Signal-Opti

Rémi Barraud and P J Anthony

April 2025

1 Signal

1.1 Question 2

The reconstruction observed was not that good, lacking many components of the original audio, and it was not smooth. The SRE was around 4.1225.

1.2 Question 3

The reconstruction observed was much better, and SRE was measured at around 13.123.

1.3 Question 4

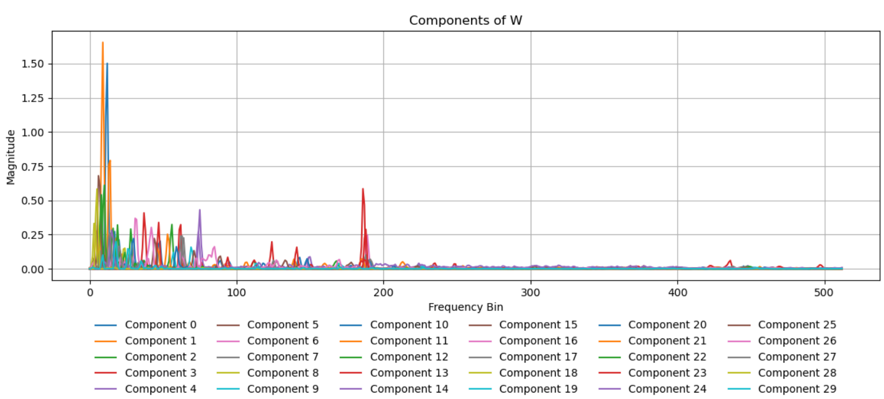

Figure 1: Components of W (columns) as a graph

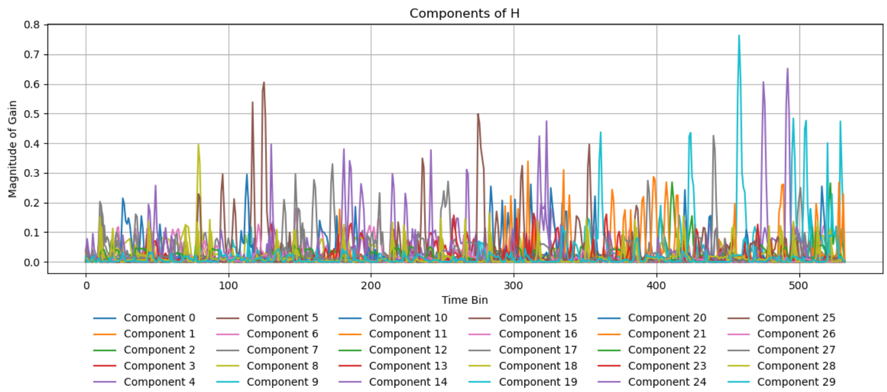

Figure 2: Components of H (rows) as a graph

We split the different tracks recognizing that violins have a high degree of harmonics, so there would be multiple peaks, with a medium spread. These peaks also are of higher frequency. In practice, we did this by calculating the standard deviation of the peaks, along with its spacings and filtering out the different components, and adjusting the different thresholds as necessary.

We obtained the following spectrograph, which mostly contains the components that form the second peak in the original graph.

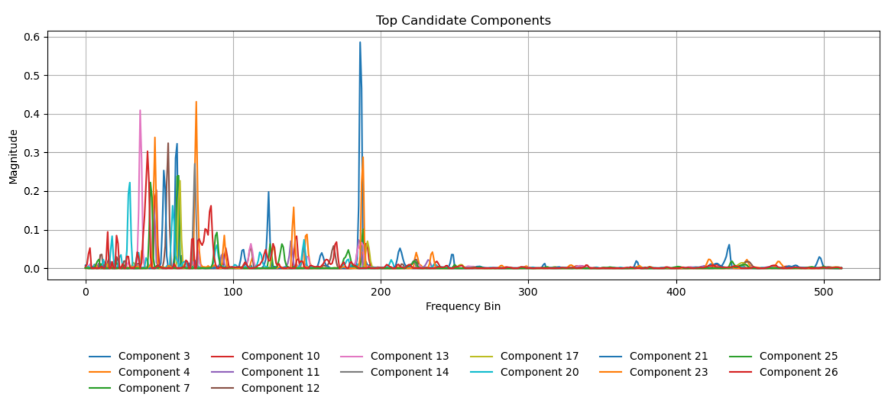

Figure 3: Components of the Violin

1.4 Question 5

In real life, we could possibly use a combination of different techniques, such as calculating the mean frequencies of each component, comparing with real life measured frequencies. We could also use machine learning or other techniques to cluster the different components together to belong to a particular instrument.

