Randomization

DS-GA 1013 / MATH-GA 2824 Mathematical Tools for Data Science

https://cims.nyu.edu/~cfgranda/pages/MTDS_spring19/index.html

Carlos Fernandez-Granda

Motivating applications

Gaussian random variables

Randomized dimensionality reduction

Compressed sensing

Dimensionality reduction

Data with a large number of features can be difficult to analyze

Data modeled as vectors in $\mathbb{R}^p$ (p very large)

Aim: Reduce dimensionality of representation

SVD provides optimal subspace for dimensionality reduction

Problem: Computationally expensive + must see dataset beforehand

What if we compute inner products with some random vectors?

Dimensionality reduction for visualization

Motivation: Visualize high-dimensional features projected onto 2D or 3D

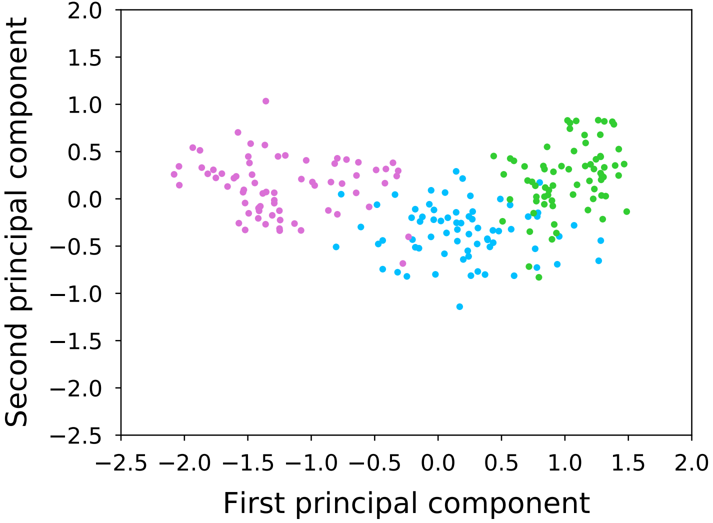
Example:

Seeds from three different varieties of wheat: Kama, Rosa and Canadian

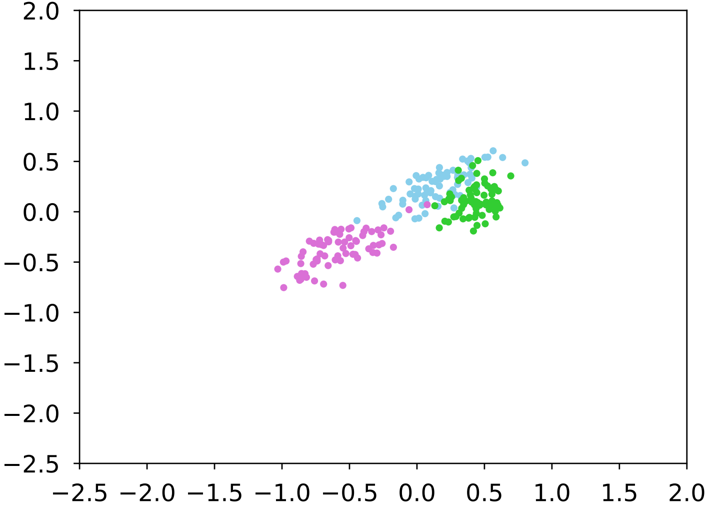
Features:

- Area
- Perimeter
- Compactness
- Length of kernel
- Width of kernel
- Asymmetry coefficient
- Length of kernel groove

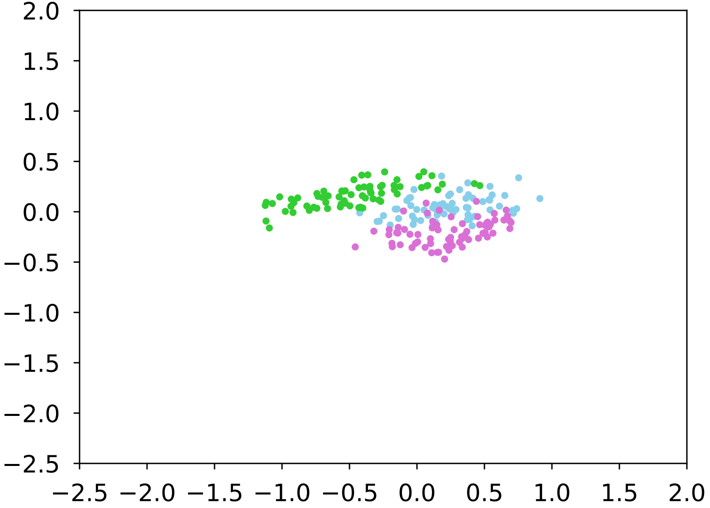
Projection onto two first PDs



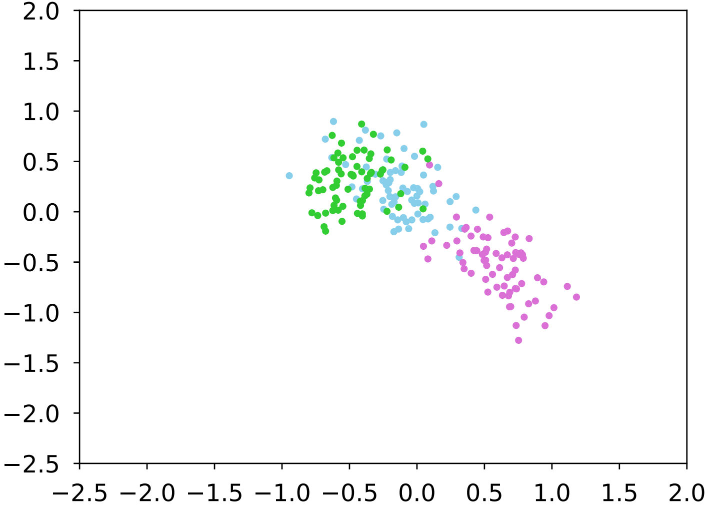
Projection onto two random vectors



Projection onto two random vectors



Projection onto two random vectors



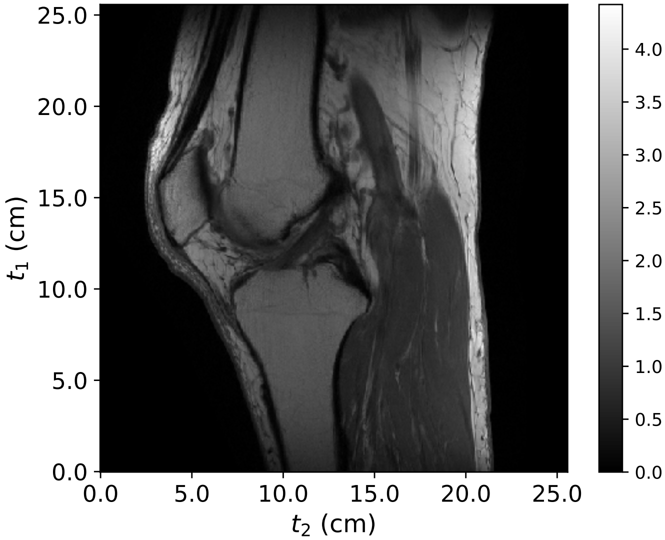
Compressed sensing in MRI

Important goal in MRI: reduce scan time

Can be achieved by measuring less frequency coefficients

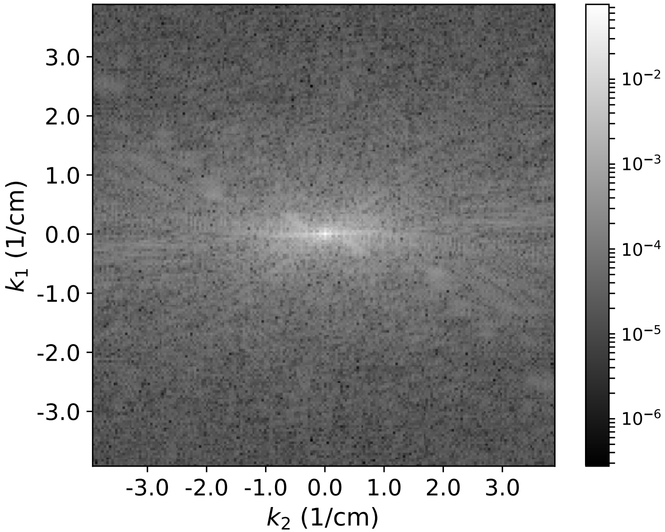
What happens if we undersample in the Fourier domain?

MR image

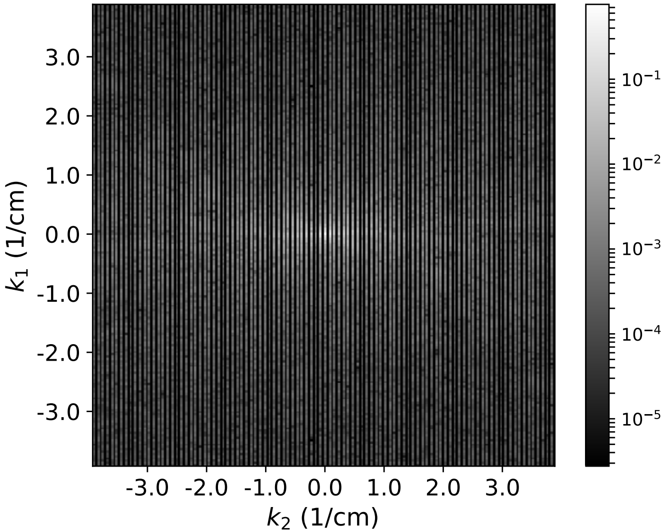


25.0
20.0
15.0
10.0
5.0
0.0
t_1 (cm)
0.0
5.0
10.0
15.0
20.0
25.0
t_2 (cm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0

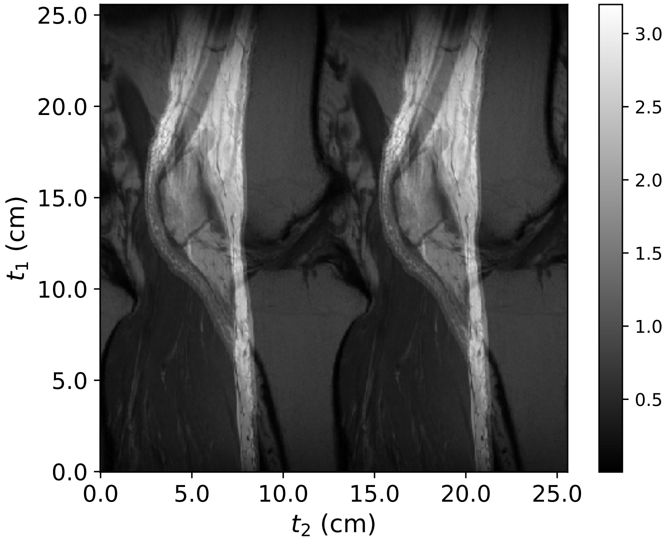
Fourier coefficients



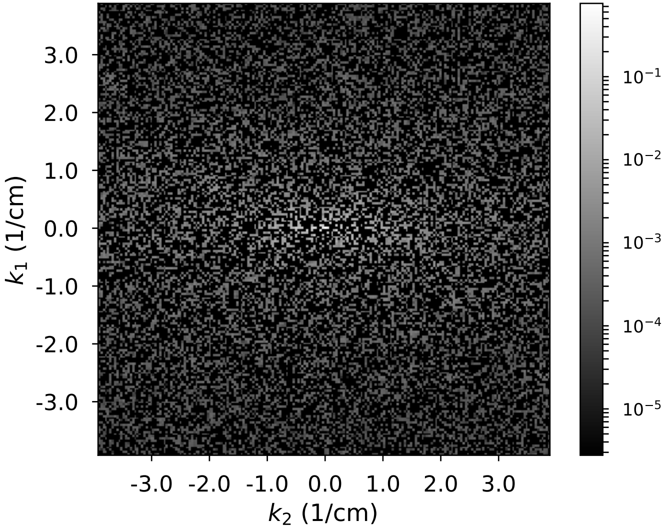
x2 regular undersampling



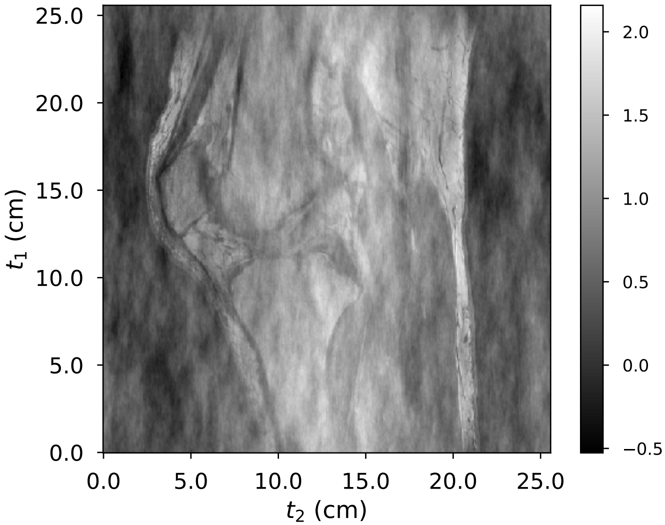
Recovered image



x2 random undersampling



Recovered image



Motivating applications

Gaussian random variables

Randomized dimensionality reduction

Compressed sensing

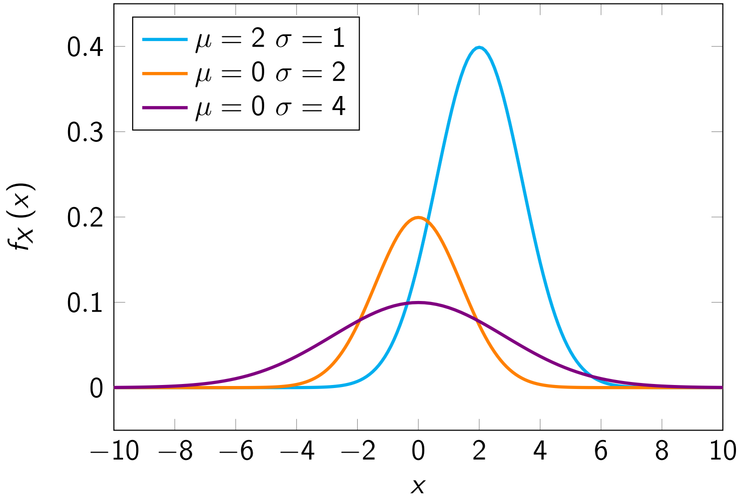
Gaussian random variables

The pdf of a Gaussian or normal random variable with mean μ and standard deviation σ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A standard Gaussian has $\mu := 0$ and $\sigma := 1$

