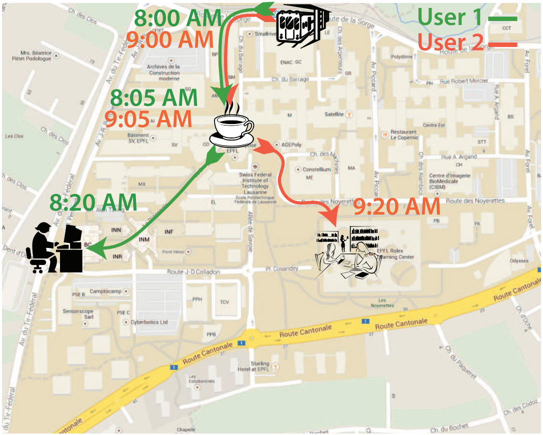
- Anonymized user data is not anonymous given auxiliary information, e.g., Netflix prize dataset [Narayanan, Shmatikov 2008]

- Users are uniquely identifiable from small set of observations
  - 87% Americans uniquely identified given ZIP code, birthdate, and sex [Sweeney 2000]
  - 95% of mobile users in a country are uniquely identified from four spatio-temporal points [Montjoye et al 2013]

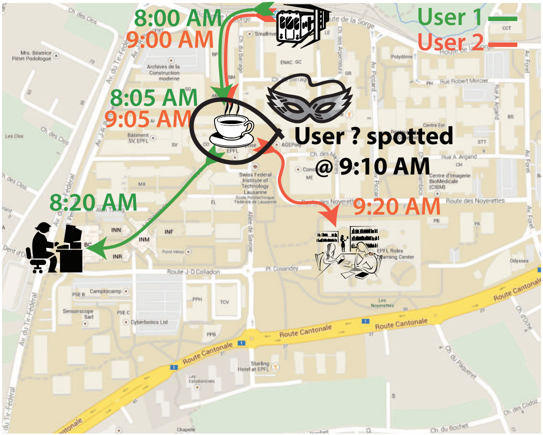
De-anonymization



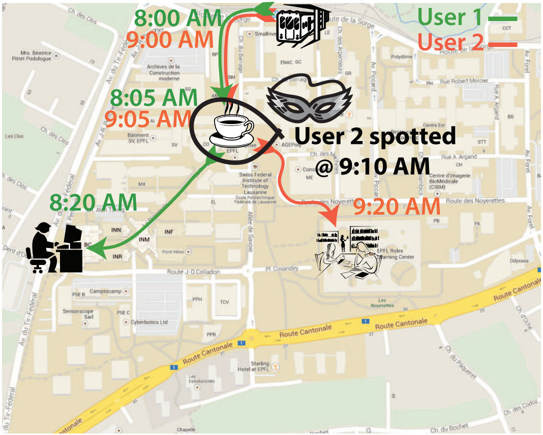
De-anonymization



De-anonymization



De-anonymization



- Uniqueness of trajectories ⇒ Easily de-anonymized

Privacy of Statistics

- Sometimes statistics of data are sufficient, e.g.,
  - Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
  - # visits to particular restaurants (for popularity surveys)
  - # tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)

Privacy of Statistics

- Sometimes statistics of data are sufficient, e.g.,
  - Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
  - $\#$ visits to particular restaurants (for popularity surveys)
  - $\#$ tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)
- Question: How private are anonymized statistics?

Privacy of Statistics

- Sometimes statistics of data are sufficient, e.g.,
  - Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
  - $\#$ visits to particular restaurants (for popularity surveys)
  - $\#$ tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)
- Question: How private are anonymized statistics?
  - What is the optimal de-anonymization strategy given independent auxiliary information?

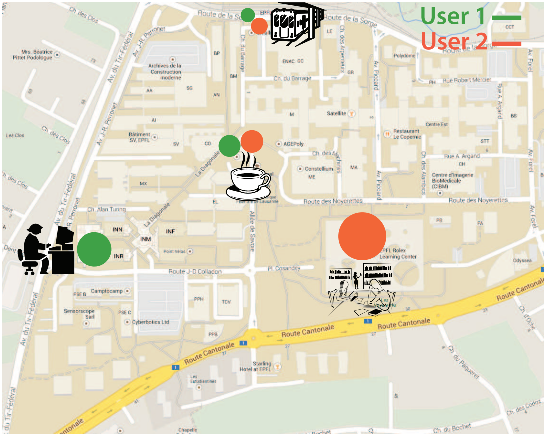
Privacy of Statistics

- Sometimes statistics of data are sufficient, e.g.,
  - Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
  - # visits to particular restaurants (for popularity surveys)
  - # tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)
- Question: How private are anonymized statistics?
  - What is the optimal de-anonymization strategy given independent auxiliary information?
  - Focus on histograms

