# Compressed Sensing – Sheet 12

Ferdinand Vanmaele

October 31, 2022

**Exercise 1** (Programming project part 1).

- The generation of sensors is done in `sheet12_ex1/sensors.py`. To make the matrices work with the assumptions of CoSaMP, OMP, SP and MP as defined in the lecture, the columns are normalized.
  - $m$ rows are drawn uniformly at random (without repetition) by generating a random permutation $S(m) \to S(m)$.
- Random $s$-sparse vectors $x \in \mathbb{R}^n$ with $1 \leq s \leq m$ are constructed in `sheet12_ex1/main.py`. The `generate_problems()` function takes the problem dimension $n$ and bounds for $s$, as well as the number repetitions for each sparsity value.
- Algorithms are run in `sheet12_ex1/main.py`. Every algorithm is implemented in its own function[1] and can be enabled or disabled selectively.

  *Remark.* The following modifications were made:

  - A tolerance of $10^{-8}$ was chosen te reduce oscillations for different sparsity levels.
  - Iterative hard thresholding was implemented with a variable step size [TBDR12],

  $$g := A^* r^{(k)}, \quad T := \begin{cases} \operatorname{supp}(\mathcal{T}(g, s)) & \text{if } k > 0 \\ \operatorname{supp}(\mathcal{T}(x^{(0)}, s)) & \text{if } k = 0 \end{cases}, \quad \mu := \frac{\|g_T\|_2^2}{\|A_T g_T\|_2^2}$$

  instead of $\mu = 1$. In the experiments performed, this was done to avoid divergence with $\mu = 1$.

- Let $\mathrm{RE}(\hat{x}) := \|x - \hat{x}\|_2 / \|x\|_2$. We consider successful recovery of a vector of sparsity $1 \leq s \leq m$ as the maximum $s$, such that

  $$s_{\max} := \max_s \left\{ \forall s' \leq s : \frac{1}{100} \sum_{i \in [100]} \mathrm{RE}(\hat{x}_{s',i}) < 10^{-6} \right\}$$

  holds. Summary of the trials, with $n = 128$, $m = 2^7$:

[1] This is a bit repetitive, especially for closely related algorithms such as CoSaMP and SP. However, every different algorithm to be tested is self-contained in this way. Outside of this exercise, I would likely summarize some functions and add additional parameters.

| Matrix | | Algorithm | $s_{\max}$ | $\approx \sum$ RE |
|---|---|---|---|---|
| **Random** $(A)$ | 1 | $\ell_1$-min. (BP) | $s = 17$ | $10^{-9}$ |
| | 2 | OMP | $s = 13$ | $10^{-15}$ |
| | 3 | MP | $s = 5$ | $10^{-9}$ |
| | 4 | IHT | $s = 10$ | $10^{-9}$ |
| | 5 | CoSaMP | $s = 14$ | $10^{-15}$ |
| | 6 | BT | $s = 0$ | – |
| | 7 | HTP | $s = 10$ | $10^{-15}$ |
| | 8 | SP | $s = 15$ | $10^{-15}$ |
| **Fourier** $(F)$ | 1 | $\ell_1$-min. (BP) | $s = 40$ | $10^{-10}$ |
| | 2 | OMP | $s = 34$ | $10^{-15}$ |
| | 3 | MP | $s = 10$ | $10^{-7}$ |
| | 4 | IHT | $s = 26$ | $10^{-9}$ |
| | 5 | CoSaMP | $s = 25$ | $10^{-15}$ |
| | 6 | BT | $s = 0$ | – |
| | 7 | HTP | $s = 26$ | $10^{-15}$ |
| | 8 | SP | $s = 31$ | $10^{-15}$ |

- Conclusions:
  - Methods using orthogonal projection recover the solution at a high accuracy (close to machine precision) for $1 \leq s$ sufficiently small. IHT is still competitive in terms of $s_{\max}$, with a higher recovery level.
  - The recovery error increased steeply from $s_{\max}$ to higher levels $s > s_{\max}$. Some trials also had oscillatory behavior.
  - The Fourier matrix was recovered for higher sparsity levels for all algorithms except Basic Thresholding. The coherence of this matrix is much lower than the Gaussian one ($\approx 0.08$ vs. $\approx 1.00$) so this is unsurprising.
  - $\ell_1$-minimization had constant performance for all trials. Methods using orthogonal projection had their CPU time increase exponentially. (See Figure 3 and 4)

