Report | HW 2

CS754 - Advanced Image Processing

Omkar Shirpure (22B0910)
Krish Rakholiya (22B0927)
Suryansh Patidar (22B1036)

Contents

Declaration: The work submitted is our own, and we have adhered to the principles of academic honesty while completing and submitting this work. We have not referred to any unauthorized sources, and we have not used generative AI tools for the work submitted here.

1 Q1

1

$$\|\Phi(x^* - x)\|_{\ell_2} \leq \|\Phi x^* - y\|_{\ell_2} + \|y - \Phi x\|_{\ell_2} \leq 2\epsilon. \tag{1}$$

We prove this step by step:

1. **Define x^* as the solution to the optimization problem**:
   The reconstruction of x is obtained by solving the convex optimization problem:

   $$\min_{x'} \|x'\|_{\ell_1} \quad \text{subject to} \quad \|\Phi x' - y\|_{\ell_2} \leq \epsilon. \tag{2}$$

   This means that x^* satisfies the constraint:

   $$\|\Phi x^* - y\|_{\ell_2} \leq \epsilon. \tag{3}$$

2. **Apply the triangle inequality**:
   We want to bound $\|\Phi(x^* - x)\|_{\ell_2}$. Adding and subtracting y, we get:

   $$\|\Phi(x^* - x)\|_{\ell_2} = \|\Phi x^* - \Phi x\|_{\ell_2} = \|\Phi x^* - y + y - \Phi x\|_{\ell_2}. \tag{4}$$

   Using the **triangle inequality**:

   $$\|\Phi(x^* - x)\|_{\ell_2} \leq \|\Phi x^* - y\|_{\ell_2} + \|y - \Phi x\|_{\ell_2}. \tag{5}$$

3. **Use the feasibility condition**:
   Since x^* satisfies $\|\Phi x^* - y\|_{\ell_2} \leq \epsilon$, and assuming $y = \Phi x + z$ with $\|z\|_{\ell_2} \leq \epsilon$, we get:

   $$\|y - \Phi x\|_{\ell_2} = \|z\|_{\ell_2} \leq \epsilon. \tag{6}$$

4. **Final bound**:
   Substituting these bounds:

   $$\|\Phi(x^* - x)\|_{\ell_2} \leq \epsilon + \epsilon = 2\epsilon. \tag{7}$$

This completes the proof.

2

The l_2-norm and l_∞-norm are related by:

$$|v|\ell_2 \leq \sqrt{k}|v|\ell_\infty \tag{8}$$

for any vector supported on at most indices.

Here, is supported on at most indices (since it is a subset of a sparse vector).

Applying this inequality with , we get:

$$|h_{T_j}|\ell_2 \leq \sqrt{s}|hT_j|\ell\infty. \tag{9}$$

By the definition of the -norm, we know:

$$|v|\ell\infty \leq \frac{|v|_{\ell_1}}{\text{number of nonzero entries in } v}. \tag{10}$$

Since has at most nonzero entries, we apply this bound:

$$|h_{T_j}|\ell\infty \leq \frac{|h_{T_{j-1}}|_{\ell_1}}{s}. \tag{11}$$

Substituting the bound on into the first inequality:

$$|h_{T_j}|\ell_2 \leq \sqrt{s} \cdot \frac{|hT_{j-1}|\ell_1}{s} = s^{-1/2}|hT_{j-1}|_{\ell_1}. \tag{12}$$

This final expression shows that the l_2-norm of h_{T_j} is controlled by the l_1-norm of $h_{T(j-1)}$.

3

For any subset of at most indices, we use the standard inequality:

$$|h_{T_j}|\ell_2 \leq s^{-1/2}|hT_j|_{\ell_1}. \tag{13}$$

This follows because the -norm and -norm satisfy:

$$|v|\ell_2 \leq \sqrt{k}|v|\ell_\infty, \quad \text{and} \quad |v|\ell\infty \leq \frac{|v|_{\ell_1}}{k}, \tag{14}$$

for a vector with at most nonzero entries.Combining these gives:

$$|h_{T_j}|\ell_2 \leq s^{-1/2}|hT_j|_{\ell_1}. \tag{15}$$

Summing over all , we obtain:

$$\sum_{j\geq 2}|h_{T_j}|\ell_2 \leq s^{-1/2}\sum j \geq 2|h_{T_j}|_{\ell_1}. \tag{16}$$