2 Optimization

2.1 Question 6a

Calcul of f gradients. $f(W, H) =< S - WH|S - WH >$ Thus :

$$\begin{aligned} f(W+w, H) &=< S - (W+w)H|S - (W+w)H > \\ &= f(W, H)+ < S - WH| - wH > + < -wH|S - WH > +o(w^2) \\ &= f(W, H)+ < -2(S - WH)H^\top|w > +o(w^2) \end{aligned}$$

Hence :

$$\nabla_W f(W, H) = 2(WH - S)H^T$$

By the exact same reasoning :

$$\nabla_H f(W, H) = 2W^T(WH - S)$$

2.2 Question 6b

Let H be fixed. $W \mapsto S - WH$ is an affine function, hence convex. By the triangular inequality, $A \mapsto ||A||^2$ is convex. By composition, $W \mapsto f(W, H)$ is convex. The same reasoning allows to conclude:

$$\forall W, \quad H \mapsto f(W, H) \text{ is convex.}$$

2.3 Question 6c

Let's suppose $k = 1, m = n = 2$. For $i, j \in \{1, 2\}$, Let :

$$E_i = \begin{pmatrix} \delta_{1,i} \\ \delta_{2,i} \end{pmatrix}, E_{i,j} = (\delta_{k,i}\delta_{l,j})_{1\leq k,l\leq 2}$$

Let

$$S = E_{1,2} + E_{2,1} \quad (W_a, H_a) = (E_1, E_2^T), (W_b, H_b) = (E_2, E_1^T)$$

We have :

$$W_a H_a = W_b H_b = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = S$$

For more praticity let the notation $\bar{\lambda} = 1 - \lambda$ if a certain $\lambda \in [0, 1]$ is defined, and the function

$$\varphi : \lambda \in [0, 1] \mapsto f\left(\lambda(W_a, H_b) + \bar{\lambda}(W_b, H_b)\right)$$

Since $S = \lambda S + \bar{\lambda} S$:

$$\begin{aligned}(\lambda W_a + \bar{\lambda} W_b)(\lambda H_a + \bar{\lambda} H_b) &= \lambda(W_a H_a) + \bar{\lambda}(W_b H_b) + \lambda\bar{\lambda}(W_a H_b + W_b H_a) \\ &= S + \lambda\bar{\lambda} I_2\end{aligned}$$

Hence :

$$\varphi(\lambda) = 2\lambda^2\bar{\lambda}^2 > \lambda\varphi(0) + \bar{\lambda}\varphi(1) = 0$$

This proves f is not convex. If $m > 2$ or $n > 2$, we use this proof setting all coordonates of S, W, and H of indice strictly greater than 2 to 0.

2.4 Question 2d

let $D := \frac{1}{2} I_n$. If (W, H) is a solution of the problem, since $f(W, H) = f(WD, D^{-1}H)$, $(WD, D^{-1}H)$ is a solution too. The solution cannot be unique.
Let's now prove a solution exists. The idea is to double-extract a saquence that realises the inf.
Let $(\tilde{W}_n, \tilde{H}_n)_{n\in\mathbb{N}} \in (M_{n,k}(\mathbb{R}) \times M_{k,m}(\mathbb{R}))^{\mathbb{N}}$ a sequence such as :

$$f(\tilde{W}_n, \tilde{H}_n) \to \inf_{W,H} f(W, H)$$

for $n \in \mathbb{N}$, let $D = ||W_n|| I_n$ The saquence defined as:

$$\forall n, W_n := \tilde{W}_n D^{-1}, H_n := D\tilde{H}_n$$

Also realises :

