MAP and MMSE Estimation

Ahmed Ibrahim
School of Electrical and Computer Engineering
Purdue University
ibrahim1@purdue.edu
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Description of the pursuit algorithm was on a deterministic level

Claim: The pursuit algorithms correspond to an approximation of MAP

To prove our claim, we give a clear and formal definition of the stochastic model generating the sparse representation vector

This leads to deriving a MMSE estimator as well

MMSE approximation is also possible
Outline

- Motivation
- **Stochastic Model & Estimation Goals**
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
The Measurement Vector

\[ y = HAx + e \]

Estimation Goal is to recover \( x \) or \( z \) based on \( y \) and the statistical information of \( x \) and \( e \).
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Background on MAP Estimator

\[
P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad \Rightarrow \quad \hat{x}^{MAP} = \max_x P(x|y)
\]

\[
\hat{x}^{MAP} = \max_x P(x|y) = \max_x P(y|x)P(x)
\]

\[
\hat{x}^{MAP} = \max_x P(y|x)
\]
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- **Background on MMSE**
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Background on MMSE Estimator

\[
\text{MSE}_x = E \left( \|\hat{x} - x\|_2^2 | y \right) = \int x \|\hat{x} - x\|_2^2 P(x|y)dx
\]

\[
\frac{\partial \text{MSE}_x}{\partial \hat{x}} = 2 \int (\hat{x} - x) P(x|y)dx = 0.
\]

\[
\hat{x}^{\text{MMSE}} = \frac{\int xP(x|y)dx}{\int P(x|y)dx} = \int xP(x|y)dx = E(x|y)
\]
Background on MWMSE Estimation

\[
WMSE_x = E \left( \| W(\hat{x} - x) \|_2^2 | y \right) \\
= \int_x \| W(\hat{x} - x) \|_2^2 P(x|y) dx
\]

\[
\frac{\partial WMSE_x}{\partial \hat{x}} = 2W^T W \int_x (\hat{x} - x) P(x|y) dx = 0
\]

Minimizing

\[
E(\| x - \hat{x} \|_2^2 | y)
\]

Minimizing

\[
E(\| Ax - \hat{z} \|_2^2 | y)
\]

Minimizing

\[
E(x|y)
\]

\[
AE(x|y)
\]

Does this have a unique solution for the minimizer??!
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Although Infeasible (as it assumes knowledge of the support $s$)

- Core Ingredient for practical estimators
- Simple and convenient closed-form expression
- Gives a reference performance quality to compare against
- Easy to derive and reveals some interesting insights
As the support is known, the measurement vector representation will change to

$$y = HA_s x_s + e$$

$x_s$ a vector of length $|s|$ containing those unknowns

$$x_s \sim \mathcal{N}(0, \sigma^2 x I)$$

$A_s$ is a sub-matrix of size $n \times k$
The Oracle Estimation

Q1: is the product of two Gaussian rv's a Gaussian rv?
Answer: NO

Q2: is the product of two Gaussian pdfs a Gaussian Pdf?
Answer: proportional to a Gaussian with a mean = sum of scaled original means

The Oracle Estimation

\[P(x_s|y) = \frac{P(y|x_s)P(x_s)}{P(y)}\]

\[P(x_s) = \frac{1}{(2\pi)^{s/2}\sigma_x^s} \exp \left\{-\frac{x_s^T x_s}{2\sigma_x^2}\right\}\]

\[P(y|x_s) = \frac{1}{(2\pi)^{q/2}\sigma_e^q} \exp \left\{-\frac{||HA_s x_s - y||_2^2}{2\sigma_e^2}\right\}\]

\[P(x_s|y) \propto \exp \left\{-\frac{x_s^T x_s}{2\sigma_x^2} - \frac{||HA_s x_s - y||_2^2}{2\sigma_e^2}\right\}\]
The Oracle Estimation

\[
\hat{x}_{s}^{MAP-oracle} = \max_{x_s} P(x_s|y)
\]

\[
= \max_{x_s} \exp \left\{ -\frac{x_s^T x_s}{2\sigma_x^2} - \frac{||HA_s x_s - y||_2^2}{2\sigma_e^2} \right\}
\]

\[
= \left( \frac{1}{\sigma_e^2} A_s^T H^T H A_s + \left( \frac{1}{\sigma_x^2} I \right)^{-1} \right)^{-1} \frac{1}{\sigma_e^2} A_s^T H^T y
\]

