Topics in MMSE Estimation for Sparse Approximation*

Michael Elad

The Computer Science Department

The Technion – Israel Institute of technology

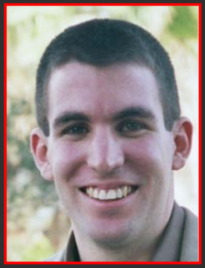
Haifa 32000, Israel

Workshop: Sparsity and Computation

Hausdorff center of Mathematics
University of Bonn
June 7-11, 2010

Joint work with

Irad Yavneh



Matan Protter

Javier Turek

The CS Department, The Technion

Part I - Motivation

Denoising By Averaging Several Sparse Representations

Sparse Representation Denoising

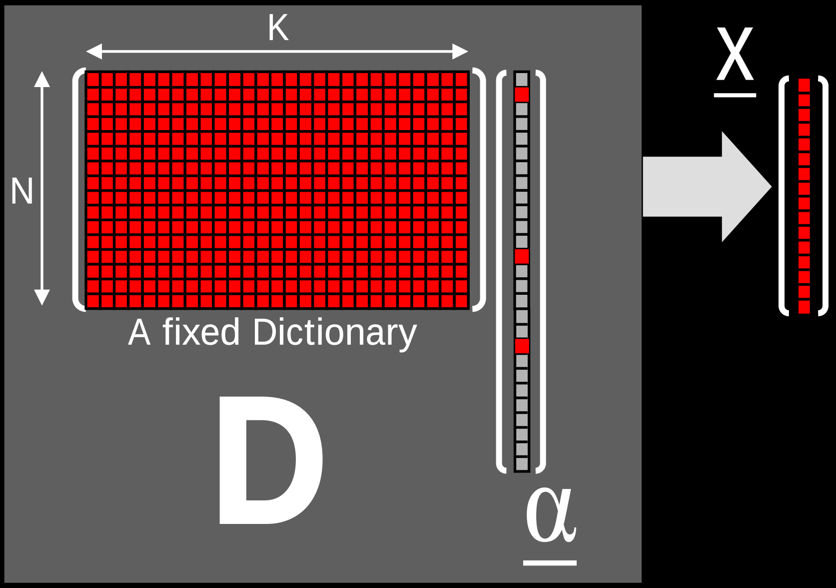
- Sparse representation modeling:
- Assume that we get a noisy measurement vector

$$\underline{y} = \underline{x} + \underline{v} = \mathbf{D}\underline{\alpha} + \underline{v}$$

where is AWGN

- Our goal – recovery of x (or $\underline{\alpha}$).
- The common practice – Approximate the solution of

$$\min_{\underline{\alpha}} \left\|\underline{\alpha}\right\|_0^0 \text{ s.t. } \left\|\mathbf{D}\underline{\alpha} - \underline{y}\right\|_2^2 \le \varepsilon^2$$



Orthogonal Matching Pursuit

OMP finds one atom at a time for approximating the solution of $\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2^2 \le \varepsilon^2$

Initialization

$n = 0,\ \underline{\alpha}^0 = 0$

$\underline{r}^0 = \underline{y} - \mathbf{D}\underline{\alpha}^0 = \underline{y}$

and $S^0 = \{\ \}$

$n = n + 1$

Main Iteration

1. Compute $E(i) = \min_z \left\| z \cdot \underline{d}_i - \underline{r}^{n-1} \right\|$ for $1 \le i \le K$
2. Choose i_0 s.t. $\forall 1 \le i \le K,\ E(i_0) \le E(i)$
3. Update S^n : $S^n = S^{n-1} \cup \{i_0\}$
4. LS : $\underline{\alpha}^n = \min_{\underline{\alpha}} \|\mathbf{D}\underline{\alpha} - \underline{y}\|$ s.t. $\text{supp}\{\underline{\alpha}\} = S^n$
5. Update Residual : $\underline{r}^n = \underline{y} - \mathbf{D}\underline{\alpha}^n$

$\|\underline{r}^n\|_2 \le \varepsilon$

No → $n = n + 1$

Yes → Stop

Using several Representations

Consider the denoising problem

$$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2^2 \le \varepsilon^2$$

and suppose that we can find a group of J candidate solutions

$$\{\underline{\alpha}_j\}_{j=1}^{J}$$

such that

$$\forall j \quad \begin{Bmatrix} \|\underline{\alpha}_j\|_0^0 << N \\ \|\mathbf{D}\underline{\alpha}_j - \underline{y}\|_2^2 \le \varepsilon^2 \end{Bmatrix}$$

Basic Questions:

- **What** could we do with such a set of competing solutions in order to better denoise $\underline{y}$?
- **Why** should this help?
- **How** shall we practically find such a set of solutions?

- Relevant work: [Leung & Barron ('06)] [Larsson & Selen ('07)] [Schintter et. al. (`08)] [Elad and Yavneh ('08)] [Giraud ('08)] [Protter et. al. ('10)] …

Generating Many Representations

Our* Answer: Randomizing the OMP

Initialization

$n=0,\ \underline{\alpha}^0=0$

$\underline{r}^0=\underline{y}-\mathbf{D}\underline{\alpha}^0=\underline{y}$

and $S^0=\{\ \}$

$n=n+1$

Main Iteration

1. Compute $E(i)=\min_z \left\| z\cdot\underline{d}_i-\underline{r}^{n-1}\right\|$ for $1\le i\le K$
2. Choose i_0 s.t. $\forall 1\le i\le K,\ E(i_0)\le E(i)$
3. Update S^n : $S^n=S^{n-1}\cup\{i_0\}$
4. LS : $\underline{\alpha}^n=\min_{\underline{\alpha}}\left\|\mathbf{D}\underline{\alpha}-\underline{y}\right\|$ s.t. $\text{supp}\{\underline{\alpha}\}=S^n$
5. Update Residual : $\underline{r}^n=\underline{y}-\mathbf{D}\underline{\alpha}^n$

$\left\|\underline{r}^n\right\|_2\le\varepsilon$

No

* Larsson and Schnitter propose a more complicated and deterministic tree pruning method

Generating Many Representations

Our* Answer: Randomizing the OMP

Initialization

$n = 0,\ \underline{\alpha}^0 = 0$

$\underline{r}^0 = \underline{y} - \mathbf{D}\underline{\alpha}^0 = \underline{y}$

and $S^0 = \{\ \}$

$n = n + 1$

Main Iteration

1. Compute $E(i) = \min_z \left\| z \cdot \underline{d}_i - \underline{r}^{n-1} \right\|$ for $1 \le i \le K$
2. Choose i_0 with probability $\propto \exp\{-c \cdot E(i)\}$
3.
4.
5.

