Lecture Notes 8: Matrix Completion

Professor: Zhihua Zhang *Scribe:Yuxi Zhang, Ruotian Luo*

8 Matrix Completion

8.1 Problem Background

Let $\mathbf{Z} = (z_{ij})$ denote user i rates movie j with ranking $z_{ij} \in \{1, 2, 3, 4, 5\}$. However, some movies are not rated by any people, i.e. z_{ij} is missing. Our purpose is to complete the ranking matrix. Let $\mathbf{X}$ be the $m \times n$ complete matrix and $\mathbf{\Omega} = \{(i, j)|z_{ij} \text{ is observed}\}$. We can reasonably suppose that $\mathbf{X}$ is a low rank matrix.
Our purpose is to minimize the error between $\mathbf{Z}$ and $\mathbf{X}$ with the constraint that $\mathbf{X}$ is low rank, i.e.

$$\begin{aligned} &\min_{\mathbf{X}} \frac{1}{2} \sum_{(i,j)\in\Omega} (z_{ij} - x_{ij})^2 \\ &\text{s.t. } \mathbf{X} \text{ a low rank matrix} \end{aligned}$$

Since $\mathbf{X}$ is a low rank matrix, we cannot merely use $\mathbf{Z} \approx \mathbf{U}^T\mathbf{X}$ as an approximation of $\mathbf{Z}$. We need to add a sparse matrix $\mathbf{S}$, i.e. $\mathbf{Z} = \mathbf{X} + \mathbf{S}$, for a matrix can always be represented as the sum of a low rank matrix and a sparse matrix. Then our objective function is:

$$\begin{aligned} &\min_{\mathbf{X},\mathbf{S}} \frac{1}{2} \sum_{(i,j)\in\mathbf{\Omega}} (z_{ij} - x_{ij} - s_{ij})^2 \\ &\text{s.t. } \mathbf{X} \text{ is a low rank matrix} \\ &\qquad \mathbf{S} \text{ is a sparse matrix} \end{aligned}$$

We use convex functions *Nuclear norm* and *1-norm* to approximate the rank and the number of non-zero elements respectively. The formulation is

$$\min_{\mathbf{X},\mathbf{S}} \frac{1}{2}\|\mathbf{Z} - \mathbf{X} - \mathbf{S}\|_F^2 + \lambda_1\|\mathbf{X}\|_* + \lambda_2\|\mathbf{S}\|_1 \tag{1}$$

However, this objective function is not differentiable. Therefore, we need to introduce the definition of *directional derivative* and *subgradient*.

8.2 Directional derivative

Definition 8.1. *Let $f : \mathbf{E} \to \mathbf{R}$, the directional derivative of a function f at $\hat{\mathbf{x}}$ in a direction $\mathbf{d} \in \mathbf{E}$ is*

$$f'(\hat{\mathbf{x}}; \mathbf{d}) = \lim_{t \downarrow 0} \frac{f(\hat{\mathbf{x}} + \mathbf{d}) - f(\hat{\mathbf{x}})}{t} \tag{2}$$

when the limit exists. When the directional derivative $f'(\hat{\mathbf{x}}; \mathbf{d})$ is actually linear in $\mathbf{d}$, that is $f'(\hat{\mathbf{x}}; \mathbf{d}) = \langle \mathbf{a}, \mathbf{d} \rangle$ for some element $\mathbf{a}$ of $\mathbf{E}$. Then, we say f is (Gâteanx) differentiable at $\hat{x}$ with (Gâteanx) derivative $\nabla f(\hat{\mathbf{x}}) = \mathbf{a}$. If f is differentiable at every point in $\mathbf{E}$, then we simply say f is differentiable (on $\mathbf{E}$).

Example 8.1. $f(\mathbf{X}) = \log |\mathbf{X}|, \mathbf{X} \in \mathbf{S}^n_{++}$, *find* $f'(\mathbf{X}; \mathbf{Y})$, *where* $\mathbf{Y} \in \mathbf{S}^n_{++}$.

