Local Differential Privacy for Physical Sensor Data and Sparse Recovery

Audra McMillan and Anna C. Gilbert

Department of Mathematics
University of Michigan

October 4, 2017
The right to privacy in your own home

It has long been accepted that you have the right to privacy in your own home. Laws protecting this right date back as early as 1604 and in the US today, this right is interpreted via the Fourth Amendment.

Before the information age:
The right to privacy in your own home

It has long been accepted that you have the right to privacy in your own home. Laws protecting this right date back as early as 1604 and in the US today, this right is interpreted via the Fourth Amendment.

Before the information age:

I’d like to know what you are doing inside your house
The right to privacy in your own home

It has long been accepted that you have the right to privacy in your own home. Laws protecting this right date back as early as 1604 and in the US today, this right is interpreted via the Fourth Amendment.

Before the information age:

I'd like to know what you are doing inside your house.

Do you have a warrant or probable cause?
The right to privacy in your own home

It has long been accepted that you have the right to privacy in your own home. Laws protecting this right date back as early as 1604 and in the US today, this right is interpreted via the Fourth Amendment.

Before the information age:

I’d like to know what you are doing inside your house

Do you have a warrant or probable cause?

No
The right to privacy in your own home

It has long been accepted that you have the right to privacy in your own home. Laws protecting this right date back as early as 1604 and in the US today, this right is interpreted via the the Fourth Amendment.

Before the information age:

I’d like to know what you are doing inside your house

Do you have a warrant or probable cause?

No

Then you can’t come in
Motivation

The right to privacy in your own home

It has long been accepted that you have the right to privacy in your own home. Laws protecting this right date back as early as 1604 and in the US today, this right is interpreted via the Fourth Amendment.

Before the information age:

<table>
<thead>
<tr>
<th>Officer</th>
<th>Child</th>
</tr>
</thead>
<tbody>
<tr>
<td>I'd like to know what you are doing inside your house.</td>
<td></td>
</tr>
<tr>
<td>Do you have a warrant or probable cause?</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Then you can't come in</td>
<td></td>
</tr>
<tr>
<td>Okay</td>
<td></td>
</tr>
</tbody>
</table>
Privacy in the Information Age

Things are a lot murkier...
Privacy in the Information Age

Things are a lot murkier...

Smart buildings monitor energy efficiency, but what are they really tracking?

Space utilisation technology aims to make offices more efficient and people friendly but there are concerns around privacy.
Privacy in the Information Age

Things are a lot murkier...

**Guardian sustainable business Smart cities**

Smart buildings monitor energy efficiency, but what are they really tracking?

Space utilisation technology aims to make offices more efficient and people friendly but there are concerns around privacy

**Who’s Watching? Privacy Concerns Persist as Smart Meters Roll Out**

By Christina Nunez, For National Geographic News

Published December 14, 2012
Privacy in the Information Age

Things are a lot murkier...

Guardian sustainable business Smart cities

Smart buildings monitor energy efficiency, but what are they really tracking?

Space utilisation technology aims to make offices more efficient and people friendly but there are concerns around privacy

Who’s Watching? Privacy Concerns Persist as Smart Meters Roll Out

By Christina Nunez, For National Geographic News

PUBLISHED DECEMBER 14, 2012

Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Shared

By MAGGIE ASTOR JULY 25, 2017
Privacy in the Information Age

Things are a lot murkier...

Who’s Watching? Privacy Concerns Persist as Smart Meters Roll Out

By Christina Nunez, For National Geographic News
PUBLISHED DECEMBER 14, 2012

Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Shared

By MAGGIE ASTOR JULY 25, 2017
Privacy in the Information Age

Things are a lot murkier...

Smart buildings monitor energy efficiency, but what are they really tracking?

Who’s Watching? Privacy Concerns Persist as Smart Meters Roll Out

New police radars can 'see' inside homes

Your Roomba May Be Collecting Data That Could Be Shared
Things are a lot murkier...

Who really owns your Internet of Things data?

In a world where more and more objects are coming online and vendors are getting involved in the supply chain, how can you keep track of what's yours and what's not?
Privacy in the Information Age

Things are a lot murkier...