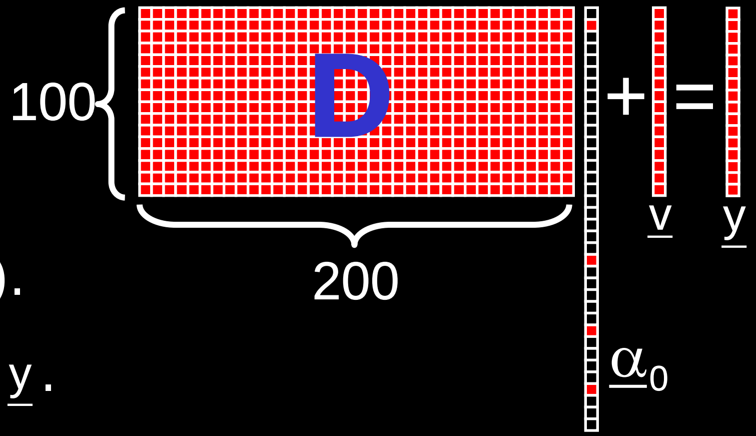
For now, lets set the parameter c manually for best performance. Later we shall define a way to set it automatically

$\left\| \underline{r}^n \right\|_2 \le \varepsilon$

No → $n = n + 1$

Yes → Stop

Lets Try



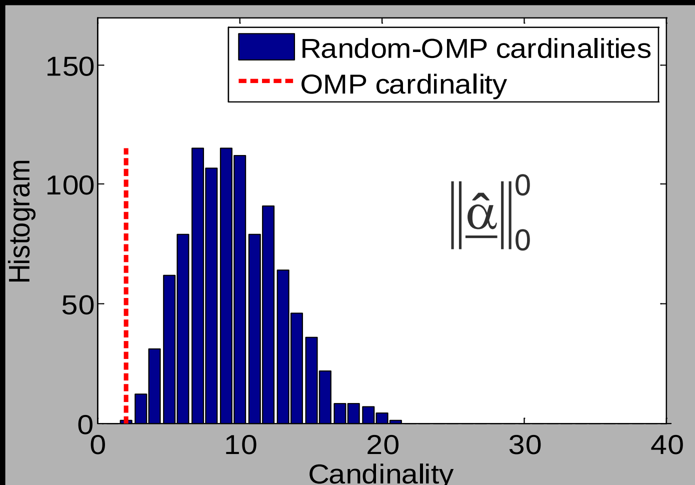
Proposed Experiment :

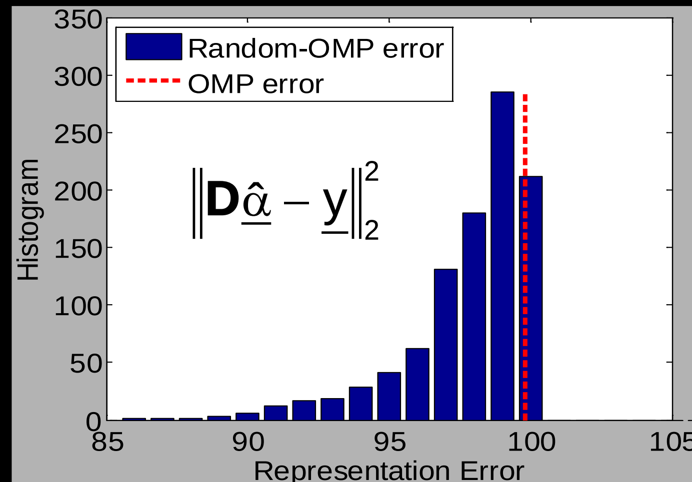
- Form a random **D**.
- Multiply by a sparse vector $\underline{\alpha}_0$ ($\|\underline{\alpha}_0\|_0^0 = 10$).
- Add Gaussian iid noise (σ=1) and obtain $\underline{y}$.
- Solve the problem

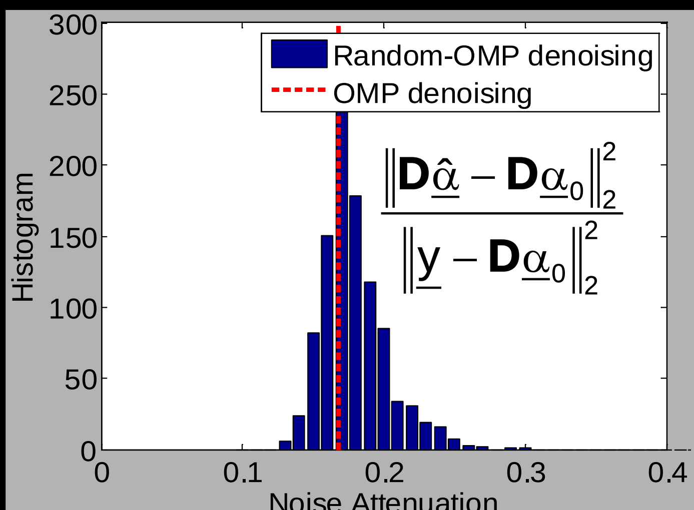
  $$\min_{\underline{\alpha}} \|\underline{\alpha}\|_0^0 \text{ s.t. } \|\mathbf{D}\underline{\alpha} - \underline{y}\|_2^2 \le 100$$

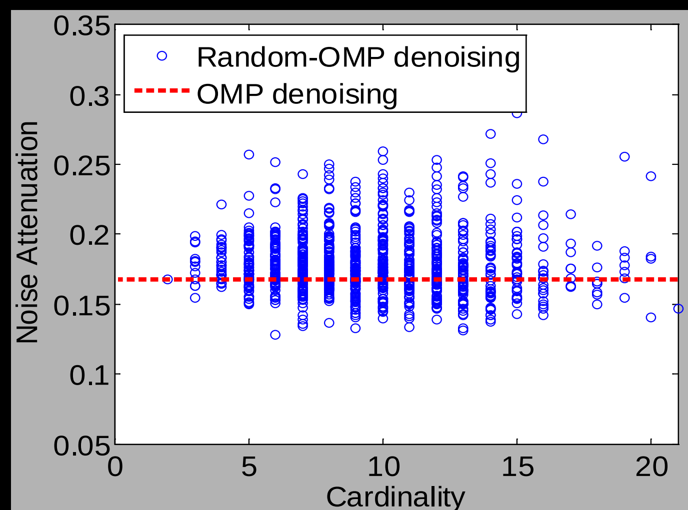
  using OMP, and obtain $\underline{\alpha}^{OMP}$.
- Use RandOMP and obtain $\left\{\underline{\alpha}_j^{RandOMP}\right\}_{j=1}^{1000}$.
- Lets look at the obtained representations ...

Some Observations









We see that

- The OMP gives the sparsest solution
- Nevertheless, it is not the most effective for denoising.
- The cardinality of a representation does not reveal its efficiency.

The Surprise … (to some of us)