$$f(W_n, H_n) \to \inf_{W,H} f(W, H)$$

and W_n is of norm one for all n. Thus, W_n belongs to the unit sphere of $M_{n,k}(\mathbb{R})$ which is a compact. Hence, $(W_n)_n$ possesses an accumulaiton point W_∞ and an extraction $(W_{\varphi(n)})_n$ that tends toward W_∞. Since f is continuous:

$$f(W_\infty, H_{\varphi(n)}) \to \inf_{W,H} f(W, H)$$

Let's decompose H_n on the kernel of W_∞:
Let $\mathcal{K} := \mathrm{Ker}(W_\infty)$ and $\mathcal{G}$ a supplementary of $\mathcal{K}$. For n let $(u_n, v_n) \in \mathcal{K} \times \mathcal{G}$
Such as: $H_{\varphi(n)} = u_n + v_n$
By definition of $\mathcal{G}$,

$$\alpha := \min_{v\in S_{\mathcal{G}}(0,1)} ||W_\infty v|| > 0$$

Hence:

$$\sqrt{f(W_\infty, H_{\varphi(n)})} = \sqrt{f(W_\infty, v_n)} \geq \alpha||v_n|| - ||S||$$

Since $f(W_\infty, v_n)$ admits a finite limit, $(v_n)_n$ is bounded.
Hence, $(v_n)_n$ realises:

$$f(W_\infty, v_n) \to \min_{W,H} f(W,H) \exists N, \forall n \geq N, ||v_n|| \leq \frac{1}{\alpha} \min_{W,H} f(W,H)$$

Hence $(v_n)_n$ possesses an accumulation value v_∞ such as, by continuity of f :

$$f(W_\infty, v_\infty) = \min_{W,H} f(W,H)$$

The problem admits a solution.

2.5 Question 7a

According to question 6, and since f is of class $\mathcal{C}^\infty$, $\phi : \alpha_W \mapsto f(g(\alpha_W), H)$ (where $g : \alpha_W \mapsto W - \alpha_W \nabla_W f(W,H)$) is a convex and differentiable funciton.

$$\begin{aligned}
\phi'(\alpha_W) &= df_W(g(\alpha_W), H) \cdot g'(\alpha_W) = < \nabla_W f(g(\alpha_W), H) | - \nabla_W f(W,H) > \\
&= - < 2(g(\alpha_W)H - S)H^T | \nabla_W f(W,H) > \\
&= -2 < (WH - \alpha_W \nabla_W f(W,H)H - S)H^T | \nabla_W f(W,H) > \\
&= 2 < (\alpha_W \nabla_W f(W,H) - W)H + S | \nabla_W f(W,H)H >
\end{aligned}$$

Hence :

$$\begin{aligned}
\phi'(\alpha_W) = 0 \Leftrightarrow \alpha_W &= \frac{< WH - S | \nabla_W f(W,H)H >}{< \nabla_W f(W,H)H | \nabla_W f(W,H)H >} \\
&= \frac{< (WH - S)H^T | (WH - S)H^T >}{||\nabla_W f(W,H)H||^2} \\
&= \frac{||(WH - S)H^T||^2}{||\nabla_W f(W,H)H||^2} \\
&= \frac{||\nabla_W f(W,H)||^2}{2||\nabla_W f(W,H)H||^2}
\end{aligned}$$

Hence: If $\nabla_W f(W,H)H \neq 0$:

$$\alpha_W = \frac{||\nabla_W f(W,H)||^2}{2||\nabla_W f(W,H)H||^2} \geq 0$$

(α_W is the unique minimum of ϕ and is reached in the admissible set).
Else: ϕ is constant, and α_W can be set to 0. Similarly $\alpha_H = \frac{||\nabla_H f(W,H)||^2}{2||W \nabla_H f(W,H)||^2}$

