Weighted Compressed Sensing and Rank Minimization

California Institute of Technology

Samet Oymak
Amin Khajehnejad
Babak Hassibi

Presented in ICASSP 2011,
Prague

Outline

- Compressed Sensing with Prior Info
- Related Work
- Problem Definition
- Contributions
  - A simple and explicit rule for weighted – CS
    - Analysis works for arbitrary number of partitions
  - A proposal for weighted RM
- Technical Details

Compressed Sensing

- Typical compressed sensing setup:
  - $x_0 \in \mathbb{R}^n$ is a sparse vector, $A \in \mathbb{R}^{m \times n}$ is the measurement matrix and $y_0 = Ax_0 \in \mathbb{R}^m$ are the measurements.
  - Aim is to recover x_0 in the case $m << n$.
  - ℓ_1 – minimization:

$$\min \|x\|_1 \quad \text{subject to} \quad Ax = y_0$$

  - A great amount of work is dedicated to sparse recovery.
    - Fast algorithms
    - Signals with certain structure
    - Noise analysis

Compressed Sensing

- Typical compressed sensing setup:
  - $x_0 \in \mathbb{R}^n$ is a sparse vector, $A \in \mathbb{R}^{m \times n}$ is the measurement matrix and $y_0 = Ax_0 \in \mathbb{R}^m$ are the measurements.
  - Aim is to recover x_0 in the case $m << n$.
  - ℓ_1 – minimization:

$$\min \|x\|_1 \quad \text{subject to} \quad Ax = y_0$$

  - A great amount of work is dedicated to sparse recovery.
    - Fast algorithms
    - Signals with certain structure
    - Noise analysis

Non-uniform sparsity

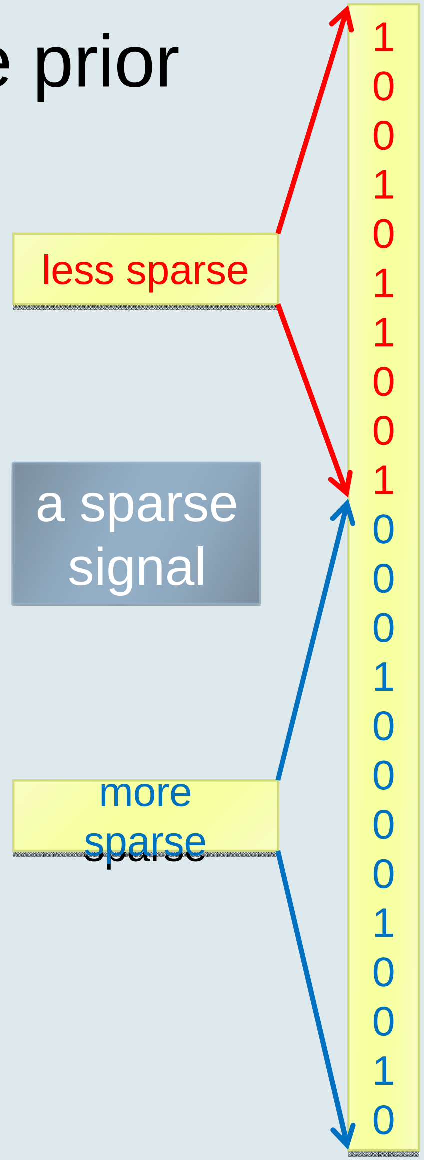
- In many applications, we have some prior knowledge on support of the signal.
  - Natural images
  - Biomedical images
  - Network monitoring data
  - DNA micro-arrays

a sparse signal

$$\begin{bmatrix}1\\0\\0\\1\\0\\1\\1\\0\\0\\1\\0\\0\\0\\1\\0\\0\\0\\0\\1\\0\\0\\1\\0\end{bmatrix}$$

Non-uniform sparsity

- In many applications, we have some prior knowledge on support of the signal.
  - Natural images
  - Biomedical images
  - Network monitoring data
  - DNA micro-arrays

