Deep learning and inverse problems **Exercise 2**:

Problem 1.1

$y = [b_0, e_0, \cdots, e_{n-1}] \cdot x = \overset{\mathbb{R}^n}{b_0} \cdot \overset{\mathbb{R}}{x_0} + e_0 \cdot x_1 + \cdots + e_{n-1} \cdot x_n$, with $x = [x_0, x_1, \cdots, x_n]^T$

case 1) : $x_0 = 0$.

$\Rightarrow$ $y = e_0 \cdot x_1 + \cdots + e_{n-1} \cdot x_n$

since $[e_0, e_1, \cdots, e_{n-1}]$ are linearly independent basis.

$\Rightarrow$ simply $y = [x_1, x_2, \cdots, x_n]^T$

as we know the support set of x

$\Rightarrow$ so we need $m = s$ measurements to recover every s-sparse vector x for probability 1.

(exactly the number of equations, where x is non-zero)

$\Rightarrow$ the largest sparsity of s is n. $(y \in \mathbb{R}^n)$

case 2) : $x_0 \neq 0$.

$$y = \left[\begin{array}{c|cccc} & 1 & 0 & \cdots & 0 \\ b_0 & 0 & 1 & & 0 \\ & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & \cdots & 1 \end{array} \right] \cdot \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix}$$

x is s-sparse and $x_0 \neq 0$

$\Rightarrow$ x has $(n+1) - s - 1 = n - s$ zero positions.

$$y = \begin{bmatrix} b_{0,0} & 1 & 0 & \cdots & 0 \\ \cancel{b_{0,1}} & \cancel{0} & 1 & & \cancel{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \cancel{b_{0,n-1}} & \cancel{0} & 0 & \cdots & \cancel{1} \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\Rightarrow$ the corresponding $(n-s)$ columns in the matrix will not contribute to building equations.

an example : when $x_1, x_n = 0$, then the two basis and the two entries of b_0 can be simply ignored.

$\Rightarrow$ The equations will be based on the rest of $(n+1) - (n-s) = 1+s$ columns.

$\Rightarrow$ To recover every s-sparse vector x with probability 1,

we need to prove that the remaining $(1+s)$ columns are linearly independent for always.

$\hat{=}$ Equivalently, this means that the b_0 vector cannot contain any zero in the positions where the entries of x are also zeros. (*)

e.g. $\begin{bmatrix} b_{0,0} \\ 0 \\ b_{0,2} \\ \vdots \\ 0 \end{bmatrix} = b_{0,0} \cdot e_0 + b_{0,2} \cdot e_2 + \cdots$

i.e., in this case the b_0 and the standard basis are not linearly independent.

$\Rightarrow$ we cannot guarantee this condition (*), since b_0 is just a random vector.

$\Rightarrow$ when $x_0 \neq 0$, the recovery of every s-sparse x would not be with probability 1.

$\Rightarrow$ The largest s is n when $x_0 = 0$. If $x_0 \neq 0$, no guarantee for 100% recovery of every s-sparse x.

Problem 1.2

$\langle e_i, b_j \rangle = \sum_{k=0}^{n-1} e_{k,i} \cdot b_{k,j} = b_{ij} \sim \mathcal{N}(0, \frac{1}{n})$

According to the Gaussian Tail Inequality :

for $X \sim \mathcal{N}(\mu, \delta^2)$, then $P[|X - \mu| > \varepsilon] \leq 2 \cdot e^{-\varepsilon^2/(2\delta^2)}$

This is equivalent to the definition that is recalled in the question for $\delta^2 = 1$.

$\Rightarrow$ In our case the $\langle e_i, b_j \rangle = b_{ij} \sim \mathcal{N}(0, \frac{1}{n})$

then $P[|b_{ij} - 0| > \beta'] \leq 2 \cdot e^{-\beta'^2/(2 \cdot (\frac{1}{n})^2)}$

$P[|b_{ij}| > \beta'] \leq 2 \cdot e^{-\frac{1}{2}(\beta' n)^2}$

Problem 1.3

case 1) : $|\langle b_i, e_j\rangle|$ for $i, j \in \{0, 1, \cdots, n-1\}$.

$$\langle b_i, e_j\rangle = b_{i,j} \sim \mathcal{N}(0, \tfrac{1}{n})$$

from Problem 1.2 : $P[\,|b_{i,j}| > \beta_1] \leq 2 \cdot e^{-\frac{1}{2}(\beta_1 n)^2} \overset{!}{=} \delta.$

$$e^{-\frac{1}{2}(\beta_1 n)^2} = \tfrac{1}{2}\cdot\delta.$$

$$-\tfrac{1}{2}(\beta_1 n)^2 = \ln(\tfrac{1}{2}\cdot\delta)$$

$$(\beta_1 n)^2 = \ln((\tfrac{1}{2}\cdot\delta)^{-2})$$

$$\beta_1 = \tfrac{1}{n}\cdot\sqrt{\ln((\tfrac{1}{2}\delta)^{-2})}, \quad \text{for } \beta_1 > 0.$$

$\Rightarrow$ with probability at least $1-\delta$, the $|\langle b_i, e_j\rangle|$ for all i, j is smaller than this β_1.

