Understanding Random Linear Estimation using a Conditional CLT

Galen Reeves
Duke University

Les Houches — March 1st, 2017
compressed sensing / random linear estimation

\[ Y^m = A^m X^n + W^m \]

IID Gaussian noise
IID Gaussian matrix
IID signal prior

\[ X_i \sim P_X \]
compressed sensing / random linear estimation

\begin{align*}
\text{measurement} & \quad \text{signal} \quad \text{noise} \\
Y^m & = A^m X^n + W^m
\end{align*}

IID Gaussian noise
IID Gaussian matrix
IID signal prior

\[ X_i \sim P_X \]

“classical” linear regression

\[ P_X = \mathcal{N}(\mu, \sigma^2) \]
compressed sensing / random linear estimation

\[ Y^m = A^m X^n + W^m \]

measurement \hspace{1cm} signal \hspace{1cm} noise

IID Gaussian noise
IID Gaussian matrix
IID signal prior

\[ X_i \sim P_X \]

“classical” linear regression \hspace{1cm} \[ P_X = \mathcal{N}(\mu, \sigma^2) \]

robust statistics \hspace{1cm} \[ P_X = (1 - \epsilon)P_{X_0} + \epsilon H \]
compressed sensing / random linear estimation

\[
\begin{align*}
Y^m &= A^m X^n + W^m \\
\end{align*}
\]

IID Gaussian noise
IID Gaussian matrix
IID signal prior

\[
X_i \sim P_X
\]

“classical” linear regression

\[
P_X = \mathcal{N}(\mu, \sigma^2)
\]

robust statistics

\[
P_X = (1 - \epsilon)P_{X_0} + \epsilon H
\]

CDMA

\[
P_X = \frac{1}{2} \delta_{-\mu} + \frac{1}{2} \delta_{+\mu}
\]
compressed sensing / random linear estimation

\[
\begin{align*}
\text{measurement} & : Y^m \\
\text{signal} & : A^m X^n \\
\text{noise} & : W^m
\end{align*}
\]

IID Gaussian noise
IID Gaussian matrix
IID signal prior

\[
X_i \sim P_X
\]

“classical” linear regression

\[
P_X = \mathcal{N}(\mu, \sigma^2)
\]

robust statistics

\[
P_X = (1 - \epsilon)P_{X_0} + \epsilon H
\]

CDMA

\[
P_X = \frac{1}{2} \delta_{-\mu} + \frac{1}{2} \delta_{+\mu}
\]

compressed sensing

\[
P_X = (1 - \epsilon)\delta_0 + \epsilon H
\]
support recovery in compressed sensing

[Reeves - Gastpar ’10]
support recovery in compressed sensing

[Reeves - Gastpar ’12]

- Linear MMSE
- AMP
- Soft Thresholding
- AMP – MMSE
- Maximum Likelihood

[Donoho, Maleki, & Montanari ’09]
[Bayati & Montanari ’11]
Reeves - Gastpar ’12

SNR (dB)

Sampling Rate

measurement rate

signal-to-noise ratio (dB)
support recovery in compressed sensing

[Reeves - Gastpar ’12]

- Linear MMSE
- AMP
- AMP – MMSE
- Maximum Likelihood
- Exact boundary obtained using (non-rigorous) replica method
  - [Guo & Verdu ’05]
- Linear estimation + thresholding
- ML upper bound
- Not Achievable

SNR (dB) vs. measurement rate

AMP – Soft Thresholding

[Guo & Verdu ’05]
compressed sensing / random linear estimation

measurement \( Y^m \) \( \quad \) signal \( A^m \) \( \quad \) noise \( X^n \) \( + \) \( W^m \)
compressed sensing / random linear estimation

\[ Y^m = A^m X^n + W^m \]

measurement

signal

noise

measurement ratio

\[ \delta = \frac{m}{n} \]
compressed sensing / random linear estimation

\[
\begin{align*}
Y^m & \quad \quad A^m \\
\quad & = \quad \quad X^n \\
\quad & \quad \quad W^m
\end{align*}
\]

measurement  
signal  
noise  

\[
\delta = \frac{m}{n}
\]

mutual information

\[
\mathcal{I}_n(\delta) = \frac{1}{n} I(X^n; Y^m | A^m)
\]

minimum mean square error (MMSE)

\[
\mathcal{M}_n(\delta) = \frac{1}{n} \text{mmse}(X^n | Y^m, A^m)
\]
compressed sensing / random linear estimation

\[ Y^m \] measurements generate according to

\[ A^m \] signal

\[ X^n \] noise

\[ W^m \] measurements generate according to

\[ \delta = \frac{m}{n} \]

mutual information

\[ I_n(\delta) = \frac{1}{n} I(X^n; Y^m | A^m) \]

minimum mean square error (MMSE)

\[ M_n(\delta) = \frac{1}{n} \text{mmse}(X^n | Y^m, A^m) \]
replica symmetric (RS) prediction

[physics Kabashima ‘03, Tanaka ’04] [general result Guo & Verdu ’05]

- define single-letter functions

\[ I_X(s) = I(X; \sqrt{s}X + W) \]
\[ \text{mmse}_X(s) = \text{mmse}(X \mid \sqrt{s}X + W) \]
\[ X \sim P_X \]
\[ W \sim N(0, 1) \]
replica symmetric (RS) prediction

[physics Kabashima ‘03, Tanaka ’04]  [general result Guo & Verdu ’05]

• define single-letter functions

\[ I_X(s) = I(X; \sqrt{s}X + W) \quad X \sim P_X \]
\[ \text{mmse}_X(s) = \text{mmse}(X | \sqrt{s}X + W) \quad W \sim N(0, 1) \]

• asymptotic mutual information and MMSE given by

\[ I_{\text{RS}}(\delta) = \min_{s > 0} \left\{ I_X(s) + \frac{\delta}{2} \left[ \log \left( \frac{\delta}{s} \right) + \frac{s}{\delta} - 1 \right] \right\} \]
\[ M_{\text{RS}}(\delta) = \text{mmse}_X(s^*) \quad \text{where } s^* \text{ attains minimum} \]
rigorous results in some cases
rigorous results in some cases

- Gaussian signal => yes
  - [Verdu & Shamai ’99] [Tse & Hanly ‘99]
- binary signal => yes for sparse matrices
  - [Montanari & Tse ’06]
- binary signal => replica solution always an upper bound
  - [Korada & Macris ’10]
- arbitrary signal => yes for some settings* (AMP is optimal)
  - [Donoho, Maleki, & Montanari ’09]
  - [Bayati & Montanari ’11]
rigorous results in some cases