Who really owns your Internet of Things data?
In a world where more and more objects are coming online and vendors are getting involved in the supply chain, how can you keep track of what's yours and what's not?

Key Issues in Perspective:
SMART METERS and DATA PRIVACY

The electric power industry is modernizing the nation’s electric grid. Using advanced technologies, electric companies are building a smart grid that will deliver more reliable power to customers across the country and allow two-way communication between customers and their electric companies.

Installing smart meters is an important step in building the smart grid. These advanced meters enable customers to track their power usage and learn more about the way they use electricity. This will help customers better manage their electricity usage in the future.
Illinois grapples with question of who owns energy data

Key Issues in Perspective:

SMART METERS and DATA PRIVACY

The electric power industry is modernizing the nation’s electric grid. Using advanced technologies, electric companies are building a smart grid that will deliver more reliable power to customers across the country and allow two-way communication between customers and their electric companies.

Installing smart meters is an important step in building the smart grid. These advanced meters enable customers to track their power usage and learn more about the way they use electricity. This will help customers better manage their electricity usage in the future.
Things are a lot murkier...

Illinois grapples with question of who owns energy data

Written by Karl Lydersen
August 28, 2014

Don't miss the latest headlines. Sign up for our Daily Digests.

Key Issue: SMART

The Question of Who Owns the Data Is About to Get a Lot Trickier

Barb Darrow
Apr 06, 2016

The issue of data ownership is about to get a lot more complicated.

Installing smart meters is an important step in building the smart grid. These advanced meters enable customers to track their power usage and learn more about the way they use electricity. This will help customers better manage their electricity usage in the future.
Thermal Data

Thermal data is collected in smart buildings by building managers primarily to control HVAC systems and minimise energy consumption.

The data also inherently contains a lot of private information about the tenants of the building.
Thermal Data

Thermal data is collected in smart buildings by building managers primarily to control HVAC systems and minimise energy consumption 🔄

The data also inherently contains a lot of private information about the tenants of the building ⚠️

Example:

Tenant: Laura

Building manager: Olivia
Thermal data is collected in smart buildings by building managers primarily to control HVAC systems and minimise energy consumption.

The data also inherently contains a lot of private information about the tenants of the building.

Example:

Laura isn't supposed to have a dog and doesn't want the thermal data to reveal that she does.

Tenant: Laura  Building manager: Olivia
Thermal Data

Thermal data is collected in smart buildings by building managers primarily to control HVAC systems and minimise energy consumption.

The data also inherently contains a lot of private information about the tenants of the building.

Example:

Laura isn’t supposed to have a dog and doesn’t want the thermal data to reveal that she does.

Olivia wants to be able to locate heat sources in the building.

Tenant: Laura

Building manager: Olivia
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

The noisy sensor measurements

The recovered heat distribution

How close are these? How much noise is needed to maintain "privacy"? What algorithm should we use to recover? In what way are these close?
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

How close are these?

How much noise is needed to maintain "privacy"?

What algorithm should we use to recover?

In what way are these close?
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

The noisy sensor measurements

How close are these? How much noise is needed to maintain “privacy”? What algorithm should we use to recover? In what way are these close?
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

Each sensor adds noise

Recovery

The recovered heat distribution

The noisy sensor measurements

How close are these?

How much noise is needed to maintain "privacy"?

What algorithm should we use to recover?

In what way are these close?
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

How close are these?

The noisy sensor measurements

Recovery

How much noise is needed to maintain "privacy"?

What algorithm should we use to recover?

In what way are these close? 
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

The noisy sensor measurements

The recovered heat distribution

How close are these?

How much noise is needed to maintain “privacy”?

Recovery

time elapses
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

time elapses

The noisy sensor measurements

The recovered heat distribution

How close are these?

How much noise is needed to maintain “privacy”?

What algorithm should we use to recover?
The Heat Source Problem (1D)

The initial heat distribution

The true sensor measurements

The noisy sensor measurements

The recovered heat distribution

Motivation

In what way are these close?

How much noise is needed to maintain “privacy”?