Privacy of Statistics

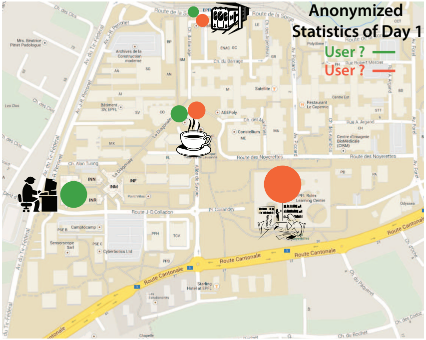
- Sometimes statistics of data are sufficient, e.g.,
  - Fraction of time spent in particular locations or websites (useful for ad-placements, infrastructure planning)
  - $\#$ visits to particular restaurants (for popularity surveys)
  - $\#$ tweets/blog comments on a particular topic/ containing a particular word (for targeted ads)
- Question: How private are anonymized statistics?
  - What is the optimal de-anonymization strategy given independent auxiliary information?
  - Focus on histograms
  - How unique are user location statistics?

Example: Location Histograms



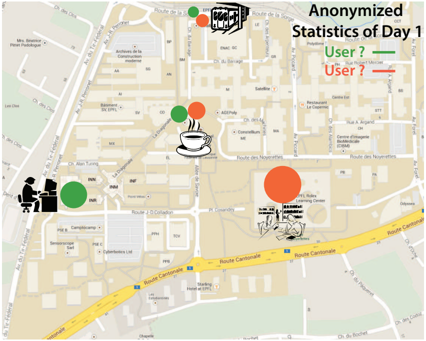
De-anonymizing as Matching

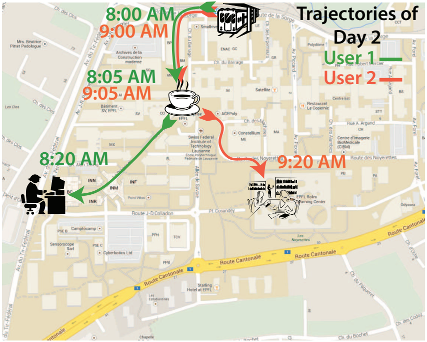
- Given: Anonymized Statistics of K users



De-anonymizing as Matching

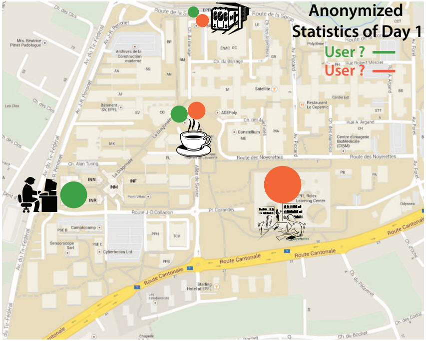
- Given: Anonymized Statistics of K users' data and Auxiliary Information about independently generated data

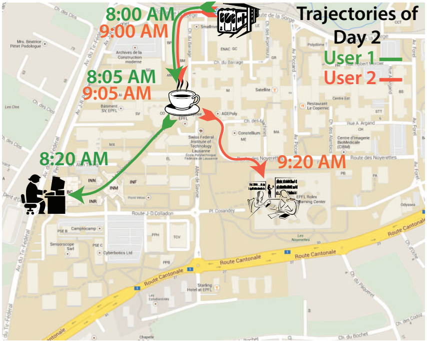




De-anonymizing as Matching

- Given: Anonymized Statistics of *K* users' data and Auxiliary Information about independently generated data





- Task: Match Auxiliary Information to the correct Anonymized Statistics

Data model and notation

- Let $x_{\pi(i)}$ and y_i denote length-n data vectors of user i

Data model and notation

- Let $x_{\pi(i)}$ and y_i denote length-n data vectors of user i
    - where $\pi : \{1, 2, \ldots, K\} \mapsto \{1, 2, \ldots, K\}$ is an unknown permutation

Data model and notation

- Let $x_{\pi(i)}$ and y_i denote length-n data vectors of user i
  - where $\pi : \{1, 2, \ldots, K\} \mapsto \{1, 2, \ldots, K\}$ is an unknown permutation
- User's data: i.i.d. according to unknown law from finite alphabet Z

