# Matrix Completion

Given matrix $A$, $m \times n$, only some entries are observed $A_{ij}$, $(i,j) \in \Omega$, the objective is to fill in the missing entries. This can be used to predict user preferences such as user rating for unseen movies. The objective is to

$$\min_{X \in \mathbb{R}^{m \times m}} \frac{1}{2} \sum_{(i,j) \in \Omega} (A_{ij} - X_{ij})^2 + \lambda \|X\|_* \tag{1}$$

where $\|X\|_*$ is the nuclear norm of $X$ and is given by

$$\|X\|_* = \sum_{i=1}^{r} \sigma_i(X)$$

where $r = rank(X)$ and $\sigma_i$ is the $i$th singular value.
To solve this first lets define a projection operator onto the observed set

$$[P_\Omega(X)]_{ij} = \begin{cases} X_{ij} & (i,j) \in \Omega \\ 0 & \text{otherwise} \end{cases}$$

# Matrix Completion

The objective is to minimize the following function

$$f(X) = \frac{1}{2}\|P_\Omega(A) - P_\Omega(X)\|_F^2 + \lambda\|X\|_*$$

which has the same form as $f(\mathbf{x}) + c(\mathbf{x})$. Now projection function is convex and the Frobenius norm is differentiable and convex and the nuclear norm is convex but not differentiable. Here we can apply generalized gradient descent. The gradient is $\nabla g(X) = -(P_\Omega(A) - P_\Omega(X))$ and the prox function is

$$\text{prox}_t(X) = \arg\min_{Z\in\mathbb{R}^{m\times n}} \frac{1}{2t}\|X - Z\|_F^2 + \lambda\|Z\|_*$$

If we select $\text{prox}_t(X) = Z$ then $Z$ should satisfy

$$0 \in Z - X + \lambda t.\delta\|Z\|_* \tag{2}$$

We recap discussion from quiz 2, problem 3 which was about the restricted case where $X$ and $Z$ were restricted to be diagonal matrices and the solution was obtained simply using ISTA (Lasso)

# Quiz 2, Problem 3: The simpler case of Matrix Completion

Consider solving the following problem of determining proximal (diagonal matrix $Z \in \Re^{n\times n}$) solution to $Z = \text{prox}_t(X)$ with the non-differentiable function, called the nuclear norm $c(Z) = \|Z\|_*$. Here $X \in \Re^{n\times n}$ is also a diagonal matrix but is fixed (in an iteration). More specifically,

$$\text{prox}_t(X) = \arg \min_{Z\in\mathbb{R}^{n\times n}, Z \text{ is diagonal}} \frac{1}{2}\|X - Z\|_F^2 + \lambda\|Z\|_*$$

Here $\|Z\|_*$ is the nuclear norm and equals the sum of its singular values. For a diagonal matrix $Z$, $\|Z\|_*$ equals the sum of absolute value of its diagonal elements and can be assumed to be a convex non-differentiable function.

The solution can be **EITHER worked out along same lines as LASSO OR simply recovered from the homework problem** and its solution presented in the class by **vectorizing** the non-zero (diagonal) elements of $X$ and $Z$ to vectors $\mathbf{x}$ and $\mathbf{z}$ respectively such that $X_{ii} = \mathbf{x}_i$ and $Z_{jj} = \mathbf{z}_j$ (and also setting $t = 1/2$) and applying proximal descent step on Lasso:

# Quiz 2, Problem 3: The simpler case of Matrix Completion (contd.)

Recap from class from the **Iterative Soft Thresholding Algorithm:**

- Compute $\mathbf{w}^{(k+1)} = \underset{\mathbf{w}}{argmin}||\mathbf{w} - \widehat{\mathbf{w}}^{(\mathbf{k+1})}||_{\mathbf{2}}^{\mathbf{2}} + \lambda\mathbf{t}||\mathbf{w}||_{\mathbf{1}}$ by:
  1. If $\widehat{w}_i^{(k+1)} > \lambda t/2$, then $w_i^{(k+1)} = -\lambda t/2 + \widehat{w}_i^{(k+1)}$
  2. If $\widehat{w}_i^{(k+1)} < -\lambda t/2$, then $w_i^{(k+1)} = \lambda t/2 + \widehat{w}_i^{(k+1)}$
  3. $0$ otherwise.