By definition, are disjoint subsets of (the complement of), meaning:

$$\sum_{j\geq 2}|h_{T_j}|\ell_1 \leq |hT_0^c|_{\ell_1}. \tag{17}$$

Thus, we conclude:

$$\sum_{j\geq 2}|h_{T_j}|\ell_2 \leq s^{-1/2}|hT_0^c|_{\ell_1}. \tag{18}$$

4,5

By the **triangle inequality**, we have:

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} = \left\| \sum_{j \geq 2} h_{T_j} \right\|_{\ell_2} \leq \sum_{j \geq 2} \|h_{T_j}\|_{\ell_2}.$$

This step follows because the ℓ_2-norm satisfies the subadditivity property:

$$\|a + b\|_{\ell_2} \leq \|a\|_{\ell_2} + \|b\|_{\ell_2}.$$

Since $h_{(T_0 \cup T_1)^c}$ is the sum of disjoint terms h_{T_j} (for $j \geq 2$), we apply the triangle inequality to get the sum of their norms.

From the earlier derived inequality:

$$\|h_{T_j}\|_{\ell_2} \leq s^{-1/2} \|h_{T_j}\|_{\ell_1}.$$

Summing over all $j \geq 2$, we obtain:

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq s^{-1/2} \sum_{j \geq 2} \|h_{T_j}\|_{\ell_1}.$$

Since h_{T_j} are disjoint subsets of $h_{T_0^c}$, it follows that:

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_1} \leq \|h_{T_0^c}\|_{\ell_1}.$$

Thus, we conclude:

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq s^{-1/2} \|h_{T_0^c}\|_{\ell_1}.$$

Using this in Step 1, we get:

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq s^{-1/2} \|h_{T_0^c}\|_{\ell_1}.$$

6

The given inequality:

$$\|x\|_{\ell_1} \geq \|x + h\|_{\ell_1} = \sum_{i \in T_0} |x_i + h_i| + \sum_{i \in T_0^c} |x_i + h_i|$$

$$\geq \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}$$

can be justified as follows:

Since $x + h$ is the minimizer of the ℓ_1**-norm minimization problem**, it must have a smaller or equal norm than x. That is,

$$\|x + h\|_{\ell_1} \leq \|x\|_{\ell_1}.$$

Thus, we begin with:

$$\|x\|_{\ell_1} \geq \|x + h\|_{\ell_1}.$$

By definition of the ℓ_1-norm,

$$\|x + h\|_{\ell_1} = \sum_{i \in T_0} |x_i + h_i| + \sum_{i \in T_0^c} |x_i + h_i|.$$

We separate this into the contributions from T_0 (the indices of the largest s coefficients of x) and T_0^c (the remaining indices).

Using the triangle inequality:

$$|a + b| \geq |a| - |b|,$$

we get:

$$\sum_{i \in T_0} |x_i + h_i| \geq \sum_{i \in T_0} |x_i| - \sum_{i \in T_0} |h_i| = \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1}.$$

Similarly,

$$\sum_{i \in T_0^c} |x_i + h_i| \geq \sum_{i \in T_0^c} |h_i| - \sum_{i \in T_0^c} |x_i| = \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}.$$

Substituting these into the earlier expression:

$$\|x + h\|_{\ell_1} \geq (\|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1}) + (\|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}).$$

Since $\|x\|_{\ell_1} \geq \|x + h\|_{\ell_1}$, this gives:

$$\|x\|_{\ell_1} \geq \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}.$$

7

The given inequality:

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}$$

can be justified step by step using the definition of the best s-term approximation and properties of the ℓ_1-norm.