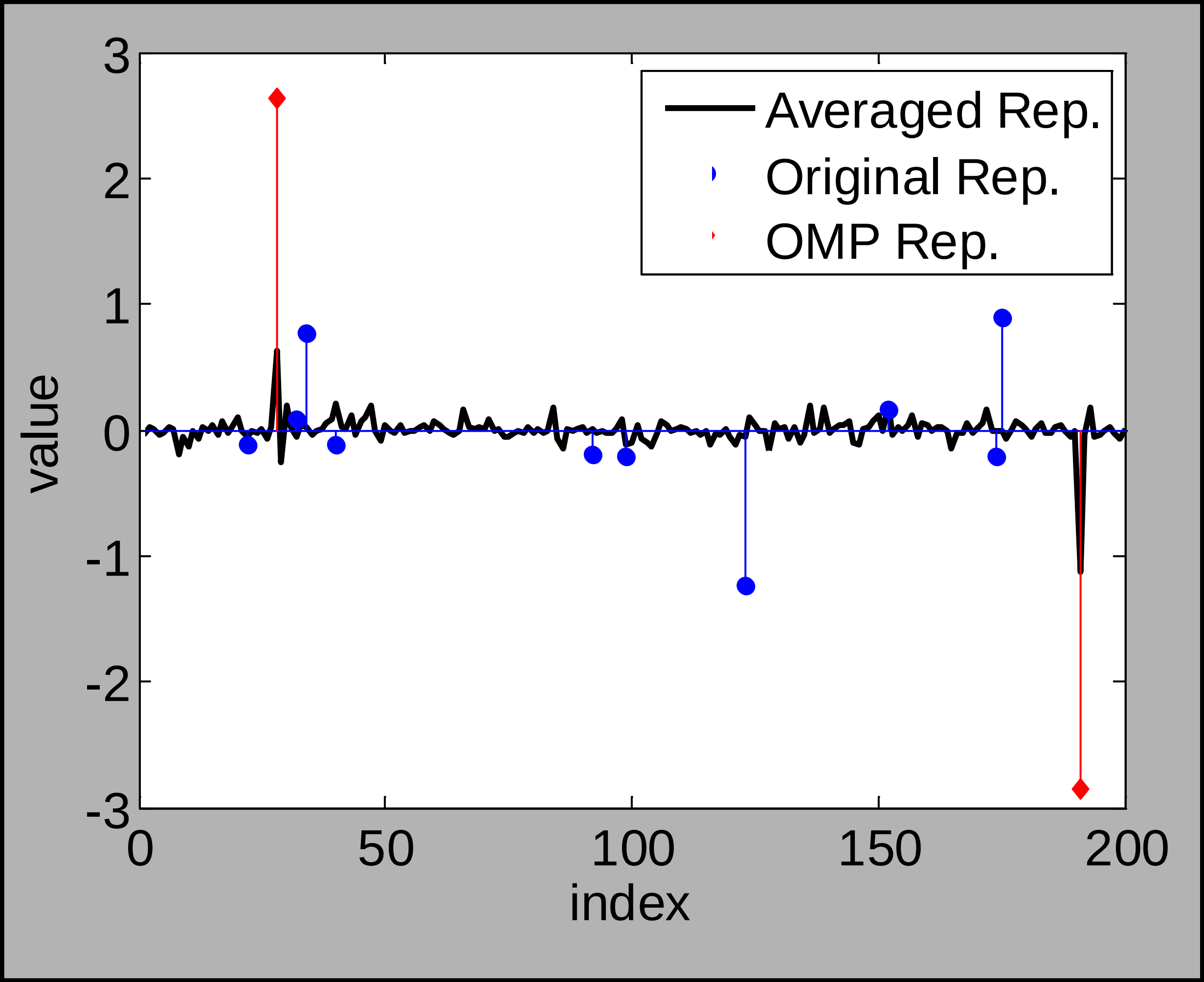
Lets propose the average

$$\underline{\hat{\alpha}} = \frac{1}{1000}\sum_{j=1}^{1000} \underline{\alpha}_j^{\text{RandOMP}}$$

as our representation

This representation **IS NOT SPARSE AT ALL** but its noise attenuation is:

0.06 (OMP gives 0.16)



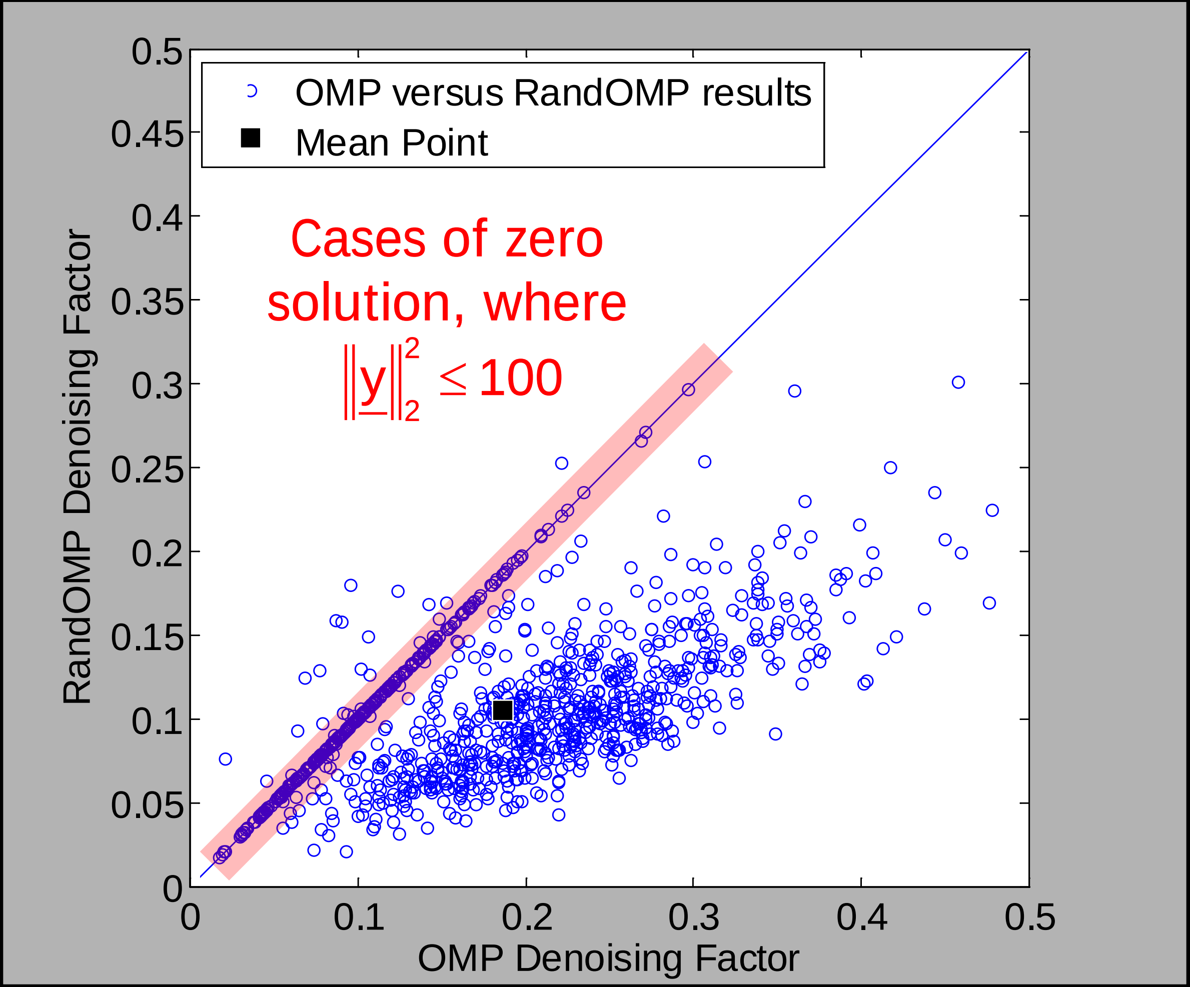
Repeat this Experiment ...