# Quiz 2, Problem 3: The simpler case of Matrix Completion (contd.)

Applying the same idea (or deriving along similar lines as discussed in class) with $\textit{vectorized}(X) = \mathbf{x} = \widehat{\mathbf{w}}^{(\mathbf{k+1})}$ and $\textit{vectorized}(Z) = \mathbf{z} = \mathbf{w}^{(k+1)}$

- Compute $Z = \underset{\textit{diagonal } Z}{\textit{argmin}} ||Z - X||_2^2 + 2\lambda||Z||_*$ by:
  1. If $X_{ii} > \lambda$, then $Z_{ii} = -\lambda + X_{ii}$
  2. If $X_{ii} < -\lambda$, then $Z_{ii} = \lambda + X_{ii}$
  3. $0$ otherwise.

Now if we let $Z$ and $X$ be arbitrary non-diagonal matrices but use the Singular Value Decomoposion for $Z$ as $Z = U\Sigma V^T$ then

$$\delta||Z||_* = \{UV^T + W : W \in \mathbb{R}^{m\times n}, \|W\| \leq 1, U^T W = 0, WV = 0\}$$

. We will see that our solution is a natural generalization of the case of diagonal $Z$ (in Quiz 2, problem 3).

# Matrix Completion

If we let $Z = U\Sigma_\lambda V^T$ then equation 2 holds, here $X = U\Sigma V^T$ is the SVD and $\Sigma_\lambda$ is given by the diagonal matrix

$$(\Sigma_\lambda)_{ii} = \max\{\Sigma_{ii} - \lambda, 0\}$$

This is true because $X - Z = \lambda UV^T \in \delta\|Z\|_*$. Thus the prox function can be written as

$$\text{prox}_t(X) = S_{\lambda t}(X) = U\Sigma_\lambda V^T$$

and the generalized gradient update step is

$$X^+ = S_{\lambda t}(X + t(P_\Omega(A) - P_\Omega(X)))$$

since $\|\nabla g(Y) - \nabla g(X)\|_F = \|P_\Omega(A) - P_\Omega(X)\|_F \leq \|Y - X\|_F$ the Lipschitz constant is $L = 1$ thus the step size can be picked as $t = 1$ leading to

$$X^+ = S_\lambda(P_\Omega(A) + X - P_\Omega(X)) = S_\lambda(P_\Omega(A) + P_\Omega^\perp(X))$$

where $P_\Omega(X) + P_\Omega^\perp(X) = X$ This is called the **soft-impute** algorithm for matrix completion.

# Matrix Completion

In the case of matrix completion, acceleration and even backtracking can hurt performance. The matrix completion problem is described in Lecture 8. Briefly, Given a matrix A, only some entries $(i, j) \in \Omega$ of which are visible to you, you want to fill in the rest of entries, while keeping the matrix low rank. We solve,

$$\min_{X} \frac{1}{2} \|P_\Omega(A) - P_\Omega(\mathbf{x})\|_F^2 + \lambda \|X\|_*$$

where $\|X\|_* = \sum_{i=1}^{r} \sigma_i(\mathbf{x})$ is the nuclear norm, r is the rank of X and $P_\Omega(\cdot)$ is the projection operator,

$$\left[P_\Omega(\mathbf{x})\right]_{ij} = \begin{cases} X_{ij} & (i,j) \in \Omega \\ 0 & (i,j) \notin \Omega \end{cases}$$

the gradient descent updates, also known as the *soft-impute* algorithm are

$$X^+ = S_\lambda(P_\Omega(A) + P_\Omega^\perp(\mathbf{x}))$$

where $S_\lambda(\cdot)$ is the matrix soft-thresholding operator which requires the SVD to compute as $S_\lambda(\mathbf{x}) = U\Sigma_\lambda V^T$ where $(\Sigma_\lambda)_{ii} = \max\{\Sigma_{ii} - \lambda, 0\}$ . Calculating the SVD can be expensive and can cost upto $O(mn^2)$ operations.