- The vector x_s is defined as the **best s-term approximation** of x, which means that it consists of the s largest magnitude entries of x, while the rest are set to zero.
- This implies:
$$\|x_{T_0^c}\|_{\ell_1} = \|x - x_s\|_{\ell_1}.$$
Since T_0 contains the indices of the largest s coefficients of x, everything outside T_0 belongs to T_0^c, so this definition holds.
- Since $x + h$ is the minimizer of the ℓ_1-norm, we have:
$$\|x\|_{\ell_1} \geq \|x + h\|_{\ell_1}.$$
- Expanding the ℓ_1-norm:
$$\|x + h\|_{\ell_1} = \sum_{i \in T_0} |x_i + h_i| + \sum_{i \in T_0^c} |x_i + h_i|.$$

Using the standard inequality:

$$|a + b| \geq |a| - |b|,$$

we obtain:

$$\sum_{i \in T_0} |x_i + h_i| \geq \sum_{i \in T_0} |x_i| - \sum_{i \in T_0} |h_i| = \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1}.$$

Similarly, for the complement T_0^c:

$$\sum_{i \in T_0^c} |x_i + h_i| \geq \sum_{i \in T_0^c} |h_i| - \sum_{i \in T_0^c} |x_i| = \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}.$$

Thus,

$$\|x + h\|_{\ell_1} \geq (\|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1}) + (\|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}).$$

Since $\|x\|_{\ell_1} \geq \|x + h\|_{\ell_1}$, we get:

$$\|x\|_{\ell_1} \geq \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}.$$

Rearranging:

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}.$$

8

We need to justify the following inequality:

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq \|h_{T_0}\|_{\ell_2} + 2e_0, \quad \text{where } e_0 = s^{-1/2}\|x - x_s\|_{\ell_1}.$$

We proceed step by step using previously established inequalities and the **Cauchy-Schwarz inequality**.

From the previous justification (Equation 11):

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq s^{-1/2}\|h_{T_0^c}\|_{\ell_1}.$$

And from Equation (12):

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}.$$

Substituting this into the previous inequality:

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq s^{-1/2}(\|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}).$$

The **Cauchy-Schwarz inequality** states:

$$\|v\|_{\ell_1} \leq \sqrt{k}\|v\|_{\ell_2}, \quad \text{for any vector } v \text{ supported on at most } k \text{ indices.}$$

Since h_{T_0} has at most s nonzero entries, applying this to h_{T_0} gives:

$$\|h_{T_0}\|_{\ell_1} \leq s^{1/2}\|h_{T_0}\|_{\ell_2}.$$

Thus, we substitute this into the previous equation:

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq s^{-1/2}(s^{1/2}\|h_{T_0}\|_{\ell_2} + 2\|x_{T_0^c}\|_{\ell_1}).$$

Simplifying,

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq \|h_{T_0}\|_{\ell_2} + 2s^{-1/2}\|x_{T_0^c}\|_{\ell_1}.$$

By definition, the best s-term approximation error is:

$$e_0 = s^{-1/2}\|x - x_s\|_{\ell_1} = s^{-1/2}\|x_{T_0^c}\|_{\ell_1}.$$

Substituting this into the last inequality:

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq \|h_{T_0}\|_{\ell_2} + 2e_0.$$

9

We need to justify the following inequality:

$$|\langle \Phi h_{T_0 \cup T_1}, \Phi h \rangle| \leq \|\Phi h_{T_0 \cup T_1}\|_{\ell_2} \|\Phi h\|_{\ell_2} \leq 2\epsilon \sqrt{1+\delta_{2s}} \|h_{T_0 \cup T_1}\|_{\ell_2}.$$

The **Cauchy-Schwarz inequality** states:

$$|\langle a, b \rangle| \leq \|a\|_{\ell_2} \|b\|_{\ell_2}.$$

Applying this to the inner product $\langle \Phi h_{T_0 \cup T_1}, \Phi h \rangle$, we get:

$$|\langle \Phi h_{T_0 \cup T_1}, \Phi h \rangle| \leq \|\Phi h_{T_0 \cup T_1}\|_{\ell_2} \|\Phi h\|_{\ell_2}.$$

From Equation (9), we have:

$$\|\Phi h\|_{\ell_2} \leq 2\epsilon.$$

Substituting this into the previous inequality:

$$|\langle \Phi h_{T_0 \cup T_1}, \Phi h \rangle| \leq \|\Phi h_{T_0 \cup T_1}\|_{\ell_2} \cdot 2\epsilon.$$

The **Restricted Isometry Property (RIP)** states that for any s-sparse vector v:

$$(1-\delta_s)\|v\|_{\ell_2}^2 \leq \|\Phi v\|_{\ell_2}^2 \leq (1+\delta_s)\|v\|_{\ell_2}^2.$$