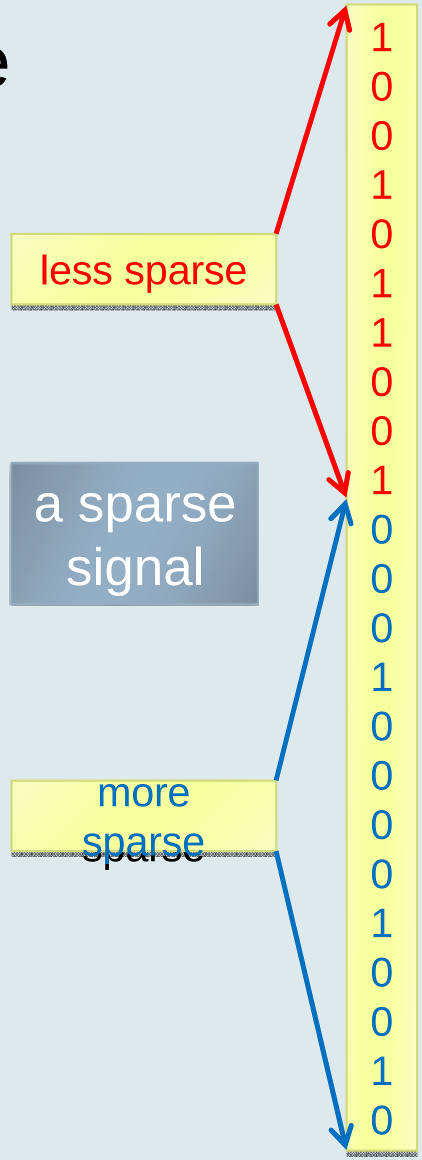


Non-uniform sparsity

- In many applications, we have some prior knowledge on support of the signal.
  - Natural images
  - Biomedical images
  - Network monitoring data
  - DNA micro-arrays
- Modify the algorithm to account for t

Algorithm: Assume $\{S_i\}_{i=1}^{t}$ partitions $\{1,2,\ldots,n\}$ into t disjoint sets. Then weighted ℓ_1 – minimization is given by:

$$\min \sum_{i=1}^{t} w_i \| x_{S_i} \|_1 \quad \text{subject to} \quad Ax = y_0$$



Problem Setting

- For each $i \le t$, let fraction of nonzero entries on S_i p_i be where $0 \le p_i \le 1$.
- We assume the knowledge of $\{p_i\}_{i=1}^{t}$ and $\{S_i\}_{i=1}^{t}$.
- Aim is to minimize number of measurements m while recovering x_0 w.h.p.

> Main question: Given $\{p_i\}_{i=1}^{t}$ and $\{S_i\}_{i=1}^{t}$ how to choose the weights $\{w_i\}_{i=1}^{t}$ to be able to recover x_0 from fewest number of measurements?

Problem Setting

- For each $i \le t$, let fraction of nonzero entries on S_i p_i be where $0 \le p_i \le 1$.
- We assume the knowledge of $\{p_i\}_{i=1}^{t}$ and $\{S_i\}_{i=1}^{t}$.
- Aim is to minimize number of measurements m while recovering x_0 w.h.p.

> Main question: Given $\{p_i\}_{i=1}^{t}$ and $\{S_i\}_{i=1}^{t}$ how to choose the weights $\{w_i\}_{i=1}^{t}$ to be able to recover x_0 from fewest number of measurements?

- Previously analyzed in a number of papers.
  - Vaswani and Lu, “Modified – CS: …” (2010)
  - Khajejnejad et al. “Weighted ℓ_1 – minimization …” (2009)

Our approach

- We assume Gaussian measurements.
  - Entries of $A \in \mathbb{R}^{m \times n}$ is i.i.d. Gaussian
- Analysis is performed via a new technique introduced by Recht et al. for RMP.
  - "Null space conditions and thresholds for RM" (2008)
- Unlike previous results, we are able to find a simple weighting rule for arbitrary number of partitions.
  - Our results suggest: $w_i = 1 - p_i$
  - Consistent with "punish the sparser" intuition
  - An alternative to exhaustive search
- Tradeoff: Analysis is suboptimal.