Solution:

$$\begin{aligned}
f'(\mathbf{X}, \mathbf{Y}) &= \lim_{t \downarrow 0} \frac{\log |(\mathbf{X} + t\mathbf{Y})| - \log |\mathbf{X}|}{t} \\
&= \lim_{t \downarrow 0} \frac{\log |\mathbf{X}(\mathbf{I} + t\mathbf{X}^{-1}\mathbf{Y})| - \log |\mathbf{X}|}{t} \\
&= \lim_{t \downarrow 0} \frac{\log |\mathbf{X}| + \log |\mathbf{I} + t\mathbf{X}^{-1}\mathbf{Y}| - \log |\mathbf{X}|}{t} \\
&= \lim_{t \downarrow 0} \frac{\log |\mathbf{I} + t\mathbf{X}^{-1}\mathbf{Y}|}{t} \\
&= \lim_{t \downarrow 0} \frac{\sum_{i=1}^n \log (1 + t\lambda_i(\mathbf{X}^{-1}\mathbf{Y}))}{t} \\
&= \lim_{t \downarrow 0} \sum_{i=1}^n \frac{\lambda_i(\mathbf{X}^{-1}\mathbf{Y})}{1 + t\lambda_i(\mathbf{X}^{-1}\mathbf{Y})} \\
&= \sum_{i=1}^n \lambda_i(\mathbf{X}^{-1}\mathbf{Y}) \\
&= \mathrm{tr}(\mathbf{X}^{-1}\mathbf{Y}) \\
&= \langle \mathbf{X}^{-1}, \mathbf{Y} \rangle
\end{aligned}$$

where $\lambda_i(\mathbf{X}^{-1}\mathbf{Y})$ as the i-th eigenvalue of $\mathbf{X}^{-1}\mathbf{Y}$. It can be prove that $\lambda_i(\mathbf{X}^{-1}\mathbf{Y}) > 0$.

Proof. Since $\mathbf{X}^{-1}\mathbf{Y} = \mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{\frac{1}{2}}$, we have $\mathbf{X}^{-1}\mathbf{Y} \sim \mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}}$, i.e. $\mathbf{X}^{-1}\mathbf{Y}$ and $\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}}$ have the same eigenvalues. Consider that $\mathbf{X}$ and $\mathbf{Y}$ are positive definite symmetric matrices, then $\mathbf{X}^{-\frac{1}{2}} = (\mathbf{X}^{-\frac{1}{2}})^T \neq \mathbf{0}$ and for any $\mathbf{z} \neq \mathbf{0}$, $\mathbf{z}^T\mathbf{Y}\mathbf{z} > \mathbf{0}$. Hence $\mathbf{z}^T\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}}\mathbf{z} > 0$ and $\mathbf{X}^{-\frac{1}{2}}\mathbf{Y}\mathbf{X}^{-\frac{1}{2}}$ is positive definite. □

Example 8.2. $f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}^T\mathbf{X}), \mathbf{A} \in \mathbb{R}^{p \times m}, \mathbf{X} \in \mathbb{R}^{p \times n}$, *find* $f'(\mathbf{X}; \mathbf{Y})$

Solution:

$$
\begin{aligned}
f'(\mathbf{X},\mathbf{Y}) &= \lim_{t\downarrow 0} \frac{\mathrm{tr}(\mathbf{A}^T(\mathbf{X}+t\mathbf{Y})) - \mathrm{tr}(\mathbf{A}^T\mathbf{X})}{t} \\
&= \lim_{t\downarrow 0} \frac{\mathrm{tr}(\mathbf{A}^T\mathbf{X}) + \mathrm{tr}(t\mathbf{A}^T\mathbf{Y})) - \mathrm{tr}(\mathbf{A}^T\mathbf{X})}{t} \\
&= \mathrm{tr}(\mathbf{A}^T\mathbf{Y}) \\
&= \langle \mathbf{A}, \mathbf{Y} \rangle
\end{aligned}
$$

Example 8.3. *$f(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^{-1})$, find $f'(\mathbf{X},\mathbf{Y})$*