Gaussian random variables



Linear transformation of Gaussian

If $\mathbf{x}$ is a Gaussian random variable with mean μ and standard deviation σ, then for any $a, b \in \mathbb{R}$

$$\mathbf{y} := a\mathbf{x} + b$$

is a Gaussian random variable with mean $a\mu + b$ and standard deviation $|a|\,\sigma$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$F_{\mathbf{y}}(y)$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$F_{\mathbf{y}}(y) = \mathrm{P}(\mathbf{y} \leq y)$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned} F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\ &= \mathrm{P}(a\mathbf{x} + b \leq y) \end{aligned}$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned} F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\ &= \mathrm{P}(a\mathbf{x} + b \leq y) \\ &= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \end{aligned}$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned}
F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\
&= \mathrm{P}(a\mathbf{x} + b \leq y) \\
&= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \\
&= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x
\end{aligned}$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned}
F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\
&= \mathrm{P}(a\mathbf{x} + b \leq y) \\
&= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \\
&= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x \\
&= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi} a\sigma} e^{-\frac{(w-a\mu-b)^2}{2a^2\sigma^2}} \, \mathrm{d}w \qquad \text{change of variables } w = ax + b
\end{aligned}$$

Proof

Let $a > 0$ (proof for $a < 0$ is very similar), to

$$\begin{aligned}
F_{\mathbf{y}}(y) &= \mathrm{P}(\mathbf{y} \leq y) \\
&= \mathrm{P}(a\mathbf{x} + b \leq y) \\
&= \mathrm{P}\left(\mathbf{x} \leq \frac{y-b}{a}\right) \\
&= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x \\
&= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi} a\sigma} e^{-\frac{(w-a\mu-b)^2}{2a^2\sigma^2}} \, \mathrm{d}w \qquad \text{change of variables } w = ax + b
\end{aligned}$$

Differentiating with respect to y:

$$f_{\mathbf{y}}(y) = \frac{1}{\sqrt{2\pi} a\sigma} e^{-\frac{(w-a\mu-b)^2}{2a^2\sigma^2}}$$

Gaussian random vector

A Gaussian random vector $\vec{\mathbf{x}}$ is a random vector with joint pdf

$$f_{\vec{\mathbf{x}}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

where $\vec{\mu} \in \mathbb{R}^d$ is the mean and $\Sigma \in \mathbb{R}^{d \times d}$ the covariance matrix

A standard Gaussian vector has $\vec{\mu} := 0$ and $\Sigma := I$

Uncorrelation implies independence

If the covariance matrix is diagonal,

$$\Sigma_{\vec{\mathbf{x}}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix},$$

the entries are independent

Proof

$$\Sigma_{\vec{\mathbf{x}}}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_d^2} \end{bmatrix}$$

$$|\Sigma| = \prod_{i=1}^{d} \sigma_i^2$$

Proof

$$f_{\vec{\mathbf{x}}}\left(\vec{x}\right)$$

Proof

$$f_{\vec{\mathbf{x}}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right)$$

Proof

$$
\begin{aligned}
f_{\vec{\mathbf{x}}}(\vec{x}) &= \frac{1}{\sqrt{(2\pi)^d\,|\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T\,\Sigma^{-1}(\vec{x}-\vec{\mu})\right) \\
&= \prod_{i=1}^{d} \frac{1}{\sqrt{(2\pi)}\sigma_i} \exp\left(-\frac{(\vec{x}_i-\mu_i)^2}{2\sigma_i^2}\right)
\end{aligned}
$$

Proof

$$
\begin{aligned}
f_{\vec{\mathbf{x}}}\left(\vec{x}\right) &= \frac{1}{\sqrt{\left(2\pi\right)^{d}\left|\Sigma\right|}} \exp\left(-\frac{1}{2}\left(\vec{x}-\vec{\mu}\right)^{T}\Sigma^{-1}\left(\vec{x}-\vec{\mu}\right)\right) \\
&= \prod_{i=1}^{d} \frac{1}{\sqrt{(2\pi)}\sigma_i} \exp\left(-\frac{\left(\vec{x}_i-\mu_i\right)^2}{2\sigma_i^2}\right) \\
&= \prod_{i=1}^{d} f_{\vec{\mathbf{x}}_i}\left(\vec{x}_i\right)
\end{aligned}
$$

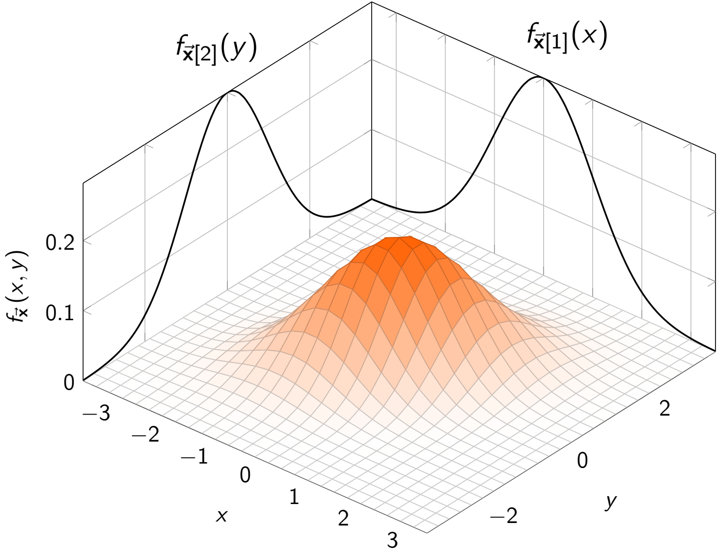
Linear transformations

Let $\vec{\mathbf{x}}$ be a Gaussian random vector of dimension d with mean $\vec{\mu}$ and covariance matrix Σ

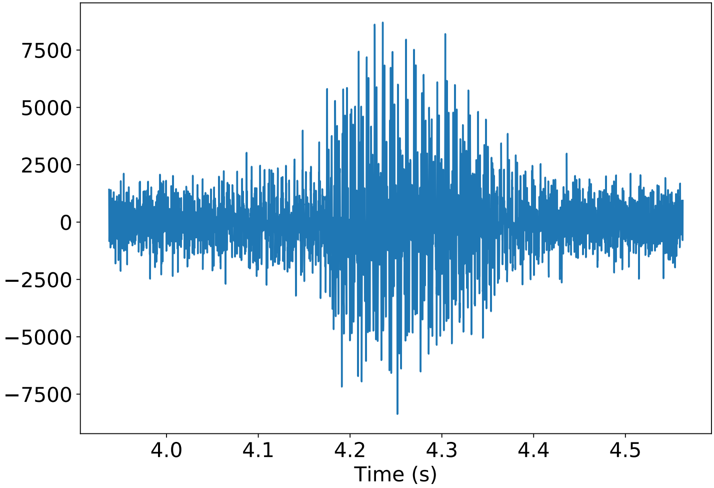
For any matrix $A \in \mathbb{R}^{m \times d}$ and $\vec{b} \in \mathbb{R}^m$ $\vec{\mathbf{y}} = A\vec{\mathbf{x}} + \vec{b}$ is Gaussian with mean $A\vec{\mu} + \vec{b}$ and covariance matrix $A\Sigma A^T$ (as long as it is full rank)

This is why Fourier and wavelet coefficients of Gaussian noise are also Gaussian noise

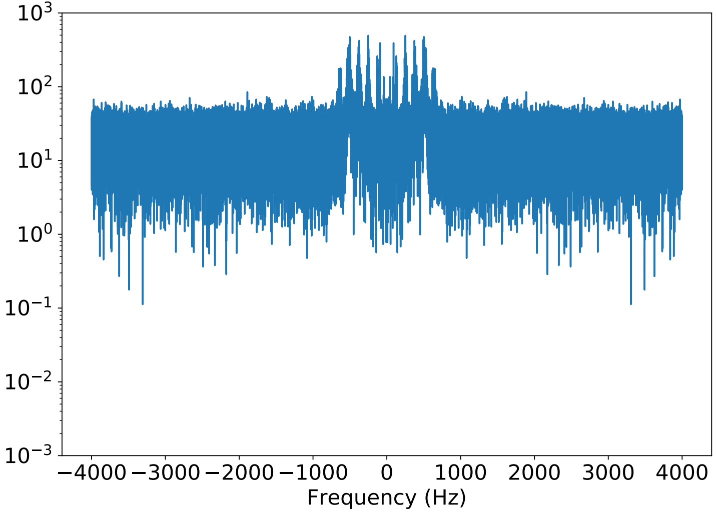
Subvectors are also Gaussian



Audio data

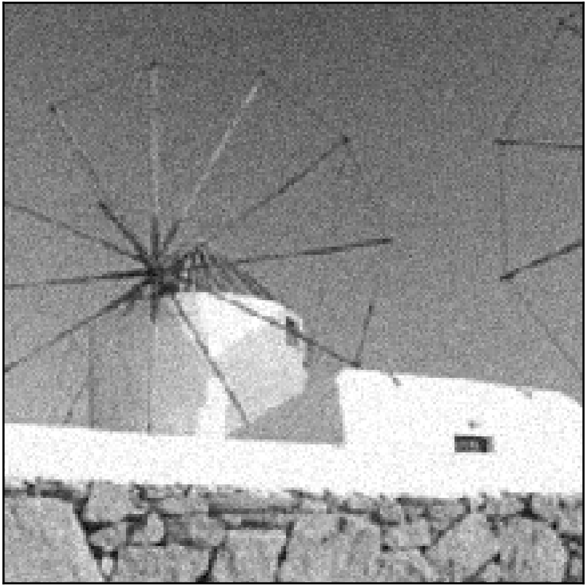


DFT

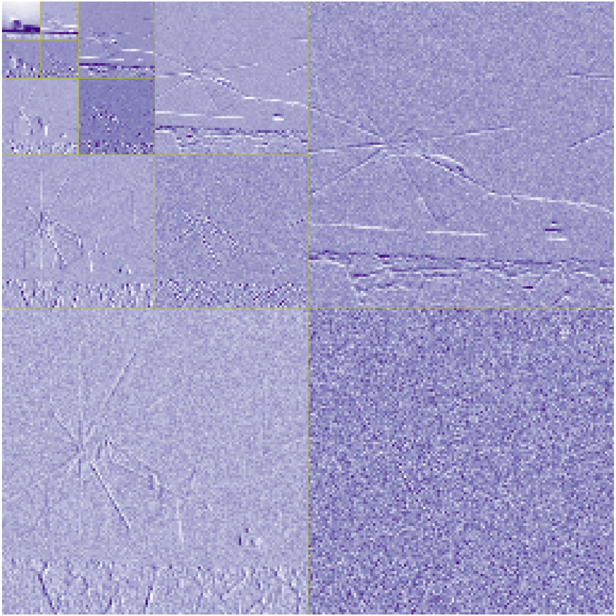


10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
−4000 −3000 −2000 −1000 0 1000 2000 3000 4000
Frequency (Hz)

Noisy image



Wavelet coefficients



Direction of iid standard Gaussian vectors

If the covariance matrix of a Gaussian vector $\vec{x}$ is I, then $\vec{x}$ is isotropic

It does not favor any direction

For any orthogonal matrix $U\vec{x}$ has the same distribution, Gaussian with mean and covariance matrix

Direction of iid standard Gaussian vectors

If the covariance matrix of a Gaussian vector $\vec{x}$ is I, then $\vec{x}$ is isotropic

It does not favor any direction

For any orthogonal matrix $U\vec{x}$ has the same distribution, Gaussian with mean $U\vec{0} = \vec{0}$ and covariance matrix

Direction of iid standard Gaussian vectors

If the covariance matrix of a Gaussian vector $\vec{x}$ is I, then $\vec{x}$ is isotropic

It does not favor any direction

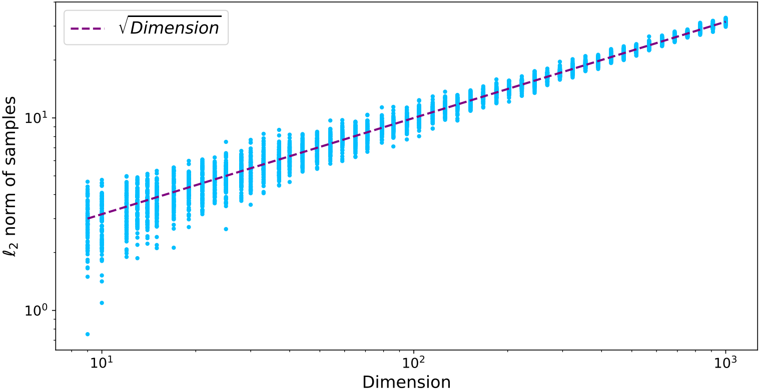
For any orthogonal matrix $U\vec{x}$ has the same distribution, Gaussian with mean $U\vec{0} = \vec{0}$ and covariance matrix $UIU^T = UU^T = I$

Magnitude of iid standard Gaussian vectors

In low dimensions joint pdf is mostly concentrated around the origin

What about in high dimensions?