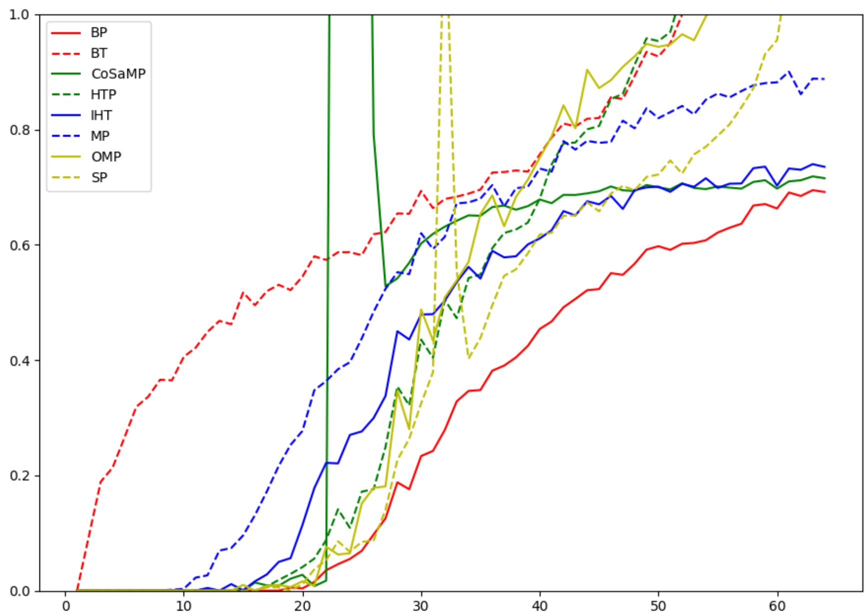


Figure 1: Relative recovery error ($y$-axis) over 100 trials per sparsity $1 \leq s \leq m$ ($x$-axis) for **Gaussian matrix** $A \in \mathbb{R}^{64\times 128}$ with normalized columns.

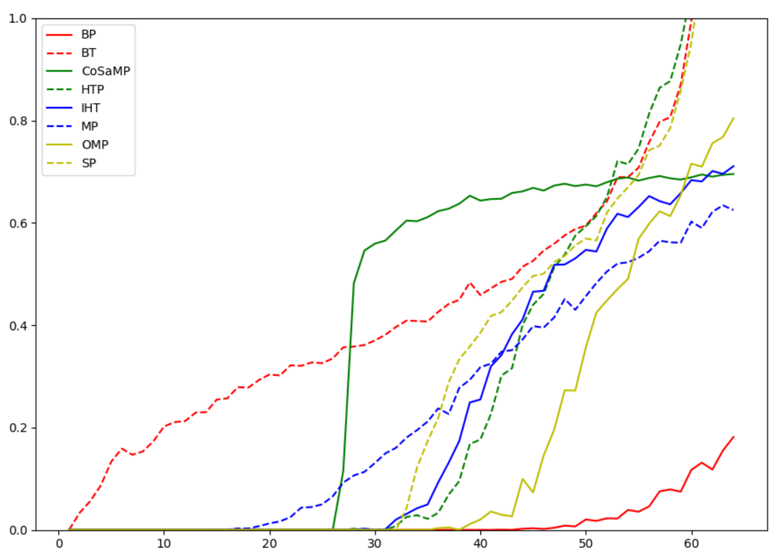


Figure 2: Relative recovery error ($y$-axis) averaged over 100 trials, by sparsity level $1 \leq s \leq m$ ($x$-axis) for **partial Fourier matrix** $F \in \mathbb{R}^{64\times 128}$ with normalized columns.

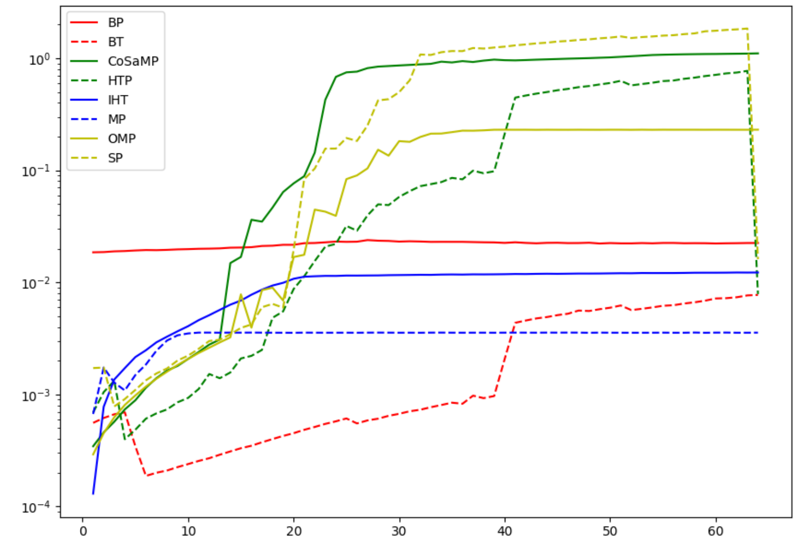


Figure 3: Combined CPU time in seconds per iteration ($y$-axis) averaged over 100 trials, by sparsity level $1 \leq s \leq m$ for **Gaussian matrix** $A \in \mathbb{R}^{64\times 128}$ with normalized columns.