Solution:

$$
\begin{aligned}
f'(\mathbf{X},\mathbf{Y}) &= \lim_{t\downarrow 0} \frac{\mathrm{tr}((\mathbf{X}+t\mathbf{Y})^{-1}) - \mathrm{tr}(\mathbf{X}^{-1})}{t} \\
&= \lim_{t\downarrow 0} \frac{\mathrm{tr}((\mathbf{X}(\mathbf{I}+t\mathbf{Y}))^{-1}) - \mathrm{tr}(\mathbf{X}^{-1})}{t} \\
&= \lim_{t\downarrow 0} \frac{\mathrm{tr}((\mathbf{I}+t\mathbf{Y})^{-1}\mathbf{X}^{-1}) - \mathrm{tr}(\mathbf{X}^{-1})}{t} \\
&\text{(using the Taylor expansion } (\mathbf{I}+t\mathbf{Y})^{-1} = \mathbf{I} - t\mathbf{X}^{-1}\mathbf{Y} + t^2(\mathbf{X}^{-1}\mathbf{Y})^2 - \dots) \\
&= \lim_{t\downarrow 0} \frac{\mathrm{tr}(\mathbf{X}^{-1}) - t\mathrm{tr}(\mathbf{X}^{-1}\mathbf{Y}\mathbf{X}^{-1}) + t^2\mathrm{tr}((\mathbf{X}^{-1}\mathbf{Y})^2\mathbf{X}^{-1}) - \dots - \mathrm{tr}(\mathbf{X}^{-1})}{t} \\
&= -\mathrm{tr}(\mathbf{X}^{-1}\mathbf{Y}\mathbf{X}^{-1}) \\
&= -\mathrm{tr}(\mathbf{X}^{-2}\mathbf{Y})
\end{aligned}
$$

8.3 Subgradient

Definition 8.2. *If function f is convex and proper (which means dom$f = \{\mathbf{x} \in \mathbf{E} | f(\mathbf{x}) < \infty\}$ is non-empty), $\boldsymbol{\phi}$ is said to be subgradient of $f(\mathbf{x})$ at $\hat{\mathbf{x}}$, if it satisfies $\langle \boldsymbol{\phi}, \mathbf{x} - \hat{\mathbf{x}} \rangle \leq f(\mathbf{x}) - f(\hat{\mathbf{x}})$ for all $\mathbf{x} \in \mathbf{E}$.*

Proposition 8.1. *For any convex proper function $f : \mathbf{E} \to (-\infty, \infty)$, the point $\hat{\mathbf{x}}$ is a global minimizer of f iff the condition $0 \in \partial f(\hat{\mathbf{x}})$ holds.*

Recall the definition of general matrix norms:

Definition 8.3. *For $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\|\mathbf{A}\|$ is a function of $\mathbf{A}$ which satisfies the following conditions.*

1. $\|\mathbf{A}\| \geq 0$
2. $\|\mathbf{A}\| = 0$ *iff* $\mathbf{A} = 0$
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$

4. $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$

According to the definition, it's easy to derive that matrix norm is convex:

$$\|\alpha\mathbf{A} + (1-\alpha)\mathbf{B}\| \leq \alpha\|\mathbf{A}\| + (1-\alpha)\|\mathbf{B}\|$$

Addition condition:

Definition 8.4. *A matrix norm $\|\cdot\|$ is called consistent if:*

$$\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|.$$

We here consider a kind of norm function which satisfy $\|\mathbf{U}^T\mathbf{AV}\| = \|\mathbf{A}\|$, $\mathbf{U}$,$\mathbf{V}$ are orthogonal matrices. $\mathbf{U}^T\mathbf{U} = \mathbf{UU}^T = \mathbf{I}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{VV}^T = \mathbf{I}$.
$\mathbf{A} = \mathbf{U\Sigma V}^T$, $\mathbf{A}$ is $m \times n$ matrix, $\mathbf{U}$ is $m \times m$ matrix; $\mathbf{\Sigma}$ is $m \times n$; $\mathbf{V}$ is $n \times n$.