ℓ_2 norm of samples



χ^2 random variable

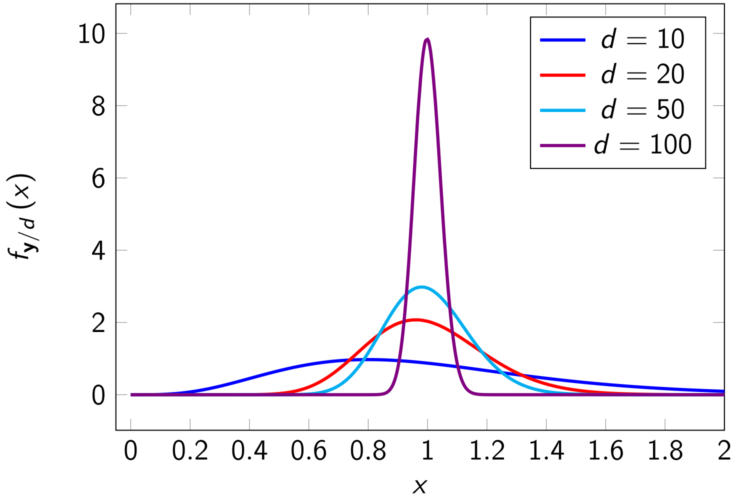
χ^2 (chi squared) random variable with d degrees of freedom

$$\mathbf{y} := \sum_{i=1}^{d} \mathbf{x}_i^2$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_d$ are standard Gaussians

Equal to squared ℓ_2 norm of d-dimensional standard Gaussian vector

Squared ℓ_2 norm divided by d



Mean

$$\mathrm{E}\left(||\vec{\mathbf{x}}||_2^2\right)$$

Mean

$$\mathrm{E}\left(||\vec{x}||_2^2\right) = \mathrm{E}\left(\sum_{i=1}^{d} \vec{x}[i]^2\right)$$

Mean

$$
\begin{aligned}
\mathrm{E}\left(||\vec{x}||_2^2\right) &= \mathrm{E}\left(\sum_{i=1}^{d} \vec{x}[i]^2\right) \\
&= \sum_{i=1}^{d} \mathrm{E}\left(\vec{x}[i]^2\right)
\end{aligned}
$$

Mean

$$
\begin{aligned}
\mathrm{E}\left(||\vec{x}||_2^2\right) &= \mathrm{E}\left(\sum_{i=1}^{d} \vec{x}[i]^2\right) \\
&= \sum_{i=1}^{d} \mathrm{E}\left(\vec{x}[i]^2\right) \\
&= d
\end{aligned}
$$

Variance

$$\mathrm{E}\left(\left(||\vec{\mathbf{x}}||_2^2\right)^2\right)$$

Variance

$$\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) = \mathrm{E}\left(\left(\sum_{i=1}^{d} \vec{x}[i]^2\right)^2\right)$$

Variance

$$\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{d} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{d}\sum_{j=1}^{d} \vec{x}[i]^2 \vec{x}[j]^2\right)
\end{aligned}$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{\mathsf{x}}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{d} \vec{\mathsf{x}}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{d}\sum_{j=1}^{d} \vec{\mathsf{x}}[i]^2 \vec{\mathsf{x}}[j]^2\right) \\
&= \sum_{i=1}^{d}\sum_{j=1}^{d} \mathrm{E}\left(\vec{\mathsf{x}}[i]^2 \vec{\mathsf{x}}[j]^2\right)
\end{aligned}
$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{d} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{d}\sum_{j=1}^{d} \vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{d}\sum_{j=1}^{d} \mathrm{E}\left(\vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{d} \mathrm{E}\left(\vec{x}[i]^4\right) + 2\sum_{i=1}^{d-1}\sum_{j=i+1}^{d} \mathrm{E}\left(\vec{x}[i]^2\right)\mathrm{E}\left(\vec{x}[j]^2\right)
\end{aligned}
$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(\|\vec{x}\|_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{d} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{d}\sum_{j=1}^{d} \vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{d}\sum_{j=1}^{d} \mathrm{E}\left(\vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{d} \mathrm{E}\left(\vec{x}[i]^4\right) + 2\sum_{i=1}^{d-1}\sum_{j=i+1}^{d} \mathrm{E}\left(\vec{x}[i]^2\right)\mathrm{E}\left(\vec{x}[j]^2\right) \\
&= 3d + d(d-1) \quad \text{4th moment of standard Gaussian equals 3}
\end{aligned}
$$

Variance

$$
\begin{aligned}
\mathrm{E}\left(\left(||\vec{x}||_2^2\right)^2\right) &= \mathrm{E}\left(\left(\sum_{i=1}^{d} \vec{x}[i]^2\right)^2\right) \\
&= \mathrm{E}\left(\sum_{i=1}^{d}\sum_{j=1}^{d} \vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{d}\sum_{j=1}^{d} \mathrm{E}\left(\vec{x}[i]^2 \vec{x}[j]^2\right) \\
&= \sum_{i=1}^{d} \mathrm{E}\left(\vec{x}[i]^4\right) + 2\sum_{i=1}^{d-1}\sum_{j=i+1}^{d} \mathrm{E}\left(\vec{x}[i]^2\right)\mathrm{E}\left(\vec{x}[j]^2\right) \\
&= 3d + d(d-1) \quad \text{4th moment of standard Gaussian equals 3} \\
&= d(d+2)
\end{aligned}
$$

Variance

$$\begin{aligned}\mathrm{Var}\left(\|\vec{x}\|_2^2\right) &= \mathrm{E}\left(\left(\|\vec{x}\|_2^2\right)^2\right) - \mathrm{E}\left(\|\vec{x}\|_2^2\right)^2 \\ &= d(d+2) - d^2 = 2d\end{aligned}$$

Relative standard deviation around mean scales as $\sqrt{2/d}$

Geometrically, probability density concentrates close to surface of a sphere with radius $\sqrt{d}$

Non-asymptotic tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension d

For any $\epsilon > 0$

$$P\left(d\left(1-\epsilon\right) < ||\vec{x}||_2^2 < d\left(1+\epsilon\right)\right) \geq 1 - \frac{2}{d\epsilon^2}$$

Markov's inequality

Let x be a nonnegative random variable

For any positive constant $a > 0$,

$$\mathrm{P}\left(\mathsf{x} \geq a\right) \leq \frac{\mathrm{E}\left(\mathsf{x}\right)}{a}$$

Proof

Define the indicator variable $1_{\mathsf{x} \geq a}$

$$\mathsf{x} - a\,1_{\mathsf{x} \geq a} \geq 0$$

Proof

Define the indicator variable $1_{\mathsf{x} \geq a}$

$$\mathsf{x} - a\, 1_{\mathsf{x} \geq a} \geq 0$$

$$\mathrm{E}\left(\mathsf{x}\right) \geq a\, \mathrm{E}\left(1_{\mathsf{x} \geq a}\right) = a\, \mathrm{P}\left(\mathsf{x} \geq a\right)$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\mathrm{P}\left(|\mathbf{y} - d| \geq d\epsilon\right)$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\mathrm{P}\left(|\mathbf{y} - d| \geq d\epsilon\right) = \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq d^2\epsilon^2\right)$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\begin{aligned}
\mathrm{P}\left(|\mathbf{y} - d| \geq d\epsilon\right) &= \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq d^2\epsilon^2\right) \\
&\leq \frac{\mathrm{E}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2\right)}{d^2\epsilon^2} \qquad \text{by Markov's inequality}
\end{aligned}$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$\begin{aligned}
\mathrm{P}\left(|\mathbf{y} - d| \geq d\epsilon\right) &= \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq d^2\epsilon^2\right) \\
&\leq \frac{\mathrm{E}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2\right)}{d^2\epsilon^2} \qquad \text{by Markov's inequality} \\
&= \frac{\mathrm{Var}\left(\mathbf{y}\right)}{d^2\epsilon^2}
\end{aligned}$$

Chebyshev bound

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$,

$$
\begin{aligned}
\mathrm{P}\left(|\mathbf{y} - d| \geq d\epsilon\right) &= \mathrm{P}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2 \geq d^2\epsilon^2\right) \\
&\leq \frac{\mathrm{E}\left(\left(\mathbf{y} - \mathrm{E}\left(\mathbf{y}\right)\right)^2\right)}{d^2\epsilon^2} \qquad \text{by Markov's inequality} \\
&= \frac{\mathrm{Var}\left(\mathbf{y}\right)}{d^2\epsilon^2} \\
&= \frac{2}{d\epsilon^2}
\end{aligned}
$$

Non-asymptotic Chernoff tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension d

For any $\epsilon > 0$

$$P\left(d\left(1-\epsilon\right) < ||\vec{x}||_2^2 < d\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{d\epsilon^2}{8}\right)$$

Proof

Let $\mathbf{y} := ||\vec{\mathbf{x}}||_2^2$. The result is implied by

$$P(\mathbf{y} > d(1+\epsilon)) \leq \exp\left(-\frac{d\epsilon^2}{8}\right)$$

$$P(\mathbf{y} < d(1-\epsilon)) \leq \exp\left(-\frac{d\epsilon^2}{8}\right)$$

Proof

Fix $t > 0$

$P(\mathbf{y} > a)$

Proof

Fix $t > 0$

$$P(\mathbf{y} > a) = P(\exp(t\mathbf{y}) > \exp(at))$$

Proof

Fix $t > 0$

$$\begin{aligned} P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \\ &\leq \exp(-at)\,\mathrm{E}(\exp(t\mathbf{y})) \qquad \text{by Markov's inequality} \end{aligned}$$

Proof

Fix $t > 0$

$$
\begin{aligned}
P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \\
&\leq \exp(-at)\,\mathrm{E}(\exp(t\mathbf{y})) \qquad \text{by Markov's inequality} \\
&\leq \exp(-at)\,\mathrm{E}\left(\exp\left(\sum_{i=1}^{d} t\mathbf{x_i}^2\right)\right)
\end{aligned}
$$

Proof

Fix $t > 0$

$$
\begin{aligned}
P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \\
&\leq \exp(-at)\,\mathrm{E}(\exp(t\mathbf{y})) \qquad \text{by Markov's inequality} \\
&\leq \exp(-at)\,\mathrm{E}\left(\exp\left(\sum_{i=1}^{d} t\mathbf{x_i}^2\right)\right) \\
&\leq \exp(-at)\prod_{i=1}^{d}\mathrm{E}\left(\exp\left(t\mathbf{x_i}^2\right)\right) \quad \text{by independence of } \mathbf{x_1}, \ldots, \mathbf{x_d}
\end{aligned}
$$

Proof

Lemma (by direct integration)

$$\mathrm{E}\left(\exp\left(t\mathsf{x}^2\right)\right) = \frac{1}{\sqrt{1-2t}}$$

Equivalent to controlling higher-order moments since

$$\begin{aligned}
\mathrm{E}\left(\exp\left(t\mathsf{x}^2\right)\right) &= \mathrm{E}\left(\sum_{i=0}^{\infty} \frac{\left(t\mathsf{x}^2\right)^i}{i!}\right) \\
&= \sum_{i=0}^{\infty} \frac{\mathrm{E}\left(t^i\left(\mathsf{x}^{2i}\right)\right)}{i!}.
\end{aligned}$$