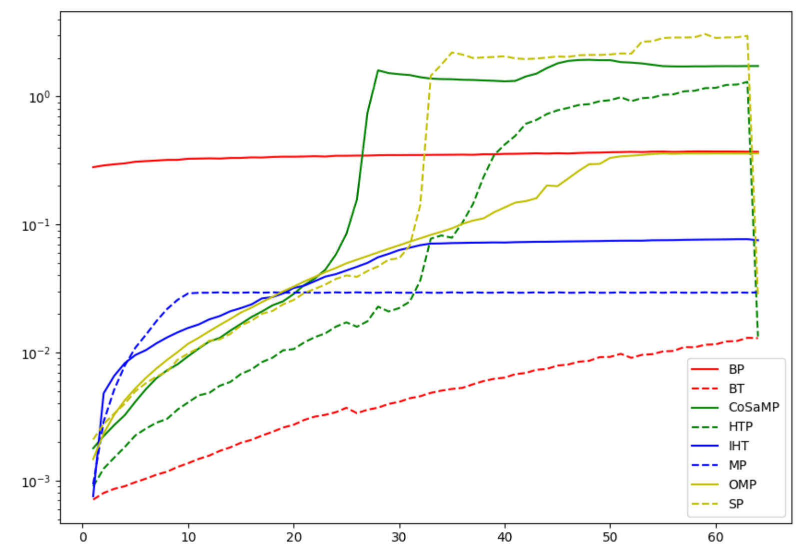


Figure 4: Combined CPU time in seconds per iteration ($y$-axis) averaged over 100 trials, by sparsity level $1 \leq s \leq m$ for **partial Fourier matrix** $F \in \mathbb{R}^{64\times 128}$ with normalized columns.

**Exercise 2.** The robust face recognition problem is given by

$$\min_{w \in \mathbb{R}^{m+n}} \frac{1}{2}\|Bw - b\|_2^2 + \lambda\|w\|_1$$

where

- $\lambda \in \mathbb{R}_{>0}$ is a regularization parameter,
- $B = [A \quad I] \in \mathbb{R}^{m\times(m+n)}$ with $A \in \mathbb{R}^{m\times n}$, $I \in \mathbb{R}^{m\times m}$ a highly correlated dictionary of $m$ face images $v \in \mathbb{R}^n$, stacked column-wise;
- $I$ the identity matrix $\in \mathbb{R}^{n\times n}$, $b = Ax + e$ a face image to be recognized not in the dictionary $A$, with $e \in \mathbb{R}^n$ an unknown error;
- $w$ the minimization variable.

The term $F(w) := \frac{1}{2}\|Bw - b\|_2^2$ is differentiable,

$$\nabla F(w) = B^T(Bw - b) \tag{1}$$

with Lipschitz constant $L := \|B^T B\|_F$,

$$\begin{aligned} \|B^T(Bw-b) - B^T(Bz-b)\|_2 &= \|B^TBw - B^Tb - B^TBz + B^Tb\|_2 \\ &= \|B^TBw - B^TBz\|_2 \\ &\leq \|B^TB\|_F\|w-z\|_2. \end{aligned}$$

The norm $\|w\|_1$ is proper, convex and lsc and has as proximal operator **soft shrinkage**, given component-wise by

$$\left[\mathrm{Prox}_{\gamma\|\cdot\|}(w)\right]_i = \begin{cases} w_i + \gamma & \text{if } w_i < \gamma, \\ 0 & \text{if } -\gamma \leq w_i \leq \gamma, \\ w_i - \gamma & \text{if } w_i > \gamma. \end{cases}$$

Multiplying by $\lambda > 0$ we have:

$$\begin{aligned} \left[\mathrm{Prox}_{\gamma R}(w)\right]_i &= \lambda^{-1}\left[\mathrm{Prox}_{\lambda^2\gamma\|\cdot\|}(\lambda w)\right]_i \\ &= \begin{cases} w_i + \lambda\gamma & \text{if } w_i < \lambda\gamma, \\ 0 & \text{if } -\lambda\gamma \leq w_i \leq \lambda\gamma, \\ w_i - \lambda\gamma & \text{if } w_i > \lambda\gamma. \end{cases} \end{aligned}$$

The non-zero entries of $x_*$ in the solution $w_* = [x_*^T, e_*^T]^T$ of problem (1) represent the person the image belongs to.

*Remark.* When we consider images without noise, we have the optimization problem

$$\min_{x \in \mathbb{R}^m} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1$$

with gradient $\nabla F(w) = A^T(Ax - b)$ and proximal operator as above. The non-zero entries of the solution $x_*$ represent the person the image belongs to.