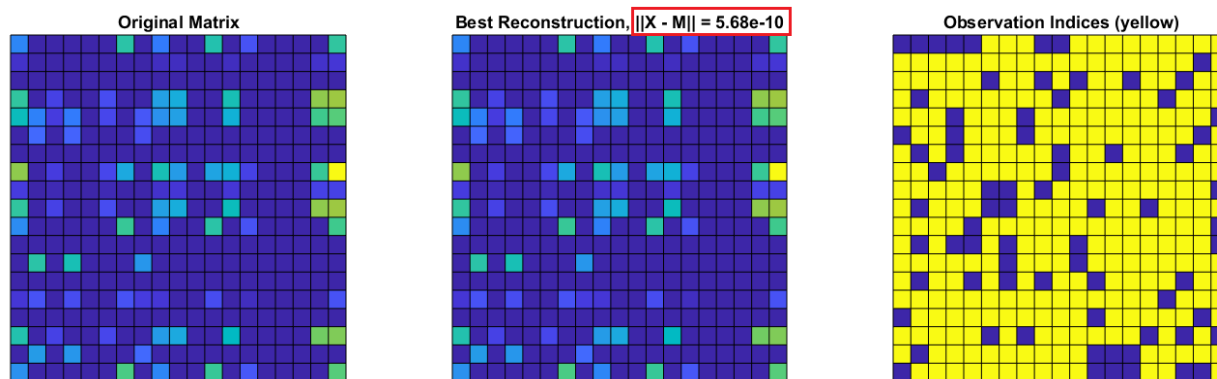
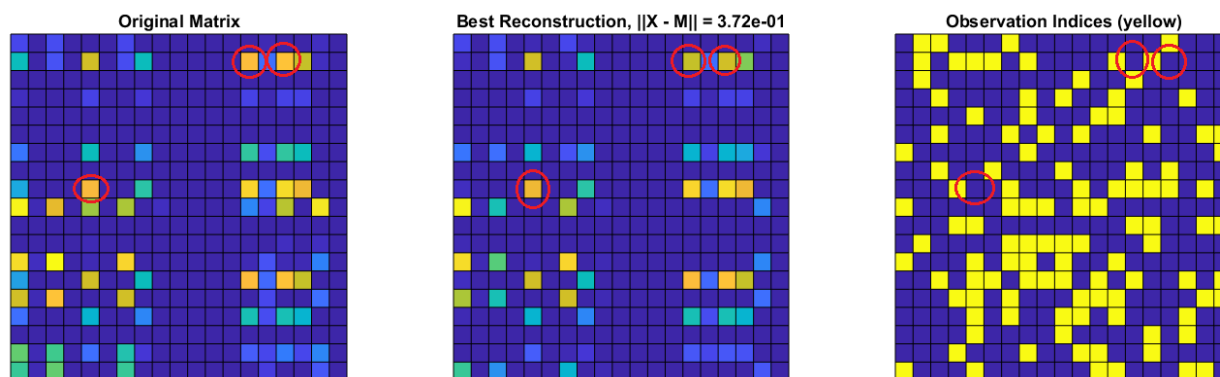


If we observe enough of the entries (80% in this example), we get perfect reconstruction for sparse matrices with relatively high probability (perfect reconstruction roughly half the time) using nuclear norm minimization

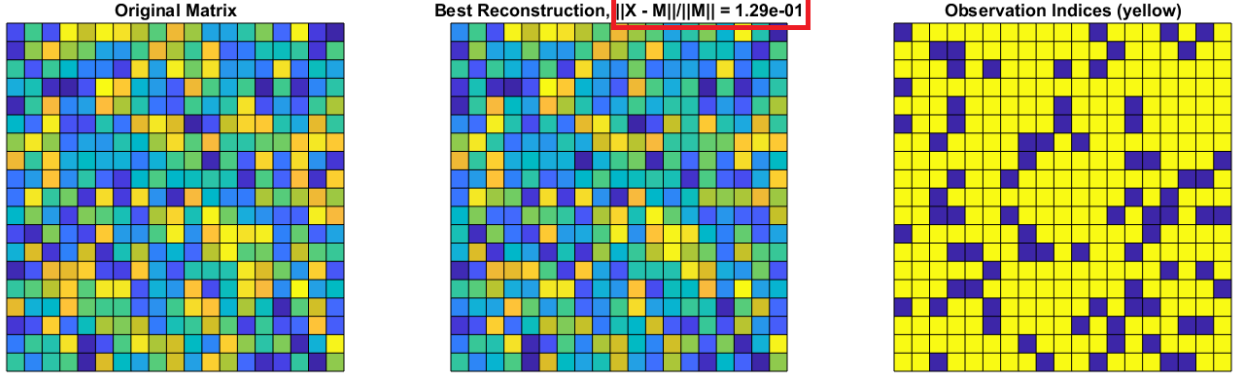
$$\min \|X\|_* \quad \text{subject to} \quad P_\Omega(X) = P_\Omega(M).$$