Proof

Fix $t > 0$

$$
\begin{aligned}
P(\mathbf{y} > a) &\leq \exp(-at) \prod_{i=1}^{d} \mathrm{E}\left(\exp\left(t\mathbf{x_i}^2\right)\right) \\
&= \frac{\exp(-at)}{(1-2t)^{\frac{d}{2}}}
\end{aligned}
$$

Proof

Setting $a := d\left(1+\epsilon\right)$ and

$$t := \frac{1}{2} - \frac{1}{2\left(1+\epsilon\right)},$$

we conclude

$$\begin{aligned} P\left(\mathbf{y} > d\left(1+\epsilon\right)\right) &\leq \left(1+\epsilon\right)^{d} 2\exp\left(-\frac{d\epsilon}{2}\right) \\ &\leq \exp\left(-\frac{d\epsilon^2}{8}\right) \end{aligned}$$

Projection onto a fixed subspace

Probability density is isotropic and has variance d

Projection onto fixed k-dimensional subspace should capture fraction of variance equal to k/d

Variance of projection should be k

Projection onto a fixed subspace

Let $\mathcal{S}$ be a k-dimensional subspace of $\mathbb{R}^d$ and $\vec{\mathbf{x}}$ a d-dimensional standard Gaussian vector

$\mathcal{P}_{\mathcal{S}}\,(\vec{\mathbf{x}}) = UU^T\vec{\mathbf{x}}$ is not a Gaussian vector

Covariance:

$$\Sigma_{\mathcal{P}_{\mathcal{S}}\,(\vec{\mathbf{x}})} = UU^T\Sigma_{\vec{\mathbf{x}}}UU^T$$

Projection onto a fixed subspace

Let $\mathcal{S}$ be a k-dimensional subspace of $\mathbb{R}^d$ and $\vec{\mathbf{x}}$ a d-dimensional standard Gaussian vector

$\mathcal{P}_{\mathcal{S}}\left(\vec{\mathbf{x}}\right) = UU^T\vec{\mathbf{x}}$ is not a Gaussian vector

Covariance:

$$\begin{aligned}\Sigma_{\mathcal{P}_{\mathcal{S}}\left(\vec{\mathbf{x}}\right)} &= UU^T\Sigma_{\vec{\mathbf{x}}}UU^T \\ &= UU^T\end{aligned}$$

Projection onto a fixed subspace

Let $\mathcal{S}$ be a k-dimensional subspace of $\mathbb{R}^d$ and $\vec{\mathbf{x}}$ a d-dimensional standard Gaussian vector

$\mathcal{P}_{\mathcal{S}}\left(\vec{\mathbf{x}}\right) = UU^T\vec{\mathbf{x}}$ is not a Gaussian vector

Covariance:

$$\begin{aligned}\Sigma_{\mathcal{P}_{\mathcal{S}}\left(\vec{\mathbf{x}}\right)} &= UU^T\Sigma_{\vec{\mathbf{x}}}UU^T \\ &= UU^T\end{aligned}$$

Not full rank

Projection onto a fixed subspace

Coefficients $U^T\vec{\mathbf{x}}$ are a Gaussian vector with covariance

$$\Sigma_{U^T\vec{\mathbf{x}}} = U^T \Sigma_{\vec{\mathbf{x}}} U = U^T U = I$$

Projection onto a fixed subspace

Coefficients $U^T \vec{x}$ are a Gaussian vector with covariance

$$\Sigma_{U^T \vec{x}} = U^T \Sigma_{\vec{x}} U = U^T U = I$$

We have

$$\begin{aligned} \left|\left|\mathcal{P}_{\mathcal{S}}\left(\vec{x}\right)\right|\right|_2^2 &= \left(UU^T \vec{x}\right)^T UU^T \vec{x} \\ &= \left|\left|U^T \vec{x}\right|\right|_2^2 \end{aligned}$$

Projection onto a fixed subspace

Coefficients $U^T\vec{\mathbf{x}}$ are a Gaussian vector with covariance

$$\Sigma_{U^T\vec{\mathbf{x}}} = U^T \Sigma_{\vec{\mathbf{x}}} U = U^T U = I$$

We have

$$\begin{aligned} \left|\left|\mathcal{P}_{\mathcal{S}}\ (\vec{\mathbf{x}})\right|\right|_2^2 &= (UU^T\vec{\mathbf{x}})^T UU^T\vec{\mathbf{x}} \\ &= \left|\left|U^T\vec{\mathbf{x}}\right|\right|_2^2 \end{aligned}$$

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < \left|\left|\mathcal{P}_{\mathcal{S}}\ (\vec{\mathbf{x}})\right|\right|_2^2 < k\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{k\epsilon^2}{8}\right)$$

Linear regression

To analyze the performance of the least-squares estimator we assume a linear model with additive iid Gaussian noise

$$\vec{y}_{\text{train}} := X_{\text{train}} \vec{\beta}_{\text{true}} + \vec{z}_{\text{train}}$$

The LS estimator equals

$$\vec{\beta}_{\text{LS}} := \arg\min_{\vec{\beta}} \; ||\vec{y}_{\text{train}} - X_{\text{train}} \vec{\beta}||_2$$

Training error

The training error is the projection of the noise onto the orthogonal complement of the column space of X_{train}

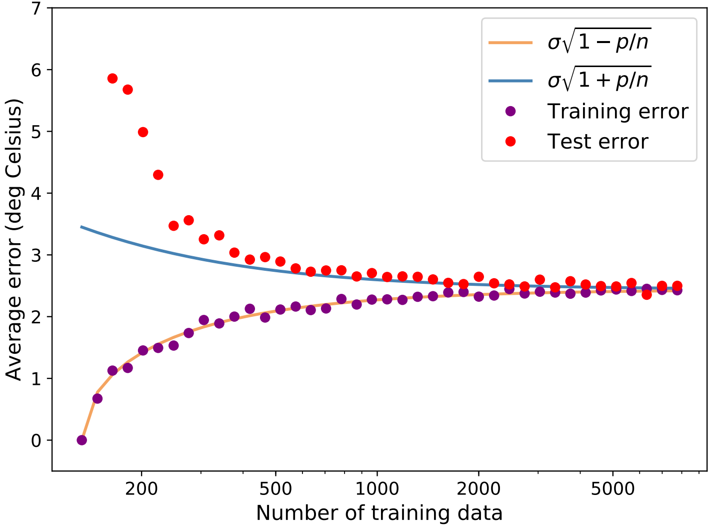
$$\vec{y}_{\text{train}} - \vec{y}_{\text{LS}} = \mathcal{P}_{\text{col}(X_{\text{train}})^{\perp}} \, \vec{z}_{\text{train}}$$

Dimension of orthogonal complement of $\text{col}(X_{\text{train}})$ equals $n - p$

$$\text{Training RMSE} := \sqrt{\frac{||\vec{y}_{\text{train}} - \vec{y}_{\text{LS}}||_2^2}{n}}$$

$$\approx \sigma \sqrt{1 - \frac{p}{n}}$$

Temperature prediction via linear regression



Motivating applications

Gaussian random variables

Randomized dimensionality reduction

Compressed sensing

Randomized linear maps

We use Gaussian matrices as randomized linear maps from $\mathbb{R}^d$ to $\mathbb{R}^k$, $k < d$

Each entry is sampled independently from standard Gaussian

Question: Do we preserve distances between points in set?

Equivalently, are any fixed vectors in the null space?

Fixed vector

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

If $\vec{v} \in \mathbb{R}^d$ is a deterministic vector with unit ℓ_2 norm, then $\mathbf{A}\vec{v}$ is a k-dimensional standard Gaussian vector

Fixed vector

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

If $\vec{v} \in \mathbb{R}^d$ is a deterministic vector with unit ℓ_2 norm, then $\mathbf{A}\vec{v}$ is a k-dimensional standard Gaussian vector

Proof:

$(\mathbf{A}\vec{v})[i]$, $1 \leq i \leq k$ is Gaussian with mean zero and variance

$$\mathrm{Var}\left(\mathbf{A}_{i,:}^T \vec{v}\right) = \vec{v}^T \Sigma_{\mathbf{A}_{i,:}} \vec{v}$$

Fixed vector

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

If $\vec{v} \in \mathbb{R}^d$ is a deterministic vector with unit ℓ_2 norm, then $\mathbf{A}\vec{v}$ is a *k*-dimensional standard Gaussian vector

Proof:

$(\mathbf{A}\vec{v})[i]$, $1 \leq i \leq k$ is Gaussian with mean zero and variance

$$\begin{aligned}
\mathrm{Var}\left(\mathbf{A}_{i,:}^T \vec{v}\right) &= \vec{v}^T \Sigma_{\mathbf{A}_{i,:}} \vec{v} \\
&= \vec{v}^T I \vec{v} \\
&= ||\vec{v}||_2^2 = 1
\end{aligned}$$

Fixed vector

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

If $\vec{v} \in \mathbb{R}^d$ is a deterministic vector with unit ℓ_2 norm, then $\mathbf{A}\vec{v}$ is a k-dimensional standard Gaussian vector

Proof:

$(\mathbf{A}\vec{v})[i]$, $1 \leq i \leq k$ is Gaussian with mean zero and variance

$$\begin{aligned}
\operatorname{Var}\left(\mathbf{A}_{i,:}^T \vec{v}\right) &= \vec{v}^T \Sigma_{\mathbf{A}_{i,:}} \vec{v} \\
&= \vec{v}^T I \vec{v} \\
&= ||\vec{v}||_2^2 = 1
\end{aligned}$$

$\mathbf{A}_{i,:}$, $1 \leq i \leq k$ are all independent

Non-asymptotic Chernoff tail bound

Let $\vec{x}$ be an iid standard Gaussian random vector of dimension k

For any $\epsilon > 0$

$$P\left(k\left(1-\epsilon\right) < ||\vec{x}||_2^2 < k\left(1+\epsilon\right)\right) \geq 1 - 2\exp\left(-\frac{k\epsilon^2}{8}\right)$$

Fixed vector

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

For any $\vec{v} \in \mathbb{R}^d$ with unit norm and any $\epsilon \in (0, 1)$

$$\sqrt{1-\epsilon} \leq \left|\left|\frac{1}{\sqrt{k}}\mathbf{A}\vec{v}\right|\right|_2 \leq \sqrt{1+\epsilon}$$

with probability at least $1 - 2\exp\left(-k\epsilon^2/8\right)$

Distance between two vectors

The result implies that if we fix two vectors $\vec{x}_1$ and $\vec{x}_2$ and define $\vec{y} := \vec{x}_2 - \vec{x}_1$ then

$$\sqrt{1-\epsilon}\,||y||_2 \leq \left|\left|\frac{1}{\sqrt{k}}\mathbf{A}y\right|\right|_2 \leq \sqrt{1+\epsilon}\,||y||_2$$

with high probability (just set $\vec{v} := \vec{y} / \, ||y||_2$)

What about distances between a set of vectors?

Johnson-Lindenstrauss lemma

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

Let $\vec{x}_1, \ldots, \vec{x}_p \in \mathbb{R}^d$ be any fixed set of p deterministic vectors

For any pair $\vec{x}_i, \vec{x}_j$ and any $\epsilon \in (0, 1)$

$$(1-\epsilon) \left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2^2 \leq \left|\left| \frac{1}{\sqrt{k}} \mathbf{A} \vec{x}_i - \frac{1}{\sqrt{k}} \mathbf{A} \vec{x}_j \right|\right|_2^2 \leq (1+\epsilon) \left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2^2$$

with probability at least $\frac{1}{p}$ as long as

$$k \geq \frac{16 \log(p)}{\epsilon^2}$$

Johnson-Lindenstrauss lemma

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

Let $\vec{x}_1, \ldots, \vec{x}_p \in \mathbb{R}^d$ be any fixed set of p deterministic vectors

For any pair $\vec{x}_i, \vec{x}_j$ and any $\epsilon \in (0, 1)$

$$(1-\epsilon) \left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2^2 \leq \left|\left|\frac{1}{\sqrt{k}}\mathbf{A}\vec{x}_i - \frac{1}{\sqrt{k}}\mathbf{A}\vec{x}_j\right|\right|_2^2 \leq (1+\epsilon) \left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2^2$$

with probability at least $\frac{1}{p}$ as long as

$$k \geq \frac{16 \log(p)}{\epsilon^2}$$

No dependence on d!