Data model and notation

- Let $x_{\pi(i)}$ and y_i denote length-n data vectors of user i
  - where $\pi : \{1, 2, \ldots, K\} \mapsto \{1, 2, \ldots, K\}$ is an unknown permutation
- User's data: i.i.d. according to unknown law from finite alphabet Z
- Let p_{x_i} be empirical distribution of vector x_i:

$$p_{x_i}(z) = \frac{1}{n} \sum_{j=1}^{n} \mathcal{I}\{x_i^j = z\}, \quad z \in \mathrm{Z}$$

where

$$x_i = (x_i^1, x_i^2, \ldots, x_i^n)$$

Problem statement

- Statistics revealed: (Anonymized) Empirical distribution of users' x data streams, i.e., the set

$$\{p_{x_1}, p_{x_2}, \ldots, p_{x_K}\}$$

is revealed, but not the user identities, i.e., π is unknown

Problem statement

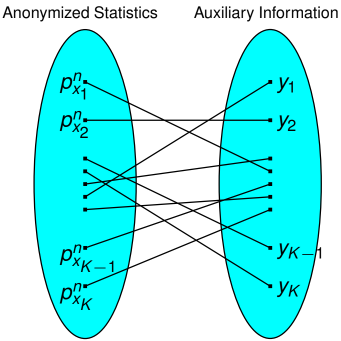
- Statistics revealed: (Anonymized) Empirical distribution of users' x data streams, i.e., the set

$$\{p_{x_1}, p_{x_2}, \ldots, p_{x_K}\}$$

  is revealed, but not the user identities, i.e., π is unknown

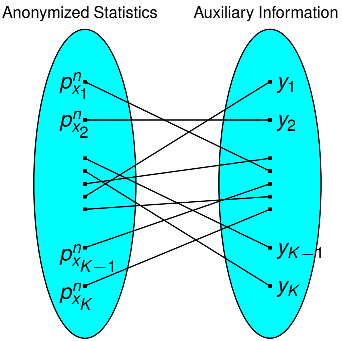
- Auxiliary information: y data streams of the users, as well as the identities
  - Note: y_i's are independent of x_i's

Matching Problem



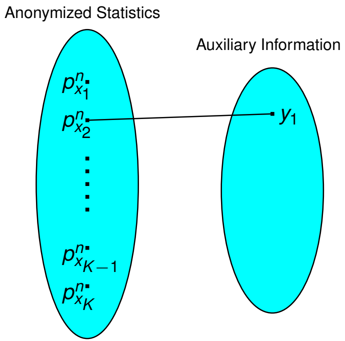
- All strings $\{x_i\}$ and $\{y_i\}$ have equal lengths

Matching Problem



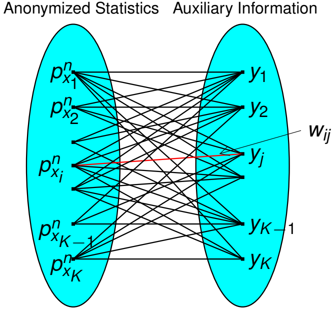
- All strings $\{x_i\}$ and $\{y_i\}$ have equal lengths
- Can be viewed as hypothesis testing problem with $M = K!$ composite hypotheses

Matching Problem



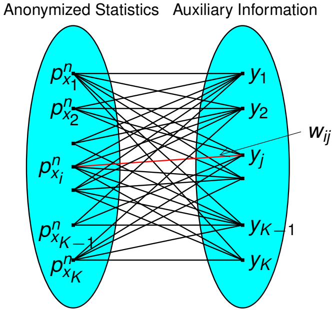
- Related problem: only one y string given; studied by Gutman ('89)

Potential solution: "ML"



$$w_{ij} = D(p^n_{x_i} \| \tfrac{1}{2}(p^n_{x_i} + p^n_{y_j})) + D(p^n_{y_j} \| \tfrac{1}{2}(p^n_{x_i} + p^n_{y_j}))$$

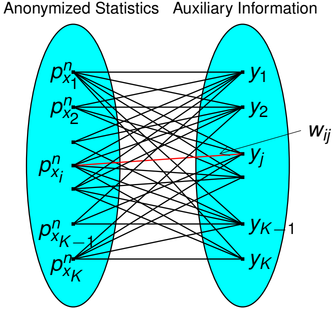
Potential solution: "ML"



$$w_{ij} = D\left(p^n_{x_i} \,\middle\|\, \tfrac{1}{2}(p^n_{x_i} + p^n_{y_j})\right) + D\left(p^n_{y_j} \,\middle\|\, \tfrac{1}{2}(p^n_{x_i} + p^n_{y_j})\right)$$