- What about the MMSE-ORACLE minimizer ??
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Return to our main assumption where we have no knowledge about the support

\[ P(x|y) = \sum_{s \in \Omega} P(x|s, y)P(s|y) \]

\[ \hat{x}^{MAP} = \arg \max_x P(x|y) = \arg \max_{s \in \Omega, x_s} P(x_s|s, y)P(s|y) \]

This means that we search for the support and the values of the entries in the support in one process

The MAP Estimation
The MAP Estimation

\[ \hat{x}^{MAP} = \arg \max_x P(x|y) = \arg \max_{s \in \Omega, x_s} P(x_s|s, y)P(s|y) \]

\[ P(x_s|s, y)P(s|y) = \frac{P(y|s, x_s)P(x_s|s)}{P(y|s)} \cdot \frac{P(y|s)P(s)}{P(y)} \]

\[ = \frac{P(y|s, x_s)P(x_s|s)P(s)}{P(y)}. \]

\[ \hat{x}^{MAP} = \arg \max_x P(x|y) \]

\[ = \arg \max_{x, s} P(y|s, x_s)P(x_s|s)P(s). \]

\[ P(y|s, x_s) = \frac{1}{(2\pi)^{q/2}\sigma_e^q} \exp \left\{ -\frac{\|HA_s x_s - y\|^2}{2\sigma_e^2} \right\} \]

\[ P(x_s|s) = \frac{1}{(2\pi)^{k/2}\sigma_x^k} \exp \left\{ -\frac{\|x_s\|^2}{2\sigma_x^2} \right\} \]
The MAP Estimation

\[ \hat{x}^{\text{MAP}} = \arg \min_{x, s} \left\{ \frac{||HA_s x_s - y||_2^2}{2\sigma_e^2} + \frac{||x_s||_2^2}{2\sigma_x^2} + k \log(\sqrt{2\pi\sigma_x}) - \log(P(s)) \right\} \]

\[ \hat{x}^{\text{MAP}} = \arg \max_x P(x|y) \]
\[ = \arg \max_{x, s} P(y|s, x_s)P(x_s|s)P(s) \]

\[ P(y|s, x_s) = \frac{1}{(2\pi)^{q/2}\sigma_e^q} \exp \left\{ -\frac{||HA_s x_s - y||_2^2}{2\sigma_e^2} \right\} \]

\[ P(x_s|s) = \frac{1}{(2\pi)^{k/2}\sigma_x^k} \exp \left\{ -\frac{||x_s||_2^2}{2\sigma_x^2} \right\} \]
The MAP Estimation

\[ \hat{x}^{MAP} = \arg \min_{x,s} \left\{ \frac{\|HA_s x_s - y\|^2}{2\sigma_e^2} + \frac{\|x_s\|^2}{2\sigma_x^2} + k \log(\sqrt{2\pi\sigma_x}) - \log(P(s)) \right\} \]

\[ P(s) = Const \cdot \exp(-\alpha|s|) \]

\[ \hat{x}^{MAP} = \arg \min_{x,s} \left\{ \frac{\|HA_s x_s - y\|^2}{2\sigma_e^2} + \frac{\|x_s\|^2}{2\sigma_x^2} + (\alpha + \log(\sqrt{2\pi\sigma_x}))|s| \right\} \]

\[ |s| = \|x\|_0 \]

\[ \hat{x}^{MAP} = \arg \min_{x} \left\{ \frac{\|Hx - y\|^2}{2\sigma_e^2} + \frac{\|x\|^2}{2\sigma_x^2} + (\alpha + \log(\sqrt{2\pi\sigma_x}))\|x\|_0 \right\} \]
The MAP Estimation

\[ \hat{x}_{MAP} = \arg \min_{x, s} \left\{ \frac{||HA_s x_s - y||^2}{2\sigma_e^2} + \frac{||x_s||^2}{2\sigma_x^2} + k \log(\sqrt{2\pi\sigma_x}) - \log(P(s)) \right\} \]

\[ P(s) = \delta(|s| - k) \]

\[ \hat{x}_{MAP} = \arg \min_{x, s} \left\{ \frac{||HA_s x_s - y||^2}{2\sigma_e^2} + \frac{||x_s||^2}{2\sigma_x^2} + k \log(\sqrt{2\pi\sigma_x}) - \log(P(s)) \right\} \]

\[ \hat{x}_s^* = \left( \frac{1}{\sigma_e^2} A_s^T H^T H A_s + \frac{1}{\sigma_x^2} I \right)^{-1} \left( \frac{1}{\sigma_e^2} A_s^T H^T y \right) \]