Proof

Aim: Control action of **A** the normalized differences

$$\vec{v}_{ij} := \frac{\vec{x}_i - \vec{x}_j}{\left|\left|\vec{x}_i - \vec{x}_j\right|\right|_2}$$

Our event of interest is the intersection of the events

$$\mathcal{E}_{ij} = \left\{ k\left(1-\epsilon\right) < \left|\left|\mathbf{A}\vec{v}_{ij}\right|\right|_2^2 < k\left(1+\epsilon\right) \right\} \quad 1 \leq i < p,\ i < j \leq p$$

Proof

Aim: Control action of $\mathbf{A}$ the normalized differences

$$\vec{v}_{ij} := \frac{\vec{x}_i - \vec{x}_j}{||\vec{x}_i - \vec{x}_j||_2}$$

Our event of interest is the intersection of the events

$$\mathcal{E}_{ij} = \left\{ k\left(1-\epsilon\right) < ||\mathbf{A}\vec{v}_{ij}||_2^2 < k\left(1+\epsilon\right) \right\} \quad 1 \leq i < p, \; i < j \leq p$$

Is it equal to $\prod_{i,j} \mathcal{E}_{ij}$?

Fixed vector

Let $\mathbf{A}$ be a $k \times d$ matrix with iid standard Gaussian entries

For any $\vec{v} \in \mathbb{R}^d$ with unit norm and any $\epsilon \in (0, 1)$

$$\sqrt{1-\epsilon} \leq \left|\left|\frac{1}{\sqrt{k}}\mathbf{A}\vec{v}\right|\right|_2 \leq \sqrt{1+\epsilon}$$

with probability at least $1 - 2\exp\left(-k\epsilon^2/8\right)$

This implies

$$\mathrm{P}\left(\mathcal{E}_{ij}^c\right) \leq \frac{2}{p^2} \quad \text{if } k \geq \frac{16\,\log\left(p\right)}{\epsilon^2}$$

Union bound

For any events $S_1, S_2, \ldots, S_n$ in a probability space

$$\mathrm{P}\left(\cup_i S_i\right) \leq \sum_{i=1}^{n} \mathrm{P}\left(S_i\right).$$

Proof

By the union bound

$$\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right)$$

Proof

By the union bound

$$\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) = 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right)$$

Proof

By the union bound

$$\begin{aligned}
\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) &= 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \\
&\geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right)
\end{aligned}$$

Proof

By the union bound

$$\begin{aligned}
\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) &= 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \\
&\geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right)
\end{aligned}$$

Number of events $\mathcal{E}_{ij}$?

Proof

By the union bound

$$
\begin{aligned}
\mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) &= 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \\
&\geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right) \\
&\geq 1 - \frac{p\,(p-1)}{2}\,\frac{2}{p^2}
\end{aligned}
$$

Number of events $\mathcal{E}_{ij}$? $\binom{p}{2} = p\,(p-1)\,/2$

Proof

By the union bound

$$\begin{aligned} \mathrm{P}\left(\bigcap_{i,j} \mathcal{E}_{ij}\right) &= 1 - \mathrm{P}\left(\bigcup_{i,j} \mathcal{E}_{ij}^c\right) \\ &\geq 1 - \sum_{i,j} \mathrm{P}\left(\mathcal{E}_{ij}^c\right) \\ &\geq 1 - \frac{p\left(p-1\right)}{2} \frac{2}{p^2} \\ &\geq \frac{1}{p} \end{aligned}$$

Number of events $\mathcal{E}_{ij}$? $\binom{p}{2} = p\left(p-1\right)/2$

Nearest-neighbor classification

Training set of points and labels $\{\vec{x}_1, l_1\}, \ldots, \{\vec{x}_n, l_n\}$

To classify a new data point $\vec{y} \in \mathbb{R}^d$, find

$$i^* := \arg\min_{1 \leq i \leq n} ||\vec{y} - \vec{x}_i||_2 ,$$

and assign l_{i^*} to $\vec{y}$

Cost: $\mathcal{O}(dnp)$ to classify p new points

Nearest neighbors in random subspace

Use a $k \times d$ iid standard Gaussian matrix to project onto k-dimensional space

Cost:

- dkn operations to project training set
- dkp operations to project test set
- knp to perform nearest-neighbor classification

Much faster!

Face recognition

Training set: 360 64×64 images from 40 different subjects (9 each)

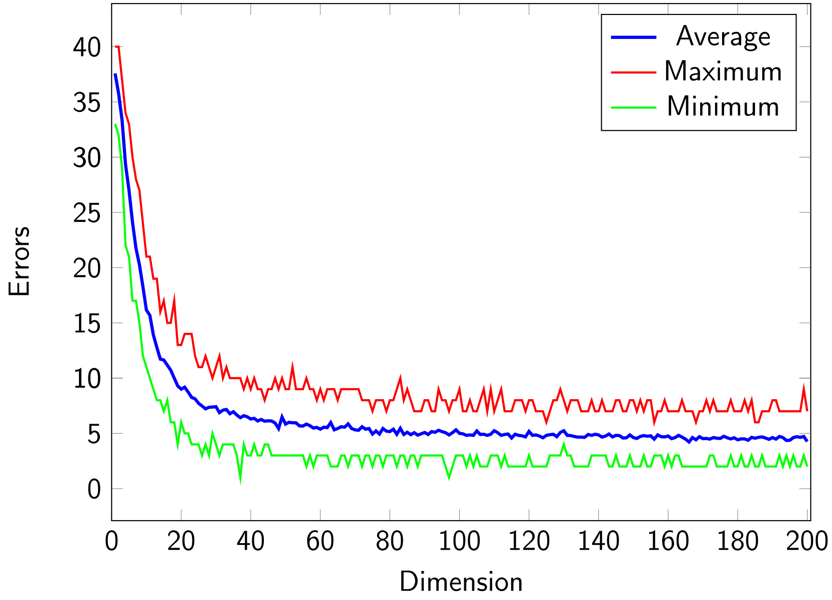
Test set: 1 new image from each subject

We model each image as a vector in $\mathbb{R}^{4096}$ ($d = 4096$)

To classify we:

1. Project onto random a k-dimensional subspace
2. Apply nearest-neighbor classification using the ℓ_2-norm distance in $\mathbb{R}^k$

Performance



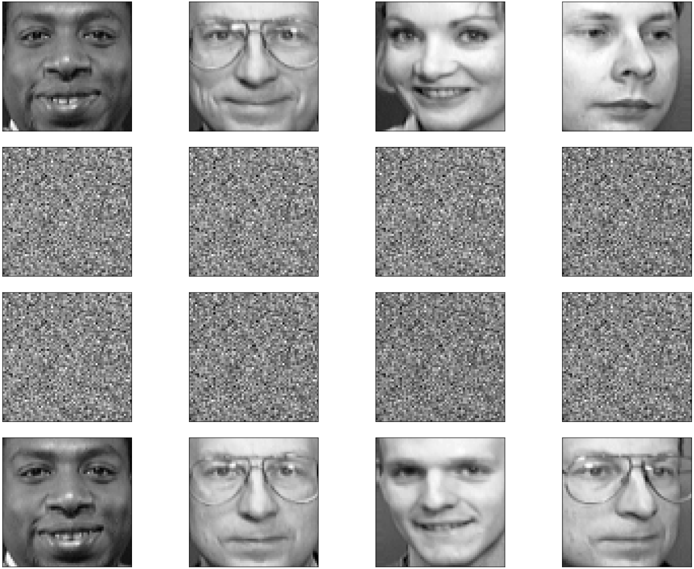
Nearest neighbor in $\mathbb{R}^{50}$

Test image

Projection

Closest projection

Corresponding image



Motivating applications

Gaussian random variables

Randomized dimensionality reduction

Compressed sensing

Compressed sensing

Goal: Recovering signals from small number of data

Arbitrary vector of dimension d cannot be recovered from $m < d$ linear measurements

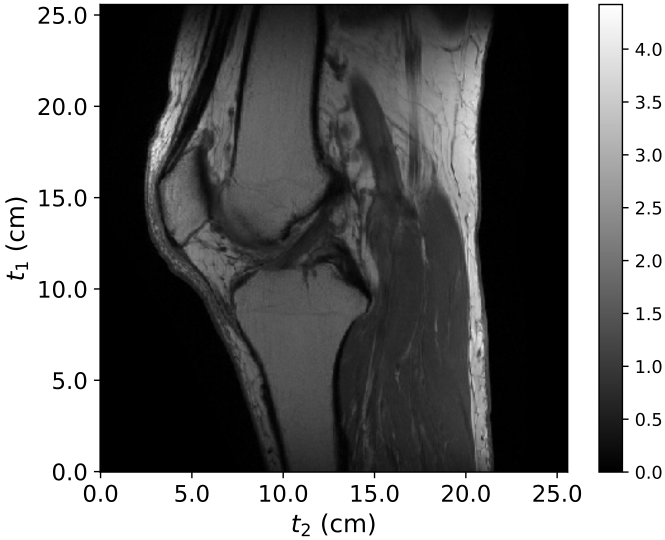
However, signals of interest are highly structured

For example, images are sparse in wavelet basis

If signal is parametrized by $s < m$ parameters, recovery may be possible

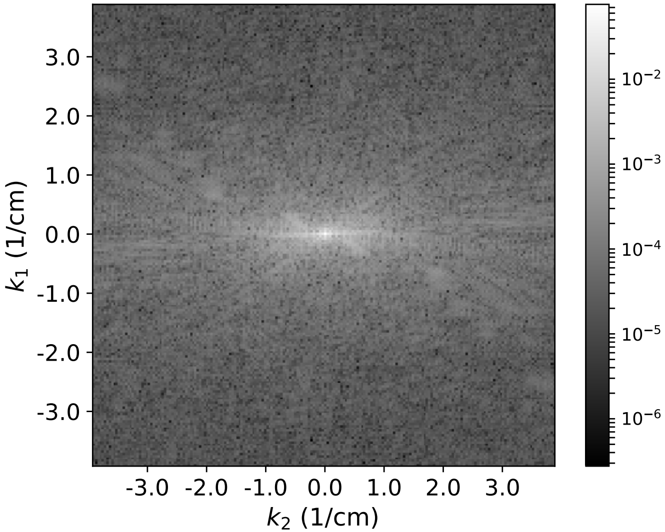
We focus on simplified problem: recovering sparse vectors

MR image

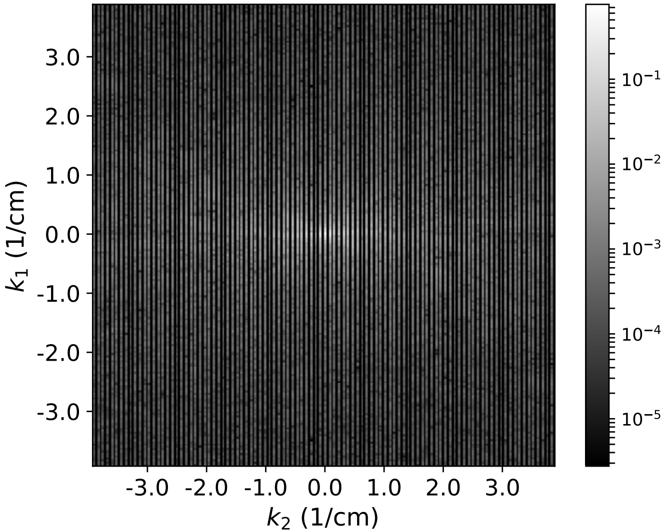


25.0
20.0
15.0
10.0
5.0
0.0
t_1 (cm)
0.0
5.0
10.0
15.0
20.0
25.0
t_2 (cm)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0