What algorithm should we use to recover?
Structure:

1. The Heat Source Localisation Problem
2. Differential Privacy
3. General ill-conditioned matrices
4. Basis Pursuit Denoising
5. Error Metric
6. Theoretical Bounds
The Heat Source Problem (1D)

Inverse Problem Set-up

\[ \frac{\partial u}{\partial t} = \mu \nabla^2 u \]

\( u(x, t) \) is the temperature at location \( x \) at time \( t \)

\( f(x) = \) initial temperature (bounded support)

Consider the Cauchy problem for the heat equation

\( u(x, 0) = f(x) \quad \forall x \in \mathbb{R} \) and

\( \lim_{|x| \to \infty} u(x, t) = 0 \quad \forall t \in [0, \infty). \)

Discrete Solution to Inverse Problem

\[ f(x) = \sum_{i=1}^{n} (f_0)_i \delta\left( \frac{i}{n} - x \right) \]

\[ g(x, t) = \frac{1}{\sqrt{4\pi \mu t}} e^{-\frac{|x|^2}{4\mu t}} \]

\[ y = Af_0 \quad \text{where} \quad A_{ij} = g\left( \frac{i}{n} - \frac{j}{m}, t \right). \]

\[ y_i = u\left( \frac{i}{m}, t \right) \]

\( n = \) number of possible source locations

\( m = \) number of sensors
The Heat Source Recovery Problem

- Initial source vector
- True sensor measurements
- Heat kernel matrix
- Private algorithm (noise addition)
- Estimated source vector
- Recovery algorithm

\( f_0 \xrightarrow{A} \{ y_1, \ldots, y_m \} \xrightarrow{\mathcal{B}} \{ \tilde{y}_1, \ldots, \tilde{y}_m \} \xrightarrow{R} \hat{f} \)
“Local” Privacy

Each sensor adds noise to its measurement to ensure privacy before being transmitted to the building manager. There are a number of reasons we consider this setting:

- The data aggregator (Olivia) may be the person we don’t trust.
“Local” Privacy

Each sensor adds noise to its measurement to ensure privacy before being transmitted to the building manager. There are a number of reasons we consider this setting:

- The data aggregator (Olivia) may be the person we don’t trust.
- Thermal sensor measurements are usually noisy already. This noise is often modeled as Gaussian.
“Local” Privacy

Each sensor adds noise to its measurement to ensure privacy before being transmitted to the building manager. There are a number of reasons we consider this setting:

- The data aggregator (Olivia) may be the person we don’t trust.
- Thermal sensor measurements are usually noisy already. This noise is often modeled as Gaussian.
- There is less need to protect the data during transmission if it is already private.
Earth Mover Distance (EMD)

The EMD is a measure of how "geographically similar" two distributions are.
Earth Mover Distance (EMD)

The EMD is a measure of how “geographically similar” two distributions are.

The EMD between two probability distributions is the amount of work (mass × distance) required to transition between the two distributions.

For example, the following two vectors are close in the EMD.
The private algorithm $B$ takes as input a sensor measurement $y_i^{f_0}$ and outputs a noisy (private) version of the sensor measurement.

Recall our goal is two-fold:

- Allow Olivia to estimate approximately where heat sources are in the building.
- Stop Laura from getting into trouble for having a dog.
The private algorithm $B$ takes as input a sensor measurement $y_{i^0}$ and outputs a noisy (private) version of the sensor measurement.

Recall our goal is two-fold:

- **Allow Olivia to estimate approximately where heat sources are in the building.**
- **Stop Laura from getting into trouble for having a dog.**

We can’t hide the existence of a heat source in the building because then the data would be useless to Olivia. Instead we inject enough uncertainty into the *location* of the dog that Laura can’t be identified as the owner.
Two source vectors $f_0$ and $f'_0$ are $\alpha$-adjacent if $\text{EMD}(f_0, f'_0) \leq \alpha$. For example, the following two source vectors are $1/5$-adjacent.

\[
\begin{array}{cccccc}
0 & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 1 \\
\end{array}
\quad
\begin{array}{cccccc}
0 & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 1 \\
\end{array}
\]