When we sample much less (30% in this example), we no longer get perfect reconstruction, but we do preserve some of the structure of the matrix. Note the entries circled in red: these are large entries in the original matrix and, though they are unobserved, the reconstruction matches them fairly well. [Of course, the reconstruction also misses some large unobserved entries.]



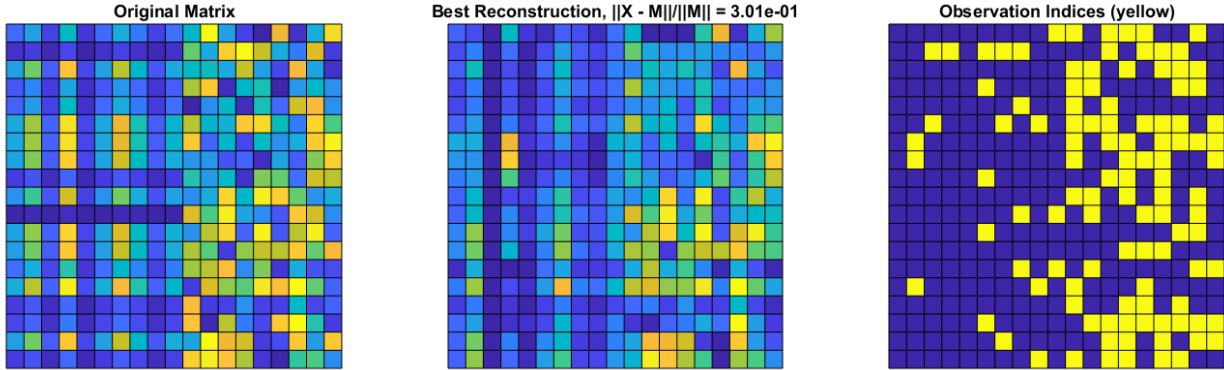
Moving to dense matrices, we cannot expect perfect reconstruction even when observing 80% of entries.



We want to consider matrices with certain columns correlated. As a first attempt, we let the first ten columns be "perfectly correlated" (i.e., they are all constant multiples of a single vector so that their correlation matrix is a  $10 \times 10$  matrix where every entry is 1). In this case, we can selectively sample the entries. Here we sample 30% of the entire matrix (120 entries), but we take only 20% (24 entries) of these samples from the correlated columns and the remaining 80% (96 entries) from the uncorrelated columns. Note, we are still using the nuclear norm minimization

$$\min \|X\|_* \quad \text{subject to} \quad P_\Omega(X) = P_\Omega(M).$$

This selective sampling does worse than if we just sample evenly in the same scenario. This should be expected since we haven't actually used the correlation yet.



We tried to think of ways to incorporate the correlation into the problem. A first thought might be to change the functional to incorporate the correlation matrix itself, perhaps:

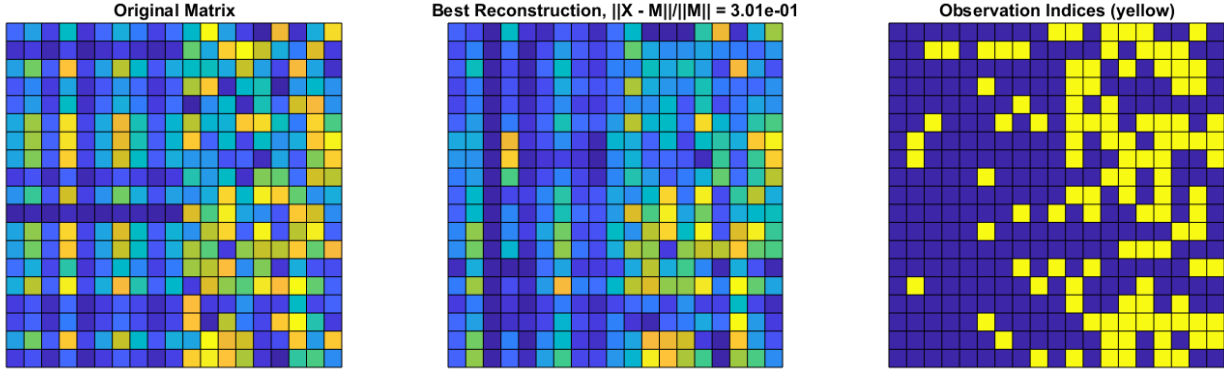
$$\min \|X\|_* + \gamma \|\text{Corr}(X_\tau) - \text{Corr}(M_\tau)\|_F^2 \quad \text{subject to} \quad P_\Omega(X) = P_\Omega(M)$$