Applying this to $h_{T_0 \cup T_1}$, we get:

$$\|\Phi h_{T_0 \cup T_1}\|_{\ell_2} \leq \sqrt{1+\delta_{2s}} \|h_{T_0 \cup T_1}\|_{\ell_2}.$$

Substituting this bound into our previous result:

$$|\langle \Phi h_{T_0 \cup T_1}, \Phi h \rangle| \leq 2\epsilon \cdot \sqrt{1+\delta_{2s}} \|h_{T_0 \cup T_1}\|_{\ell_2}.$$

10

We need to justify the following inequality, which follows from **Lemma 2.1**:

$$|\langle \Phi h_{T_0}, \Phi h_{T_j} \rangle| \leq \delta_{2s} \|h_{T_0}\|_{\ell_2} \|h_{T_j}\|_{\ell_2}.$$

Lemma 2.1 states that for any two vectors x, x' supported on disjoint sets of sizes at most s and s', the following inequality holds:

$$|\langle \Phi x, \Phi x' \rangle| \leq \delta_{s+s'} \|x\|_{\ell_2} \|x'\|_{\ell_2}.$$

- This is derived using the **Restricted Isometry Property (RIP)** and the **parallelogram identity**.

- The parallelogram identity states:

$$\|a+b\|_{\ell_2}^2 + \|a-b\|_{\ell_2}^2 = 2\|a\|_{\ell_2}^2 + 2\|b\|_{\ell_2}^2.$$

- Applying RIP to this identity provides an upper bound on the inner product of two sparse vectors.
- The vectors h_{T_0} and h_{T_j} are supported on **disjoint** subsets of size at most s.
- By setting $s = s'$, we directly apply Lemma 2.1:

$$|\langle \Phi h_{T_0}, \Phi h_{T_j} \rangle| \leq \delta_{2s} \|h_{T_0}\|_{\ell_2} \|h_{T_j}\|_{\ell_2}.$$

11

We need to show that:

$$\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2} \leq \sqrt{2} \|h_{T_0 \cup T_1}\|_{\ell_2}.$$

By definition, the ℓ_2-norm of a vector v is:

$$\|v\|_{\ell_2} = \sqrt{\sum_i v_i^2}.$$

Since T_0 and T_1 are **disjoint** subsets, we have:

$$\|h_{T_0 \cup T_1}\|_{\ell_2}^2 = \sum_{i \in T_0} h_i^2 + \sum_{i \in T_1} h_i^2.$$

Thus, we can rewrite:

$$\|h_{T_0 \cup T_1}\|_{\ell_2} = \sqrt{\|h_{T_0}\|_{\ell_2}^2 + \|h_{T_1}\|_{\ell_2}^2}.$$

We now apply the **inequality between sums and squares**, which states that for any two nonnegative numbers a, b:

$$a + b \leq \sqrt{2(a^2 + b^2)}.$$

Setting $a = \|h_{T_0}\|_{\ell_2}$ and $b = \|h_{T_1}\|_{\ell_2}$, we get:

$$\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2} \leq \sqrt{2(\|h_{T_0}\|_{\ell_2}^2 + \|h_{T_1}\|_{\ell_2}^2)}.$$

Since we previously established:

$$\|h_{T_0}\|_{\ell_2}^2 + \|h_{T_1}\|_{\ell_2}^2 = \|h_{T_0 \cup T_1}\|_{\ell_2}^2,$$

this simplifies to:

$$\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2} \leq \sqrt{2} \|h_{T_0 \cup T_1}\|_{\ell_2}.$$

12

We need to justify the following inequality:

$$(1-\delta_{2s})\|h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|h_{T_0\cup T_1}\|_{\ell_2}^2 (2\epsilon\sqrt{1+\delta_{2s}} + \sqrt{2}\delta_{2s}\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}).$$

- The **Restricted Isometry Property (RIP)** states that for any s-sparse vector v:

$$(1-\delta_s)\|v\|_{\ell_2}^2 \le \|\Phi v\|_{\ell_2}^2 \le (1+\delta_s)\|v\|_{\ell_2}^2.$$

Applying this to $h_{T_0\cup T_1}$, which has sparsity at most $2s$, we obtain:

$$(1-\delta_{2s})\|h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2.$$

- From **Lemma 2.1**, we have:

$$|\langle \Phi h_{T_0\cup T_1}, \Phi h\rangle| \le 2\epsilon\sqrt{1+\delta_{2s}}\|h_{T_0\cup T_1}\|_{\ell_2}.$$