- Dictionary (random) of size N=100, K=200
- True support of $\underline{\alpha}$ is 10
- σ_x=1 and ε=10
- We run OMP for denoising.
- We run RandOMP J=1000 times and average

$$\hat{\underline{\alpha}} = \frac{1}{J}\sum_{j=1}^{J} \underline{\alpha}_j^{RandOMP}$$

- Denoising is assessed by

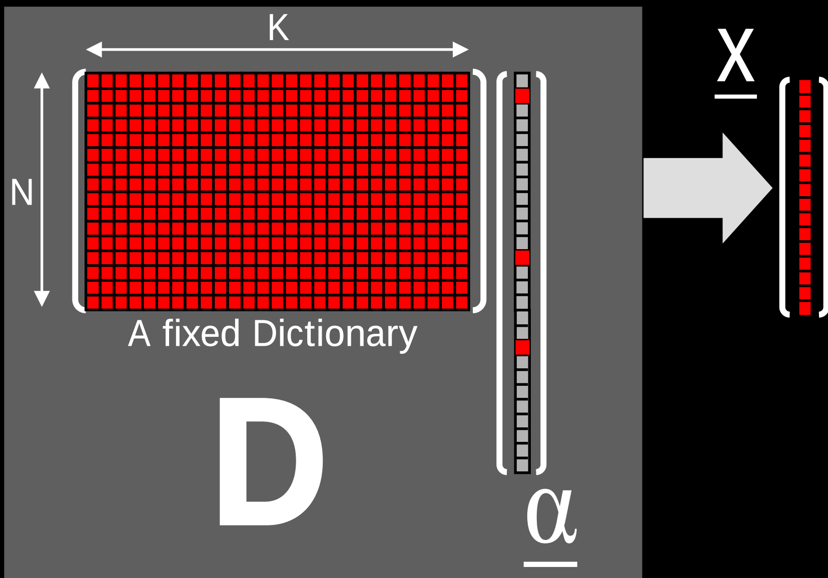
$$\frac{\left\|\mathbf{D}\hat{\underline{\alpha}} - \mathbf{D}\underline{\alpha}_0\right\|_2^2}{\left\|\underline{y} - \mathbf{D}\underline{\alpha}_0\right\|_2^2}$$



Part II - Explanation

It is Time to be More Precise

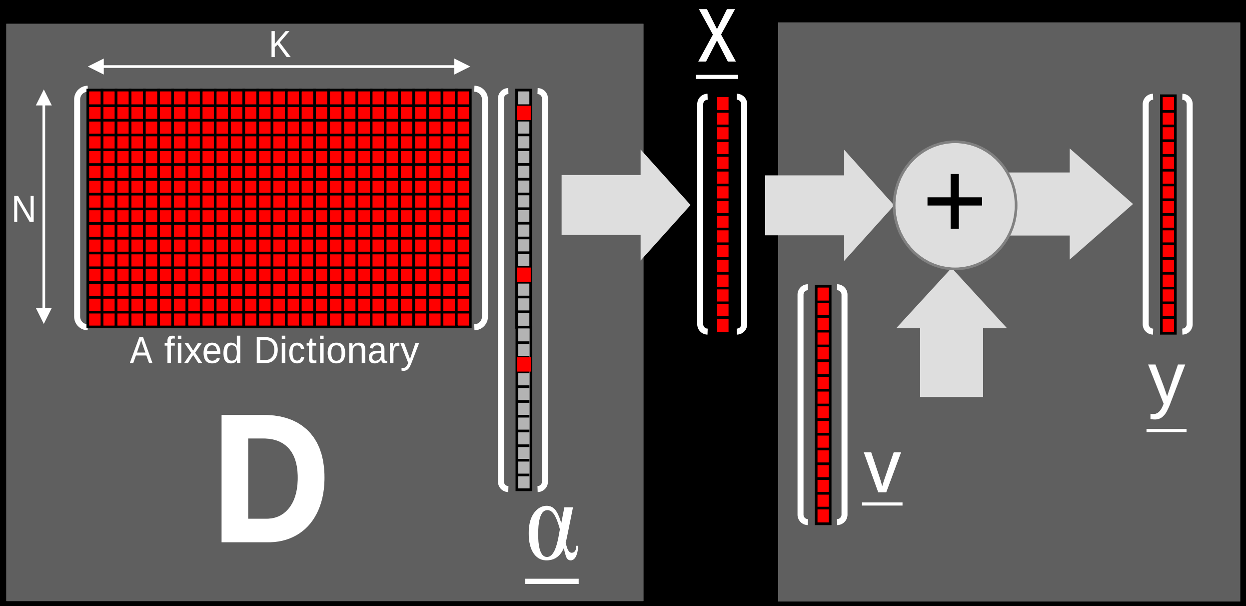
Our Signal Model



- **D** is fixed and known.
- Assume that $\underline{\alpha}$ is built by:
  - Choosing the support s with probability P(s) from all the 2^K possibilities Ω.
  - **Lets assume that P**$(i \in S) = P_i$ **are drawn independently.**
  - Choosing the $\underline{\alpha}_s$ coefficients using iid Gaussian entries $N\left(0, \sigma_x^2\right)$.
- The ideal signal is $\underline{x} = \mathbf{D}\underline{\alpha} = \mathbf{D}_s\underline{\alpha}_s$.

The p.d.f. $P(\underline{\alpha})$ and $P(\underline{x})$ are clear and known

Adding Noise



Noise Assumed:

The noise $\underline{v}$ is additive white Gaussian vector with probability $P_v(\underline{v})$

$$P(\underline{y}|\underline{x}) = C \cdot \exp\left\{-\frac{\|\underline{x}-\underline{y}\|^2}{2\sigma^2}\right\}$$

The conditional p.d.f.'s $P(\underline{y}|\underline{\alpha})$, $P(\underline{\alpha}|\underline{y})$, and even $P(\underline{y}|s)$, $P(s|\underline{y})$, are all clear and well-defined (although they may appear nasty).

The Key – The Posterior $P(\underline{\alpha}|\underline{y})$

We have access to $P\left(\underline{\alpha}\,|\,\underline{y}\right)$

MAP*

$$\hat{\underline{\alpha}}^{MAP} = \underset{\underline{\alpha}}{\text{ArgMax}}\; P(\underline{\alpha}\,|\,\underline{y})$$

Oracle
known
support s

$$\hat{\underline{\alpha}}^{oracle}$$

MMSE

$$\hat{\underline{\alpha}}^{MMSE} = E\left\{\underline{\alpha}\,|\,\underline{y}\right\}$$

* Actually, there is a delicate problem with this definition, due to the unavoidable mixture of continuous and discrete PDF's. The solution is to estimate the MAP's support S.

Lets Start with The Oracle*

$$P(\underline{\alpha} \mid \underline{y}, s) = P(\underline{\alpha}_s \mid \underline{y})$$

$$P(\underline{y} \mid \underline{\alpha}_s) \propto \exp\left\{-\frac{\|\underline{y} - \mathbf{D}_s \underline{\alpha}_s\|^2}{2\sigma^2}\right\}$$

$$P(\underline{\alpha}_s) \propto \exp\left\{-\frac{\|\underline{\alpha}_s\|^2}{2\sigma_x^2}\right\}$$

$$\Rightarrow P(\underline{\alpha}_s \mid \underline{y}) \propto \exp\left\{-\frac{\|\underline{y} - \mathbf{D}_s \underline{\alpha}_s\|^2}{2\sigma^2} - \frac{\|\underline{\alpha}_s\|^2}{2\sigma_x^2}\right\}$$

$$\underline{\hat{\alpha}}_s = \left[\frac{1}{\sigma^2}\mathbf{D}_s^T \mathbf{D}_s + \frac{1}{\sigma_x^2}\mathbf{I}\right]^{-1} \frac{1}{\sigma^2}\mathbf{D}_s^T \underline{y}$$

Comments:

- This estimate is both the MAP and MMSE.
- The oracle estimate of $\underline{x}$ is obtained by multiplication by $\mathbf{D}_s$.