where  $\tau$  represents the collection of columns that are known to be correlated. However, the map  $X \mapsto \text{Corr}(X)$  is non-linear and so that map  $X \mapsto \|\text{Corr}(X_\tau) - \text{Corr}(M_\tau)\|_F^2$  is non-convex. If we wanted to do this, we would need some more sophisticated non-convex optimization routines. Instead, it may be better to use some portion of the samples to

produce an estimation  $\tilde{M}_\tau$  to  $M_\tau$  and then insert a fidelity term in order to match  $X_\tau$  to  $\tilde{M}_\tau$  and combine this with nuclear norm minimization:

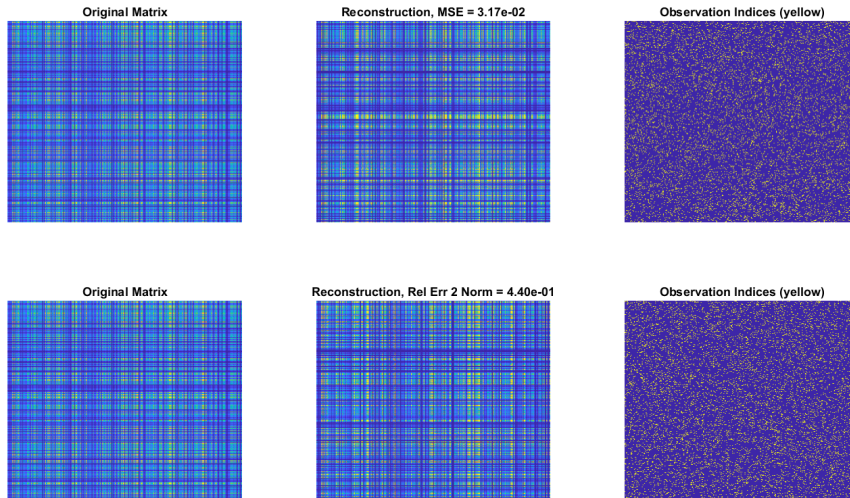
$$\min \|X\|_* + \gamma \|X_\tau - \tilde{M}_\tau\|_F^2 \quad \text{subject to} \quad P_\Omega(X) = P_\Omega(M).$$

This idea shows some promise. Indeed assuming we can produce  $\tilde{M}_\tau$  with some amount of accuracy, we can sample at a much lower rate from the correlated columns since they are accounted for by the fidelity term. Indeed, we did a small parameter sweep and it suggested that taking  $\tilde{\gamma} = 1$  and taking all the samples from the uncorrelated columns is ideal (here we assume we have used 40 samples to construct  $\tilde{M}_\tau$  and then take the remaining 80 samples from the uncorrelated portion).



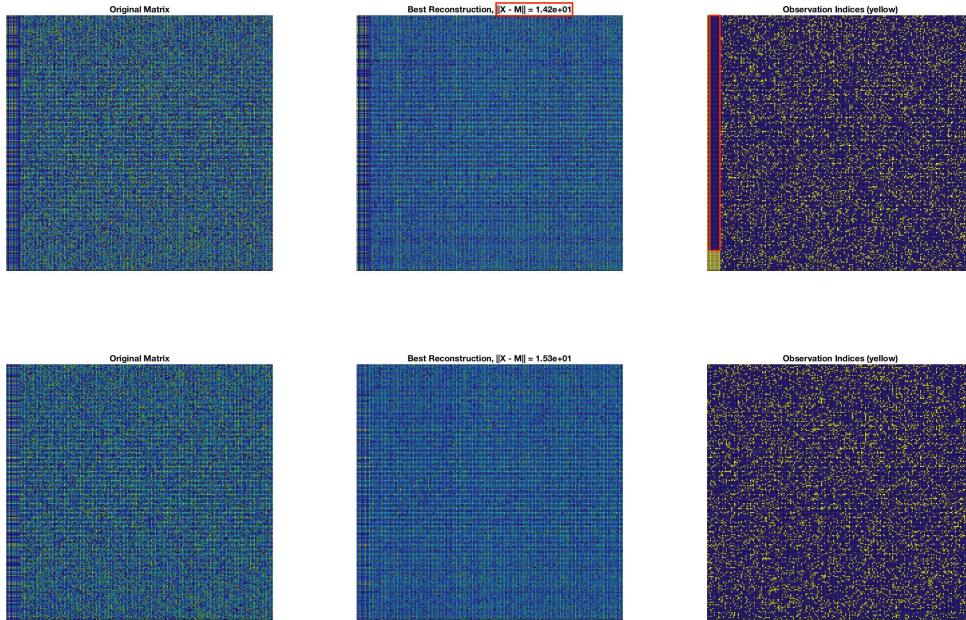
Next we are focusing on how to accurately reconstruct  $M_\tau$  with relatively few samples, given that we know that the columns are correlated (or an easier problems: given that we know exactly how the columns are correlated; i.e., we know  $\text{Corr}(M_\tau)$  exactly).