- Since $h = h_{T_0\cup T_1} + \sum_{j\ge 2} h_{T_j}$, we rewrite:

$$\|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 = \langle \Phi h_{T_0\cup T_1}, \Phi h\rangle - \sum_{j\ge 2}\langle \Phi h_{T_0\cup T_1}, \Phi h_{T_j}\rangle.$$

- Using the **triangle inequality**, we get:

$$\|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|h_{T_0\cup T_1}\|_{\ell_2}^2 (2\epsilon\sqrt{1+\delta_{2s}} + \sqrt{2}\delta_{2s}\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}).$$

13

We need to justify the following inequality:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\epsilon + \rho s^{-1/2}\|h_{T_0^c}\|_{\ell_1},$$

where:

$$\alpha = \frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}, \quad \rho = \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}.$$

$\|h_{T_j}\|\ell_2$ From **Equation (10)**, we have:

$$\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2} \le s^{-1/2}\|h_{T_0^c}\|_{\ell_1}.$$

By substituting this result into the previous inequality:

$$(1-\delta_{2s})\|h_{T_0\cup T_1}\|_{\ell_2}^2 \le \|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 \le 2\epsilon\sqrt{1+\delta_{2s}}\|h_{T_0\cup T_1}\|_{\ell_2} + \sqrt{2}\delta_{2s}\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}.$$

Using the **Restricted Isometry Property (RIP)**, we have:

$$\|\Phi h_{T_0\cup T_1}\|_{\ell_2}^2 \le (1+\delta_{2s})\|h_{T_0\cup T_1}\|_{\ell_2}^2.$$

Thus, the inequality becomes:

$$(1-\delta_{2s})\|h_{T_0\cup T_1}\|_{\ell_2}^2 \le 2\epsilon\sqrt{1+\delta_{2s}}\|h_{T_0\cup T_1}\|_{\ell_2} + \sqrt{2}\delta_{2s}s^{-1/2}\|h_{T_0^c}\|_{\ell_1}.$$

Dividing both sides by $(1-\delta_{2s})$, we get:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \frac{2\epsilon\sqrt{1+\delta_{2s}}}{1-\delta_{2s}} + \frac{\sqrt{2}\delta_{2s}s^{-1/2}\|h_{T_0^c}\|_{\ell_1}}{1-\delta_{2s}}.$$

Define:

$$\alpha = \frac{2\sqrt{1+\delta_{2s}}}{1-\delta_{2s}}, \quad \rho = \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}.$$

Substituting these values, we get:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\epsilon + \rho s^{-1/2}\|h_{T_0^c}\|_{\ell_1}.$$

14

We need to justify the following inequality:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le (1-\rho)^{-1}(\alpha\epsilon + 2\rho e_0),$$

where $e_0 = s^{-1/2}\|x-x_s\|_{\ell_1}$.

From the previous step, we established:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\epsilon + \rho s^{-1/2}\|h_{T_0^c}\|_{\ell_1}.$$

Now, using "Equation (12)":

$$\|h_{T_0^c}\|_{\ell_1} \le \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}.$$

Since we know from "Cauchy-Schwarz":

$$\|h_{T_0}\|_{\ell_1} \le s^{1/2}\|h_{T_0}\|_{\ell_2},$$

we get:

$$\|h_{T_0^c}\|_{\ell_1} \le s^{1/2}\|h_{T_0}\|_{\ell_2} + 2\|x_{T_0^c}\|_{\ell_1}.$$

Dividing by $s^{1/2}$ and using $e_0 = s^{-1/2}\|x-x_s\|_{\ell_1}$, we obtain:

$$s^{-1/2}\|h_{T_0^c}\|_{\ell_1} \le \|h_{T_0}\|_{\ell_2} + 2e_0.$$

Substituting this into the previous bound for $\|h_{T_0\cup T_1}\|_{\ell_2}$:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\epsilon + \rho(\|h_{T_0}\|_{\ell_2} + 2e_0).$$

Rearrange the inequality:

$$\|h_{T_0\cup T_1}\|_{\ell_2} - \rho\|h_{T_0}\|_{\ell_2} \le \alpha\epsilon + 2\rho e_0.$$