Strategy

- Main steps:
  1. Find the condition for recovery of x_0
  2. Analyze the condition with the new framework
  3. Find the optimal weights

Strategy

- Main steps:
  1. Find the condition for recovery of x_0
  2. Analyze the condition with the new framework
  3. Find the optimal weights

Lemma (Null Space Condition)

Let K be support set of x_0. Then x_0 can be recovered via weighted ℓ_1 if and only if for all $\mathbf{z} \in N(A)$ we have:

$$F(\mathbf{z},\mathbf{w}) = \sum_{i=1}^{t} w_i \Big(\sum_{j \in S_i \cap K} sgn(x_j) z_j + \sum_{l \in S_i \cap K^c} |z_l| \Big) > 0$$

Strategy

- Main steps:
  1. Find the condition for recovery of x_0
  2. Analyze the condition with the new framework
     1. We need to make sure $F(\mathbf{z},\mathbf{w})$ is positive for all $\mathbf{z}$.
  3. Find the optimal weights

Lemma (Null Space Condition)

Let K be support set of x_0. Then x_0 can be recovered via weighted ℓ_1 if and only if for all $\mathbf{z} \in \mathcal{N}(A)$ we have:

$$F(\mathbf{z},\mathbf{w}) = \sum_{i=1}^{t} w_i \Big(\sum_{j \in S_i \cap K} sgn(x_j) z_j + \sum_{l \in S_i \cap K^c} |z_l| \Big) > 0$$

Strategy

- Main steps:
  1. Find the condition for recovery of x_0
  2. Analyze the condition with the new framework
     1. We need to make sure $F(\mathbf{z},\mathbf{w})$ is positive for all $\mathbf{z}$.
  3. Find the optimal weights

Theorem (Lower bounding $F(\mathbf{z},\mathbf{w})$)
Let $\alpha_i = |S_i|/n$ for all $i \le t$ and $\mu = m/n$. Then $F(\mathbf{z},\mathbf{w})$ is positive for all $\mathbf{z} \in \mathrm{N}(\mathbf{A})$ if the following holds:

$$f(\mathbf{w}) = c_0 \sum_{i=1}^{t} w_i \alpha_i (1 - p_i) - ((1-\mu) \sum_{i=1}^{t} \mathrm{w}_i^2 \alpha_i)^{1/2} > 0$$

where $c_0 = \sqrt{2/\pi}$.

Proof idea

Theorem (Lower bounding $F(\mathbf{z},\mathbf{w})$)

Let $\alpha_i = |S_i|/n$ for all $i \le t$ and $\mu = m/n$. Then $F(\mathbf{z},\mathbf{w})$ is positive for all $\mathbf{z} \in \mathcal{N}(\mathbf{A})$ if the following holds:

$$f(\mathbf{w}) = c_0 \sum_{i=1}^{t} w_i \alpha_i (1 - p_i) - ((1-\mu) \sum_{i=1}^{t} w_i^2 \alpha_i)^{1/2} > 0$$

where $c_0 = \sqrt{2/\pi}$.

- Proof is based on
  - Gordon’s inequalities on Gaussian processes
    - “Some inequalities for Gaussian processes …” (1985)
  - Concentration for Lipschitz functions of Gaussians

Proof idea

Theorem (Lower bounding $F(\mathbf{z},\mathbf{w})$)

Let $\alpha_i = |S_i|/n$ for all $i \le t$ and $\mu = m/n$. Then $F(\mathbf{z},\mathbf{w})$ is positive for all $\mathbf{z} \in \mathrm{N}(\mathbf{A})$ if the following holds:

$$f(\mathbf{w}) = c_0 \sum_{i=1}^{t} w_i \alpha_i (1-p_i) - ((1-\mu)\sum_{i=1}^{t} \mathrm{w}_i^2 \alpha_i)^{1/2} > 0$$

where $c_0 = \sqrt{2/\pi}$.

- Proof is based on
  - Gordon's inequalities on Gaussian processes

In particular, we try to estimate

$$\mathrm{E}\left[\inf_{\|\mathbf{v}\|_2=1} \| \sum_{i=1}^{p} v_i \mathbf{g_i} \|_1 \right]$$

where $\{\mathbf{g}_i\}_{i=1}^{p}$ are i.i.d. Gaussian vectors which corresponds to a basis in $\mathrm{N}(A)$.

Optimal weights

- Main steps:
  1. Find the condition for recovery of x_0
  2. Analyze the condition with the new framework
     1. We need to make sure $F(\mathbf{z}, \mathbf{w})$ is positive for all $\mathbf{z}$.
  3. Find the optimal weights

Theorem (Optimal weights)

Given $\{p_i\}_{i=1}^t$ and $\{S_i\}_{i=1}^t$, in order to minimize m while keeping $f(\mathbf{w})$ positive, we need to choose

$$w_i = 1 - p_i \quad \text{for all} \quad i \leq t$$

- Interestingly, weighting rule does not depend on $|S_i|$.