To use this expression, we sweep over all possible \( s \) and evaluate the Oracle Estimator penalty for each of them and choose the one leading to the least penalty !!!!
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
Approximating the MAP Estimator

\[ \hat{x}_s^* = \left( \frac{1}{\sigma_e^2} A_s^T H^T H A_s + \frac{1}{\sigma_x^2} I \right)^{-1} \frac{1}{\sigma_e^2} A_s^T H^T y \]

\( k = 1 \)

\( \tilde{a}_i \) is column \( i \) in the matrix \( HA \)

\[ \text{Val}(i) = \frac{\tilde{a}_i^T y}{\frac{\|\tilde{a}_i\|_2^2}{\sigma_e^2} + \frac{1}{\sigma_x^2}}. \]
Assuming that $k=1$, we find the column maximizing the term

$$\text{Val}(i) = \frac{||\tilde{a}_i^T y||_2^2}{\frac{||\tilde{a}_i||_2^2}{\sigma_e^2} + \frac{1}{\sigma_x^2}}.$$

If the columns $\tilde{a}_i$ are of the same norm, this is just maximizing the absolute inner product of the measurement vector with each column in HA.

Can propose a greedy algorithm that accumulates the elements of $k$ one by one.

Leads to a variant of the Matching-Pursuit algorithm, thus, we can regard OMP (or any of its variants) as an approximation of the MAP Estimation problem.
Approximating the MAP Estimator

- It should be noted that the OMP is slightly different from the greedy algorithm that emerges here.

- Choosing the second entry will be through maximizing

\[
\text{Val}(i_2) = \begin{bmatrix} \tilde{a}_{i_1}^T y & \tilde{a}_{i_2}^T y \end{bmatrix} \left[ \frac{\|\tilde{a}_{i_1}\|_2^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} \frac{\tilde{a}_{i_1}^T \tilde{a}_{i_2}}{\sigma_e^2} + \frac{\|\tilde{a}_{i_2}\|_2^2}{\sigma_e^2} + \frac{1}{\sigma_x^2} \right]^{-1} \begin{bmatrix} \tilde{a}_{i_1}^T y \\ \tilde{a}_{i_2}^T y \end{bmatrix}
\]

- This is different from computing the residual from \( y \) after removal of the first column and repeating.
Outline

- Motivation
- Stochastic Model & Estimation Goals
- Background on MAP
- Background on MMSE
- The Oracle Estimation
- The MAP Estimation
- Approximating the MAP Estimator
- The MMSE Estimation
The MMSE Estimation

- As mentioned earlier

\[ \hat{x}_{MMSE}^{MMSE} = \frac{\int x P(x|y)dx}{\int P(x|y)dx} = \int x P(x|y)dx = E(x|y) \]

- Can be written to include prior of the support

\[ \hat{x}_{MMSE}^{MMSE} = E(x|y) = \int x P(x|y)dx \]

\[ = \sum_{s \in \Omega} P(s|y) \int_{x} x P(x|s, y)dx \]
The MMSE Estimation

\[
\hat{X}^{MMSE} = \sum_{s \in \Omega} P(s|y) \left( \frac{1}{\sigma_e^2} A_s^T H^T H A_s + \frac{1}{\sigma_x^2} I \right)^{-1} \frac{1}{\sigma_e^2} A_s^T H^T y
\]

- The MMSE is a weighted average of many oracle estimators, each weighted by its likelihood to be correct \(P(s|y)\).
- How to develop an expression for the probability \(P(s|y)\)?!
The MMSE Estimation

\[ P(s|y) = \frac{P(y|s)P(s)}{P(y)} \]

\( P(s) \) is constant over all the supports \( s \in \mathcal{O} \) such that \( |s| = k \).

\( P(y) \) can be considered as a normalizing factor.

\[ P(y|s) = \int_{x_s} P(y|s, x_s) P(x_s|s) \, dx_s \]

\( P(x_s|s) \) is known to be Gaussian, \( \mathcal{N}(0, \sigma_x^2 I) \).