- ML permutation $\Leftrightarrow$ Min-Weight Matching
  - $\Rightarrow$ easy to compute

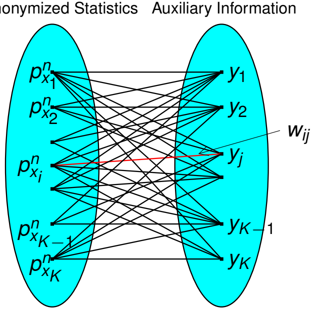
Potential solution: "ML"



$$w_{ij}=D(p^n_{x_i}\|\tfrac{1}{2}(p^n_{x_i}+p^n_{y_j}))+D(p^n_{y_j}\|\tfrac{1}{2}(p^n_{x_i}+p^n_{y_j}))$$

- ML permutation $\Leftrightarrow$ Min-Weight Matching
  - $\Rightarrow$ easy to compute
  - Is this optimal?

Potential solution: “ML”

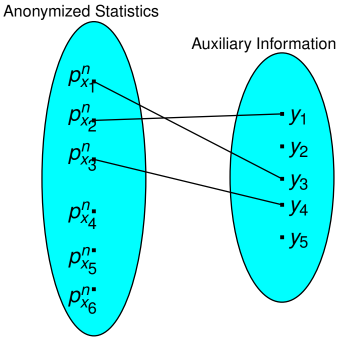


$$w_{ij} = D(p^n_{x_i} \| \tfrac{1}{2}(p^n_{x_i} + p^n_{y_j})) + D(p^n_{y_j} \| \tfrac{1}{2}(p^n_{x_i} + p^n_{y_j}))$$

- ML permutation $\Leftrightarrow$ Min-Weight Matching
  - $\Rightarrow$ easy to compute
  - A slight variant of ML is optimal!

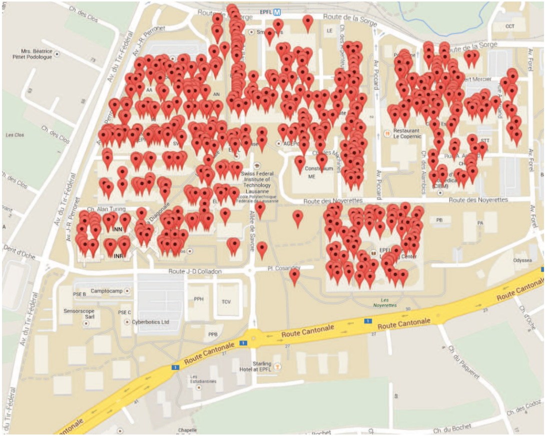
Generalizations

- Different sets of distinct users observed in two days
  - Can be handled provided # common users known



Experiment: Mobility traces on EPFL campus

Obtained from Wi-Fi connections



Dataset

- Anonymized mobility traces of $K \approx 1000$ users on EPFL campus measured on Mondays for two weeks
  - x_i in first week and y_i in second week

Dataset

- Anonymized mobility traces of $K \approx 1000$ users on EPFL campus measured on Mondays for two weeks
  - x_i in first week and y_i in second week
- Alphabet size $|Z| = 933$ access points

Dataset

- Anonymized mobility traces of $K \approx 1000$ users on EPFL campus measured on Mondays for two weeks
    - x_i in first week and y_i in second week
- Alphabet size $|Z| = 933$ access points
- Data-length $n = 28800$ seconds (8 hours)

Results

Dataset	# users (K)	Matching K users in second week (fraction of correct matches)
Mondays	1154	52.9%

Results

Dataset	# users (K)	Matching K users in second week (fraction of correct matches)	Matching 1 random user in second week (prob of success)
Mondays	1154	52.9%	44.5%

- Info on more users on second day $\Rightarrow$ Higher accuracy

Results

Dataset	# users (K)	Matching K users in second week (fraction of correct matches)	Matching 1 random user in second week (prob of success)
Mondays	1154	52.9%	44.5%
Mondays and Tuesdays	1047	70.5%	53.5%

- Info on more users on second day ⇒ Higher accuracy
- More days ⇒ Higher accuracy
- Conclusion: Simple anonymization is not effective