• arbitrary signal $\Rightarrow$ consistent with analysis of support recovery
  ‣ [Reeves & Gastpar ’12]

• arbitrary signal $\Rightarrow$ consistent with analysis of phase transitions
  ‣ [Wu & Verdu ’12] [Reeves & Donoho ‘13]

• arbitrary signal & spatially coupled matrix $\Rightarrow$ yes
  ‣ [Krzakala, Mezard, Sausset, Sun, Zdeborova ’13]
  ‣ [Donoho, Javanmard, Montanari ’13]

• Bernoulli-Gaussian signal $\Rightarrow$ maybe correct
  ‣ arxiv papers [Huleihel & Merhav ’13, ’15, ‘16]
replica symmetric prediction is exact

[Reeves & Pfister ISIT ’16]

• Result holds for any IID signal distribution that has:
  ‣ bounded forth moment
  ‣ “single-crossing property” (can be verified numerically)

• Full paper available on arXiv
replica symmetric prediction is exact

[Reeves & Pfister ISIT ’16]

• Result holds for any IID signal distribution that has:
  ‣ bounded forth moment
  ‣ “single-crossing property” (can be verified numerically)

• Full paper available on arXiv

• Related result with different assumptions and different proof
  ‣ proof technique applies generally to other problems

  [Barbier, Dia, Macris, Krzakla Allerton ‘16]
$\frac{1}{2} \log(1 + \mathcal{M}(\delta))$
MMSE vs measurements

\[ \frac{1}{2} \log(1 + M(\delta)) \]

replica symmetric MMSE

MSE of AMP
MMSE vs measurements

\[ \frac{1}{2} \log(1 + M(\delta)) \]

\[ \delta \]

\[ \text{MSE of AMP} \]

\[ \text{fixed-point information curve} \]

\[ \left\{ \left( \frac{1}{2} \log(1 + M), \delta \right) : M = \text{mmse}_X \left( \frac{\delta}{1 + M} \right) \right\} \]

\[ \text{replica symmetric MMSE} \]
MMSE vs measurements

\[ \frac{1}{2} \log(1 + \mathcal{M}(\delta)) ]

fixed-point information curve
\[ \{(\frac{1}{2} \log(1 + \mathcal{M}), \delta) : \mathcal{M} = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}} \right) \} \]

replica symmetric MMSE

\[ \mathcal{M} = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}} \right) \]
G. Reeves is with the Department of Electrical and Computer Engineering and the Department of Statistical Science, Duke University (e-mail: galen.reeves@duke.edu)

\[
\frac{1}{2} \log (1 + \mathcal{M}(\delta)) = \text{replica symmetric MMSE}
\]

fixed-point information curve
\[
\left\{ \left( \frac{1}{2} \log (1 + \mathcal{M}), \delta \right) : \mathcal{M} = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}} \right) \right\}
\]

fixed-point constraint
\[
\mathcal{M}_{RS}(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}_{RS}(\delta)} \right)
\]
MMSE vs measurements

\[
\frac{1}{2} \log(1 + \mathcal{M}(\delta))
\]

replica symmetric MMSE

fixed-point information curve
\[
\left\{ \left( \frac{1}{2} \log(1 + \mathcal{M}), \delta \right) : \mathcal{M} = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}} \right) \right\}
\]

fixed-point constraint
\[
\mathcal{M}_{RS}(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}_{RS}(\delta)} \right)
\]

integral constraint
\[
\mathcal{I}_{RS}(\delta) = \int_{0}^{\delta} \frac{1}{2} \log(1 + \mathcal{M}_{RS}(\delta')) \, d\delta'
\]
MMSE vs measurements

\[ \frac{1}{2} \log(1 + \mathcal{M}(\delta)) \]

fixed-point information curve

\[ \left\{ \left( \frac{1}{2} \log(1 + \mathcal{M}), \delta \right) : \mathcal{M} = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}} \right) \right\} \]

fixed-point constraint

\[ \mathcal{M}_{RS}(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}_{RS}(\delta)} \right) \]

integral constraint

\[ \mathcal{I}_{RS}(\delta) = \int_{0}^{\delta} \frac{1}{2} \log(1 + \mathcal{M}_{RS}(\delta')) \, d\delta' \]

MMSE jumps when areas are equal

replica symmetric MMSE
MMSE vs measurements

\[ \frac{1}{2} \log(1 + \mathcal{M}(\delta)) \]

fixed-point information curve
\[ \{ \left( \frac{1}{2} \log(1 + \mathcal{M}), \delta \right) : \mathcal{M} = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}} \right) \} \]

fixed-point constraint
\[ \mathcal{M}_{RS}(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}_{RS}(\delta)} \right) \]

integral constraint
\[ \mathcal{I}_{RS}(\delta) = \int_0^\delta \frac{1}{2} \log(1 + \mathcal{M}_{RS}(\delta')) \, d\delta' \]
key ideas in proof

show constraints hold almost everywhere in large system limit

\[ M_n(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + M_n(\delta)} \right) \]

\[ I_n(\delta) = \int_0^\delta \frac{1}{2} \log(1 + M_n(\delta')) d\delta' \]
key ideas in proof

show constraints hold almost everywhere in large system limit

\[ M_n(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + M_n(\delta)} \right) \]

\[ I_n(\delta) = \int_0^\delta \frac{1}{2} \log(1 + M_n(\delta')) d\delta' \]

use boundary conditions to localize the jump
(requires single-crossing property)
**key ideas in proof**

show constraints hold almost everywhere in large system limit

\[ M_n(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + M_n(\delta)} \right) \]

\[ J_n(\delta) = \int_0^\delta \frac{1}{2} \log(1 + M_n(\delta')) \, d\delta' \]

use boundary conditions to localize the jump (requires single-crossing property)
main theorems

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy
\[
\int_0^\delta \left| \mathcal{I}'_n(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| \, d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]
for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0,1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies
\[
\int_0^\delta \left| \mathcal{M}_n(\gamma) - \text{mmse}_X\left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| \, d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]
for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0,1) \) is a universal constant.
main theorems

Gaussian matrix and IID signal with finite 4th moment

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

\[
\int_0^\delta \left| \mathcal{I}'_n(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0,1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

\[
\int_0^\delta \left| \mathcal{M}_n(\gamma) - \text{mmse}_X \left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0,1) \) is a universal constant.
main theorems