2.6 Question 8b

let $X \in \mathbb{R}_+^{m\times k}$. $P_1(X)$ exists because it is the projection of X on the convex closed subset $\mathbb{R}_+^{m\times k}$ of the Hilbert space $\mathbb{R}^{m\times k}$. Thus, $P_1(X)$ is the unique element of $\mathbb{R}_+^{m\times k}$ that verifies:

$$\forall y \in \mathbb{R}_+^{m\times k}, < P_1(x) - x | P_1(x) - y > \leq 0$$

Let then $y_0 = \max\{X, 0\} \in \mathbb{R}_+^{m\times k}$

We have, with $\mathcal{I}_- = \{(i,j) \mid x_{ij} < 0\}$, and $(E_\gamma)_{\gamma\in[|1,m|]\times[|1,k|]}$ the canonic base of $R^{m\times k}$:

$$y_0 - x = \sum_{\gamma\in\mathcal{I}_-} |x_\gamma| E_\gamma$$

$$y_0 = \sum_{\gamma\notin\mathcal{I}_-} |x_\gamma| E_\gamma$$

Hence :

$$\langle y_0 - x | y_0 \rangle = 0$$

And $\forall y \in \mathbb{R}_+^{m\times k}$

$$\begin{aligned}
\langle y_0 - x | y_0 - y \rangle &= -\langle y_0 - x | y \rangle \\
&= -\sum_{\gamma\in\mathcal{I}_-} y_\gamma |x_\gamma| \leq 0
\end{aligned}$$

Because $y \in \mathbb{R}_+^{m\times k}$. Thus, $P_1(X) = \frac{X+|X|}{2} = \max\{X, 0\}$ where $X \in \mathbb{R}^{m\times k}$ and $P_2(X) = \frac{X+|X|}{2} = \max\{X, 0\}$ where $X \in \mathbb{R}^{k\times n}$.

2.7 Question 8a

The Lipschitz constant has to satisfy

$$\|\nabla_W f(W_1, H) - \nabla_W f(W_2, H)\| \leq L_W \|W_1 - W_2\|$$

Given that $\nabla_W f(W, H) = -2(S - WH)H^\top$, and with the following:

$$\begin{aligned}
\|(W_1 - W_2)HH^\top\|^2 &= \sum_i^m \|HH^\top (W_1 - W_2)_i\|^2 \\
&\leq \sum_i^m \|HH^\top\|_{op}^2 \|(W_1 - W_2)_i\|^2 \\
&\leq \|HH^\top\|_{op}^2 \| \|W_1 - W_2\|^2
\end{aligned}$$

where we consider $(W_1 - W_2)_i$ to be the transpose of the ith row of $(W_1 - W_2)$ and by the definition of the Frobenius norm, we get the first equality. By the definition of the operator norm, we have the second inequality. Thus, $L_W \leq 2\|HH^\top\|_{op}$, as L_W is the smallest such constant.

Similarly, since $\|HH^\top\|_{op}$ is an eigenvalue of $HH^\top$ (the greatest one), there exists an eigenvector w associated with the eigenvalue. Furthermore, we can find $W_1, W_2 \geq 0$ such that $(W_1 - W_2) \in \mathbb{R}^{m \times k}$ and $\forall i\ (W_1 - W_2)_i = w$.

$$
\begin{aligned}
\|(W_1 - W_2)HH^\top\|^2 &= \sum_i^m \|HH^\top w\|^2 \\
&= \|HH^\top\|_{op}^2 \sum_i^m \|w\|^2 \\
&= \|HH^\top\|_{op}^2 \|(W_1 - W_2)\|^2
\end{aligned}
$$

Thus, we have that $L_W \geq 2\|HH^\top\|_{op}$ and thus, we have equality. Similarly, we can find that $L_H = 2\|W^\top W\|_{op}$. That was the best Lipschitz constant, but its calculation requires to diagonalize HH^T.

$$
\|(W_1 - W_2)HH^\top\|^2 \leq \|HH^\top\|^2 \|(W_1 - W_2)\|^2
$$

Thus, $\|HH^\top\| \geq \|HH^\top\|_{op}$, which is clear from sub-multiplicity of the norm, and similarly for L_H, $\|W^\top W\| \geq \|W^\top W\|_{op}$. To avoid the expensive calculation of the eigenvalues, we can use these values instead.

2.8 Question 8b

$$
\begin{aligned}
\operatorname*{Prox}_{\delta_S}(X) &= \operatorname*{argmin}_{Y \in \mathbb{R}^{m \times n}} \{\delta_S(Y) + \|X - Y\|^2\} && (1) \\
&= \operatorname*{argmin}_{Y \in S} \{\|X - Y\|^2\} && (2) \\
&= P_S(X) && (3)
\end{aligned}
$$

This is because if $Y \notin S$, the function $\delta_S(Y) + \|X - Y\|^2 = \infty$, thus our search space is reduced to when $Y \in S$, and then $\delta_S(Y) + \|X - Y\|^2 = \|X - Y\|^2$, giving us the definition of a projection on S.