$$\begin{aligned} \|\mathbf{U}^T\mathbf{U\Sigma V}^T\mathbf{V}\| &= \|\mathbf{A}\|, \\ \|\mathbf{A}\| &= \|\mathbf{\Sigma}\|. \end{aligned}$$

Now we consider the uniqueness of SVD.

$$\mathbf{A} = \mathbf{U\Sigma V}^T, \mathbf{U} \in \mathbb{R}^{m\times n}, \mathbf{\Sigma} \in \mathbb{R}^{n\times n}, \mathbf{V} \in \mathbb{R}^{n\times n},$$

where $\mathbf{\Sigma}$ is unique if we order the singular value. If all the singular values are different, then $\mathbf{U}$,$\mathbf{V}$ is unique with ± 1. In the case $\sigma_1 = \sigma_2$, we have

$$\begin{aligned} \mathbf{\Sigma} &= \begin{bmatrix} \mathbf{\Sigma}_1\mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\sigma_3 \cdots \sigma_n) \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{Q}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{\Sigma} \begin{bmatrix} \mathbf{Q}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \end{aligned}$$

where $\mathbf{QQ}^T = \mathbf{I}$. Then we have

$$\mathbf{U\Sigma V}^T = \mathbf{U} \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{\Sigma} \begin{bmatrix} \mathbf{Q}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \mathbf{V}^T.$$

In general, for any $\mathbf{U}_i$ and $\mathbf{V}_i$ that satisfy $\mathbf{A} = \mathbf{U}_i\mathbf{\Sigma V}_i^T$, we can rewrite in the following form as

$$\mathbf{A} = \mathbf{UQ}_i\mathbf{\Sigma}(\mathbf{VQ}_i)^T.$$

where $\mathbf{Q}_i\mathbf{Q}_i^T = \mathbf{I}$.

Definition 8.5 (The Schatten p-norm)**.** *Let $\boldsymbol{\sigma} = (\sigma_1 \cdots \sigma_n)^T$ be the vector contain singular values of $\mathbf{A}$, then the Schatten p-norm of $\mathbf{A}$ is $\|\mathbf{A}\|_p = \|\boldsymbol{\sigma}\|_p$.*

There are three examples of Schatten p-norm:

- $p = 1$, $\|\mathbf{A}\|_* = \sum_{i=1}^n \sigma_i$.
- $p = 2$, $\|\mathbf{A}\|_F = \sum_{i=1}^n \sigma_i^2$.
- $p = \infty$, $\|\mathbf{A}\|_\infty = \sigma_1$. We call σ_1 the spectrum radius, and $\|\mathbf{A}\|_\infty$ is also called spectral norm.

Lemma 8.1. *Let $\mathbf{A}$ and $\mathbf{R}$ be given $m \times n$ matrices, $\|\cdot\|$ is Schatten p-norm, $\phi(\cdot)$ is the corresponding norm on singular vector, then there is a SVD of $\mathbf{A}$ such that*

$$\lim_{t \downarrow 0} \frac{\|\mathbf{A} + t\mathbf{R}\| - \|\mathbf{A}\|}{t} = \max_{\mathbf{d} \in \partial\phi(\boldsymbol{\sigma})} \sum_{i=1}^n d_i \mathbf{U}_i^T \mathbf{R} \mathbf{V}_i,$$

where

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \sum_{i=1}^n \sigma_i \mathbf{U}_i \mathbf{V}_i^T,$$

$$\boldsymbol{\sigma} = (\sigma_1 \cdots \sigma_n)^T,$$

$$\mathbf{d} = (d_1 \cdots d_n)^T.$$

Recall the definition of subdifferential, here gives an equivalent definition:

Definition 8.6. *For $\|\cdot\|$, $\mathbf{A} \in \mathbb{R}^{m \times n}$*

$$\partial\|\mathbf{A}\| = \left\{ \mathbf{G} : \|\mathbf{B}\| \geq \|\mathbf{A}\| + \mathrm{tr}((\mathbf{B} - \mathbf{A})^T \mathbf{G}),\ for\ all\ \mathbf{B} \in \mathbb{R}^{m \times n} \right\}.$$