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

\[
\int_0^\delta \left| \mathcal{I}_n'(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| \, d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

\[
\int_0^\delta \left| \mathcal{M}_n(\gamma) - \text{mmse}_X \left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| \, d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.
main theorems

constant depend on 4th moment if signal distribution

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

\[
\int_0^\delta \left| \mathcal{I}'_n(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

\[
\int_0^\delta \left| \mathcal{M}_n(\gamma) - \text{mmse}_X \left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.
**main theorems**

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

\[
\int_0^\delta \left| \mathcal{I}_n'(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

\[
\int_0^\delta \left| \mathcal{M}_n(\gamma) - \text{mmse}_X \left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.
main theorems

universal constant $1/2 < r < 1$

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

$$
\int_0^\delta \left| I_n'(\gamma) - \frac{1}{2} \log(1 + M_n(\gamma)) \right| \, d\gamma \leq C_{B,\delta} \cdot n^{-r},
$$

for all $n \in \mathbb{N}$ and $\delta \in \mathbb{R}_+$ where $r \in (0, 1)$ is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

$$
\int_0^\delta \left| M_n(\gamma) - \text{mmse}_X\left(\frac{\gamma}{1 + M_n(\gamma)}\right) \right| \, d\gamma \leq C_{B,\delta} \cdot n^{-r},
$$

for all $n \in \mathbb{N}$ and $\delta \in \mathbb{R}_+$ where $r \in (0, 1)$ is a universal constant.
Theorem 2. Under Assumptions 1 and 2, the MI and MMSE functions satisfy
\[
\int_0^\delta \left| \mathcal{I}_n'(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]
for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.

Theorem 3. Under Assumptions 1 and 2, the MMSE function satisfies
\[
\int_0^\delta \left| \mathcal{M}_n(\gamma) - \operatorname{mmse}_X \left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]
for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.
main theorems

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

\[
\int_0^\infty \left| \mathcal{I}_n'(\gamma) - \frac{1}{2} \log(1 + \mathcal{M}_n(\gamma)) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

\[
\int_0^\infty \left| \mathcal{M}_n(\gamma) - \text{mmse}_X \left( \frac{\gamma}{1 + \mathcal{M}_n(\gamma)} \right) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0, 1) \) is a universal constant.
main theorems

**Theorem 2.** Under Assumptions 1 and 2, the MI and MMSE functions satisfy

\[
\int_{0}^{\delta} \left| I'_n(\gamma) - \frac{1}{2} \log(1 + M_n(\gamma)) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0,1) \) is a universal constant.

**Theorem 3.** Under Assumptions 1 and 2, the MMSE function satisfies

\[
\int_{0}^{\delta} \left| M_n(\gamma) - \text{mmse}_X \left( \frac{\gamma}{1 + M_n(\gamma)} \right) \right| d\gamma \leq C_{B,\delta} \cdot n^{-r},
\]

for all \( n \in \mathbb{N} \) and \( \delta \in \mathbb{R}_+ \) where \( r \in (0,1) \) is a universal constant.
proof of theorems
proof of theorems

\[ I_n(\delta) = \int_0^\delta \frac{1}{2} \log(1 + \mathcal{M}_n(\delta')) \, d\delta' \]

\[ \mathcal{M}_n(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + \mathcal{M}_n(\delta)} \right) \]
compressed sensing model

\[ Y^m = A^m X^n + W^m \]
compressed sensing model

\[ Y^m = A^m X^n + W^m \]
compressed sensing model

\[
\begin{align*}
Y^m & = A^m X^n + W^m \\
\bar{X}^n & = X^n - \mathbb{E}\left[ X^n \mid \tilde{Y}^m, \tilde{A}^m \right]
\end{align*}
\]
compressed sensing model

\[
\begin{bmatrix}
Y^m \\
\end{bmatrix} = \begin{bmatrix}
\cdot \\
\end{bmatrix}
\begin{bmatrix}
A^m \\
\end{bmatrix}
\begin{bmatrix}
X^n \\
\end{bmatrix} + \begin{bmatrix}
W^m \\
\end{bmatrix}
\]

signal error

\[
\bar{X}^n = X^n - \mathbb{E}[X^n | \tilde{Y}^m, \tilde{A}^m]
\]

posterior variance

\[
V_m = \frac{1}{n} \mathbb{E}[\|\bar{X}\|^2 | Y^m, A^m]
\]
compressed sensing model

\[
Y^m = A^m X^n + W^m
\]

signal error

\[
\bar{X}^n = X^n - \mathbb{E}[X^n | \tilde{Y}^m, \tilde{A}^m]
\]

posterior variance

\[
V_m = \frac{1}{n} \mathbb{E}[\|\bar{X}\|^2 | Y^m, A^m]
\]

MMSE sequence

\[
M_m = \frac{1}{n} \mathbb{E}[\|\bar{X}^n\|^2]
\]
compressed sensing model

\[ Y^m = A^m X^n + W^m \]

signal error

\[ \tilde{X}^n = X^n - \mathbb{E}[X^n | \tilde{Y}^m, \tilde{A}^m] \]

posterior variance

\[ V_m = \frac{1}{n} \mathbb{E}[\|\tilde{X}\|^2 | Y^m, A^m] \]

MMSE sequence

\[ M_m = \frac{1}{n} \mathbb{E}[\|\tilde{X}^n\|^2] \]

MI sequence

\[ I_m = I(X^n; Y^m | A^m) \]
add a new measurement
add a new measurement

\[
\begin{align*}
Y^m & \quad A^m & \quad X^n & \quad W^m \\
\hline
\text{=} & \hline
\hline
\end{align*}
\]

subtract conditional expectation

\[
\bar{Y}_{m+1} = Y_{m+1} - \mathbb{E}[Y_{m+1} \mid Y^m, A^m, A_{m+1}]
\]
add a new measurement

\[
\begin{bmatrix}
Y^m \\
A^m \\
X^n \\
W^m
\end{bmatrix} = \begin{bmatrix}
\vdots \\
\vdots \\
\vdots \\
\vdots
\end{bmatrix} + \begin{bmatrix}
\vdots \\
\vdots \\
\vdots \\
\vdots
\end{bmatrix}
\]

subtract conditional expectation

\[
\bar{Y}_{m+1} = Y_{m+1} - \mathbb{E}[Y_{m+1} \mid Y^m, A^m, A_{m+1}]
\]