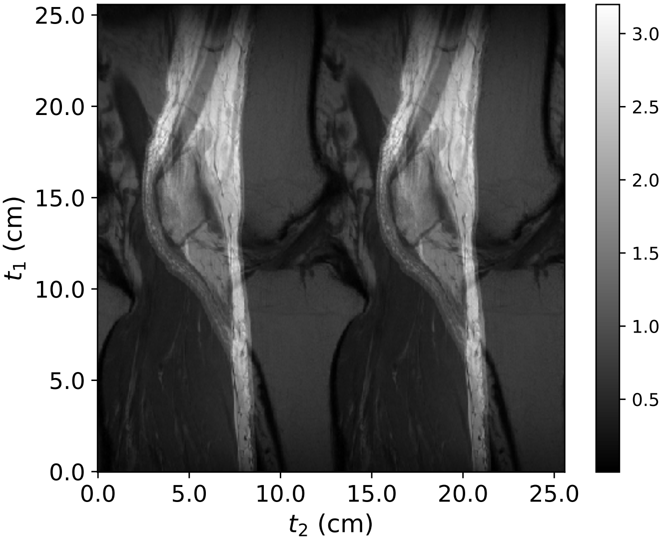
Fourier coefficients



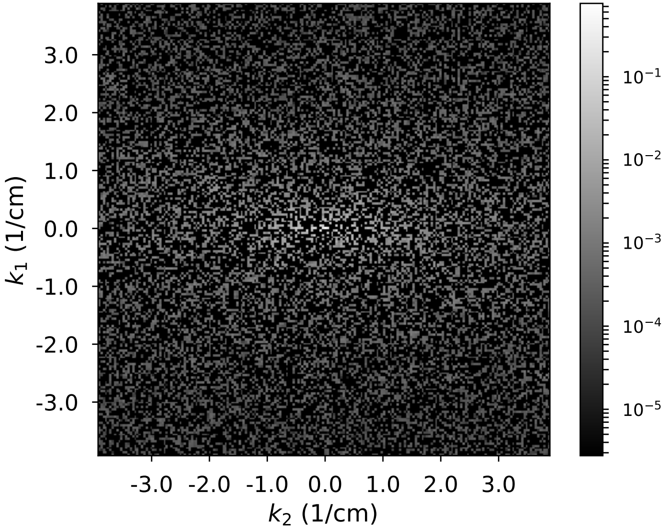
x2 regular undersampling



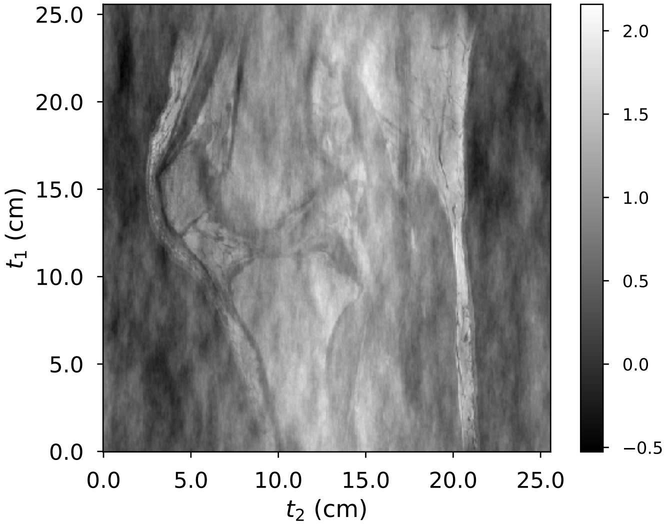
Recovered image



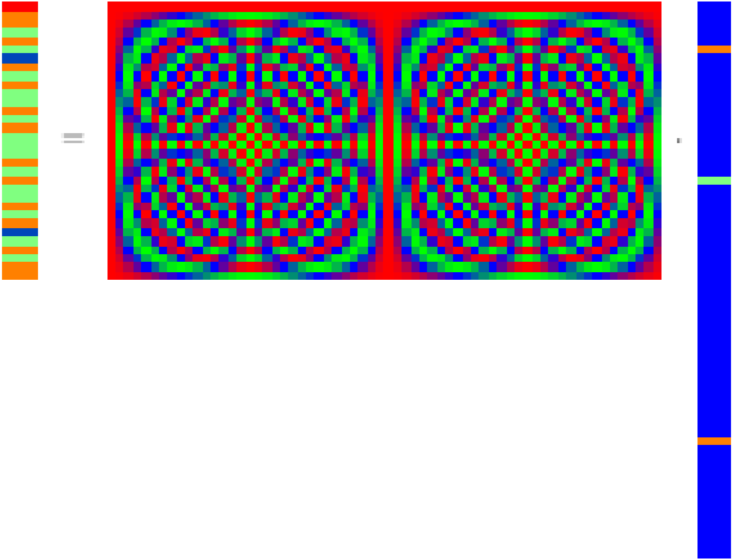
x2 random undersampling



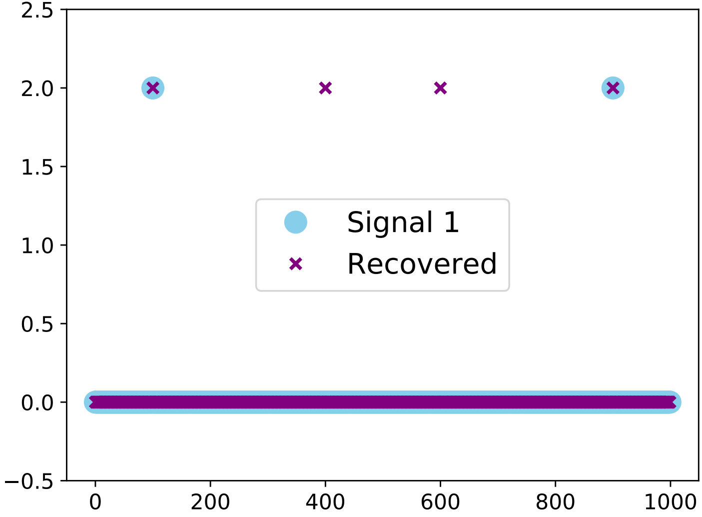
Recovered image



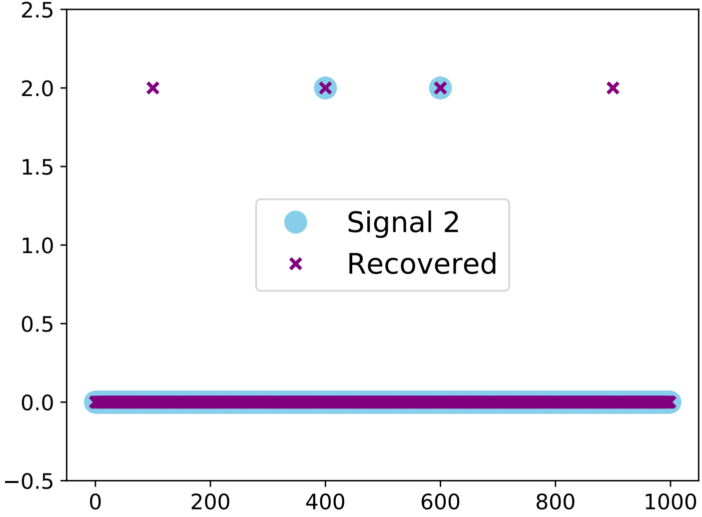
DFT regular undersampling



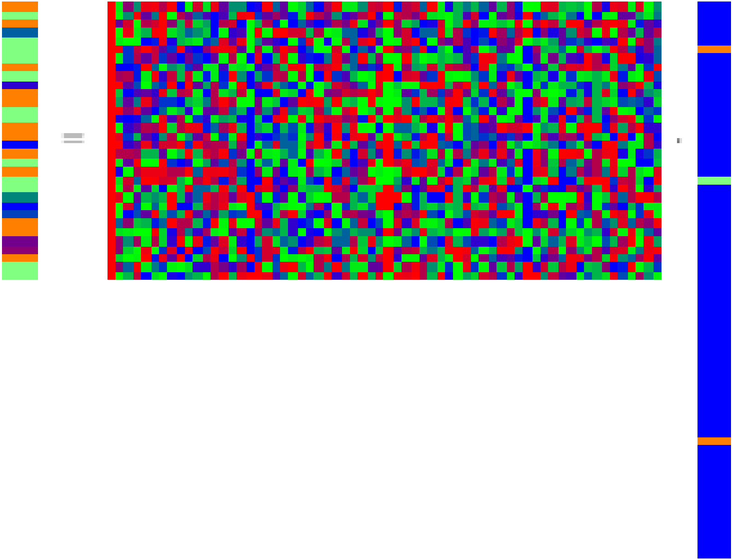
DFT regular undersampling



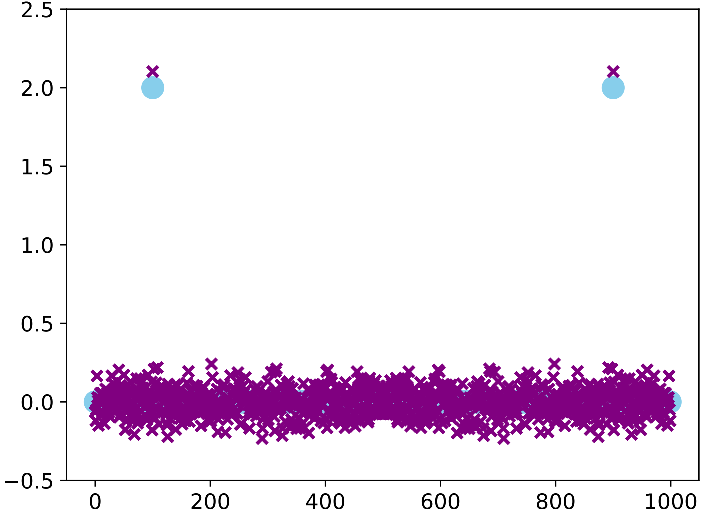
DFT regular undersampling



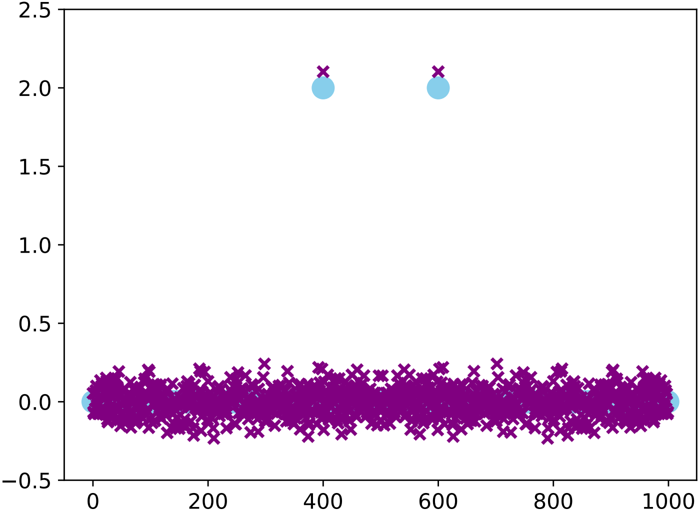
DFT random undersampling



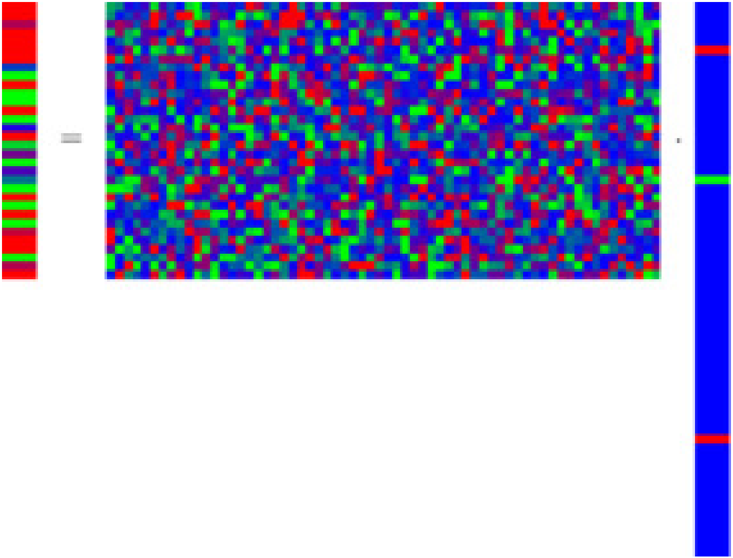
DFT random undersampling



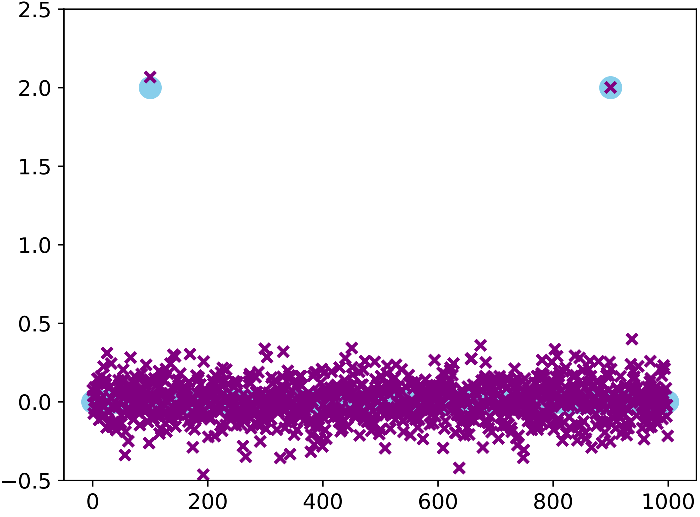
DFT random undersampling



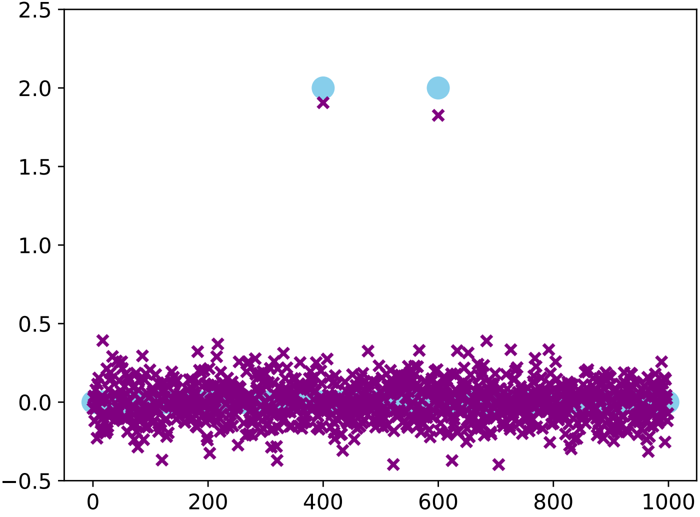
Gaussian measurements



Gaussian measurements



Gaussian measurements



Restricted-isometry property

Different sparse vectors should never produce similar data

If two s-sparse vectors $\vec{x}_1$, $\vec{x}_2$ are far, then $A\vec{x}_1$, $A\vec{x}_2$ should be far