2.9 Question 8c

In this case when $g = \delta_{W \geq 0}$ and $h = \delta_{H \geq 0}$, we have essentially the same algorithm as A-PGD, except that the constants c and d are not optimal, and are based on hyper-parameters like γ_W, γ_H. PALM is thus, a generalization of A-PGD, which allows us to add regularization terms to reduce the L^1 norm for example.

2.10 Question 8a

$$\begin{aligned}\operatorname*{Prox}_{\mu g}(X) &= \operatorname*{argmin}_{W\in\mathbb{R}^{m\times n}}\{\delta_{W\geq 0}(W) + \lambda_W \|\text{vec}(W)\|_1 + \frac{1}{2\mu}\|X-W\|_F^2\} \\ &= \operatorname*{argmin}_{W\geq 0}\{\lambda_W \|\text{vec}(W)\|_1 + \frac{1}{2\mu}\|X-W\|_F^2\}\end{aligned}$$

This is a strongly convex optimization problem, and thus there exists a unique minimum. By KKT conditions, we can set the constraints $c_{ij} = -W_{ij} \leq 0$, $\lambda_{ij} \geq 0$, $\lambda_{ij}c_{ij}(W^*) \geq 0$.

Then by KKT and taking the ij component:

$$0 \in \partial f(W^*)_{ij} + \sum \lambda_{ij}\partial c_{ij}(W^*) \tag{4}$$

$$0 \in \lambda_W \partial\|\text{vec}(\cdot)\|_1(W^*) + \frac{\partial}{\partial W_{ij}}\left(\frac{1}{2\mu}\|X-W\|^2\right)\Bigg|_{W=W^*} + \lambda_{ij}\partial c_{ij}(W^*) \tag{5}$$

$$0 \in \lambda_W \partial|W_{ij}^*| - \frac{1}{\mu}(X-W^*)_{ij} + \lambda_{ij}\partial c_{ij}(W^*) \tag{6}$$

$$0 \in \lambda_W \partial|W_{ij}^*| - \frac{1}{\mu}(X-W^*)_{ij} - \lambda_{ij} \tag{7}$$

$$0 \in \mu\lambda_W \partial|W_{ij}^*| - X_{ij} + W_{ij}^* - \mu\lambda_{ij} \tag{8}$$

If $W_{ij}^* \neq 0$ then $W_{ij}^* > 0$ and $\lambda_{ij} = 0$. Furthermore $\partial|W_{ij}^*| = \{1\}$.

$$0 \in \mu\lambda_W \partial|W_{ij}^*| - X + W^* - \mu\lambda_{ij} \tag{9}$$

$$0 = \mu\lambda_W \ - X_{ij} + W_{ij}^* \tag{10}$$

$$W_{ij}^* = X_{ij} - \mu\lambda_W \tag{11}$$

Otherwise, if $W_{ij}^* = 0$ then $\lambda_{ij} \geq 0$. Furthermore $\partial|W_{ij}^*| = [-1, 1]$.

$$0 \in \mu\lambda_W \partial|W_{ij}^*| - X + W^* - \mu\lambda_{ij} \tag{12}$$

$$0 \in \mu\lambda_W[-1,1] - X_{ij} - \mu\lambda_{ij} \tag{13}$$

$$\lambda_{ij} \in [-\lambda_W - \frac{X_{ij}}{\mu}, \lambda_W - \frac{X_{ij}}{\mu}] \tag{14}$$

But $\lambda_{ij} \geq 0$, thus $\lambda_W - \frac{X_{ij}}{\mu} \geq 0 \iff X_{ij} \leq \mu\lambda_W$.
Therefore we have

$$W_{ij}^* = \begin{cases} X_{ij} - \mu\lambda_W & \text{if } X_{ij} > \mu\lambda_W \\ 0 & \text{otherwise} \end{cases}$$

Thus, we can take $\text{Prox}_{\mu g}(X) = \max\{X - \mu\lambda_W \mathbb{1}, 0\}$, where $\forall (i,j)\ \mathbb{1}_{ij} = 1$ and $\text{Prox}_{\mu h}(X) = \max\{X - \mu\lambda_H \mathbb{1}, 0\}$.

