Deep learning and inverse problems **Exercise 3**:

Problem 1.1

$\tilde{x} = w_1 \cdot x_1 + w_2 \cdot x_2 + \cdots + w_p \cdot x_p$, with $W = \{w_1, \cdots, w_p\} \subset \mathbb{R}^n$

if $\tilde{x}$ has k blocks of constants with lengths $\ell_1, \ell_2, \cdots, \ell_k$ respectively,

$$\Rightarrow \tilde{x} = \begin{bmatrix} \left.\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}\right\} \ell_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot x_1 + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \left.\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}\right\} \ell_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot x_2 + \cdots + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \left.\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}\right\} \ell_k \end{bmatrix} \cdot x_k + \underbrace{w_{k+1} \cdot x_{k+1} + \cdots + w_p \cdot x_p}_{\substack{w_{k+1}, \cdots, w_p \text{ can actually be any} \\ \text{basis } \mathbb{R}^n \text{ multiplied by } x_{k+1}, \cdots, x_p = 0.}}$$

hence, W is an overcomplete dictionary such that $x_s = [x_1, x_2, \cdots, x_p]^T$ is k-sparse.

$\Rightarrow$ an example of W can be:

$$W = \left\{ \begin{bmatrix} 1/\sqrt{\ell_1} \\ \vdots \\ 1/\sqrt{\ell_1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1/\sqrt{\ell_2} \\ \vdots \\ 1/\sqrt{\ell_2} \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1/\sqrt{\ell_k} \\ \vdots \\ 1/\sqrt{\ell_k} \end{bmatrix}, w_{k+1}, \cdots, w_n \right\} \subset \mathbb{R}^n$$

with n orthonormal basis $w_1, w_2, \cdots, w_n$ (ONB)

so that $\tilde{x} = W \cdot x_s$, $x_s \in \mathbb{R}^n$ is k-sparse

Problem 1.2

$$y = A \cdot \tilde{x} = A \cdot W \cdot x_s \quad \in \mathbb{R}^m$$

for ℓ_0 minimization, a necessary and sufficient condition to recover every k-sparse vector x_s from $y = A \cdot W \cdot x_s$ is that every set of $2k$ column of $A \cdot W$ is linearly independent.

$\Rightarrow$ $m = 2k$ linear measurements are necessary (with any method).

Problem 1.3

- Choice of a_i: Rows of a Gaussian random matrix (with i.i.d $\mathcal{N}(0, \frac{1}{m})$).
- Algorithm: (Candes, Romberg, Tao)

  if $\begin{cases} x_s \text{ is } k\text{-sparse in ONB } W, \\ A \text{ contains the rows of a Gaussian random matrix} \end{cases}$

  $\Rightarrow$ the reconstruction of $\hat{x}_1(y) = \underset{x_s}{\operatorname{argmin}} \|x_s\|_1$ s.t. $A \cdot W \cdot x_s = y$ succeeds with high probability.
- Number of measurements: $m \geq C \cdot k \cdot \log(n)$. with constant C.

Problem 2

From RIP: x is s-sparse

$$A \cdot x = [a_1, a_2, \cdots, a_s] \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} \nwarrow \text{non-zero elements}$$

A is μ-incoherent $\Rightarrow \begin{cases} A \text{ has unit norm columns}: \|a_i\| = 1 \text{ for } \forall a_i \text{ of } A. \\ |\langle a_i, a_j \rangle| \leq \mu \text{ for all pairs of columns } i \neq j \text{ of } A. \end{cases}$

$$\|Ax\|_2^2 = x^T \cdot A^T \cdot A x$$
$$= x^T \cdot \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_s^T \end{bmatrix} \cdot [a_1, a_2, \cdots, a_s] \cdot x = x^T \cdot \underbrace{\begin{bmatrix} 1 & a_1^T \cdot a_2 & \cdots & a_1^T \cdot a_s \\ \vdots & \ddots & & \vdots \\ a_s^T \cdot a_1 & & \cdots & 1 \end{bmatrix}}_{B} \cdot x$$

Then, if λ_{min} and λ_{max} are the minimal/maximal eigenvalues of matrix B

$$\Rightarrow x^T \cdot \lambda_{min} \cdot x \leq x^T \cdot B \cdot x \leq x^T \cdot \lambda_{max} \cdot x \quad (*)$$

Proof: $B = B^T \overset{\text{EVD}}{=} Q^T \cdot \Lambda Q$, with $Q^T = [q_1, q_2, \cdots, q_s]$ orthonormal.

Then, we use the orthonormal basis Q to represent vector $x = \sum_{i=1}^{s} q_i \cdot d_i$,

$$\Rightarrow x^T \cdot B \cdot x = x^T \cdot B \cdot \left(\sum_{i=1}^{s} q_i \cdot d_i\right)$$
$$= x^T \cdot Q^T \cdot \Lambda \cdot \begin{bmatrix} q_1^T \\ \vdots \\ q_s^T \end{bmatrix} \cdot \left(\sum_{i=1}^{s} q_i \cdot d_i\right)$$
$$\underset{\text{Q is orthonormal}}{=} [d_1, d_2, \cdots, d_s] \cdot \underset{n \times n}{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_s \end{bmatrix}} \cdot \overset{n \times 1}{\begin{bmatrix} d_1 \\ \vdots \\ d_s \end{bmatrix}}$$
$$= \sum_{k=1}^{s} \lambda_k \cdot S_k^2$$
$$\Rightarrow \sum_{k=1}^{s} \lambda_{min} \cdot d_k^2 \leq \sum_{k=1}^{s} \lambda_k \cdot d_k^2 \leq \sum_{k=1}^{s} \lambda_{max} \cdot d_k^2$$
$$\Rightarrow x^T \cdot \lambda_{min} \cdot x \leq x^T \cdot B \cdot x \leq x^T \cdot \lambda_{max} \cdot x. \quad \blacksquare$$

Then, use the Geshgorin's circle theorem:

$$|\lambda_{max} - b_{ii}| \leq \sum_{j \neq i} |b_{ij}|$$
$$|\lambda_{max} - 1| \leq \sum_{j \neq i} |b_{ij}| \overset{\text{A is } \mu\text{-incoherent}}{\leq} (s-1) \cdot \mu$$
$$\lambda_{max} \leq (s-1) \cdot \mu + 1.$$
$$\Rightarrow x^T \cdot \lambda_{max} \cdot x \leq x^T \cdot [(s-1) \cdot \mu + 1] \cdot x$$
$$\Rightarrow \|A \cdot x\|_2^2 = x^T \cdot B \cdot x \overset{(*)}{\leq} x^T \cdot \lambda_{max} \leq x \leq x^T \cdot [(s-1) \cdot \mu + 1] \cdot x$$
$$\Rightarrow \left| \|Ax\|_2^2 - \|x\|_2^2 \right| \leq x^T [(s-1)\mu] \cdot x = \|x\|_2^2 \cdot \epsilon_s$$
$$\Rightarrow \epsilon_s = (s-1) \cdot \mu.$$

Limitation: It requires a small μ (incoherent measurement matrix) to maintain the robustness of RIP.