* When s is known

The MAP Estimation

$$\hat{\underline{s}}^{MAP} = \underset{s\in\Omega}{\text{ArgMax}}\ P(s\,|\,\underline{y}) = \underset{s\in\Omega}{\text{ArgMax}} \frac{P(\underline{y}\,|\,s)P(s)}{P(\underline{y})}$$

$$P(\underline{y}\,|\,s) = \int_{\underline{\alpha}_s} P(\underline{y}\,|\,s,\underline{\alpha}_s)P(\underline{\alpha}_s)d\underline{\alpha}_s$$

$$= \ldots.$$

$$\propto \sigma_x^{-|s|}\cdot\exp\left\{\frac{\underline{h}_s^T\mathbf{Q}_s^{-1}\underline{h}_s}{2} - \frac{\log(\text{de}}{2}\right.$$

Based on our prior for generating the support

$$P(s) = \prod_{i\in s} P_i \prod_{j\notin s}(1-P_j)$$

$$\hat{\underline{s}}^{MAP} = \underset{s\in\Omega}{\text{ArgMax}}\ \exp\left\{\frac{\underline{h}_s^T\mathbf{Q}_s^{-1}\underline{h}_s}{2} - \frac{\log(\det(\mathbf{Q}_s))}{2}\right\}\prod_{i\in s}\frac{P_i}{\sigma_x}\prod_{j\notin s}(1-P_j)$$

The MAP Estimation

$$\underline{\hat{s}}^{MAP} = \underset{s\in\Omega}{\text{ArgMax}} \left\{ \frac{\underline{h}_s^T \mathbf{Q}_s^{-1} \underline{h}_s}{2} - \frac{\log(\det(\mathbf{Q}_s))}{2} + \sum_{i\in s} \log\left(\frac{P_i}{\sigma_x}\right) + \sum_{j\notin s} \log\left(1 - P_j\right) \right\}$$

Implications:

- The MAP estimator requires to test all the possible supports for the maximization. For the found support, the oracle formula is used.
- In typical problems, this process is impossible as there is a combinatorial set of possibilities.
- This is why we rarely use exact MAP, and we typically replace it with approximation algorithms (e.g., OMP).

The MMSE Estimation

$$\underline{\hat{\alpha}}^{MMSE} = E\{\underline{\alpha} \mid \underline{y}\} = \sum_{s\in\Omega} P(s \mid \underline{y}) \cdot E\{\underline{\alpha} \mid \underline{y}, s\}$$

This is the oracle for s, as we have seen before

$$E\{\underline{\alpha} \mid \underline{y}, s\} = \underline{\hat{\alpha}}_s = \mathbf{Q}_s^{-1}\underline{h}_s$$

$$P(s \mid \underline{y}) \propto P(s) \cdot P(\underline{y} \mid s) = \ldots$$

$$\propto \exp\left\{\frac{\underline{h}_s^T \mathbf{Q}_s^{-1} \underline{h}_s}{2} - \frac{\log(\det(\mathbf{Q}_s))}{2}\right\} \prod_{i\in s} \frac{P_i}{\sigma_x} \prod_{j\notin s} (1 - P_j)$$

$$\underline{\hat{\alpha}}^{MMSE} = \sum_{s\in\Omega} P(s \mid \underline{y}) \cdot \underline{\hat{\alpha}}_s$$

The MMSE Estimation

$$\underline{\hat{\alpha}}^{MMSE} = E\{\underline{\alpha} \mid \underline{y}\} = \sum_{s \in \Omega} P(s \mid \underline{y}) \cdot E\{\underline{\alpha} \mid \underline{y}, s\}$$

Implications:

$$\underline{\hat{\alpha}}^{MMSE} = \sum_{s \in \Omega} P(s \mid \underline{y}) \cdot \underline{\hat{\alpha}}_s$$

- The best estimator (in terms of L_2 error) is a weighted average of many sparse representations!!!
- As in the MAP case, in typical problems one cannot compute this expression, as the summation is over a combinatorial set of possibilities. We should propose approximations here as well.

The Case of |s|=1 and P_i=P

$$P(s\,|\,\underline{y}) \propto \exp\left\{\frac{\underline{h}_s^T \mathbf{Q}_s^{-1} \underline{h}_s}{2} - \frac{\log(\det(\mathbf{Q}_s))}{2}\right\} \prod_{i\in s} \frac{P_i}{\sigma_x} \prod_{j\notin s} (1 - P_j)$$

$$\propto \exp\left\{\frac{1}{2\sigma^2} \cdot \frac{\sigma_x^2}{\sigma_x^2 + \sigma^2} \cdot (\underline{y}^T \underline{d}_i)^2\right\}$$

This is our c in the Random-OMP

The i-th atom in **D**

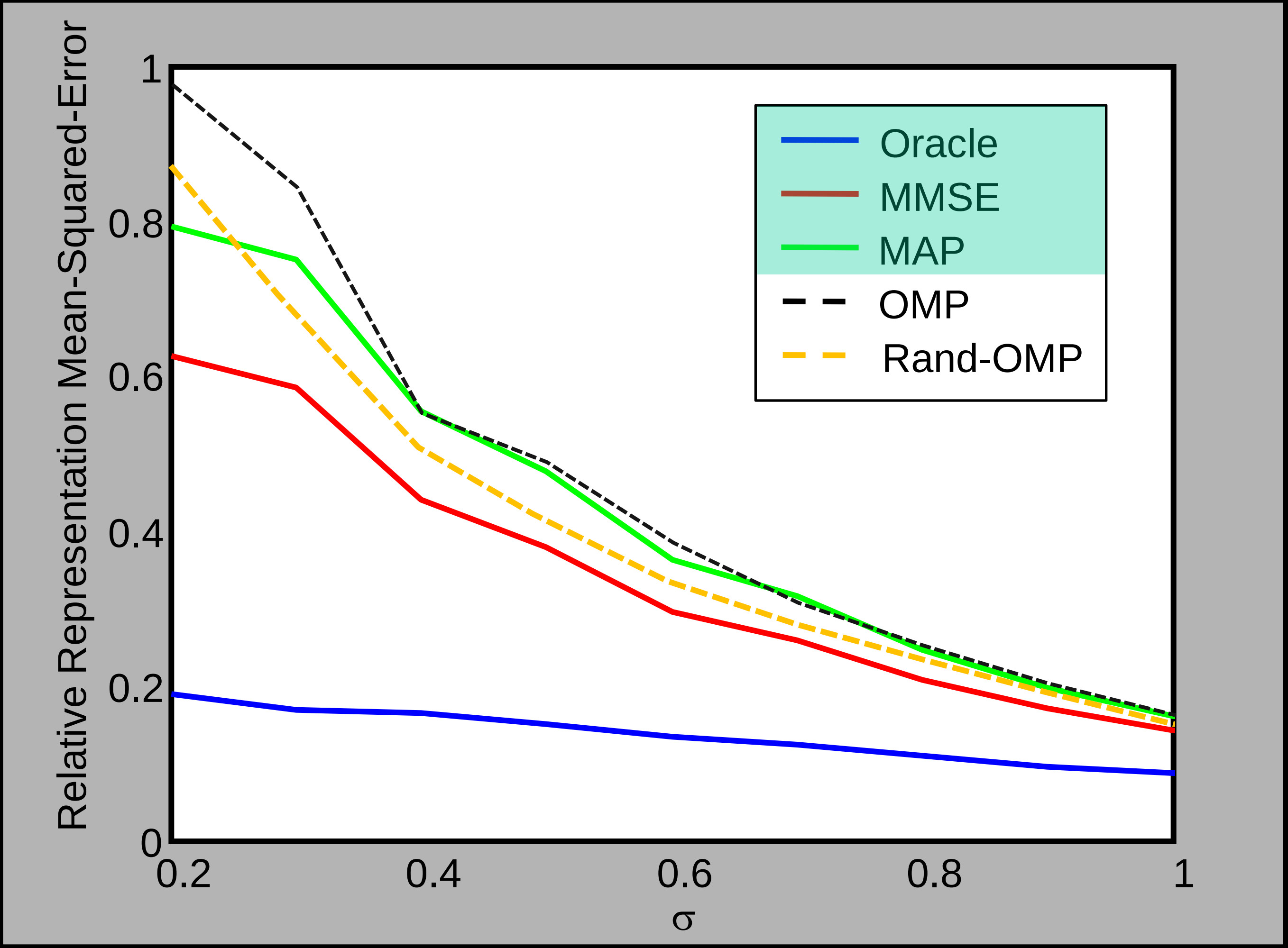
- Based on this we can propose a greedy algorithm for both MAP and MMSE:
  - MAP – choose the atom with the largest inner product (out of K), and do so one at a time, while freezing the previous ones (almost OMP).
  - MMSE – draw at random an atom in a greedy algorithm, based on the above probability set, getting close to P(s|y) in the overall draw (almost RandOMP).