A randomised algorithm $B$ is $(\epsilon, \delta, \alpha)$-differentially private if for all $\alpha$-adjacent source vectors $f_0, f'_0$ and events $E$,

\[
P(B(y_i f_0) \in E) \leq e^{\epsilon} P(B(y_i f'_0) \in E) + \delta.
\]

The smaller $\epsilon$ and $\delta$ are, the "more private" the algorithm is.
Intuitively: Suppose Olivia is going to decide whether or not to evict Laura based on the output of the algorithm $B$. If the algorithm $B$ is differentially private then the event that Laura is evicted is almost equally as likely to happen whether the dog lives with Laura or her neighbour.

That is, Olivia can’t tell from the output of the algorithm where exactly the dog is. We say then that we have protected Laura’s privacy.

Remember that the dog was just an example and this statement can be replaced with any information that we can determine about Laura from the thermal data.
How much noise? - Ill-conditionedness of the Heat Source Equation

Each sensor is going to add i.i.d. noise to its true measurement.
How much noise? - Ill-conditionedness of the Heat Source Equation

Each sensor is going to add i.i.d. noise to its true measurement.

The addition of noise makes it difficult to determine between these two settings.

sensor measurements from source at 0.45

sensor measurements from source at 0.5
Each sensor is going to add i.i.d. noise to its true measurement.

The addition of noise makes it difficult to determine between these two settings.

X Bad for exact recovery
How much noise? - Ill-conditionedness of the Heat Source Equation

Each sensor is going to add i.i.d. noise to its true measurement.

The addition of noise makes it difficult to determine between these two settings

X Bad for exact recovery

✔️ Good for privacy
How much noise? - Ill-conditionedness of the Heat Source Equation

Each sensor is going to add i.i.d. noise to its true measurement.

The addition of noise makes it difficult to determine between these two settings.

X Bad for exact recovery

✓ Good for privacy

Intuition: We should not have to add much noise to achieve privacy.
Exactly how much noise - The Gaussian mechanism
(Dwork, Smith, McSherry, Nissim 2006)

Two Gaussian distributions centred at the answer if the dog is in Laura’s apartment, and the answer if the dog is in her neighbours apartment.

Choose the variance so that they look pretty similar.

Privacy: If a sample from one of them is given to Olivia, she can’t tell which distribution it was sampled from.

Utility: Gaussians are light tailed so w.h.p. the sample will be close to the true value (the mean).
How much noise does each sensor need to add?

Restricting to $f_0 \in [0, 1]^n$ we have

$$B(y_{i}^{f_0}) = y_{i}^{f_0} + \frac{2 \log(1.25/\delta) \triangle_2(A)}{\epsilon} N(0, 1)$$

is a $(\epsilon, \delta)$-differentially private algorithm where

$$\triangle_2(A) = \max_{(f_0, f'_0) \text{ adjacent}} \|y_{i}^{f_0} - y_{i}^{f'_0}\|_2$$

$$= \max_{i \in [n]} \|A_i - A_{i+1}\|_2 = O\left(\frac{\sqrt{m\alpha}}{T^{1.5}}\right)$$
How much noise does each sensor need to add?

Restricting to \( f_0 \in [0, 1]^n \) we have

\[
\beta(y_i^{f_0}) = y_i^{f_0} + \frac{2 \log(1.25/\delta) \triangle_2(A)}{\epsilon} N(0, 1)
\]

is a \((\epsilon, \delta)\)-differentially private algorithm where

\[
\triangle_2(A) = \max_{(f_0, f'_0) \text{ adjacent}} \|y_{f_0} - y_{f'_0}\|_2
\]

\[
= \max_{i \in [n]} \|A_i - A_{i+1}\|_2 = O \left( \frac{\sqrt{m\alpha}}{T^{1.5}} \right)
\]

If you want more sensor measurements, you have to pay

If \( \alpha \) is smaller than the adjacency condition is more strict

As more time elapses, the difference between the measurements for difference source vectors decreases
Preserving Privacy for Ill-Conditioned Inverse Problems