The measurement operator should preserve distances (be an isometry) when restricted to act upon sparse vectors

Restricted-isometry property

A satisfies the restricted isometry property (RIP) with constant ϵ if

$$(1-\epsilon)\,||\vec{x}||_2 \leq ||A\vec{x}||_2 \leq (1+\epsilon)\,||\vec{x}||_2$$

for any s-sparse vector $\vec{x}$

Restricted-isometry property

A satisfies the restricted isometry property (RIP) with constant ϵ if

$$(1-\epsilon)\,||\vec{x}||_2 \leq ||A\vec{x}||_2 \leq (1+\epsilon)\,||\vec{x}||_2$$

for any s-sparse vector $\vec{x}$

If A satisfies the RIP for a sparsity level $2s$ then for any s-sparse $\vec{x}_1$, $\vec{x}_2$

$$||\vec{x}_2 - \vec{x}_1||_2$$

Restricted-isometry property

A satisfies the restricted isometry property (RIP) with constant ϵ if

$$(1-\epsilon)\,||\vec{x}||_2 \leq ||A\vec{x}||_2 \leq (1+\epsilon)\,||\vec{x}||_2$$

for any s-sparse vector $\vec{x}$

If A satisfies the RIP for a sparsity level $2s$ then for any s-sparse $\vec{x}_1$, $\vec{x}_2$

$$\begin{aligned} ||\vec{x}_2 - \vec{x}_1||_2 &= ||A\,(\vec{x}_1 - \vec{x}_2)||_2 \\ &\geq (1-\epsilon)\,||\vec{x}_2 - \vec{x}_1||_2 \end{aligned}$$

Restricted-isometry property

Deterministic matrices tend to not satisfy the RIP

It is NP-hard to check if spark or RIP hold

Random matrices satisfy RIP with high probability

We prove it for Gaussian iid matrices, ideas in proof for random Fourier matrices are similar

Restricted-isometry property for Gaussian matrices

Let $\mathbf{A} \in \mathbb{R}^{m \times d}$ be a random matrix with iid standard Gaussian entries

$\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies the RIP for a constant ϵ with probability $1 - \frac{C_2}{d}$ as long as

$$m \geq \frac{C_1 s}{\epsilon^2} \log\left(\frac{d}{s}\right)$$

for two fixed constants $C_1, C_2 > 0$

Restricted-isometry property for Gaussian matrices

Let $\mathbf{A} \in \mathbb{R}^{m \times d}$ be a random matrix with iid standard Gaussian entries

$\frac{1}{\sqrt{m}}\mathbf{A}$ satisfies the RIP for a constant ϵ with probability $1 - \frac{C_2}{d}$ as long as

$$m \geq \frac{C_1 s}{\epsilon^2} \log\left(\frac{d}{s}\right)$$

for two fixed constants $C_1, C_2 > 0$

Measurements proportional to sparsity (up to log factor)

Singular values of submatrix

Fix subset of s indices $T \subset \{1, \ldots, d\}$

Any matrix $A \in \mathbb{R}^{m \times d}$, $m < d$, satisfies

$$\sigma_s(A_T) \leq ||A\vec{x}||_2 \leq \sigma_1(A_T)$$

for all vectors $\vec{x} \in \mathbb{R}^d$ with support restricted to T

A_T is the $m \times s$ submatrix of A containing columns indexed by T

$\sigma_1(A_T)$ and $\sigma_s(A_T)$ are the largest and smallest singular value of A_T

Proof

For any vector $\vec{x} \in \mathbb{R}^d$ with support restricted to T

$$A\vec{x} = A_T \vec{x}_T$$

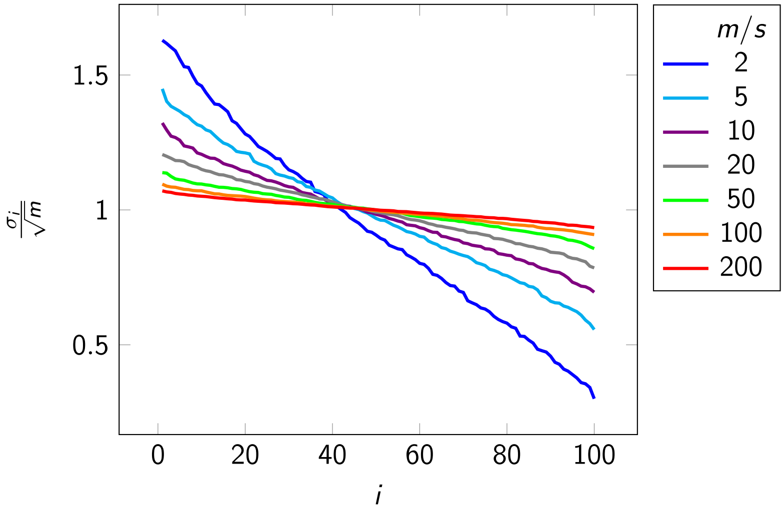
where $\vec{x}_T \in \mathbb{R}^s$ is the subvector of $\vec{x}$ that contains its nonzero entries

Proof strategy

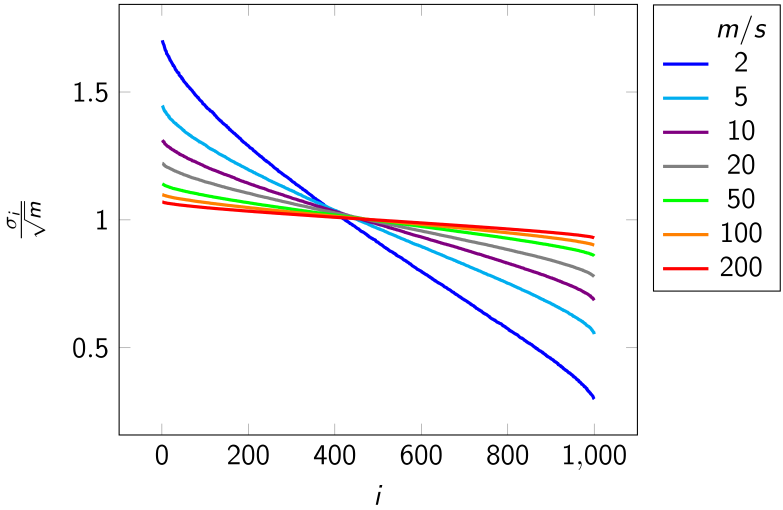
Control singular values for fixed submatrix

Apply union bound to extend bounds to all submatrices

Singular values of $m \times s$ matrix, $s = 100$



Singular values of $m \times s$ matrix, $s = 1000$



Singular values of a Gaussian matrix

For large enough m

$$\mathbf{M} \approx U \left(\sqrt{m}\, I \right) V^T = \sqrt{m}\, U V^T,$$

Standard Gaussian vectors in high dimensions are *almost* orthogonal

Singular values of a Gaussian matrix

Let $\mathbf{M}$ be a $m \times s$ matrix with iid standard Gaussian entries such that $m > s$

For any fixed $\epsilon > 0$, the singular values of $\mathbf{M}$ satisfy

$$\sqrt{m\,(1-\epsilon)} \leq \boldsymbol{\sigma}_{\mathbf{s}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq \sqrt{m\,(1+\epsilon)}$$

with probability at least $1 - 2\left(\frac{12}{\epsilon}\right)^{s} \exp\left(-\frac{m\epsilon^2}{32}\right)$

Union bound

For any events $S_1, S_2, \ldots, S_n$ in a probability space

$$\mathrm{P}\left(\cup_i S_i\right) \leq \sum_{i=1}^{n} \mathrm{P}\left(S_i\right).$$

Proof

Number of different supports of size s

Proof

Number of different supports of size s

$$\binom{d}{s} \leq \left(\frac{ed}{s}\right)^s$$

Proof

Number of different supports of size s

$$\binom{d}{s} \leq \left(\frac{ed}{s}\right)^s$$

By the union bound

$$\sqrt{1-\epsilon}\,||\vec{x}||_2 \leq \frac{1}{\sqrt{m}}\,||\mathbf{A}\vec{x}||_2 \leq \sqrt{1+\epsilon}\,||\vec{x}||_2$$

holds for any s-sparse vector $\vec{x}$ with probability at least

$$\begin{aligned}
&1-2\left(\frac{ed}{s}\right)^s\left(\frac{12}{\epsilon}\right)^s \exp\left(-\frac{m\epsilon^2}{32}\right)\\
&= 1-\exp\left(\log 2 + s + s\log\left(\frac{d}{s}\right) + s\log\left(\frac{12}{\epsilon}\right) - \frac{m\epsilon^2}{2}\right)\\
&\leq 1-\frac{C_2}{d} \qquad \text{as long as} \quad m \geq \frac{C_1 s}{\epsilon^2}\log\left(\frac{d}{s}\right)
\end{aligned}$$

Singular values of a Gaussian matrix

Let $\mathbf{M}$ be a $m \times s$ matrix with iid standard Gaussian entries such that $m > s$

For any fixed $\epsilon > 0$, the singular values of $\mathbf{M}$ satisfy

$$\sqrt{m(1-\epsilon)} \leq \boldsymbol{\sigma}_{\mathbf{s}} \leq \boldsymbol{\sigma}_{\mathbf{1}} \leq \sqrt{m(1+\epsilon)}$$

with probability at least $1 - 2\left(\frac{12}{\epsilon}\right)^s \exp\left(-\frac{m\epsilon^2}{32}\right)$

How do we prove this?

More of the same?

We need to prove that for any vector $\vec{v}$ of the s-dimensional sphere $\mathcal{S}^{s-1}$ in $\mathbb{R}^s$

$$\sqrt{m}\,(1-\epsilon) < ||\mathbf{M}\vec{v}||_2 < \sqrt{m}\,(1+\epsilon)$$

Can we prove it for a fixed vector and use the union bound?

Proof strategy

1. Consider *spread-out* finite subset $\mathcal{N}_\epsilon \subset \mathcal{S}^{s-1}$ such that any point in $\mathcal{S}^{s-1}$ is close to a point in $\mathcal{N}_\epsilon$

2. Prove bound on $\mathcal{N}_\epsilon$

3. Show that bounds hold for all points that are close to $\mathcal{N}_\epsilon$

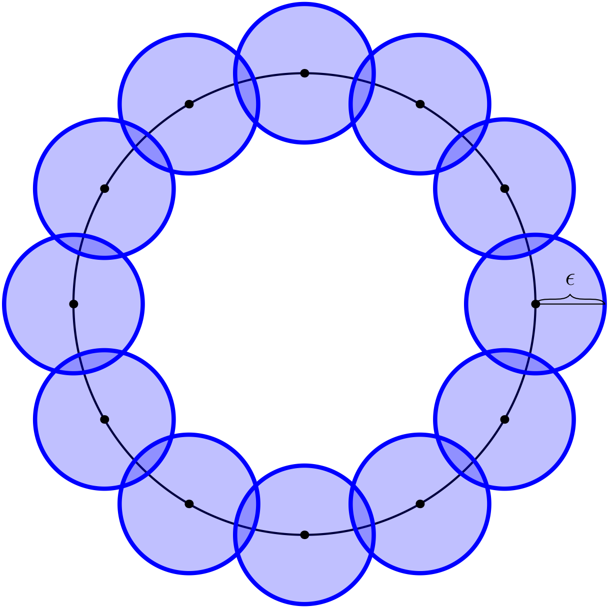
ϵ-net

An ϵ-net of a set $\mathcal{X} \subseteq \mathbb{R}^s$ is a subset $\mathcal{N}_\epsilon \subseteq \mathcal{X}$ such that for every vector $\vec{x} \in \mathcal{X}$ there exists $\vec{y} \in \mathcal{N}_\epsilon$ for which

$$||\vec{x} - \vec{y}||_2 \leq \epsilon.$$

The covering number $\mathcal{N}(\mathcal{X}, \epsilon)$ of a set $\mathcal{X}$ at scale ϵ is the minimal cardinality of an ϵ-net of $\mathcal{X}$

ϵ-net



Covering number of a sphere

The covering number of the s-dimensional sphere $\mathcal{S}^{s-1}$ at scale ϵ satisfies

$$\mathcal{N}\left(\mathcal{S}^{s-1}, \epsilon\right) \leq \left(\frac{2+\epsilon}{\epsilon}\right)^{s} \leq \left(\frac{3}{\epsilon}\right)^{s}$$