new measurement is noisy random projection of error:

\[
\bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]

\[
A_{m+1} \sim \mathcal{N}(0, \frac{1}{n} I_n) \quad \bar{X}^n \sim P_{\bar{X}^n \mid Y^m, A^m} \quad W_{m+1} \sim \mathcal{N}(0, 1)
\]
add a new measurement

\[
Y^m_{m+1} = Y_{m+1} - \mathbb{E}[Y_{m+1} | Y^m, A^m, A_{m+1}]
\]

new measurement is \textbf{noisy random projection} of error:

\[
\tilde{Y}_{m+1} = [A_{m+1}]^T \tilde{X}^n + W_{m+1}
\]

\[
A_{m+1} \sim \mathcal{N}(0, \frac{1}{n} I_n) \quad \tilde{X}^n \sim P_{\tilde{X}^n | Y^m, A^m} \quad W_{m+1} \sim \mathcal{N}(0, 1)
\]

How close is distribution of new measurement to the Gaussian distribution with same mean and variance?
moments of new measurement corresponds to signal error

\[ E[\bar{Y}_{m+1} \mid Y^m, A^m] = 0 \]

\[ E[(\bar{Y}_{m+1})^2 \mid Y^m, A^m] = V_m \]

\[ E[(\bar{Y}_{m+1})^2] = M_m \]
moments of new measurement corresponds to signal error

\[
\mathbb{E}[\bar{Y}_{m+1} \mid Y^m, A^m] = 0
\]
\[
\mathbb{E}[(\bar{Y}_{m+1})^2 \mid Y^m, A^m] = V_m
\]
\[
\mathbb{E}[(\bar{Y}_{m+1})^2] = M_m
\]

posterior nonGaussianess

\[
\Delta^P_m = \mathbb{E}\left[D\left(P_{\bar{Y}_{m+1} \mid Y^m, A^m, A_{m+1}} \mid \mathcal{N}(0, 1 + V_m)\right) \mid Y^m, A^m\right]
\]
moments of new measurement corresponds to signal error

\[ \mathbb{E}[\bar{Y}_{m+1} \mid Y^m, A^m] = 0 \]

\[ \mathbb{E}[(\bar{Y}_{m+1})^2 \mid Y^m, A^m] = V_m \]

\[ \mathbb{E}[(\bar{Y}_{m+1})^2] = M_m \]

posterior non-Gaussianess

\[ \Delta^P_m = \mathbb{E}\left[D\left(P_{\bar{Y}_{m+1}} \mid Y^m, A^m, A_{m+1} \bigg\| \mathcal{N}(0, 1 + V_m)\right) \mid Y^m, A^m\right] \]

non-Gaussianess

\[ \Delta_m = D\left(P_{\bar{Y}_{m+1}} \mid Y^m, A^m, A_{m+1} \bigg\| \mathcal{N}(0, 1 + M_m)\right) \]

\[ = \mathbb{E}[\Delta^P_m] + \frac{1}{2} \mathbb{E}\left[\log\left(\frac{1 + M_m}{1 + V_m}\right)\right] \]
increase in mutual information given by

\[ I'_m = I(X^n; Y_{m+1} \mid Y^m, A^m, A_{m+1}) \]

\[ = I(X^n; \bar{Y}_{m+1} \mid Y^m, A^m, A_{m+1}) \]
increase in mutual information given by

\[
I'_m = I(X^n; Y_{m+1} \mid Y^m, A^m, A_{m+1}) \\
= I(X^n; \bar{Y}_{m+1} \mid Y^m, A^m, A_{m+1}) \\
= \frac{1}{2} \log(1 + M_n) - \Delta_m
\]
increase in mutual information given by

\[
I_m' = I(X^n; Y_{m+1} \mid Y^m, A^m, A_{m+1})
\]

\[
= I(X^n; \tilde{Y}_{m+1} \mid Y^m, A^m, A_{m+1})
\]

\[
= \frac{1}{2} \log(1 + M_n) - \Delta_m
\]

\[
= \frac{1}{2} \log(1 + M_n) - \mathbb{E}[\Delta_m^P] - \frac{1}{2} \mathbb{E} \left[ \log \left( \frac{1 + M_m}{1 + V_m} \right) \right]
\]

increase if Gaussian  \hspace{1cm} posterior nonGaussianness  \hspace{1cm} deviation of posterior variance
increase in mutual information given by

\[
I'_m = I(X^n; Y_{m+1} \mid Y^m, A^m, A_{m+1}) \\
= I(X^n; \tilde{Y}_{m+1} \mid Y^m, A^m, A_{m+1}) \\
= \frac{1}{2} \log(1 + M_n) - \Delta_m \\
= \frac{1}{2} \log(1 + M_n) - \mathbb{E}[\Delta^P_m] - \frac{1}{2} \mathbb{E}\left[\log\left(\frac{1 + M_m}{1 + V_m}\right)\right]
\]

increase if Gaussian
posterior nonGaussianness
development of posterior variance

furthermore, measurement MMSE satisfies

\[
\frac{M_m}{1 + M_m} \leq \text{mmse}(A_{m+1}X^n \mid Y^{m+1}, A^{m+1}) \leq \frac{M_m}{1 + M_m} + C_B \sqrt{\Delta_m}
\]

[key result in Barbier et al. 2016]
proof of theorems
proof of theorems

\[
\mathcal{I}_n(\delta) = \int_0^\delta \frac{1}{2} \log(1 + M_n(\delta')) \, d\delta'
\]

\[
M_n(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + M_n(\delta)} \right)
\]
the rotation trick

IID Gaussian matrix

\[ A_{i,j} \sim \mathcal{N}(0, 1/n) \]

“rotated” matrix

\[ \tilde{A} = QA \]