Reduction in measurements

- Our analysis provides a simple intuition on how weighting is useful.
  - Let $r = n - m$ denote the *dimensionality reduction*.
- For regular ℓ_1 – minimization

$$r_{reg} = c_0^2 (1 - \sum_i \alpha_i p_i)^2$$

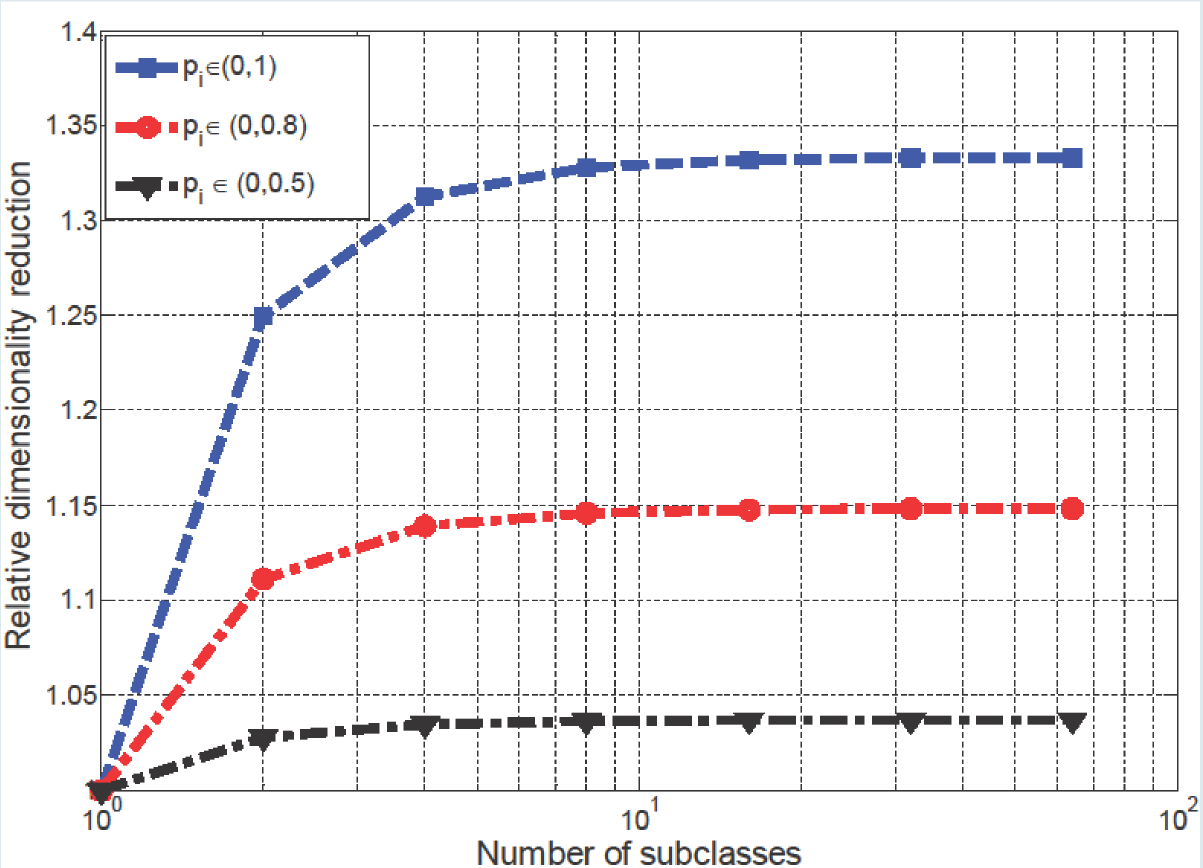
- For weighted ℓ_1 with $w_i = 1 - p_i$

$$r_{opt} = c_0^2 \sum_i (1 - p_i)^2 \alpha_i$$

- From convexity of x^2 we always have: $r_{opt} \geq r_{reg}$

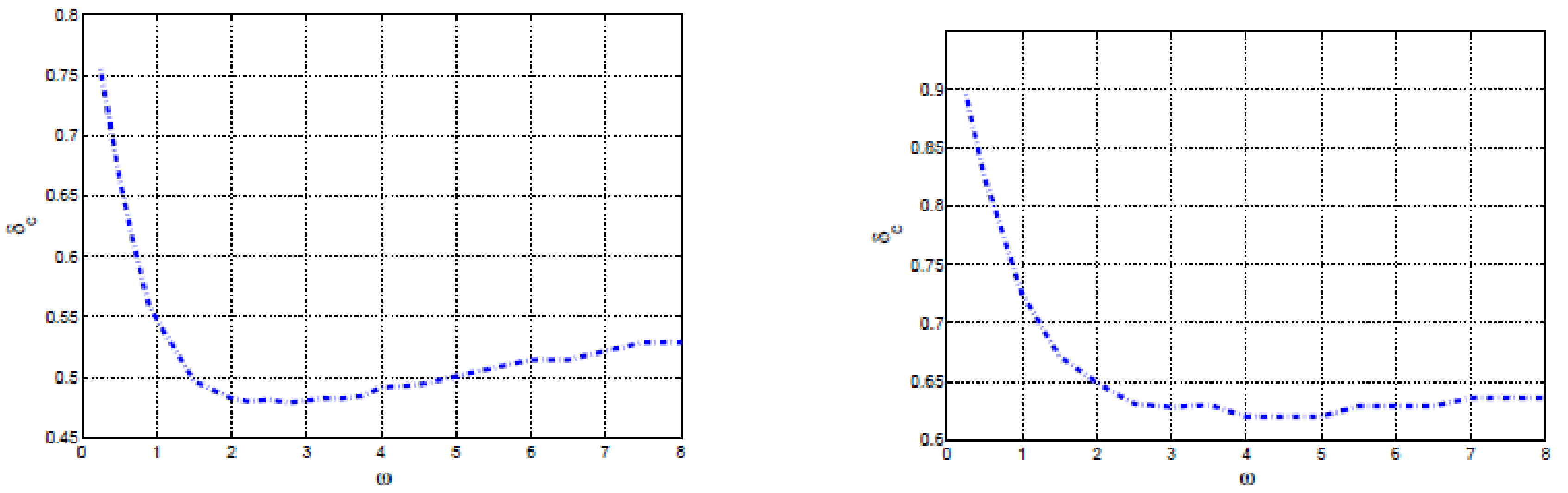
Dimensionality reduction

- Relative reduction in number of measurements (r_{opt} / r_{reg}) as we use more and more weighted sections.
  - Sparsity of partitions is uniformly spaced from 0 to 0.5, 0.8 or 1.



Comparison with previous work

- Khajehnejad et al. studied the optimal weighting for two partitions. Which w_2 / w_1 is optimal?
  - Problem becomes intractable as number of partitions increases