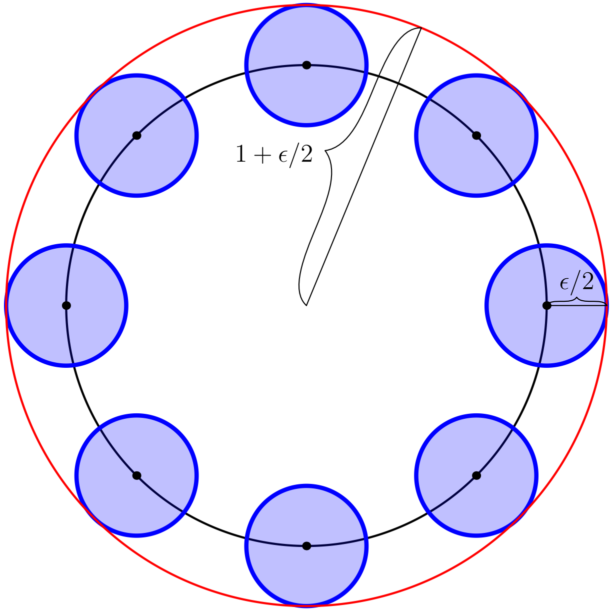
Covering number of a sphere

- Initialize $\mathcal{N}_\epsilon$ to the empty set
- Choose a point $\vec{x} \in \mathcal{S}^{s-1}$ such that

$$||\vec{x} - \vec{y}||_2 > \epsilon \quad \text{for any } \vec{y} \in \mathcal{N}_\epsilon$$

- Add $\vec{x}$ to $\mathcal{N}_\epsilon$ until there are no points in $\mathcal{S}^{s-1}$ that are ϵ away from any point in $\mathcal{N}_\epsilon$

Covering number of a sphere



Covering number of a sphere

$$\mathrm{Vol}\left(\mathcal{B}^s_{1+\epsilon/2}\left(\vec{0}\right)\right) \geq \mathrm{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^s_{\epsilon/2}\left(\vec{x}\right)\right)$$

Covering number of a sphere

$$\begin{aligned}
\mathrm{Vol}\left(\mathcal{B}^{s}_{1+\epsilon/2}\left(\vec{0}\right)\right) &\geq \mathrm{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^{s}_{\epsilon/2}\left(\vec{x}\right)\right) \\
&= \left|\mathcal{N}_\epsilon\right| \mathrm{Vol}\left(\mathcal{B}^{s}_{\epsilon/2}\left(\vec{0}\right)\right)
\end{aligned}$$

Covering number of a sphere

$$
\begin{aligned}
\operatorname{Vol}\left(\mathcal{B}^s_{1+\epsilon/2}\left(\vec{0}\right)\right) &\geq \operatorname{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^s_{\epsilon/2}\left(\vec{x}\right)\right) \\
&= \left|\mathcal{N}_\epsilon\right| \operatorname{Vol}\left(\mathcal{B}^s_{\epsilon/2}\left(\vec{0}\right)\right)
\end{aligned}
$$

By multivariable calculus

$$
\operatorname{Vol}\left(\mathcal{B}^s_r\left(\vec{0}\right)\right) = r^s \operatorname{Vol}\left(\mathcal{B}^s_1\left(\vec{0}\right)\right)
$$

Covering number of a sphere

$$\begin{aligned}\mathrm{Vol}\left(\mathcal{B}^s_{1+\epsilon/2}\left(\vec{0}\right)\right) &\geq \mathrm{Vol}\left(\cup_{\vec{x}\in\mathcal{N}_\epsilon}\mathcal{B}^s_{\epsilon/2}\left(\vec{x}\right)\right)\\ &= \left|\mathcal{N}_\epsilon\right| \mathrm{Vol}\left(\mathcal{B}^s_{\epsilon/2}\left(\vec{0}\right)\right)\end{aligned}$$

$$\mathrm{Vol}\left(\mathcal{B}^s_{r}\left(\vec{0}\right)\right) = r^s\,\mathrm{Vol}\left(\mathcal{B}^s_{1}\left(\vec{0}\right)\right)$$

so we conclude

$$\left(1+\epsilon/2\right)^s \geq \left|\mathcal{N}_\epsilon\right| \left(\epsilon/2\right)^s$$

Proof

1. We prove the bounds

$$n\left(1-\epsilon_2\right) < ||\mathbf{M}\vec{v}||_2^2 < n\left(1+\epsilon_2\right)$$

   where $\epsilon_2 := \epsilon/2$ on an $\epsilon_1 := \epsilon/4$ net of the sphere

2. We show that by the triangle inequality, this implies that the bounds hold on all the sphere

Fixed vector

Let $\mathbf{M}$ be a $a \times b$ matrix with iid standard Gaussian entries

For any $\vec{v} \in \mathbb{R}^b$ with unit norm and any $\epsilon \in (0, 1)$

$$\sqrt{a\,(1-\epsilon)} \leq ||\mathbf{M}\vec{v}||_2 \leq \sqrt{a\,(1+\epsilon)}$$

with probability at least $1 - 2\exp\left(-a\epsilon^2/8\right)$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ m\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{M}\vec{v}||_2^2 \leq m\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^c_{\vec{v},\epsilon_2}\right)$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ m\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{M}\vec{v}||_2^2 \leq m\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2}\right) \leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right)$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ m\left(1-\epsilon_2\right)||\vec{v}||_2^2 \leq ||\mathbf{M}\vec{v}||_2^2 \leq m\left(1+\epsilon_2\right)||\vec{v}||_2^2 \right\}$$

$$\begin{aligned} \mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^c_{\vec{v},\epsilon_2}\right) &\leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\ &\leq \left|\mathcal{N}_{\epsilon_1}\right| \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \end{aligned}$$

Bound on the ϵ_1-net

We define the event

$$\mathcal{E}_{\vec{v},\epsilon_2} := \left\{ m\,(1-\epsilon_2)\,||\vec{v}||_2^2 \leq ||\mathbf{M}\vec{v}||_2^2 \leq m\,(1+\epsilon_2)\,||\vec{v}||_2^2 \right\}$$

$$\begin{aligned}
\mathrm{P}\left(\cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2}\right) &\leq \sum_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\
&\leq |\mathcal{N}_{\epsilon_1}|\,\mathrm{P}\left(\mathcal{E}^c_{\vec{v},\epsilon_2}\right) \\
&\leq 2\left(\frac{12}{\epsilon}\right)^s \exp\left(-\frac{m\epsilon^2}{32}\right)
\end{aligned}$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{s-1}$

There exists $\vec{v} \in \mathcal{N}\left(\mathcal{X}, \epsilon_1\right)$ such that $\left|\left|\vec{x} - \vec{v}\right|\right|_2 \leq \epsilon/4$

$\left|\left|\mathbf{M}\vec{x}\right|\right|_2$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{s-1}$

There exists $\vec{v} \in \mathcal{N}(\mathcal{X}, \epsilon_1)$ such that $||\vec{x} - \vec{v}||_2 \leq \epsilon/4$

$$||\mathbf{M}\vec{x}||_2 \leq ||\mathbf{M}\vec{v}||_2 + ||\mathbf{M}(\vec{x} - \vec{v})||_2$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{s-1}$

There exists $\vec{v} \in \mathcal{N}\left(\mathcal{X}, \epsilon_1\right)$ such that $\left|\left|\vec{x} - \vec{v}\right|\right|_2 \leq \epsilon/4$

$$
\begin{aligned}
\left|\left|\mathbf{M}\vec{x}\right|\right|_2 &\leq \left|\left|\mathbf{M}\vec{v}\right|\right|_2 + \left|\left|\mathbf{M}\left(\vec{x} - \vec{v}\right)\right|\right|_2 \\
&\leq \sqrt{m}\left(1 + \frac{\epsilon}{2}\right) + \left|\left|\mathbf{M}\left(\vec{x} - \vec{v}\right)\right|\right|_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v}, \epsilon_2} \text{ holds}
\end{aligned}
$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{s-1}$

There exists $\vec{v} \in \mathcal{N}\left(\mathcal{X}, \epsilon_1\right)$ such that $\left|\left|\vec{x} - \vec{v}\right|\right|_2 \leq \epsilon/4$

$$\begin{aligned}
\left|\left|\mathbf{M}\vec{x}\right|\right|_2 &\leq \left|\left|\mathbf{M}\vec{v}\right|\right|_2 + \left|\left|\mathbf{M}\left(\vec{x} - \vec{v}\right)\right|\right|_2 \\
&\leq \sqrt{m}\left(1 + \frac{\epsilon}{2}\right) + \left|\left|\mathbf{M}\left(\vec{x} - \vec{v}\right)\right|\right|_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v}, \epsilon_2} \text{ holds} \\
&\leq \sqrt{m}\left(1 + \frac{\epsilon}{2}\right) + \boldsymbol{\sigma_1}\left|\left|\vec{x} - \vec{v}\right|\right|_2
\end{aligned}$$

Upper bound on the sphere

Let $\vec{x} \in \mathcal{S}^{s-1}$

There exists $\vec{v} \in \mathcal{N}\left(\mathcal{X}, \epsilon_1\right)$ such that $||\vec{x} - \vec{v}||_2 \leq \epsilon/4$

$$
\begin{aligned}
||\mathbf{M}\vec{x}||_2 &\leq ||\mathbf{M}\vec{v}||_2 + ||\mathbf{M}\left(\vec{x} - \vec{v}\right)||_2 \\
&\leq \sqrt{m}\left(1 + \frac{\epsilon}{2}\right) + ||\mathbf{M}\left(\vec{x} - \vec{v}\right)||_2 \qquad \text{assuming } \cup_{\vec{v} \in \mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v}, \epsilon_2} \text{ holds} \\
&\leq \sqrt{m}\left(1 + \frac{\epsilon}{2}\right) + \boldsymbol{\sigma_1} ||\vec{x} - \vec{v}||_2 \\
&\leq \sqrt{m}\left(1 + \frac{\epsilon}{2}\right) + \frac{\boldsymbol{\sigma_1}\epsilon}{4}
\end{aligned}
$$

Upper bound on the sphere

$$\boldsymbol{\sigma_1} \leq \sqrt{m}\left(1+\frac{\epsilon}{2}\right)+\frac{\boldsymbol{\sigma_1}\epsilon}{4}$$

$$\begin{aligned}\boldsymbol{\sigma_1} &\leq \sqrt{m}\left(\frac{1+\epsilon/2}{1-\epsilon/4}\right)\\ &= \sqrt{m}\left(1+\epsilon-\frac{\epsilon\,(1-\epsilon)}{4-\epsilon}\right)\\ &\leq \sqrt{m}\,(1+\epsilon)\end{aligned}$$

Lower bound on the sphere

$||\mathbf{M}\vec{x}||_2$

Lower bound on the sphere

$$||\mathbf{M}\vec{x}||_2 \geq ||\mathbf{M}\vec{v}||_2 - ||\mathbf{M}\left(\vec{x} - \vec{v}\right)||_2$$

Lower bound on the sphere

$$\begin{aligned} ||\mathbf{M}\vec{x}||_2 &\geq ||\mathbf{M}\vec{v}||_2 - ||\mathbf{M}(\vec{x}-\vec{v})||_2 \\ &\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - ||A(\vec{x}-\vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2} \text{ holds} \end{aligned}$$

Lower bound on the sphere

$$
\begin{aligned}
||\mathbf{M}\vec{x}||_2 &\geq ||\mathbf{M}\vec{v}||_2 - ||\mathbf{M}(\vec{x}-\vec{v})||_2 \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - ||A(\vec{x}-\vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}} \mathcal{E}^c_{\vec{v},\epsilon_2} \text{ holds} \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - \boldsymbol{\sigma_1}\,||\vec{x}-\vec{v}||_2
\end{aligned}
$$

Lower bound on the sphere

$$
\begin{aligned}
||\mathbf{M}\vec{x}||_2 &\geq ||\mathbf{M}\vec{v}||_2 - ||\mathbf{M}(\vec{x}-\vec{v})||_2 \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - ||A(\vec{x}-\vec{v})||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^{c}_{\vec{v},\epsilon_2} \text{ holds} \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - \boldsymbol{\sigma_1}||\vec{x}-\vec{v}||_2 \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - \frac{\epsilon}{4}\sqrt{m}(1+\epsilon)
\end{aligned}
$$

Lower bound on the sphere

$$
\begin{aligned}
||\mathbf{M}\vec{x}||_2 &\geq ||\mathbf{M}\vec{v}||_2 - ||\mathbf{M}\left(\vec{x}-\vec{v}\right)||_2 \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - ||A\left(\vec{x}-\vec{v}\right)||_2 \qquad \text{assuming } \cup_{\vec{v}\in\mathcal{N}_{\epsilon_1}}\mathcal{E}^{c}_{\vec{v},\epsilon_2} \text{ holds} \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - \boldsymbol{\sigma_1}\,||\vec{x}-\vec{v}||_2 \\
&\geq \sqrt{m}\left(1-\frac{\epsilon}{2}\right) - \frac{\epsilon}{4}\sqrt{m}\left(1+\epsilon\right) \\
&= \sqrt{m}\left(1-\epsilon\right)
\end{aligned}
$$