Factoring:

$$(1-\rho)\|h_{T_0\cup T_1}\|_{\ell_2} \le \alpha\epsilon + 2\rho e_0.$$

Dividing by $(1-\rho)$:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le (1-\rho)^{-1}(\alpha\epsilon + 2\rho e_0).$$

15

We need to justify the final inequality:

$$\|h\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2} + \|h_{(T_0\cup T_1)^c}\|_{\ell_2} \le 2(1-\rho)^{-1}(\alpha\epsilon + (1+\rho)e_0).$$

Since h is split into two disjoint parts:

$$\|h\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2} + \|h_{(T_0\cup T_1)^c}\|_{\ell_2}.$$

This follows from the **triangle inequality**:

$$\|a+b\|_{\ell_2} \le \|a\|_{\ell_2} + \|b\|_{\ell_2}.$$

1. **From the previous step**, we derived:

$$\|h_{T_0\cup T_1}\|_{\ell_2} \le (1-\rho)^{-1}(\alpha\epsilon + 2\rho e_0).$$

2. **From a previous bound**, we also have:

$$\|h_{(T_0\cup T_1)^c}\|_{\ell_2} \le \|h_{T_0}\|_{\ell_2} + 2e_0.$$

   Substituting $\|h_{T_0}\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2}$, we get:

$$\|h_{(T_0\cup T_1)^c}\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2} + 2e_0.$$

Summing these two bounds:

$$\|h\|_{\ell_2} \le \|h_{T_0\cup T_1}\|_{\ell_2} + \|h_{(T_0\cup T_1)^c}\|_{\ell_2}.$$

Substituting the bound for $\|h_{(T_0\cup T_1)^c}\|_{\ell_2}$:

$$\begin{aligned}\|h\|_{\ell_2} &\le \|h_{T_0\cup T_1}\|_{\ell_2} + (\|h_{T_0\cup T_1}\|_{\ell_2} + 2e_0).\\ &= 2\|h_{T_0\cup T_1}\|_{\ell_2} + 2e_0.\end{aligned}$$

Now, substituting the bound on $\|h_{T_0\cup T_1}\|_{\ell_2}$:

$$\|h\|_{\ell_2} \le 2(1-\rho)^{-1}(\alpha\epsilon + 2\rho e_0) + 2e_0.$$

Factor out the terms:

$$\|h\|_{\ell_2} \le 2(1-\rho)^{-1}(\alpha\epsilon + (1+\rho)e_0).$$

16

We need to justify the final inequality:

$$\|h\|_{\ell_1} = \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} \leq 2(1+\rho)(1-\rho)^{-1}\|x_{T_0^c}\|_{\ell_1}.$$

From "Equation (15)", we have:

$$\|h_{T_0}\|_{\ell_1} \leq \rho\|h_{T_0^c}\|_{\ell_1},$$

where:

$$\rho = \frac{\sqrt{2}\delta_{2s}}{1-\delta_{2s}}.$$

This follows from the "Restricted Isometry Property (RIP)" and the "ℓ_1-ℓ_2 norm relationship":

$$\|h_{T_0}\|_{\ell_1} \leq s^{1/2}\|h_{T_0}\|_{\ell_2} \leq s^{1/2}\|h_{T_0 \cup T_1}\|_{\ell_2}.$$

From "Equation (12)":

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}.$$

Substituting "Equation (15)" into this:

$$\|h_{T_0^c}\|_{\ell_1} \leq \rho\|h_{T_0^c}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}.$$

Rearrange:

$$(1-\rho)\|h_{T_0^c}\|_{\ell_1} \leq 2\|x_{T_0^c}\|_{\ell_1}.$$

Dividing by $(1-\rho)$:

$$\|h_{T_0^c}\|_{\ell_1} \leq \frac{2}{1-\rho}\|x_{T_0^c}\|_{\ell_1}.$$

We now sum:

$$\|h\|_{\ell_1} = \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1}.$$

Substituting $\|h_{T_0}\|_{\ell_1} \leq \rho\|h_{T_0^c}\|_{\ell_1}$:

$$\|h\|_{\ell_1} \leq (1+\rho)\|h_{T_0^c}\|_{\ell_1}.$$

Using the bound on $\|h_{T_0^c}\|_{\ell_1}$:

$$\|h\|_{\ell_1} \leq (1+\rho)\frac{2}{1-\rho}\|x_{T_0^c}\|_{\ell_1}.$$

Rearrange:

$$\|h\|_{\ell_1} \leq 2(1+\rho)(1-\rho)^{-1}\|x_{T_0^c}\|_{\ell_1}.$$

Thus, the final bound is fully justified!