Outline of the talk

1. Sampling Trajectories for Mobile Sampling
   - Classical sampling vs mobile sampling
   - Sampling trajectories
   - Optimal parallel trajectories
2. Spatial Anti-aliasing via Mobile Sensing
3. Privacy of Mobility traces
4. Recap

Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories

Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories
- Perfect reconstruction conditions; Shortest trajectories
  - Uniform set better than unions of Uniform sets for $\mathbb{R}^2$
  - Optimal design of parallel trajectory sets for $\mathbb{R}^d$

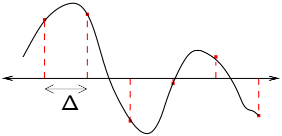
Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories
- Perfect reconstruction conditions; Shortest trajectories
  - Uniform set better than unions of Uniform sets for $\mathbb{R}^2$
  - Optimal design of parallel trajectory sets for $\mathbb{R}^d$
- Spatial anti-aliasing via time-domain filtering

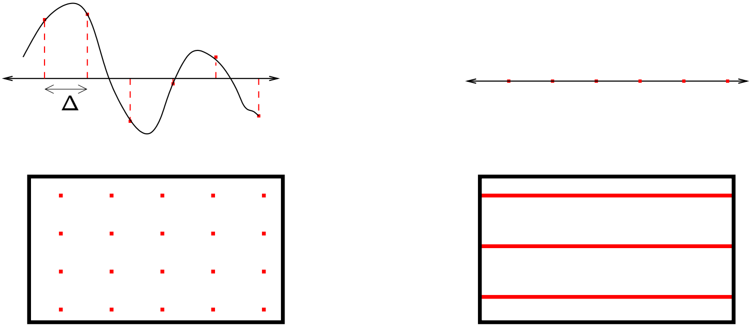
Recap

- Design of sampling trajectories for bandlimited fields
  - Notions of path density and optimal trajectories
- Perfect reconstruction conditions; Shortest trajectories
  - Uniform set better than unions of Uniform sets for $\mathbb{R}^2$
  - Optimal design of parallel trajectory sets for $\mathbb{R}^d$
- Spatial anti-aliasing via time-domain filtering
- Mobility statistics: Simple anonymization is ineffective
  - Optimal de-anonymization strategy can be identified

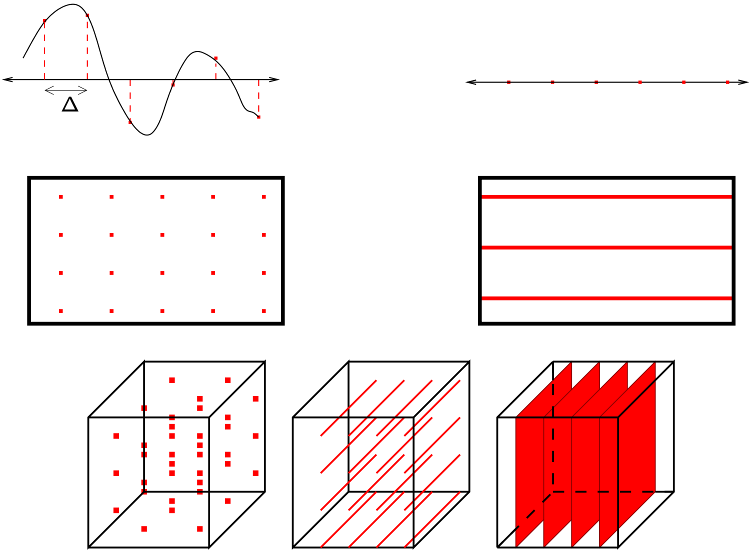
A new sampling theory



A new sampling theory



A new sampling theory



Thank You!

Questions?

Relevant publications

Mobile Sensing

JU, M. Vetterli, "*Sampling High-Dimensional Bandlimited Fields on Low-Dimensional Manifolds*" IEEE Trans. Inf. Theory, 2013

JU, M. Vetterli, "*Sampling and Reconstruction of Spatial Fields using Mobile Sensors*" IEEE Trans. on Signal Process., 2013

JU, M. Vetterli, "*On Optimal Sampling Trajectories for Mobile Sensing* " in Proc. of SampTA, 2013

K. Gröchenig, JU, M. Vetterli, J. L. Romero, "*On Minimal Trajectories for Mobile Sampling of Bandlimited Fields* " in prep., 2013

Privacy of Mobile Traces

JU, F. M. Naini, "*De-anonymizing Private Data by Matching Statistics*" Allerton conference, 2013.