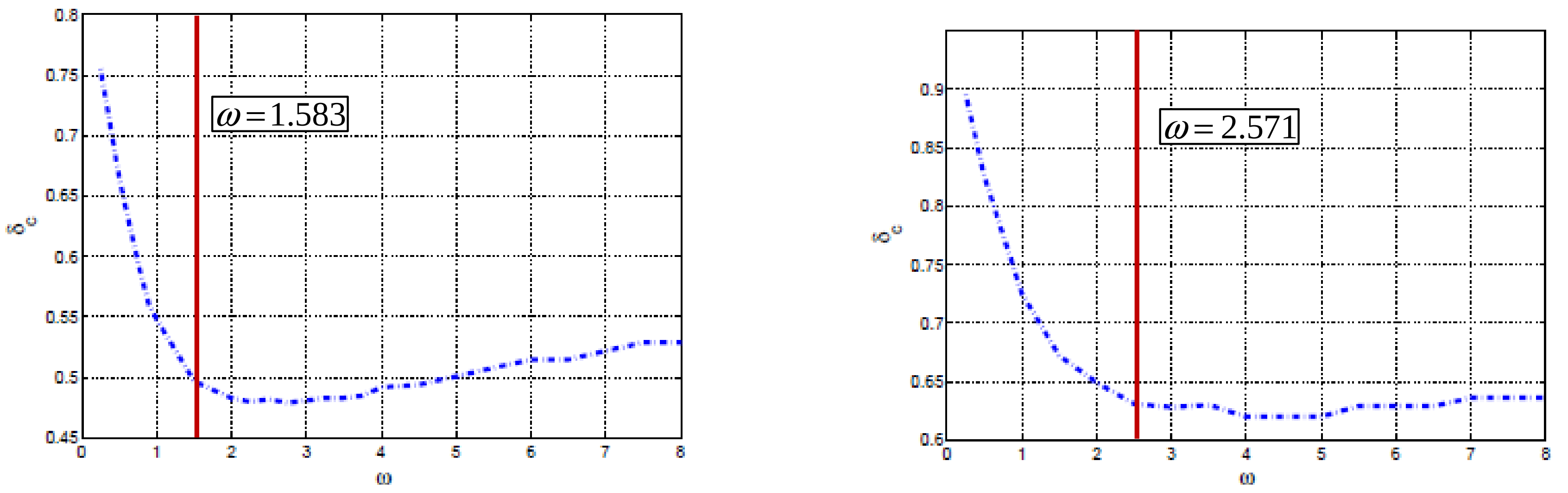


(a) $p_1 = 0.4$, $p_2 = 0.05$. (b) $p_1 = 0.65$, $p_2 = 0.1$.

Figure 2: δ_c as a function of $\omega = \frac{w_{K_2}}{w_{K_1}}$ for $\gamma_1 = \gamma_2 = 0.5$.

Comparison with previous work

- Khajehnejad et al. studied the optimal weighting for two partitions. Which w_2 / w_1 is optimal?