A problem is **ill-conditioned** if the relative error of recovery is large. That is,

\[
\kappa(A) = \left( \frac{\|A^{-1} \Delta x\|_2}{\|A^{-1} x\|_2} / \frac{\|\Delta x\|_2}{\|x\|_2} \right) = \frac{\sigma_{\text{max}}(A)}{\sigma_{\text{min}}(A)}
\]
Preserving Privacy for Ill-Conditioned Inverse Problems

A problem is ill-conditioned if the relative error of recovery is large. That is,

$$\kappa(A) = \left( \frac{\| A^{-1} \Delta x \|_2}{\| A^{-1} x \|_2} / \frac{\| \Delta x \|_2}{\| x \|_2} \right) = \frac{\sigma_{\text{max}}(A)}{\sigma_{\text{min}}(A)}$$

We’ve mentioned that the fact that the heat kernel is ill-conditioned means that we only need a small amount of noise to mask the original data. However,

there is a fundamental difference between the notion of a problem being ill-conditioned and being easily kept private.
Preserving Privacy for Ill-Conditioned Inverse Problems

A problem is **ill-conditioned** if the relative error of recovery is large. That is,

\[ \kappa(A) = \left( \frac{\|A^{-1}\triangle x\|_2}{\|A^{-1}x\|_2} \right) \frac{\|\triangle x\|_2}{\|x\|_2} = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \]

We’ve mentioned that the fact that the heat kernel is ill-conditioned means that we only need a small amount of noise to mask the original data. However,

there is a fundamental difference between the notion of a problem being ill-conditioned and being easily kept private.

**Example:** Assume \( \rho << 1 \) and consider the general inverse problem \( y = Mx \) where

\[ M = \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix} \]

Then \( \kappa(M) = 1/\rho \) is large so recovery is “difficult“ but we still need to add a significant amount of noise to mask the first coordinate.
The relationship between $\triangle_2(A)$ and $\kappa(A)$

Recall $\triangle_2(A) = \max_i \|A_i - A_{i+1}\|_2$ is proportional to the amount of noise we need to add to maintain privacy (with the Gaussian mechanism).

If $\triangle_2(A)$ is small then $A$ is “nearly rank 1“, which is to say the spectrum is dominated by the maximum singular value. The condition number $\kappa(A)$ only requires the minimum singular value to be much smaller than the maximum singular value.

If the amount of noise, $\triangle_2(A)$, we need to add to maintain privacy is small then the problem is necessarily ill-conditioned. By the previous example, a matrix may be ill-conditioned but still require a large amount of noise to maintain privacy.

Lemma
Let $M$ be a matrix such that $\|M\|_2 = 1$ and suppose the domain is $[0, 1]^n$, then

$$\triangle_2(A) \geq \frac{1}{\kappa(A) \sqrt{n}}$$
Recovery - Are the noisy measurements useful?

In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

Assume the initial source vector is sparse. That is, $\|f_0\|_0 = k$ where $k$ is small.

This gives us a whole set of tools from the field of sparse signal recovery!

Unfortunately, results in the sparse signal recovery literature often only hold if the operator is "well-behaved", which the heat kernel is not. Surprisingly, a foundational algorithm from that field is still going to work for us! The error is still going to be terrible in $\ell_1$ and $\ell_2$ norms so we have to change our error metric.

This has resulted in a lack of development in theoretical upper bounds (prior to this work and Li et al.).
Recovery - Are the noisy measurements useful?

In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

Assume the initial source vector is sparse. That is, $\|f_0\|_0 = k$ where $k$ is small.
In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

**Assume the initial source vector is sparse.** That is, $\|f_0\|_0 = k$ where $k$ is small.

- This gives us a whole set of tools from the field of sparse signal recovery!
Recovery - Are the noisy measurements useful?

In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

Assume the initial source vector is sparse. That is, $\|f_0\|_0 = k$ where $k$ is small.

- This gives us a whole set of tools from the field of sparse signal recovery!
- Unfortunately, results in the sparse signal recovery literature often only hold if the operator is “well-behaved”, which the heat kernel is not.
Recovery - Are the noisy measurements useful?