Proposition 8.2. *$\mathbf{G} \in \partial\|\mathbf{A}\|$ is equivalent to the following statements:*

1. $\|\mathbf{A}\| = \mathrm{tr}(\mathbf{G}^T \mathbf{A})$
2. $\|\mathbf{G}\|^* \leq 1$

where $\|\cdot\|^*$ is dual norm of $\|\cdot\|$, which defined as:

$$\|\mathbf{G}\|^* = \max_{\|\mathbf{B}\| \leq 1} \mathrm{tr}(\mathbf{B}^T \mathbf{G}).$$

Proof.

To get the equation, the intuition is to assign different values of $\mathbf{B}$.
Let $\mathbf{B} = \mathbf{0}$, then

$$0 \leq \|\mathbf{A}\| - \mathrm{tr}(\mathbf{A}^T \mathbf{G})$$
$$\Longrightarrow \|\mathbf{A}\| \leq \mathrm{tr}(\mathbf{A}^T \mathbf{G}).$$

Let $\mathbf{B} = 2\mathbf{A}$, then

$$\begin{aligned} 2\|\mathbf{A}\| &\geq \|\mathbf{A}\| - \text{tr}(\mathbf{A}^T\mathbf{G}) + 2\text{tr}(\mathbf{A}^T\mathbf{G}) \\ \Longrightarrow &\|\mathbf{A}\| \geq \text{tr}(\mathbf{A}^T\mathbf{G}). \end{aligned}$$

So, we get $\|\mathbf{A}\| = tr(\mathbf{A}^T\mathbf{G})$.

Using the conclusion above, we can simplify the inequation to $\|\mathbf{B}\| \geq tr(\mathbf{B}^T\mathbf{G})$.
When $\mathbf{B} = \mathbf{0}$, $\|\mathbf{G}\|^* = \mathbf{0}$.
When $\mathbf{B} \neq \mathbf{0}$,

$$\frac{\text{tr}(\mathbf{B}^T\mathbf{G})}{\|\mathbf{B}\|} \leq 1.$$

This form is equivalent to

$$\max_{\|\mathbf{B}\| \leq 1} \text{tr}(\mathbf{B}^T\mathbf{G}) \leq 1.$$

Hence $\|\mathbf{G}\|^* \leq 1$.

□

Theorem 8.1. *Let* $\mathbf{A}$ *denote an* $m \times n$ *matrix,* $\|\cdot\|$ *is Schatten p-norm, and* $\phi(\cdot)$ *is the corresponding norm on singular values, then*

$$\partial\|\mathbf{A}\| = conv\{\mathbf{U}\mathbf{D}\mathbf{V}^T, \mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \mathbf{D} = \text{diag}(\mathbf{d}), \mathbf{d} \in \partial\phi(\boldsymbol{\sigma})\},$$

where $conv(\cdot)$ *is the convex hull of a set, i.e,*
If $\mathbf{G} \in \partial\|\mathbf{A}\|$*, there exists* $\{\lambda_i\}, \lambda_i \geq 0, \sum_i \lambda_i = 1$*, that satisfies*

$$\mathbf{G} = \sum_i \lambda_i \mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^T.$$

Pay attention to the $\mathbf{U}_i$ and $\mathbf{V}_i$ here. In previous expression, $\mathbf{U}_i$ means the ith row of $\mathbf{U}$; but here, $\mathbf{U}_i$ and $\mathbf{V}_i$ is an assignment that satisfy $\mathbf{A} = \mathbf{U}_i\boldsymbol{\Sigma}\mathbf{V}_i^T$. Now, let's see $\|\mathbf{A}\|_*$, $\phi(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_1 = \sum_{i=1}^n \sigma_i$. Suppose $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, and $\mathbf{A}$ have q zero sigular values. Then let

$$\mathbf{U} = [\mathbf{U}^{(1)}, \mathbf{U}^{(2)}], \quad \mathbf{V} = [\mathbf{V}^{(1)}, \mathbf{V}^{(2)}],$$

where $\mathbf{U}^{(1)}$ and $\mathbf{V}^{(1)}$ have $n - q$ columns. Then we have

$$\partial\|\boldsymbol{\sigma}\|_1 = \{\mathbf{d} \in \mathbb{R}^n : |d_i| \leq 1, d_i = 1 \ for \ i = 1, \cdots, n - q\}.$$