  - Our prediction: $\omega = (1 - p_2)/(1 - p_1)$



(a) $p_1 = 0.4$, $p_2 = 0.05$.

(b) $p_1 = 0.65$, $p_2 = 0.1$.

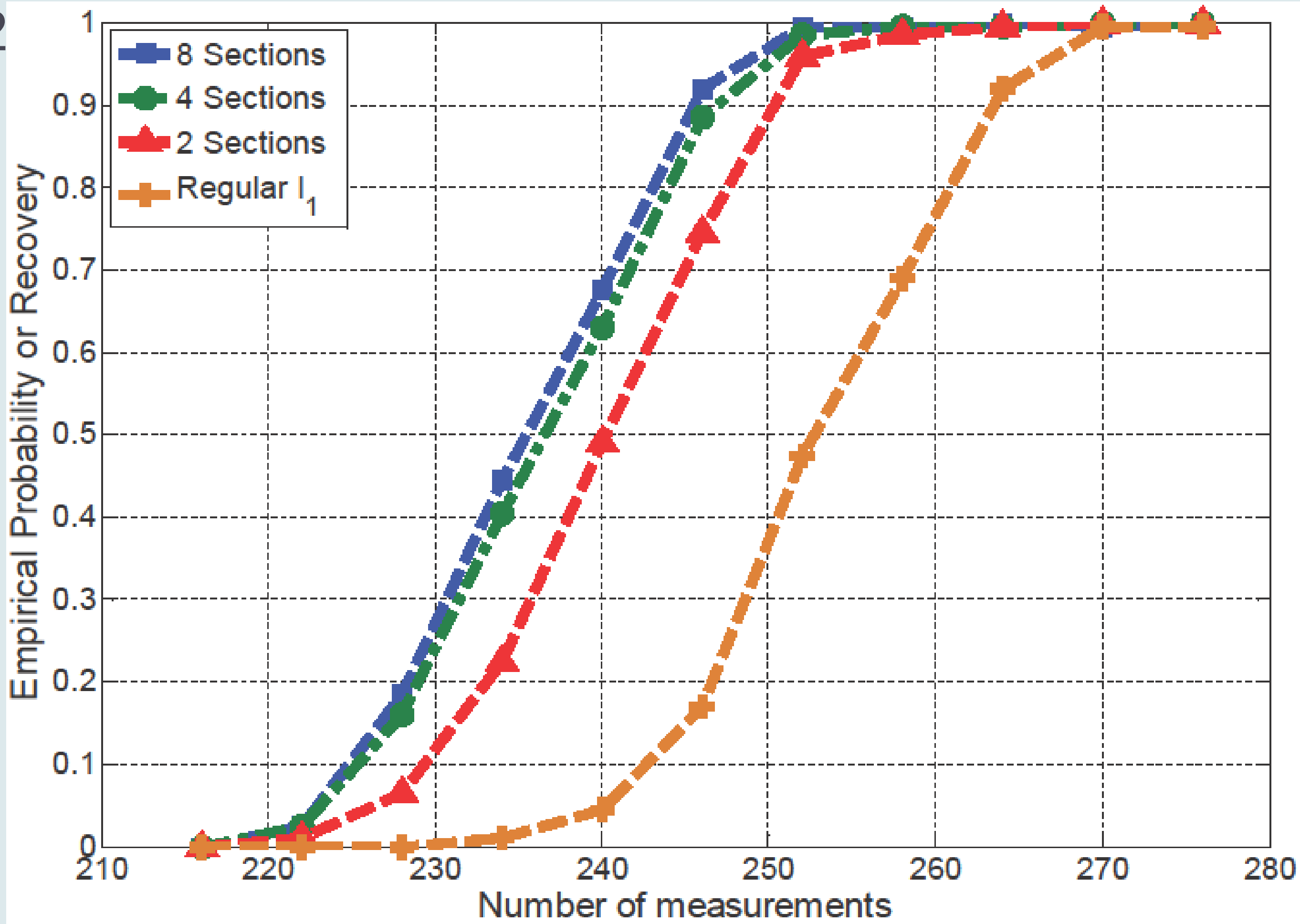
Figure 2: δ_c as a function of $\omega = \frac{w_{K_2}}{w_{K_1}}$ for $\gamma_1 = \gamma_2 = 0.5$.