Orthogonal transform Q depends only on last column of A
original model

\[
Y^m = A^m \times X^n + W^m
\]

rotated model

\[
\tilde{Y}^m = \tilde{A}^m \times X^n + \tilde{W}^m
\]
original model

\[ Y^m = A^m X^n + W^m \]

rotated model

\[ \tilde{Y}^m = \tilde{A}^m X^n + \tilde{W}^m \]
original model

\[
Y^m = A^m X^n W^m
\]

rotated model

\[
\tilde{Y}^m = \tilde{A}^m X^n \tilde{W}^m
\]

conditional distribution given everything but last measurement
original model

\[
Y^m = A^m X^n + W^m
\]

rotated model

\[
\tilde{Y}^m = \tilde{A}^m X^n + \tilde{W}^m
\]

conditional distribution given everything but last measurement

\[
\tilde{X}^{n-1} = X^{n-1} - \mathbb{E}[X^{n-1} \mid \tilde{Y}^{m-1}, \tilde{A}^{m-1}]
\]

signal error
conditional distribution given everything but last measurement

\[
\bar{X}^{n-1} = X^{n-1} - \mathbb{E}[X^{n-1} | \tilde{Y}^{m-1}, \tilde{A}^{m-1}]
\]

\[
\bar{Y}_m = Y_m - \mathbb{E}[Y_m | \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}]
\]

signal error
original model

\[
Y_m = A^m X^n W^m
\]

rotated model

\[
\tilde{Y}_m = \tilde{A}^m X^n \tilde{W}^m
\]

classical mean square error

\[
\tilde{X}^{n-1} = X^{n-1} - \mathbb{E} \left[ X^{n-1} | \tilde{Y}^{m-1}, \tilde{A}^{m-1} \right]
\]

signal error

\[
\tilde{Y}_m = Y_m - \mathbb{E} \left[ Y_m | \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right]
\]

noisy measurement of last signal entry

\[
= \sqrt{\frac{1}{n} \chi_m^2} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \tilde{X}_k + \tilde{W}_m
\]
original model

\[ Y^m = A^m X^n + W^m \]

rotated model

\[ \tilde{Y}^m = \tilde{A}^m X^n + \tilde{W}^m \]

MMSE can be expressed in terms of last signal entry
MMSE can be expressed in terms of last signal entry

\[ M_m = \text{mmse}(X_n \mid Y^m, A^m) \]
original model

\[
\begin{bmatrix}
Y^m \\
A^m \\
X^n \\
W^m
\end{bmatrix}
\]

rotated model

\[
\begin{bmatrix}
\tilde{Y}^m \\
\tilde{A}^m \\
\tilde{X}^n \\
\tilde{W}^m
\end{bmatrix}
\]

MMSE can be expressed in terms of last signal entry

\[
M_m = \text{mmse}(X_n | Y^m, A^m) = \text{mmse}(X_n | \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m)
\]
MMSE can be expressed in terms of last signal entry

\[ M_m = \text{mmse}(X_n | Y^m, A^m) = \text{mmse}(X_n | \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m) \]

\[ = \text{mmse} \left( X_n \left| \sqrt{\frac{1}{n} X_m^2} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \bar{X}_k + \tilde{W}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right. \right) \]
MMSE can be expressed in terms of last signal entry

\[ M_m = \text{mmse}(X_n \mid Y^m, A^m) = \text{mmse}(X_n \mid \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m) \]

\[ = \text{mmse}\left( X_n \mid \sqrt{\frac{1}{n}X_m^2} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \bar{X}_k + \tilde{W}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right) \]
MMSE can be expressed in terms of last signal entry

\[ M_m = \text{mmse}(X_n | Y^m, A^m) = \text{mmse}(X_n | \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m) \]

\[ = \text{mmse} \left( X_n \mid \sqrt{\frac{1}{m/n} X^2_m} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \tilde{X}_k + \tilde{W}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right) \]

close to Gaussian?
MMSE can be expressed in terms of last signal entry

\[
M_m = \text{mmse}(X_n | Y^m, A^m) = \text{mmse}(X_n | \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m)
\]

\[
= \text{mmse}
\left( X_n \left| \sqrt{\frac{1}{n} \chi^2_m} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \tilde{X}_k + \tilde{W}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right. \right)
\]

close to Gaussian?

can derive MMSE fixed-point equation

\[
\left| M_m - \text{mmse}_X \left( \frac{m/n}{1 + M_m} \right) \right| \leq C_B \sqrt{|I''_m| + \Delta_m + \frac{m}{n^2}}
\]
Introduction

1. Background

EY can derive MMSE fixed-point equation

\[ P_{\text{mmse}}(e) = \mathbb{E}(e^2) \]

\[ E_{q} \sim m, X \]

\[ X^n = \sum_{k=1}^{n-1} A_{m,k} X_k + W_m, \quad \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \]

MMSE can be expressed in terms of last signal entry

\[ M_m = \text{mmse}(X_n | Y^m, A^m) = \text{mmse}(X_n | \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m) \]

\[ = \text{mmse}\left( X_n \left| \sqrt{\frac{1}{n} X_m^2} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \tilde{X}_k + \tilde{W}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right. \right) \]

can derive MMSE fixed-point equation

\[ \left| M_m - \text{mmse}_X \left( \frac{m/n}{1 + M_m} \right) \right| \leq C_B \sqrt{I''_m} + \Delta_m + \frac{m}{n^2} \]

2nd order MI difference
original model

\[
\begin{bmatrix}
Y^m \\
A^m \\
X^n \\
W^m
\end{bmatrix} = \begin{bmatrix}
\end{bmatrix} + \begin{bmatrix}
\end{bmatrix}
\]

rotated model

\[
\begin{bmatrix}
\tilde{Y}^m \\
\tilde{A}^m \\
X^n \\
\tilde{W}^m
\end{bmatrix} = \begin{bmatrix}
\end{bmatrix} + \begin{bmatrix}
\end{bmatrix}
\]

MMSE can be expressed in terms of last signal entry

\[
M_m = \text{mmse}(X_n \mid Y^m, A^m) = \text{mmse}(X_n \mid \tilde{Y}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m)
\]

\[
= \text{mmse} \left( X_n \mid \sqrt{\frac{1}{n} X_m^2} X_n + \sum_{k=1}^{n-1} \tilde{A}_{m,k} \tilde{X}_k + \tilde{W}_m, \tilde{Y}^{m-1}, \tilde{A}^{m-1}, \tilde{A}_m \right)
\]

can derive MMSE fixed-point equation

\[
\left| M_m - \text{mmse}_X \left( \frac{m/n}{1 + M_m} \right) \right| \leq C_B \sqrt{I_m'' + \Delta_m + \frac{m}{n^2}}
\]

2nd order MI difference
proof of theorems
proof of theorems

\[ I_n(\delta) = \int_{0}^{\delta} \frac{1}{2} \log(1 + M_n(\delta')) d\delta' \]

\[ M_n(\delta) = \text{mmse}_X \left( \frac{\delta}{1 + M_n(\delta)} \right) \]
Random projections of high dimensional data

random vector

\[ U \in \mathbb{R}^n \]