We've found that singular value thresholding (SVT) works pretty well for large, low rank matrices so we may be able to use SVT for  $\tilde{M}_\tau$  if we assume that  $M_\tau$  is low-rank. In the following picture, we are using SVT to approximately reconstruct at rank 1,  $300 \times 300$  matrix. We are still not using any information about the correlation (How to incorporate this into SVT?) but even so the reconstruction is very good when we are only observing 10% of the entries.



Otherwise, we may be able to use multivariate regression to identify the correlation. In this method, given that we know that which columns are correlated, we assume that they are linearly correlated. Then, we obtain a multivariate linear regression by sampling the minimum amount of rows from the correlated columns such that the covariance matrix is positive-definite. After that, we sample from the independent columns and use these coefficients to obtain best estimates for the corresponding rows of the dependent columns. These amount of samples normally add to about 5% of observations depending on the particular observation rate, with the other 95% we use to sample the non-correlated columns. The multivariate linear regression provides us with the hidden coefficients relating the correlated columns, which we then combine with the nuclear norm minimization as additional optimization constraints. In the following objective function, assuming that  $n$  columns are correlated with each other, we treat the elements in the first column as the independent variables and the elements in the next  $n - 1$  columns as the dependent variables.

$\min \|X\|_*$  subject to  $P_\Omega(X) = P_\Omega(M)$  and  $X_{k,j} = X_{k,i} * \beta_{(i,j)_0} + \beta_{(i,j)_1} \forall i, j$  where columns  $j$  are dependent



Our preliminary results are shown above for a 200 x 200 matrix, with 10 trials for each method. As we can see, adaptive sampling (the top one) improves the reconstruction error from  $1.53 \times 10^1$  to  $1.42 \times 10^1$  in comparison to the random sampling method. Also, the red rectangle in the adaptive sampling method highlights how most of the correlated columns remains unobserved.

We can always decrease the reconstruction error in the correlated columns by sampling more from the correlated columns to obtain a better estimate for the correlation. Next, we need to loosen the assumption that we know which rows/columns are correlated, and see if we can learn this as we sample. For example, we could sample coarsely and obtain a rough estimation. Then we could examine the correlation pattern from the estimation and adjust

the sampling scheme accordingly, then repeat. Even further down the road, we can loosen the assumption that the correlations are linear.

---

Rather than use the multivariate regression, if we know that the correlated columns are low rank, we can use random sampling to try to explicitly determine their relationship. We formulate this idea mathematically. Given  $M \in \mathbb{R}^{m \times n}$  with columns  $\vec{m}_1, \dots, \vec{m}_n$ , suppose that  $\tau \subset [n]$  with  $t := |\tau| \ll n$  and let  $M_\tau \in \mathbb{R}^{m \times t}$  be the matrix whose columns are  $\vec{m}_j$  for  $j \in \tau$ . Further, suppose we know that  $M_\tau$  is low rank, i.e.  $\text{rank}(M_\tau) = k < t$ .

Our goal is to find a collection of  $k$  columns of  $M_\tau$  (we call this collection  $M_\tau^{(k)} = [\vec{m}_{j_1} \cdots \vec{m}_{j_k}]$ ) and a matrix  $B \in \mathbb{R}^{k \times (t-k)}$  such that

$$M_\tau = M_\tau^{(k)} B. \quad (*)$$

That is, we want to extract  $k$  columns which form a basis for the column space of  $M_\tau$  and we want to find the coordinates of the other columns of  $M_\tau$  in this basis.