2.11 Question 8c

The solutions we obtained from A-PGD took longer to converge but had a smaller error as compared to factorisation with PALM. However, we observed that the L1-Norm for the A-PGD solution was abnormally high at $\|W\|_1 + \|H\|_1 \approx 10^8$, (probably due to the very large number being multiplied by a small number in the matrix multiplication, thus not affecting the overall cost function). However, in PALM this was not the case as $\|W\|_1 + \|H\|_1 \approx 174$. This is due to the regularization terms λ_W and λ_H which encourage sparsity and a smaller L1-Norm. However, this has a small impact on the overall error, but it was not significant.

2.12 Question 9a

Lemma: If $M \in T_p(A)$, then $\|M\|_0 = \min\{p, \|A\|_0\}$
Suppose that $\|A\|_0 \geq p$ and $\|M\|_0 < p$. Then $\exists(i,j)$ such that $M_{ij} = 0$ and $A_{ij} \neq 0$.

$$M'_{pq} = \begin{cases} A_{ij} & \text{if } (i,j) = (p,q) \\ M_{pq} & \text{otherwise} \end{cases}$$

Then $\|M'\|_0 \leq p$ and $\|A - M'\|^2 < \|A - M\|^2$, which is a contradiction. Suppose now that $\|A\|_0 < p$ and $\|M\|_0 < \|A\|_0$. Then $\exists(i,j)$ such that $M_{ij} = 0$ and $A_{ij} \neq 0$ and by a similar M' above, we arrive at a contradiction.

Suppose that $\|A\|_0 < p$ and $\|M\|_0 > \|A\|_0$. Then $\exists(i,j)$ such that $M_{ij} \neq 0$ and $A_{ij} = 0$.

$$M'_{pq} = \begin{cases} 0 & \text{if } (i,j) = (p,q) \\ M_{pq} & \text{otherwise} \end{cases}$$

Then $\|M'\|_0 \leq p$ and $\|A - M'\|^2 < \|A - M\|^2$, which is a contradiction.
Thus, we arrive at the conclusion that if $M \in T_p(A)$, then $\|M\|_0 = \min\{p, \|A\|_0\}$

Suppose that $M \in T_p(A)$ and $\exists(i,j) \in \{\text{index of p largest values of A}\}$ such that $M_{ij} \neq A_{ij}$
Case 1: $M_{ij} \neq 0$:
Consider

$$M'_{pq} = \begin{cases} A_{ij} & \text{if } (i,j) = (p,q) \\ M_{pq} & \text{otherwise} \end{cases}$$

This leads to a contradiction, as $\|M'\| \leq p$ and $\|A - M'\|^2 < \|A - M\|^2$.
Case 2: $M_{ij} = 0$:
Consider if $\|M\|_0 = \|A\|_0$. Then there $\exists(m,n)$ such that $M_{mn} \neq 0$ and $A_{mn} = 0$

$$M'_{pq} = \begin{cases} A_{ij} & \text{if } (i,j) = (p,q) \\ 0 & \text{if } (i,j) = (m,n) \\ M_{pq} & \text{otherwise} \end{cases}$$

This leads to a contradiction, as $\|M'\| \leq p$ and $\|A - M'\|^2 < \|A - M\|^2$.

Now consider if $\|M\|_0 \neq \|A\|_0 \implies \|A\|_0 > p$ and $\|M\|_0 = p$. Then there $\exists(m, n)$ such that $M_{mn} \neq 0$ and $(m, n) \notin \{\text{index of p largest values}\}$ (otherwise $\|M\|_0 \neq p$)

$$M'_{pq} = \begin{cases} A_{ij} & \text{if } (i,j) = (p,q) \\ 0 & \text{if } (i,j) = (m,n) \\ M_{pq} & \text{otherwise} \end{cases}$$

This is essentially swapping the (i, j) index for the (m, n) index. This gives us:

$$\|A - M'\|^2 - \|A - M\|^2 = A_{mn}^2 - (A_{ij}^2 + (A_{mn} - M_{mn})^2)$$

If $A_{mn} \neq A_{ij} \implies A_{mn} < A_{ij}$, then the above inequality < 0 and we have our contradiction.