Let $\mathbf{G} = \partial\|\mathbf{A}\|$ and $\mathbf{G} = \sum_i \lambda_i \mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^T, \lambda_i \geq 0, \sum_i \lambda_i = 1$. Then for each i, we have $\mathbf{d}_i \in \partial\|\boldsymbol{\sigma}\|_1, \mathbf{D}_i = \mathrm{diag}(\mathbf{d}_i)$, and $\mathbf{A} = \mathbf{U}_i\mathbf{\Sigma}\mathbf{V}_i^T$ and

$$\begin{aligned}
\mathbf{G} &= \sum_i \lambda_i \mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^T \\
&= \sum_i \lambda[\mathbf{U}_i^{(1)}\ \mathbf{U}_i^{(2)}] \begin{bmatrix} \mathbf{I}_{n-q} & 0 \\ 0 & \mathbf{W}_i \end{bmatrix} [\mathbf{V}_i^{(1)}\ \mathbf{V}_i^{(2)}]^T \\
&= \sum_i \lambda_i \mathbf{U}_i^{(1)}{\mathbf{V}_i^{(1)}}^T + \sum_{i=1}^n \lambda_i \mathbf{U}_i^{(2)}{\mathbf{V}_i^{(2)}}^T.
\end{aligned}$$

$\mathbf{W}_i$ is diagonal and the absolute value of the elements are less than or equal to 1.
According to the uniqueness of SVD decomposition, we have

$$\mathbf{U}_i = \mathbf{U}\mathbf{Q}_i, \mathbf{V}_i = \mathbf{V}\mathbf{Q}_i^T, \mathbf{Q}_i\mathbf{Q}_i^T = \mathbf{I},$$

which implies

$$\begin{aligned}
\mathbf{G} &= \mathbf{U}^{(1)}{\mathbf{V}^{(1)}}^T + \sum_i \lambda_i \mathbf{U}^{(2)}\mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i^T{\mathbf{V}^{(2)}}^T \\
&= \mathbf{U}^{(1)}{\mathbf{V}^{(1)}}^T + \mathbf{U}^{(2)}\left(\sum_i \lambda_i \mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i^T\right){\mathbf{V}^{(2)}}^T.
\end{aligned}$$

Let $\mathbf{T} = \mathbf{U}^{(2)}\left(\sum_i \lambda_i \mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i^T\right){\mathbf{V}^{(2)}}^T$. Then

$$\mathbf{G} = \mathbf{U}^{(1)}{\mathbf{V}^{(1)}}^T + \mathbf{U}^{(2)}\mathbf{T}{\mathbf{V}^{(2)}}^T.$$

By the property of $\mathbf{W}_i$, we have

$$\begin{aligned}
\sigma_1(\mathbf{T}) &= \sigma_1\left(\sum_i \lambda_i \mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i^T\right) \\
&\leq \sum_i \lambda_i\sigma_1\left(\mathbf{Q}_i\mathbf{W}_i\mathbf{Q}_i^T\right) \\
&\leq 1
\end{aligned}$$

Finally we have

$$\partial\|\mathbf{A}\|_* = \{\mathbf{U}^{(1)}{\mathbf{V}^{(1)}}^T + \mathbf{U}^{(2)}\mathbf{T}{\mathbf{V}^{(2)}}^T,\ for\ all\ \mathbf{T} \in \mathbb{R}^{q\times q},\ \sigma_1(\mathbf{T}) \leq 1\}.$$