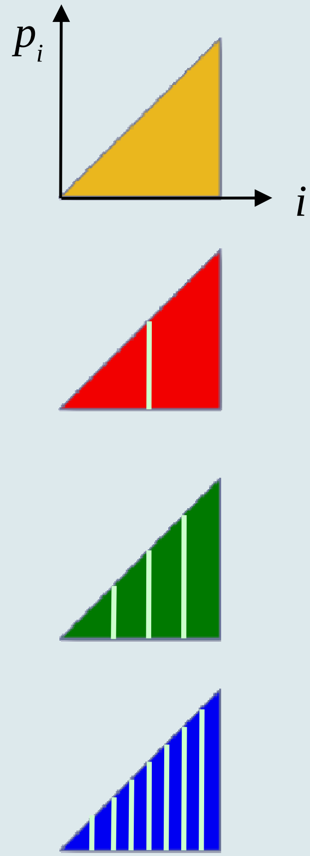
Simulation results

- Empirical recovery curves for the proposed weighting method.

  - 32 .8.

p_i





Extension to Rank Minimization

- Similar to support set of a sparse signal, singular vectors of low rank matrices are not necessarily uniformly distributed.
  - In images, generally most of the energy is concentrated in low frequencies.
  - Quantum tomography
  - Structured problems
- Some subspaces might be sparser than others.

Weighted RM

- We want to recover low rank $X_0 \in \mathbb{R}^{n_1 \times n_2}$ subject to measurements $y_0 = \mathcal{A}(X_0)$.

Algorithm: Let $X_0 \in \mathbb{R}^{n_1 \times n_2}$ and $\{S_i\}_{i=1}^{t}$ be subspaces of $\mathbb{R}^{n_1}$ orthogonal to each other with $S_1 \oplus S_2 \oplus \ldots \oplus S_t = \mathbb{R}^{n_1}$.
Then weighted NNM algorithm is given by

$$\min \| (w_1 U_1 \ \ldots \ w_t U_t)^T X \|_* \quad \text{subject to} \quad \mathcal{A}(X) = y_0$$

Weighted RM

- We want to recover low rank $X_0 \in \mathbb{R}^{n_1 \times n_2}$ subject to measurements $y_0 = \mathcal{A}(X_0)$.

> Algorithm: Let $X_0 \in \mathbb{R}^{n_1 \times n_2}$ and $\{S_i\}_{i=1}^{t}$ be subspaces of $\mathbb{R}^{n_1}$ orthogonal to each other with $S_1 \oplus S_2 \oplus \ldots \oplus S_t = \mathbb{R}^{n_1}$.
> Then weighted NNM algorithm is given by
>
> $$\min \|(w_1 U_1 \ \ldots \ w_t U_t)^T X\|_* \quad \text{subject to} \quad \mathcal{A}(X) = y_0$$

- Algorithm modifies regular nuclear norm mi $\min \|X\|_* \quad \text{subject to} \quad \mathcal{A}(X) = y_0$

Analysis

- Assume X_0 has $p_i \dim(S_i)$ singular vectors lying on S_i.
- Problem can be analyzed similarly.
  - Gaussian measurements
  - Exactly same technique
- Result is more convolved. However when $n_1 / n_2 \to 0$ we $w_i = 1 - p_i$ d:

  as optimal weights.