Otherwise, (m, n) belongs to a second set of $\{\text{index of p largest values of } A\}$ and then by the initial assumption about M we have that $M_{mn} \neq A_{mn}$ which gives us our contradiction.

Thus, $\forall(i, j) \in \{\text{index of p largest values}\}$ such that $M_{ij} = A_{ij}$ and then as any matrix with the above condition would have the same value of $\|A - M\|_F^2$ we have that the set of matrices completely defines $T_p(A)$ ie:

$$\begin{aligned} T_p(A) = \{M \in \mathbb{R}^{m \times n} \mid \forall(i,j) \in \{\text{index of p largest values}\}\ M_{ij} = A_{ij}, \\ \forall(i,j) \notin \{\text{index of p largest values}\}\ M_{ij} = 0\} \end{aligned}$$

2.13 Question 9b

If we consider a modified version of the Prox_g, since $\forall(i, j) \in I^+(A)$ we have that $\max\{A, 0\}_{ij} = A_{ij}$ and that $\forall(i, j) \in I^-(A)$ we have that $\max\{A, 0\}_{ij} = 0$

$$\begin{aligned} \text{Prox}_g(A) &= \underset{M \in \mathbb{R}^{m \times n}}{\text{argmin}} \left\{ \sum_{(i,j) \in I^+(A)} (A_{ij} - M_{ij})^2 + \sum_{(i,j) \in I^-(A)} M_{ij}^2 \;\middle|\; \|M\|_0 \leq p,\ M \geq 0 \right\} \\ &= \underset{M \in \mathbb{R}^{m \times n}}{\text{argmin}} \left\{ \sum_{(i,j) \in I^+(A)} (\max\{A, 0\} - M)_{ij}^2 + \sum_{(i,j) \in I^-(A)} (\max\{A, 0\} - M)_{ij}^2 \;\middle|\; \|M\|_0 \leq p,\ M \geq 0 \right\} \\ &= \underset{M \in \mathbb{R}^{m \times n}}{\text{argmin}} \left\{ \|\max\{A, 0\} - M\|^2 \;\middle|\; \|M\|_0 \leq p,\ M \geq 0 \right\} \\ &= T_P(\max\{A, 0\}) \end{aligned}$$

Otherwise, we can only obtain subset inclusion in one way, $T_p(\max\{A, 0\}) \subseteq \text{Prox}_g(A)$.

$$
\begin{aligned}
S &= \operatorname*{argmin}_{M\in\mathbb{R}^{m\times n}} \left\{ \|\max\{A,0\}-M\|^2 \;\middle|\; \|M\|_0 \le p,\ M\ge 0,\ \forall (i,j)\in I^-(A)\ M_{ij}=0 \right\} \\
&= \operatorname*{argmin}_{M\in\mathbb{R}^{m\times n}} \left\{ \sum_{(i,j)\in I^+(A)} (A_{ij}-M_{ij})^2 + \sum_{(i,j)\in I^-(A)} M_{ij}^2 \;\middle|\; \|M\|_0 \le p,\ M\ge 0,\ \forall (i,j)\in I^-(A)\ M_{ij}=0 \right\} \\
&= \operatorname*{argmin}_{M\in\mathbb{R}^{m\times n}} \left\{ \sum_{(i,j)\in I^+(A)} (A_{ij}-M_{ij})^2 + \sum_{(i,j)\in I^-(A)} M_{ij}^2 \;\middle|\; \|M\|_0 \le p,\ M\ge 0 \right\} \\
&= \mathrm{Prox}_g(A)
\end{aligned}
$$

$$
S' = \operatorname*{argmin}_{M\in\mathbb{R}^{m\times n}} \left\{ \|\max\{A,0\}-M\|^2 \;\middle|\; \|M\|_0 \le p,\ M\ge 0 \right\}
$$