In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

Assume the initial source vector is sparse. That is, \( \|f_0\|_0 = k \) where \( k \) is small.

- This gives us a whole set of tools from the field of sparse signal recovery!
- Unfortunately, results in the sparse signal recovery literature often only hold if the operator is “well-behaved”, which the heat kernel is not.
- Surprisingly, a foundational algorithm from that field is still going to work for us!
In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

Assume the initial source vector is sparse. That is, \( \| f_0 \|_0 = k \) where \( k \) is small.

- This gives us a whole set of tools from the field of sparse signal recovery!
- Unfortunately, results in the sparse signal recovery literature often only hold if the operator is “well-behaved”, which the heat kernel is not.
- Surprisingly, a foundational algorithm from that field is still going to work for us!
- The error is still going to be terrible in \( \ell_1 \) and \( \ell_2 \) norms so we have to change our error metric.
Recovery - Are the noisy measurements useful?

In order to make the recovery problem tractable, we are going to introduce an assumption on the initial source vector:

**Assume the initial source vector is sparse.** That is, \( \| f_0 \|_0 = k \) where \( k \) is small.

- This gives us a whole set of tools from the field of sparse signal recovery!
- Unfortunately, results in the sparse signal recovery literature often only hold if the operator is “well-behaved”, which the heat kernel is not.
- Surprisingly, a foundational algorithm from that field is still going to work for us!
- The error is still going to be terrible in \( \ell_1 \) and \( \ell_2 \) norms so we have to change our error metric.

This has resulted in a lack of development in theoretical upper bounds (prior to this work and Li et al.)
The Recovery Algorithm: Basis Pursuit Denoising

If $\tilde{y}$ is the noisy version of our measurement vector $y$ then

$$\hat{f} = \arg \min_{f \in [0,1]^n} \|f\|_1 \quad \text{s.t.} \quad \|Af - \tilde{y}\|_2 \leq \sigma \sqrt{m}$$

the $\ell_1$-norm promotes sparsity in the solution

This ensures that $f_0$ is a feasible point w.h.p.
The Recovery Algorithm: Basis Pursuit Denoising

If \( \tilde{y} \) is the noisy version of our measurement vector \( y \) then

\[
\hat{f} = \arg \min_{f \in [0,1]^n} \|f\|_1 \quad \text{s.t.} \quad \|Af - \tilde{y}\|_2 \leq \sigma \sqrt{m}
\]

- The \( \ell_1 \)-norm promotes sparsity in the solution.
- This ensures that \( f_0 \) is a feasible point w.h.p.
- We do not assume we know the sparsity a priori.
- This algorithm is computationally tractable.
- This algorithm is commonly used for sparse signal recovery when the matrix \( A \) is well-behaved. It is often used when the system is underdetermined (the matrix \( A \) is short and fat).
- The use of this algorithm for the poorly behaved heat matrix was proposed in [Li, Osher and Tsai 2014].
We added enough noise to ensure that there is uncertainty in the exact location of heat source. Thus, by design, our recovery algorithm can not perform well in norms like the $\ell_1$ and $\ell_2$ norms.
The Error Guarantee

We added enough noise to ensure that there is uncertainty in the exact location of heat source. Thus, by design, our recovery algorithm can not perform well in norms like the $\ell_1$ and $\ell_2$ norms.

We injected enough noise that we can’t tell if a heat source moves to an adjacent location. This guarantee does not hold if the heat source moves very far away in the building.
The Error Guarantee

We added enough noise to ensure that there is uncertainty in the exact location of heat source. Thus, by design, our recovery algorithm can not perform well in norms like the $\ell_1$ and $\ell_2$ norms.

We injected enough noise that we can’t tell if a heat source moves to an adjacent location. This guarantee does not hold if the heat source moves very far away in the building.