The question is how to find  $M_\tau^{(k)}$  and  $B$  while observing as little of  $M_\tau$  as possible. Notice, equation  $(*)$  is really  $t - k$  vector equations:

$$\vec{m}_\ell = b_{j_1, \ell} \vec{m}_{j_1} + b_{j_2, \ell} \vec{m}_{j_2} + \cdots + b_{j_k, \ell} \vec{m}_{j_k}, \quad \text{for } \ell \in \tau \setminus \{j_1, \dots, j_k\}.$$

However, these vectors live in  $\mathbb{R}^m$  while there are only  $k$  basis vectors. Thus we do not actually need these entire equations; rather, if we can extract a  $k \times k$  submatrix from  $M_\tau^{(k)}$  then we can solve for all the coefficients in  $B$  with only  $k^2$  observations. Unfortunately, knowing nothing about the specific entries of  $M_\tau$ , the best we can do is to randomly sample a  $k \times k$  submatrix. If it is invertible, then great, we have our  $M_\tau^{(k)}$  and we can solve for  $B$ . If it isn't invertible, we can keep those observations, but we need to re-sample to find an invertible  $k \times k$  matrix. However, so long as  $M$  is defined somewhat randomly and isn't sparse, a random  $k \times k$  submatrix will be invertible with high probability (in fact, we test this with the first  $t$  columns of  $M$  defined by  $XY$  where  $X \in \mathbb{R}^{m \times k}$  and  $Y \in \mathbb{R}^{k \times t}$  have entries that are uniform  $[0, 1]$  and thus a random  $k \times k$  submatrix of  $M_\tau$  will be invertible almost surely).

To summarize, our algorithm is:

1. Randomly sample  $I = \{i_1, \dots, i_k\} \subset [m]$  and  $J = \{j_1, \dots, j_k\} \subset \tau$ .
2. If the matrix  $(m_{ij})_{(i,j) \in I \times J}$  is invertible, then
  - (a) Define  $M_\tau^{(k)} = [\vec{m}_{j_1} \cdots \vec{m}_{j_k}]$
  - (b) Sample the remaining entries of the rows corresponding  $i_1, \dots, i_k$
  - (c) Solve for  $B$  using  $(*)$
  - (d) Break loop
3. If you reach this step, save the already observed entries and return to step 1.

Assuming that  $\tau = [t]$  (so that the first  $t$  columns are the correlated ones), this algorithm in MATLAB looks like:

```

indices1 = zeros(m,t);
DET = 0;
while DET == 0
    I = randperm(m,k);
    J = randperm(t,k);
    DET = det(M(I,J));
    indices1(i,j) = 1;
end
J1 = setdiff(1:t,j);
indices1(I,J1) = 1;
B = M(I,J) \ M(I,J1);

```

where `indices1` is keeping track of the observed entries.

Using this, we will explicitly find the relationship between the correlated columns. From here we can do one of two things. First option: we can use the relationship  $M_\tau = M_\tau^{(k)}B$  as a constraint in the minimization problem and proceed as before. In this case, our problem becomes

$$\min \|X\|_* \text{ subject to } P_\Omega(X) = P_\Omega(M) \text{ and } X_\tau = X_\tau^{(k)}B.$$

Here  $\Omega$  would no longer be a random sampling of the entries because it would include the entries we sampled in finding  $B$ . Also, we could choose to selectively sample more or less in the correlated columns (that is, there may be some advantage to selectively constructing  $\Omega$  rather than using the remaining observations randomly).

Otherwise, we could do the problem in two steps. We could first sample some entries  $\Omega_\tau$  of  $M_\tau$  and solve

$$\min \|Y\|_* \text{ subject to } P_{\Omega_\tau}(Y) = P_{\Omega_\tau}(M_\tau) \text{ and } Y = Y^{(k)}B.$$