2 Q2

C1 mean C implemented for part-A and *C2* mean that C implemented for part-B. (Both for code and below images)

(a) Original

(b) Part A (RMSE=0.013)

(a) Part B (RMSE=0.26)

(b) Part C1 (RMSE=0.013)

(a) Part C2 (RMSE=0.09)

(a) Original

(a) Part B (RMSE=0.41)

(b) Part C2 (RMSE=0.13)

d

FISTA (Fast Iterative Shrinkage-Thresholding Algorithm) is mathematically proven to be faster than ISTA due to its **accelerated gradient descent approach**. The key difference lies in the update step:

- ISTA follows a **first-order proximal gradient descent**, which has a convergence rate of $O(1/k)$.
- FISTA **incorporates a momentum term** inspired by Nesterov's acceleration, achieving a convergence rate of $O(1/k^2)$.

This improvement is achieved by introducing an auxiliary variable that combines the previous two iterates to **enhance step selection** and **increase convergence speed**. While ISTA updates the solution at each step using only the current iterate, FISTA adds a **momentum term** that considers the past iterate, effectively reducing the number of iterations required for convergence.

The key update steps in FISTA are:

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$\theta_{k+1} = \text{prox}_{\lambda/L}\left(x_k + \frac{1}{L}\nabla f(x_k)\right)$$

$$y_{k+1} = \theta_{k+1} + \frac{t_k - 1}{t_{k+1}}(\theta_{k+1} - \theta_k)$$

This acceleration significantly improves the convergence speed compared to ISTA.

3 Q3

Introduction

Group testing is a technique traditionally used in fields like medical diagnostics to identify defective items efficiently. In the paper *Neural Group Testing to Accelerate Deep Learning* by Weixin Liang and James Zou , the authors apply group testing principles to deep learning to reduce computational costs during inference. The paper is available at this link.

Specific ML/DS Problem Targeted

The paper addresses the challenge of reducing the computational burden of deep learning inference. As neural networks grow in complexity, processing each input sample individually becomes expensive. The authors propose a method inspired by group testing to test multiple input samples simultaneously, reducing the number of forward passes required and accelerating inference.

Pooling Matrix, Measurement Vector, and Unknown Signal Vector

In the context of this problem:

- **Pooling Matrix (Design Matrix)**: Determines how input samples are combined and fed into the neural network in a single forward pass. Each row represents a unique combination of input samples.
- **Measurement Vector**: Contains the outcomes of each pooled test, representing the neural network's output for each combination of input samples.
- **Unknown Signal Vector**: Represents the true labels or statuses of the individual input samples, indicating whether each sample possesses the feature of interest.

Algorithm Used to Solve the Problem

The paper introduces three neural network designs that implement group testing without adding new parameters:

- **Max Pooling Design**: Combines multiple input samples by taking the element-wise maximum, ensuring that prominent features are retained in the pooled representation.
- **Mean Pooling Design**: Averages input samples element-wise to create a pooled representation that reflects the average features of the combined samples.
- **Weighted Pooling Design**: Assigns learnable weights to each input sample before pooling, allowing the network to prioritize certain samples in the pooled representation.

These designs enable the neural network to process multiple samples in a single forward pass. If the pooled output indicates the presence of the feature of interest, the algorithm adaptively retests smaller groups or individual samples to pinpoint the exact sources. This hierarchical testing strategy significantly reduces computational costs.

Conclusion

The application of group testing in deep learning inference demonstrates how interdisciplinary approaches can enhance computational efficiency. The proposed method reduces computational costs by over 73% while maintaining or improving detection performance.