IID Gaussian matrix

\[ A \in \mathbb{R}^{k \times n} \]

random projection

\[ Y = AU \]
Random projections of high dimensional data

random vector

\[ U \in \mathbb{R}^n \]

IID Gaussian matrix

\[ A \in \mathbb{R}^{k \times n} \]

random projection

\[ Y = AU \]

distribution of \( Y \) is approx. Gaussian

\[ P_Y \]
Random projections of high dimensional data

random vector
\( U \in \mathbb{R}^n \)

IID Gaussian matrix
\( A \in \mathbb{R}^{k \times n} \)

random projection
\( Y = AU \)

distribution of \( Y \) is approx. Gaussian
\( P_Y \)

conditional distribution given \( A \) is random
\( P_{Y|A} \)
Random projections of high dimensional data

- random vector
  \( U \in \mathbb{R}^n \)

- IID Gaussian matrix
  \( A \in \mathbb{R}^{k \times n} \)

- random projection
  \( Y = AU \)

- distribution of \( Y \) is approx. Gaussian
  \( P_Y \)

- conditional distribution given \( A \) is random
  \( P_{Y|A} \)

Conditional CLT:

\[ P_{Y|A} \approx G_Y, \quad G_Y = \mathcal{N}(\mathbb{E}[Y], \text{Cov}(Y)) \]
ASYMPTOTICS OF GRAPHICAL PROJECTION PURSUIT

BY PERSI DIACONIS¹ AND DAVID FREEDMAN²

Stanford University and University of California, Berkeley

Mathematical tools are developed for describing low-dimensional projections of high-dimensional data. Theorems are given to show that under suitable conditions, most projections are approximately Gaussian.

\[
\frac{1}{n}\text{card}\{j \leq n: |\|x_j\|^2 - \sigma^2 p| > \epsilon p\}\rightarrow 0
\]

\[
\frac{1}{n^2}\text{card}\{1 \leq j, k \leq n: |x_j \cdot x_k| > \epsilon p\}\rightarrow 0.
\]

weak convergence for distributions with bounded support
conditional central limit theorem [Reeves ’16]

Consider random projection

\[ Y = A^T U + W \]

\[ A \sim \mathcal{N}(0, \frac{1}{n} I_n), \quad W \sim \mathcal{N}(0, 1) \]
conditional central limit theorem [Reeves ’16]

Consider random projection

\[ Y = ATU + W \]

\[ A \sim \mathcal{N}(0, \frac{1}{n}I_n), \quad W \sim \mathcal{N}(0, 1) \]

**Theorem:** The expected KL divergence between (random) conditional distribution of \( Y \) given \( A \) and Gaussian satisfies:

\[
\mathbb{E}_A \left[ D_{KL} \left( P_Y | A \parallel \mathcal{N}(0, \text{Var}(Y)) \right) \right] 
\leq \frac{1}{n} \mathbb{E} \left[ \|U\| - \gamma \right] + \sqrt{\frac{1}{n} \mathbb{E} \left[ \|\langle U_1, U_2 \rangle\| \right]}
\]
conditional central limit theorem [Reeves ’16]

Consider random projection

\[ Y = A^T U + W \]

\[ A \sim \mathcal{N}(0, \frac{1}{n} I_n), \quad W \sim \mathcal{N}(0, 1) \]

**Theorem:** The expected KL divergence between (random) conditional distribution of \( Y \) given \( A \) and Gaussian satisfies:

\[
\mathbb{E}_A \left[ D_{KL}(P_{Y|A} \parallel \mathcal{N}(0, \text{Var}(Y))) \right] \\
\leq \frac{1}{n} \mathbb{E} \left[ \| U \| - \gamma \right] + \sqrt{\frac{1}{n} \mathbb{E} \left[ \langle U_1, U_2 \rangle \right]}
\]
conditional central limit theorem [Reeves ’16]

Consider random projection

\[ Y = A^T U + W \]
\[ A \sim \mathcal{N}(0, \frac{1}{n} I_n), \quad W \sim \mathcal{N}(0, 1) \]

**Theorem:** The expected KL divergence between (random) conditional distribution of \( Y \) given \( A \) and Gaussian satisfies:

\[
\mathbb{E}_A \left[ D_{KL} \left( P_{Y|A} \parallel \mathcal{N}(0, \text{Var}(Y)) \right) \right] \\
\leq \frac{1}{n} \mathbb{E} \left[ \|U\| - \gamma \right] + \sqrt{\frac{1}{n} \mathbb{E} \left[ \langle U_1, U_2 \rangle \right]} \\
\leq \sqrt{\frac{1}{n^2} \text{Var}(\|U\|^2) + \frac{1}{n} \|\text{Cov}(U)\|_F}
\]
recall that new measurement is \textbf{noisy random projection} of error:

\[
\bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]
recall that new measurement is **noisy random projection** of error:

\[
\bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]

by CCLT, the posterior nonGaussianess satisfies

\[
\mathbb{E}[\Delta_m^P] = \mathbb{E}\left[D\left(P_{\bar{Y}_{m+1}|Y^m,A^m,A_{m+1}} \mid \mathcal{N}(0, 1 + V_m)\right)\right] \\
\text{(CCLT)} \quad \leq \mathbb{E}\left[|V_m - \frac{1}{n} \|ar{X}^n\|^2|\right] + \sqrt{\frac{1}{n} \mathbb{E}[\|\text{Cov}(X^n | Y^m, A^m, A_{m+1})\|_F]} 
\]
recall that new measurement is **noisy random projection** of error:

\[
    \bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]

by CCLT, the posterior nonGaussianess satisfies

\[
    \mathbb{E} \left[ \Delta_m^P \right] = \mathbb{E} \left[ D \left( P_{\bar{Y}_{m+1} | Y^m, A^m, A_{m+1}} \left\| \mathcal{N}(0, 1 + V_m) \right\| \right) \right] \\
    \leq \mathbb{E} \left[ |V_m - \frac{1}{n} \| \bar{X}^n \|^2 | \right] + \sqrt{\frac{1}{n} \mathbb{E} [\| \text{Cov}(X^n | Y^m, A^m, A_{m+1}) \|_F]} \\
\]