Comparative Results

The following results correspond to a small dictionary (10×16), where the combinatorial formulas can be evaluated as well.

Parameters:

- N,K: 10×16
- P=0.1 (varying cardinality)
- σ_x=1
- J=50 (RandOMP)
- Averaged over 1000 experiments



Part III – Diving In

A Closer Look At the Unitary Case

$$DD^T = D^T D = I$$

Few Basic Observations

Let us denote $\underline{\beta} = \mathbf{D}^T \underline{y}$

$$\mathbf{Q}_s = \frac{1}{\sigma^2}\mathbf{D}_s^T\mathbf{D}_s + \frac{1}{\sigma_x^2}\mathbf{I} = \frac{\sigma^2 + \sigma_x^2}{\sigma^2\sigma_x^2}\mathbf{I}$$

$$\underline{h}_s = \frac{1}{\sigma^2}\mathbf{D}_s^T\underline{y} = \frac{1}{\sigma^2}\underline{\beta}_s$$

$$\underline{\hat{\alpha}}_s^{oracle} = \mathbf{Q}_s^{-1}\underline{h}_s = \frac{\sigma_x^2}{\sigma^2 + \sigma_x^2}\cdot\underline{\beta}_s = c^2\cdot\underline{\beta}_s$$ (The Oracle)

Back to the MAP Estimation

$$P(S|\underline{y}) \propto \exp\left\{\frac{\underline{h}_s^T \mathbf{Q}_s^{-1} \underline{h}_s}{2} - \frac{\log(\det(\mathbf{Q}_s))}{2}\right\} \prod_{i \in s} \frac{P_i}{\sigma_x} \prod_{j \notin s} (1 - P_j)$$

$$\frac{\underline{h}_s^T \mathbf{Q}_s^{-1} \underline{h}_s}{2} = \frac{c^2}{\sigma^2} \left\| \underline{\beta}_s \right\|_2^2$$

$$\frac{\log(\det(\mathbf{Q}_s))}{2} = |S| \log \frac{1}{(1 - c^2)\sigma_x^2}$$

$$P(S|\underline{y}) \propto \prod_{i \in s} \exp\left\{\frac{c^2}{\sigma^2} \beta_i^2\right\} \frac{P_i \sqrt{1 - c^2}}{1 - P_i} = \prod_{i \in s} q_i$$

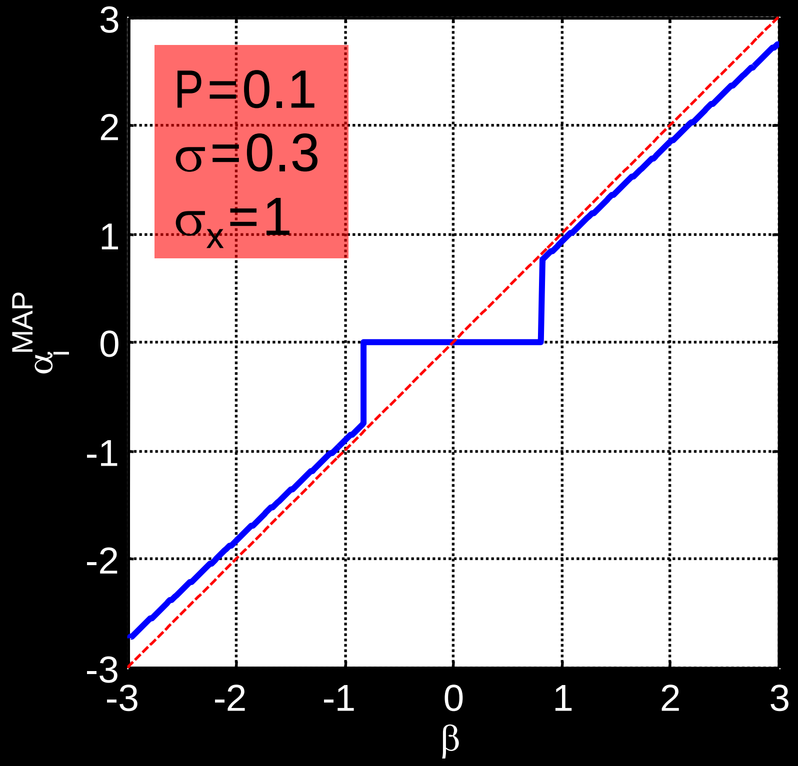
The MAP Estimator

$$q_i = \exp\left\{\frac{c^2}{\sigma^2}\beta_i^2\right\}\frac{P_i\sqrt{1-c^2}}{1-P_i}$$

$\hat{S}_{MAP}$ is obtained by maximizing the expression $$P(S \mid \underline{y}) \propto \prod_{i \in S} q_i$$

Thus, every i such that $q_i > 1$ should be in the support, which leads to

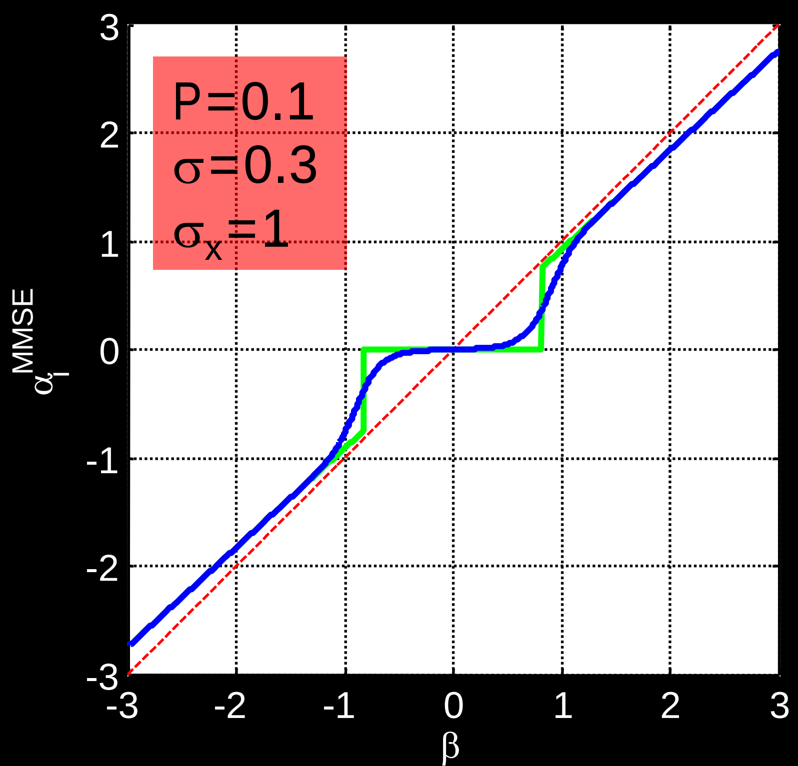
$$\hat{\alpha}_i^{MAP} = \begin{cases} c^2\beta_i & \beta_i^2 > \dfrac{2\sigma^2}{c^2}\log\dfrac{1-P_i}{P_i\sqrt{1-c^2}} \\ 0 & \text{Otherwise} \end{cases}$$