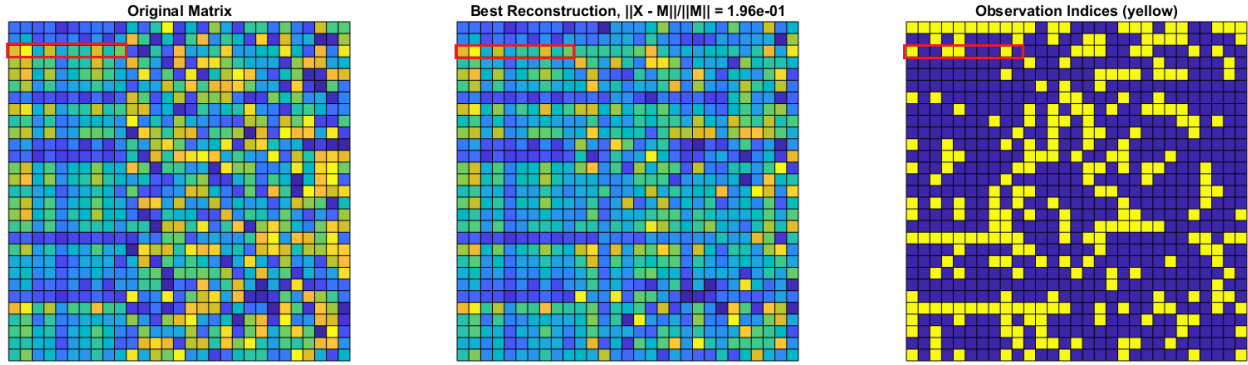
The solution  $\tilde{Y}$  to this problem will be an approximation to  $M_\tau$  (and it should be a very good approximation) so we can use it as a fidelity term in the next step and solve

$$\min \|X\|_* + \gamma \|X_\tau - \tilde{Y}\|_F^2 \text{ subject to } P_\Omega(X) = P_\Omega(M)$$

where  $\Omega$  again is a set of observations which could be selectively designed. In this case  $X_\tau$  would not carry the exact same correlation pattern as  $M_\tau$  but it should be fairly close.

Thus far, we have only taken the first approach: enforcing the relationship as a constraint in the problem. We are showing some accuracy gains in doing this. The following is a  $30 \times 30$  example where the first 10 columns form a rank 3 matrix. We show both our approach and the vanilla nuclear norm minimization approach. Notice that our approach does a bit better, both in average relative error (over 100 trials) and in best relative error. Also notice that in this case, we aren't doing anything special with  $\Omega$ . While using our approach, we are still taking roughly 1/3 of the observations from the correlated columns (which is uniform sampling since these columns take up 1/3 of the matrix) - after having selectively sampled a few entire rows in order to solve for  $B$ . One interesting feature of our approach is that, since we perfectly resolve the linear relationship between the correlated columns which is determined by  $k$  variables, if we sample  $k$  entries from a single row in  $M_\tau$ , then we perfectly

reconstruct that row (notice the row circled in red below - this is not one of the rows which we sample in its entirety, but because there happen to have been more than three observations in that row, it is perfectly reconstructed).



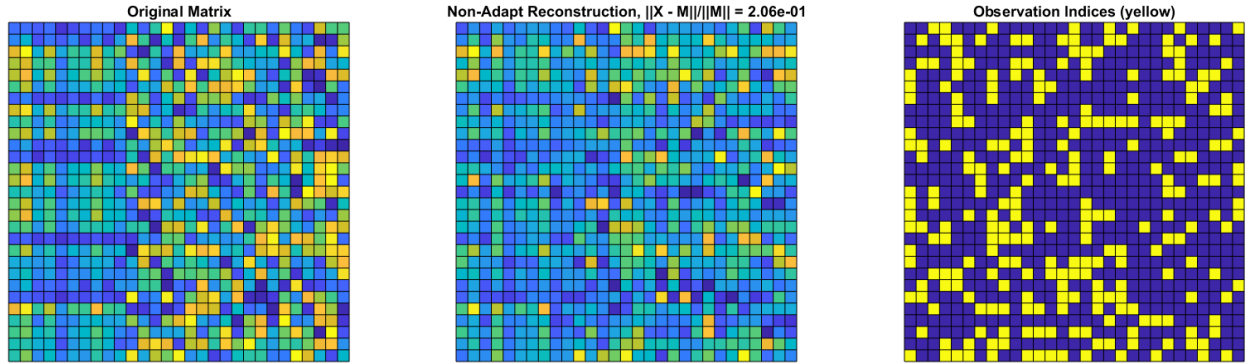
——- ADAPTIVE SAMPLE ——-

Proportion of observations from correlated columns: 0.34

Number of trials: 100

Average relative error: 0.2598

Best relative error: 0.1965



——- NON-ADAPTIVE SAMPLE ——-

Number of trials: 100

Average relative error: 0.2638

Best relative error: 0.2063

Currently, we are running a parameter sweep to find the best way to selectively design  $\Omega$ . That is, once we have solved for  $B$ , we are trying to find the best proportion of the remaining observations which should be taken from  $M_\tau$  versus  $M_{[n]\setminus\tau}$ .