(CCLT)

return to this in a moment...
add two new measurements
add two new measurements

second order different of MI

\[
I''_m = I'_{m+1} - I'_m = -I(Y_{m+1}; Y_{m+2} \mid Y^m, A^m, A_{m+1}, A_{m+2})
\]
add two new measurements

second order different of MI

\[
I''_m = I'_{m+1} - I'_m = -I(Y_{m+1}; Y_{m+2} \mid Y^m, A^m, A_{m+1}, A_{m+2})
\]

mutual information upper bounds covariance

\[
\frac{1}{n} \mathbb{E}[\| \text{Cov}(X^n \mid Y^m, A^m, A_{m+1}) \|_F] \leq C_B |I''_m|^{1/2}
\]

\[
\mathbb{E}[|V_m - \frac{1}{n} \| \bar{X}^n \|_2^2|] \leq C_B |I''_m|^{1/2}
\]
recall that new measurement is **noisy random projection** of error:

\[
\bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]

by CCLT, the posterior nonGaussianeness satisfies

\[
\mathbb{E}[\Delta^P_m] = \mathbb{E}
\left[
D
\left(P_{Y_{m+1}|Y^m, A^m, A_{m+1}} \bigg| \mathcal{N}(0, 1 + V_m)\right)\right]
\]

**(CCLT)**

\[
\leq \mathbb{E}\left[|V_m - \frac{1}{n}\|\bar{X}^n\|^2|\right] + \sqrt{\frac{1}{n}\mathbb{E}[\|\operatorname{Cov}(X^n \mid Y^m, A^m, A_{m+1})\|_F]}
\]
recall that new measurement is **noisy random projection** of error:

\[
\bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]

by CCLT, the posterior nonGaussianess satisfies

\[
\mathbb{E} \left[ \Delta^P_m \right] = \mathbb{E} \left[ D \left( P_{\bar{Y}_{m+1}} | Y^m, A^m, A_{m+1} \right) \| \mathcal{N}(0, 1 + V_m) \right] \\
\leq \mathbb{E} \left[ |V_m - \frac{1}{n} \| \bar{X}^n \| |^2 | \right] + \sqrt{\frac{1}{n} \mathbb{E}[\| \text{Cov}(X^n | Y^m, A^m, A_{m+1}) \|_F]} \\
\leq C_B |I_m''|^\frac{1}{8}
\]
recall that new measurement is **noisy random projection** of error:

\[
\bar{Y}_{m+1} = [A_{m+1}]^T \bar{X}^n + W_{m+1}
\]

by CCLT, the posterior nonGaussianess satisfies

\[
\mathbb{E}[\Delta_m^P] = \mathbb{E}\left[D\left(P_{\bar{Y}_{m+1}|Y^m, A^m, A_{m+1}} \mid \mathcal{N}(0, 1 + V_m)\right)\right]
\]

\[
\leq \mathbb{E}[|V_m - \frac{1}{n}\|\bar{X}^n\|^2|] + \sqrt{\frac{1}{n}\mathbb{E}[\|\text{Cov}(X^n \mid Y^m, A^m, A_{m+1})\|_F]}
\]

\[
\leq C_B |I''_m|^{\frac{1}{8}}
\]

the normalized summation of this term converges to zero:

\[
\frac{1}{m} \sum_{k=1}^{m}|I''_m|^{\frac{1}{8}} \leq \left| \frac{1}{m} \sum_{k=1}^{m} I''_m \right|^{\frac{1}{8}} \leq \left| \frac{1}{m} I'_1 \right|^{\frac{1}{8}} \to 0 \quad m \to \infty
\]
proof of theorems
Understanding the phase transition
Understanding the phase transition

posterior distribution of error

approx. uniform on sphere
radius = $\sqrt{\text{MMSE}}$

posterior after new measurement

$I' \approx \frac{1}{2} \log(1 + \text{MMSE})$

$M' \approx 0$
Understanding the phase transition

approx. uniform on sphere
radius $= \sqrt{\text{MMSE}}$

posterior distribution of error

no phase transition

posterior after new measurement

\[ I' \approx \frac{1}{2} \log(1 + \text{MMSE}) \]

\[ M' \approx 0 \]
Understanding the phase transition

approx. uniform on sphere
radius $= \sqrt{\text{MMSE}}$

no phase transition

non-isotropic

posterior distribution of error

posterior after new measurement

$I' \approx \frac{1}{2} \log(1 + \text{MMSE})$

$M' \approx 0$

$M' \ll 0$
Understanding the phase transition

approx. uniform on sphere
radius $= \sqrt{\text{MMSE}}$

no phase transition

non-isotropic

phase transition

posterior distribution of error

posterior after new measurement

$I' \approx \frac{1}{2} \log(1 + \text{MMSE})$

$M' \approx 0$

$M' \ll 0$
summary of proof
summary of proof

• Rigorous analysis of MMSE and MI limits for linear estimation
  ‣ Results match the replica-symmetric prediction
summary of proof

• Rigorous analysis of MMSE and MI limits for linear estimation
  ‣ Results match the replica-symmetric prediction

• Proof based on analysis of fixed-point conditions:
  ‣ increase in MI with adding new measurement
  ‣ rotation trick to establish MMSE fixed-point
  ‣ existence of limits is last step in proof
summary of proof

• Rigorous analysis of MMSE and MI limits for linear estimation
  ‣ Results match the replica-symmetric prediction

• Proof based on analysis of fixed-point conditions:
  ‣ increase in MI with adding new measurement
  ‣ rotation trick to establish MMSE fixed-point
  ‣ existence of limits is last step in proof

• Key step is conditional CLT for random projections

MMSE phase transition  posterior distribution is non-isotropic
summary of proof

• Rigorous analysis of MMSE and MI limits for linear estimation
  ‣ Results match the replica-symmetric prediction

• Proof based on analysis of fixed-point conditions:
  ‣ increase in MI with adding new measurement
  ‣ rotation trick to establish MMSE fixed-point
  ‣ existence of limits is last step in proof

• Key step is conditional CLT for random projections

• Many possibilities for extending this result…

MMSE phase transition
posterior distribution is non-isotropic
further results from proof
Conditional Central Limit Theorems for Gaussian Projections

Galen Reeves

(Submitted on 29 Dec 2016 (v1), last revised 30 Dec 2016 (this version, v2))