The MMSE Estimation

$$q_i = \exp\left\{\frac{c^2}{\sigma^2}\beta_i^2\right\}\frac{P_i\sqrt{1-c^2}}{1-P_i}$$

Some algebra . . . . and we get that



$$\hat{\alpha}_i^{MMSE} = \frac{q_i}{1+q_i}c^2\beta_i$$

g_i

This result leads to a dense representation vector. The curve is a smoothed version of the MAP one.

What About the Error ?

$$g_i = \frac{q_i}{1+q_i}$$

$$E\left\{\left\|\hat{\underline{\alpha}}^{oracle} - \alpha\right\|_2^2\right\} = \text{trace}\left\{\mathbf{Q}_s^{-1}\right\} = \ldots = \sum_{i=1}^{n} c^2\sigma^2 g_i$$

$$E\left\{\left\|\hat{\underline{\alpha}}^{MMSE} - \alpha\right\|_2^2\right\} = \sum_{s\in\Omega} P\left(s \mid \underline{y}\right)\left[\text{trace}\left\{\mathbf{Q}_s^{-1}\right\} + \left\|\hat{\underline{\alpha}}^{MMSE} - \hat{\underline{\alpha}}_s^{oracle}\right\|_2^2\right]$$

$$= \ldots = \sum_{i=1}^{n} c^2\sigma^2 g_i + \sum_{i=1}^{n} c^4\beta_i^2\left(g_i - g_i^2\right)$$

$$E\left\{\left\|\hat{\underline{\alpha}}^{MAP} - \alpha\right\|_2^2\right\} = \left\|\hat{\underline{\alpha}}^{MAP} - \hat{\underline{\alpha}}^{MMSE}\right\|_2^2 + E\left\{\left\|\hat{\underline{\alpha}}^{MMSE} - \alpha\right\|\right\}$$

$$= \ldots = \sum_{i=1}^{n} c^2\sigma^2 g_i + \sum_{i=1}^{n} c^4\beta_i^2\left(g_i + I_i^{MAP}(1 - 2g_i)\right)$$

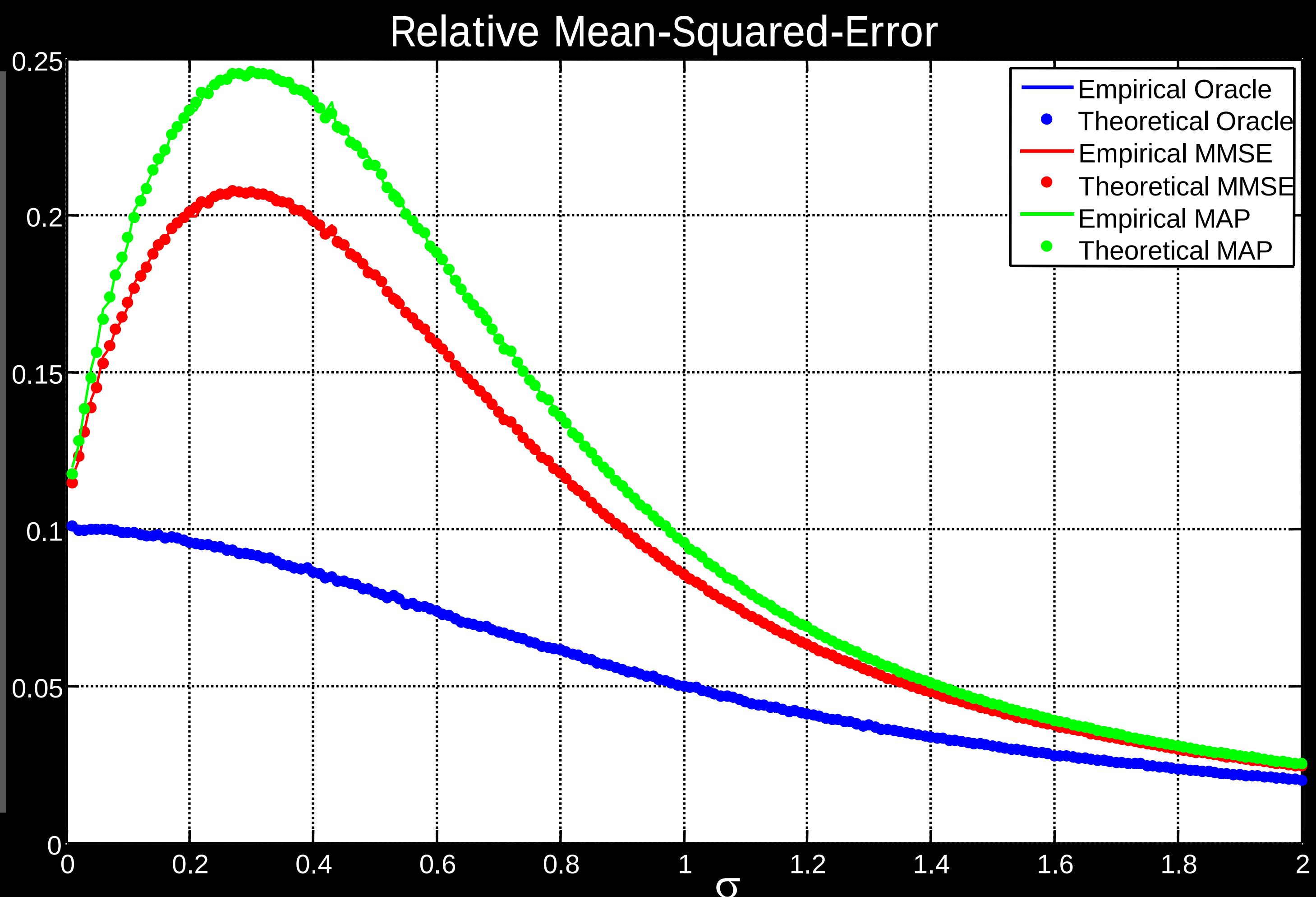
A Synthetic Experiment

The following results correspond to a dictionary of size (100×100)

Parameters:

- n,K: 100×100
- P=0.1
- σ_x=1
- Averaged over 1000 experiments

The average errors are shown relative to $n\sigma^2$



Part IV - Theory

Estimation Errors

Useful Lemma

Let (a_k, b_k) k=1,2, ... ,n be pairs of positive real numbers. Let m be the index of a pair such that

$$\forall k \quad \frac{a_k}{b_k} \le \frac{a_m}{b_m}.$$

$$\frac{\sum_{k=1}^{n} a_k}{\sum_{k=1}^{n} b_k} \le \frac{a_m}{b_m}$$

Equality is obtained only if all the ratios a_k/b_k are equal.