$\Rightarrow$ $\mu_1 = \beta_1 = \frac{1}{n}\cdot\sqrt{\ln((\frac{1}{2}\cdot\delta)^{-2})}$

case 2) : $|\langle b_i, b_j\rangle|$ for $i, j \in \{0, 1, \cdots, n-1\}$

$$\langle b_i, b_j\rangle = \left\langle \begin{bmatrix} b_{0,i} \\ b_{1,i} \\ \vdots \\ b_{n-1,i} \end{bmatrix}, \begin{bmatrix} b_{0,j} \\ b_{1,j} \\ \vdots \\ b_{n-1,j} \end{bmatrix} \right\rangle$$ with the entries of $b \sim \mathcal{N}(0, \frac{1}{n})$ and i.i.d.

① for $X \sim \mathcal{N}(\mu_X, \delta_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \delta_Y^2)$, X and Y independent, $Z = X \cdot Y$

$\Rightarrow$ $Z \sim (\mu_Z, \delta_Z^2)$ with $\mu_Z = \frac{\mu_X \cdot \delta_Y^2 + \mu_Y \cdot \delta_X}{\delta_Y^2 + \delta_X^2}$, $\delta_Z^2 = \frac{\delta_X^2 \cdot \delta_Y^2}{\delta_X^2 + \delta_Y^2}$

$\Rightarrow$ $\delta_{k,i} \cdot \delta_{k,j}$ $(k \in \{0, 1, \cdots, n-1\})$ is also Gaussian distributed.

With $\mu_{multiply} = 0$, $\delta^2_{multiply} = \frac{(\frac{1}{n})^2 \cdot (\frac{1}{n})^2}{(\frac{1}{n})^2 + (\frac{1}{n})^2} = \frac{1}{2n^2}$

② for $X \sim \mathcal{N}(\mu_X, \delta_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \delta_Y^2)$, X and Y independent, $Z = X + Y$.

$\Rightarrow$ $Z \sim (\mu_Z, \delta_Z^2)$ with $\mu_Z = \mu_X + \mu_Y$, $\delta_Z^2 = \delta_X^2 + \delta_Y^2$.

$\Rightarrow$ $\langle b_i, b_j\rangle = \sum_{k=0}^{n-1} b_{k,i} \cdot b_{k,j}$ is also Gaussian distributed.

With $\mu_{sum} = 0$. $\delta^2_{sum} = n \cdot \frac{1}{2n^2} = \frac{1}{2n}$.

$\Rightarrow$ $\langle b_i, b_j\rangle \sim \mathcal{N}(0, \frac{1}{2n})$.

from Problem 1.2 : $P[\,|\langle b_i, b_j\rangle| > \beta_2] \leq 2 \cdot e^{-\beta_2^2/2 \cdot (\frac{1}{2n})^2} = 2 \cdot e^{-2(\beta_2 \cdot n)^2} \overset{!}{=} \delta$

$\Rightarrow$ $\beta_2 = \frac{1}{n}\cdot\sqrt{\ln((\frac{1}{2}\delta)^{-\frac{1}{2}})}$, for $\beta_2 > 0$.

$\Rightarrow$ Similarily, $\mu_2 = \beta_2 = \frac{1}{n}\cdot\sqrt{\ln((\frac{1}{2}\cdot\delta)^{-\frac{1}{2}})}$

case 3) : $|\langle e_i, e_j\rangle|$ for $i, j \in \{0, 1, \cdots, n-1\}$.

because $\langle e_i, e_j\rangle = 0$. for all i, j.

$\Rightarrow$ $\mu_3 = 0$.

Pick the largest number of μ_1, μ_2, μ_3 : $\mu_1 > \mu_2 > \mu_3$

And since $\|b_i\|_2 = 1$. $\|e_j\|_2 = 1$. They are unit norm columns

$\Rightarrow$ The coherence parameter of D is $\mu = \frac{1}{n}\cdot\sqrt{-2 \cdot \ln(\frac{1}{2}\delta)}$

Problekm 1.4

- For the recovery of $y = A \cdot x$, $A \in \mathbb{R}^{m \times n}$, the lower bound of μ is : $\mu_{min} = O(\sqrt{\frac{\log n}{m}})$

  $\Rightarrow$ The maximum sparsity based on ℓ_1-minimization is $S_{max} = \frac{1}{2\mu} = \frac{1}{2}\sqrt{\frac{m}{\log(n)}}$

- This result is pessimistic, since any improvement means that the μ needs to be smaller. But the lower bound of μ shows that it would not be possible. So we cannot get a better recovery.