This paper addresses the question of when projections of a high-dimensional random vector are approximately Gaussian. This problem has been studied previously in the context of high-dimensional data analysis, where the focus is on low-dimensional projections of high-dimensional point clouds. The focus of this paper is on the typical behavior when the projections are generated by an i.i.d. Gaussian projection matrix. The main results are bounds on the deviation between the conditional distribution of the projections and a Gaussian approximation, where the conditioning is on the projection matrix. The bounds are given in terms of the quadratic Wasserstein distance and relative entropy and are stated explicitly as a function of the number of projections and certain key properties of the random vector. The proof uses Talagrand's transportation inequality and a general integral–moment inequality for mutual information. Applications to random linear estimation and compressed sensing are discussed.

\[ Z = \Theta X, \quad Y = Z + \sqrt{t}N \]
further results from proof

Conditional Central Limit Theorems for Gaussian Projections

Galen Reeves

(Submitted on 29 Dec 2016 (v1), last revised 30 Dec 2016 (this version, v2))

This paper addresses the question of when projections of a high-dimensional random vector are approximately Gaussian. This problem has been studied previously in the context of high-dimensional data analysis, where the focus is on low-dimensional projections of high-dimensional point clouds. The focus of this paper is on the typical behavior when the projections are generated by an i.i.d. Gaussian projection matrix. The main results are bounds on the deviation between the conditional distribution of the projections and a Gaussian approximation, where the conditioning is on the projection matrix. The bounds are given in terms of the quadratic Wasserstein distance and relative entropy and are stated explicitly as a function of the number of projections and certain key properties of the random vector. The proof uses Talagrand's transportation inequality and a general integral–moment inequality for mutual information. Applications to random linear estimation and compressed sensing are discussed.

\[ Z = \Theta X, \quad Y = Z + \sqrt{t}N \]

\[ W_2^2(P_{Z|\Theta}, G_Z) \leq 4tk + 4(t + \gamma)D(P_{Y|\Theta} \parallel G_Y) \]

Wasserstein distance

KL divergence
Further results from proof

Conditional Central Limit Theorems for Gaussian Projections

Galen Reeves

(Submitted on 29 Dec 2016 (v1), last revised 30 Dec 2016 (this version, v2))

This paper addresses the question of when projections of a high-dimensional random vector are approximately Gaussian. This problem has been studied previously in the context of high-dimensional data analysis, where the focus is on low-dimensional projections of high-dimensional point clouds. The focus of this paper is on the typical behavior when the projections are generated by an i.i.d. Gaussian projection matrix. The main results are bounds on the deviation between the conditional distribution of the projections and a Gaussian approximation, where the conditioning is on the projection matrix. The bounds are given in terms of the quadratic Wasserstein distance and relative entropy and are stated explicitly as a function of the number of projections and certain key properties of the random vector. The proof uses Talagrand's transportation inequality and a general integral-moment inequality for mutual information. Applications to random linear estimation and compressed sensing are discussed.

\[ Z = \Theta X, \quad Y = Z + \sqrt{t}N \]

\[ W_2^2(P_{Z|\Theta}, G_Z) \leq 4tk + 4(t + \gamma)D\left(P_{Y|\Theta} \| G_Y\right) \]

Wasserstein distance

KL divergence

\[ \mathbb{E}\left[W_2^2(P_{Z|\Theta}, G_Z)\right] \leq C' \left(n^{-\frac{1}{4}} + k n^{-\frac{2}{k+4}}\right). \]
Two-Moment Inequalities for Rényi Entropy and Mutual Information

Galen Reeves

(Submitted on 23 Feb 2017)

This paper explores some applications of a two-moment inequality for the integral of the \( r \)-th power of a function, where \( 0 < r < 1 \). The first contribution is an upper bound on the Rényi entropy of a random vector in terms of the two different moments. When one of the moments is the zeroth moment, these bounds recover previous results based on maximum entropy distributions under a single moment constraint. More generally, evaluation of the bound with two carefully chosen nonzero moments can lead to significant improvements with a modest increase in complexity. The second contribution is a method for upper bounding mutual information in terms of certain integrals with respect to the variance of the conditional density. The bounds have a number of useful properties arising from the connection with variance decompositions.
Two-Moment Inequalities for Rényi Entropy and Mutual Information

Galen Reeves

(Submitted on 23 Feb 2017)

This paper explores some applications of a two-moment inequality for the integral of the $r$-th power of a function, where $0 < r < 1$. The first contribution is an upper bound on the Rényi entropy of a random vector in terms of the two different moments. When one of the moments is the zeroth moment, these bounds recover previous results based on maximum entropy distributions under a single moment constraint. More generally, evaluation of the bound with two carefully chosen nonzero moments can lead to significant improvements with a modest increase in complexity. The second contribution is a method for upper bounding mutual information in terms of certain integrals with respect to the variance of the conditional density. The bounds have a number of useful properties arising from the connection with variance decompositions.

$$0 < r < 1$$
$$p < \frac{1-r}{r} < q$$

$$\left( \int f^r(x) \, dx \right)^{\frac{1}{r}} \leq C \left( \int |x|^p f(x) \, dx \right)^{\lambda} \left( \int |x|^q f(x) \, dx \right)^{1-\lambda},$$
Two-Moment Inequalities for Rényi Entropy and Mutual Information

Galen Reeves

(Submitted on 23 Feb 2017)

This paper explores some applications of a two-moment inequality for the integral of the \( r \)-th power of a function, where \( 0 < r < 1 \). The first contribution is an upper bound on the Rényi entropy of a random vector in terms of the two different moments. When one of the moments is the zeroth moment, these bounds recover previous results based on maximum entropy distributions under a single moment constraint. More generally, evaluation of the bound with two carefully chosen nonzero moments can lead to significant improvements with a modest increase in complexity. The second contribution is a method for upper bounding mutual information in terms of certain integrals with respect to the variance of the conditional density. The bounds have a number of useful properties arising from the connection with variance decompositions.

\[
0 < r < 1 \\
p < \frac{1 - r}{r} < q
\]

\[
\left( \int f^r(x) \, dx \right)^{\frac{1}{r}} \leq C \left( \int |x|^p f(x) \, dx \right)^{\lambda} \left( \int |x|^q f(x) \, dx \right)^{1 - \lambda},
\]

\[
p < 1 < q
\]

\[
I(X; Y) \leq C(\lambda) \sqrt{\frac{\omega(S_Y)V_{np}^{\lambda}(Y|X)V_{nq}^{1-\lambda}(Y|X)}{(q-p)}},
\]

\[
V_s(Y|X) = \int_{S_Y} \|y\|^s \text{Var}(f(y|X)) \, dy.
\]
the end!