\( y \) is also Gaussian, with \( \mathbf{H} \mathbf{A}_s x_s \) as its mean, and \( \sigma_e^2 \mathbf{I} \) as its covariance.
The MMSE Estimator

\[ P(y|s) \propto \int_{v \in \mathbb{R}^k} \exp \left\{ -\frac{\|HA_s v - y\|^2_2}{2\sigma^2_e} - \frac{\|v\|^2_2}{2\sigma^2_x} \right\} dv \]

\[ \frac{\|HA_s v - y\|^2_2}{2\sigma^2_e} + \frac{\|v\|^2_2}{2\sigma^2_x} = \frac{1}{2} (v - h_s)^T Q_s (v - h_s) - \frac{1}{2} h_s^T Q_s h_s + \frac{1}{2\sigma^2_e} \|y\|^2_2 \]

\[ Q_s = \frac{1}{\sigma^2_e} A_s^T H^T H A_s + \frac{1}{\sigma^2_x} I \]

and

\[ h_s = \frac{1}{\sigma^2_e} Q_s^{-1} A_s^T H^T y \]
The MMSE Estimator

\[ P(y|s) \propto \int_{v \in \mathbb{R}^k} \exp \left\{ -\frac{||HA_s v - y||^2_2}{2\sigma_e^2} - \frac{||v||^2_2}{2\sigma_x^2} \right\} dv \]

\[
\frac{||HA_s v - y||^2_2}{2\sigma_e^2} + \frac{||v||^2_2}{2\sigma_x^2} = \frac{1}{2} (v - h_s)^T Q_s (v - h_s) - \frac{1}{2} h_s^T Q_s h_s + \frac{1}{2\sigma_e^2} ||y||^2_2
\]
The MMSE Estimator

\[ P(y|s) \propto \int_{v \in \mathbb{R}^k} \exp \left\{ -\frac{\|HA_s v - y\|_2^2}{2\sigma_e^2} - \frac{\|v\|_2^2}{2\sigma_x^2} \right\} dv \]

\[ \frac{\|HA_s v - y\|_2^2}{2\sigma_e^2} + \frac{\|v\|_2^2}{2\sigma_x^2} = \frac{1}{2} (v - h_s)^T Q_s (v - h_s) - \frac{1}{2} h_s^T Q_s h_s + \frac{1}{2\sigma_e^2} \|y\|_2^2 \]

\[ P(y|s) \propto \exp \left\{ \frac{1}{2} h_s^T Q_s h_s - \frac{1}{2\sigma_e^2} \|y\|_2^2 \right\} \cdot \int_{v \in \mathbb{R}^k} \exp \left\{ -\frac{1}{2} (v - h_s)^T Q_s (v - h_s) \right\} dv. \]
The MMSE Estimator

\[ P(y|s) \propto \int_{v \in \mathbb{R}^k} \exp \left\{ -\frac{\|HA_s v - y\|^2}{2\sigma_e^2} - \frac{\|v\|^2}{2\sigma_x^2} \right\} dv \]

\[ = \exp \left\{ \frac{1}{2} h_s^T Q_s h_s - \frac{1}{2\sigma_e^2} \|y\|^2 \right\} \]

\[ \sqrt{(2\pi)^k \det(Q_s^{-1})} \]

\[ \int_{v \in \mathbb{R}^k} \exp \left\{ -\frac{1}{2} (v - h_s)^T Q_s (v - h_s) \right\} dv \]

\[ P(y|s) \propto \exp \left\{ \frac{1}{2} h_s^T Q_s h_s \right\} \cdot \sqrt{\det(Q_s^{-1})} \]
Evaluate each of the above terms and normalize them for all \( s \in \Omega \).

This will give us the pdf \( P(s|y) \).

The MMSE estimator will be

\[
\hat{x}_{\text{MMSE}} = \frac{\sum_{s \in \Omega} q_s \left( \frac{1}{\sigma_e^2} A_s^T H^T H A_s + \frac{1}{\sigma_x^2} I \right)^{-1} \frac{1}{\sigma_e^2} A_s^T H^T y}{\sum_{s \in \Omega} q_s}
\]
Thank You

Questions?