We are interested in this result because :

$$E\left\{\left\|\hat{\underline{\alpha}}^{oracle} - \alpha\right\|_2^2\right\} = \sum_{i=1}^{n} c^2\sigma^2 g_i$$

$$E\left\{\left\|\hat{\underline{\alpha}}^{MMSE} - \alpha\right\|_2^2\right\} = \sum_{i=1}^{n} c^2\sigma^2 g_i + \sum_{i=1}^{n} c^4\beta_i^2\left(g_i - g_i^2\right)$$

$$E\left\{\left\|\hat{\underline{\alpha}}^{MAP} - \alpha\right\|_2^2\right\} = \sum_{i=1}^{n} c^2\sigma^2 g_i + \sum_{i=1}^{n} c^4\beta_i^2\left(g_i + I_i^{MAP}(1 - 2g_i)\right)$$

This leads to ...

Theorem 1 – MMSE Error

Define $G_k = \frac{P_k\sqrt{1-c^2}}{1-P_k}$. Choose m such that $\forall k,\ G_m \le G_k$.

$$\frac{E\left\{\left\|\hat{\underline{\alpha}}^{MMSE}-\underline{\alpha}\right\|_2^2\right\}}{E\left\{\left\|\hat{\underline{\alpha}}^{oracle}-\underline{\alpha}\right\|_2^2\right\}} \le \begin{cases} 1+2\log\frac{1}{4G_m} & G_m \le \frac{e^{-2}}{4} \\ 1+\frac{2}{\sqrt{G_m e}} & \text{Otherwise} \end{cases}$$

$P_k = P = \frac{\ell}{K} << 1$

this error ratio bound becomes

$$\frac{E\left\{\left\|\hat{\underline{\alpha}}^{MMSE}-\underline{\alpha}\right\|_2^2\right\}}{E\left\{\left\|\hat{\underline{\alpha}}^{oracle}-\underline{\alpha}\right\|_2^2\right\}} \le \text{Const}\cdot\log K$$

Theorem 2 – MAP Error

Define $G_k = \frac{P_k\sqrt{1-c^2}}{1-P_k}$. Choose m such that $\forall k,\ G_m \le G_k$.

$$\frac{E\left\{\left\|\underline{\hat{\alpha}}^{MAP} - \alpha\right\|_2^2\right\}}{E\left\{\left\|\underline{\hat{\alpha}}^{oracle} - \alpha\right\|_2^2\right\}} \le \begin{cases} 1 + 2\log\frac{1}{G_m} & G_m \le e^{-1} \\ 1 + \frac{2}{G_m e} & \text{Otherwise} \end{cases}$$

$P_k = P = \frac{\ell}{K} << 1$

this error ratio bound becomes

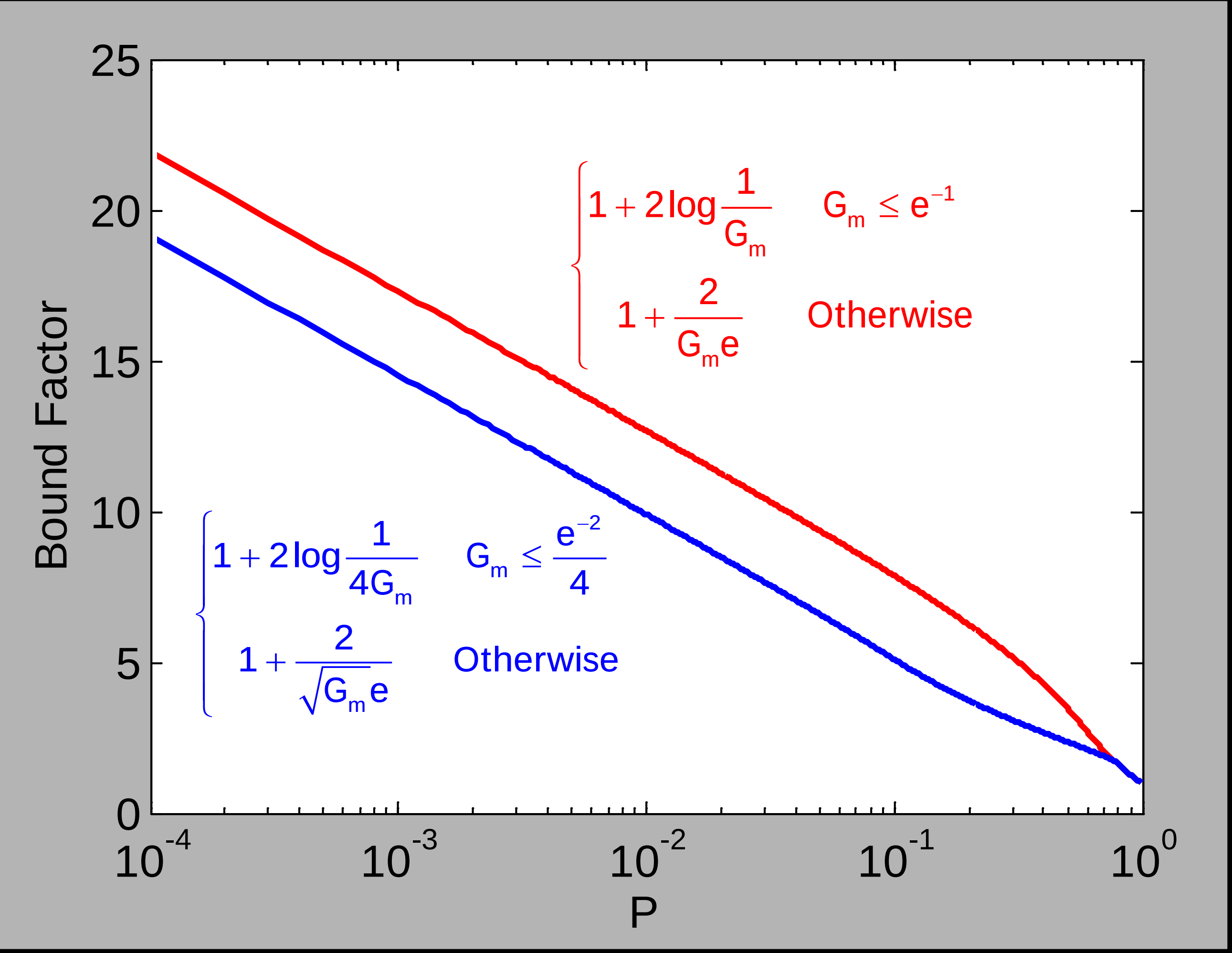
$$\frac{E\left\{\left\|\underline{\hat{\alpha}}^{MMSE} - \alpha\right\|_2^2\right\}}{E\left\{\left\|\underline{\hat{\alpha}}^{oracle} - \alpha\right\|_2^2\right\}} \le \text{Const} \cdot \log K$$

The Bounds' Factors vs. P

Parameters:

- $P=[0,1]$
- $\sigma_x=1$
- $\sigma=0.3$

Notice that the tendency of the two estimators to align for $P\to 0$ is not reflected in these bounds.



Part V – We Are Done

Summary and Conclusions

Today We Have Seen that …

Sparsity and Redundancy are used for denoising of signals/images

How ?

By finding the sparsest representation and using it to recover the clean signal

Can we do better?

Yes! Averaging several rep's lead to better denoising, as it approximates the MMSE

Unitary case?

MAP and MMSE enjoy a closed-form, exact and cheap formulae. Their error is bounded and tightly related to the oracle's error

More on these (including the slides and the relevant papers) can be found in http://www.cs.technion.ac.il/~elad