We should still be able to determine approximately where the heat sources are.
Earth Mover Distance (EMD)

The EMD gives a measure of how “geographically similar“ two distributions are. In the context of the heat source location problem, $\text{EMD}(f_0, \hat{f})$ being small means that even though we may not be able to pinpoint exactly where the heat sources are, we can say *approximately* where they are.
Earth Mover Distance (EMD)

The EMD gives a measure of how “geographically similar“ two distributions are. In the context of the heat source location problem, $\text{EMD}(f_0, \hat{f})$ being small means that even though we may not be able to pinpoint exactly where the heat sources are, we can say approximately where they are.

$\epsilon = 1, \delta = 0.1, \mu t = 0.05, n = 100$ and $m = 50.$
We need to make a few more assumptions to get a good EMD error upper bound:

- There have to be sensors within $\sqrt{2T}$ of each source.
- Measurements have to be taken quickly enough (so not too much heat escapes).
- Heat sources must be sufficiently separated.
- Sparsity.

Diagram of Assumptions
We need to make a few more assumptions to get a good EMD error upper bound:

- There have to be sensors within $\sqrt{2T}$ of each source.
- Measurements have to be taken quickly enough (so not too much heat escapes).
- Heat sources must be sufficiently separated.

Are these assumptions necessary?

- For our analysis, yes.
- For the algorithm to output a good estimate to $f_0$, it's unclear.
Formal Result Statement

Suppose that $f_0$ is a source vector, $\hat{y} = R(\tilde{y})$ and assume the following:

1. $m\sqrt{T/2} > 1$
2. $\sqrt{2T} < 1$
3. $|x_i - x_j| > \sqrt{2T} + 2A$ for some $A > 0$

then w.h.p.

$$\text{EMD} \left( \frac{f_0}{\|f_0\|_1}, \frac{\hat{f}}{\|\hat{f}\|_1} \right) \leq \min \left\{ 1, \ O \left[ \frac{1}{1 - \min\{1, C\}} \left( \frac{1}{k} \sqrt{\frac{T^{1.5}C}{\sqrt{T} + 1}} \right) + k \min\{1, C\} + \frac{T^2C}{(T + 1)k} \right] \right\}$$

where $C = \min \left\{ k, \frac{\sqrt{m\alpha \log(1/\delta)}}{T\epsilon} + \sqrt{T}ke^{-A^2/4T} \right\}$.
Upper Bound broken down by variable:

<table>
<thead>
<tr>
<th>Variable</th>
<th>EMD($\frac{f_0}{|f_0|_1}$, $\frac{\hat{f}}{|\hat{f}|_1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$O(\sqrt{te^{-A^2/4t}} + \sqrt{\alpha + \alpha})$</td>
</tr>
<tr>
<td>$m$</td>
<td>$O(1)$</td>
</tr>
<tr>
<td>$t$</td>
<td>$\min{1, O\left(1 + \frac{1}{t} + t^{2.5}e^{-A^2/4t}\right)}$</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>$O(\sqrt{te^{-A^2/4t}} + \frac{1}{\epsilon})$</td>
</tr>
</tbody>
</table>

Experiments and theoretical lower bounds suggest these are asymptotically tight if we assume $\sqrt{te^{-A^2/4t}}$ is asymptotically smaller than $\sqrt{\alpha}$ and $\frac{1}{\epsilon}$.

If $t$ is small then it’s hard to maintain privacy, if $t$ is large then it’s hard to recover. This bound is consistent with experimental results.

These upper bounds are a generalisation of work of Li, Osher and Tsai 2014 to more than one source.
What have we done?

We add Gaussian noise with standard deviation $O\left(\frac{\log(1/\delta)\alpha\sqrt{m}}{\epsilon T^{1.5}}\right)$ to maintain “privacy”.

Basis pursuit denoising $\hat{f}$ is close to $f_0$ in the EMD.
It is possible to produce locally differentially private thermal sensor measurements from which one can determine the general vicinity of the heat sources but can not infer the exact locations of the heat sources.
Future Directions

- Basis pursuit denoising is unlikely to be the best algorithm for every set of parameters. We’d like to explore what the optimal algorithm may be for different parameter settings.
- Remove sparsity and separability assumptions.
- This problem is related to “learning” means of Gaussian mixture models. Can these tools be applied there?
- Are the measurements of ill-conditioned matrices somehow “morally” private?
Thank you!