Now consider the matrix $Q^* \in S'$ and suppose that $\exists (i,j) \in I^-(A)$ such that $Q^*_{ij} \neq 0 \implies Q^*_{ij} > 0$. Then it satisfies the following:
$\forall M \in \mathbb{R}^{m\times n}, \|M\|_0 \le p,\ M \ge 0$

$$
\begin{gathered}
\|\max\{A,0\}-Q^*\|^2 \le \|\max\{A,0\}-M\|^2 \\
\sum_{(i,j)\in I^+(A)} (A_{ij}-Q^*_{ij})^2 + \sum_{(i,j)\in I^-(A)} (Q^*_{ij})^2 \le \sum_{(i,j)\in I^+(A)} (A_{ij}-M_{ij})^2 + \sum_{(i,j)\in I^-(A)} (M_{ij})^2
\end{gathered}
$$

Consider

$$
P^* = \begin{cases} 0 & \text{if } (i,j)\in I^-(A) \\ Q^*_{ij} & \text{otherwise} \end{cases}
$$

Clearly, $\|P^*\|_0 \le p$ and $P^* \ge 0$ but then

$$
\begin{gathered}
\|\max\{A,0\}-Q^*\|^2 \le \|\max\{A,0\}-P^*\|^2 \\
\sum_{(i,j)\in I^+(A)} (A_{ij}-Q^*_{ij})^2 + \sum_{(i,j)\in I^-(A)} (Q^*_{ij})^2 \le \sum_{(i,j)\in I^+(A)} (A_{ij}-P^*_{ij})^2 + \sum_{(i,j)\in I^-(A)} (P^*_{ij})^2 \\
\sum_{(i,j)\in I^-(A)} (Q^*_{ij})^2 \le \sum_{(i,j)\in I^-(A)} (P^*_{ij})^2 \\
\sum_{(i,j)\in I^-(A)} (Q^*_{ij})^2 \le 0
\end{gathered}
$$

This leads to a contradiction as $Q^*_{ij} > 0$ and thus we have $S = S'$. This implies that $T_p(\max\{A,0\}) \subseteq \mathrm{Prox}_g(A)$. These two sets are not equal as we can find for

$$
A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \; M = \begin{bmatrix} 1 & 0 \\ 0 & x \end{bmatrix}
$$

for some $p = 2,\ x \geq 0,\ M \in \text{Prox}_g(A)$ but $M \notin T_p(\max\{A, 0\})$. It is likely that the set equality is only true when $p \leq \mathbf{card}(I_+)$ and we can prove that otherwise $Tp(\max\{A, 0\})$ will always have only one element, and Prox_g will have multiple elements.

2.14 Question 9d

The parameters p_W and p_H control the sparsity of the matrices. Sparser matrices tend to limit the accuracy of the matrix factorization, and the rate of convergence, especially at really high sparsity. However, after a certain point, there are diminishing returns, and decreasing the sparsity does not increase the error much. Furthermore, we can observe that the calculations can take a shorter amount of time if `scipy.sparse_array` is used. We tested a few values of p_W and p_H and found that setting $p_W = \frac{m \times k}{2}$ and $p_H = \frac{k \times n}{2}$ has led to sufficiently good results, which was about 0.2% higher than the optimal result $p_W = m \times k$ and $p_H = k \times n$, which is an acceptable trade-off for faster calculations.

Final Results from the implementation of the algorithms:

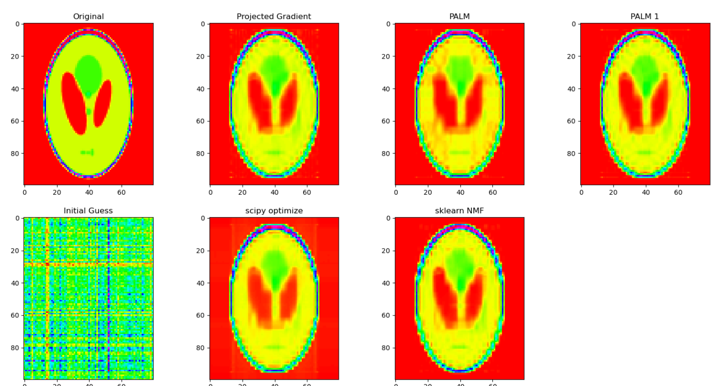

Figure 4: Optimization Results