4 Q4

(a) Condition for Any Set S with $|S| \leq s$

Let S^* be the set of indices corresponding to the s largest absolute values of v, i.e.,

$$S^* = \arg \max_{S \subseteq [n], |S|=s} \|v_S\|_1.$$

By assumption, the condition $\|v_{S^*}\|_1 \leq \|v_{\overline{S^*}}\|_1$ holds. Now, let S be any other set with $|S| \leq s$. Since S^* contains the s largest absolute values of v, we have:

$$\|v_S\|_1 \leq \|v_{S^*}\|_1.$$

By the assumption, $\|v_{S^*}\|_1 \leq \|v_{\overline{S^*}}\|_1$. Moreover, since $\overline{S}$ contains at least as many entries as $\overline{S^*}$, we have:

$$\|v_{\overline{S^*}}\|_1 \leq \|v_{\overline{S}}\|_1.$$

Combining these inequalities, we get:

$$\|v_S\|_1 \leq \|v_{S^*}\|_1 \leq \|v_{\overline{S^*}}\|_1 \leq \|v_{\overline{S}}\|_1.$$

Thus, the condition $\|v_S\|_1 \leq \|v_{\overline{S}}\|_1$ holds for any set S with $|S| \leq s$.

(b) MNSP Implies $\|v\|_1 \leq 2\sigma_{s,1}(v)$

Let $v \in \text{nullspace}(A) - \{0\}$. By the Modified Null Space Property (MNSP), for any set S with $|S| \leq s$, we have:

$$\|v_S\|_1 < \|v_{\overline{S}}\|_1.$$

Let w be the best s-sparse approximation to v in the L_1 norm, i.e.,

$$w = \arg \min_{\|z\|_0 \leq s} \|v - z\|_1.$$

Let S be the support of w, so $|S| \leq s$. By the MNSP, we have:

$$\|v_S\|_1 < \|v_{\overline{S}}\|_1.$$

The L_1 norm of v can be written as:

$$\|v\|_1 = \|v_S\|_1 + \|v_{\overline{S}}\|_1.$$

Using the MNSP inequality, we get:

$$\|v\|_1 < 2\|v_{\overline{S}}\|_1.$$

Since w is supported on S, we have $v_{\overline{S}} = (v - w)_{\overline{S}}$. Therefore:

$$\|v_{\overline{S}}\|_1 = \|(v - w)_{\overline{S}}\|_1 \leq \|v - w\|_1 = \sigma_{s,1}(v).$$

Substituting this into the previous inequality, we obtain:

$$\|v\|_1 < 2\sigma_{s,1}(v).$$

Thus, the MNSP implies $\|v\|_1 \leq 2\sigma_{s,1}(v)$.

(c) MNSP and Uniqueness of s-Sparse Solutions

We prove the two directions separately.

(i) Necessity (MNSP $\Rightarrow$ Uniqueness)

Assume A satisfies the MNSP of order s. Let x be an s-sparse vector, and let $y = Ax$. Suppose there exists another vector $x' \neq x$ such that $Ax' = y$ and $\|x'\|_1 \leq \|x\|_1$. Let $v = x' - x$. Then:

$$v \in \text{nullspace}(A), \quad v \neq 0.$$

Let S be the support of x, so $|S| \leq s$. By the MNSP, we have:

$$\|v_S\|_1 < \|v_{\overline{S}}\|_1.$$

However, since $\|x'\|_1 \leq \|x\|_1$, we have:

$$\|x + v\|_1 \leq \|x\|_1.$$

This implies:

$$\|v_S\|_1 \geq \|v_{\overline{S}}\|_1,$$

which contradicts the MNSP. Thus, x must be the unique solution.

(ii) Sufficiency (Uniqueness $\Rightarrow$ MNSP)

Assume every s-sparse vector x is the unique solution of the P_1 problem. Let $v \in \text{nullspace}(A) - \{0\}$, and let S be any set with $|S| \leq s$. Define $x = v_S$, so x is s-sparse. Since $v \in \text{nullspace}(A)$, we have:

$$A(x + v_{\overline{S}}) = Ax = y.$$

If $\|v_S\|_1 \geq \|v_{\overline{S}}\|_1$, then:

$$\|x + v_{\overline{S}}\|_1 \leq \|x\|_1,$$

which would imply that $x + v_{\overline{S}}$ is also a minimizer of the P_1 problem, contradicting the uniqueness of x. Therefore, we must have:

$$\|v_S\|_1 < \|v_{\overline{S}}\|_1.$$

Thus, A satisfies the MNSP of order s.