The Spectral Operator of Matrices

C. Ding, D. Sun, J. Sun, and K. Toh

National University of Singapore

ICCP 2012, Singapore

December 18, 2012

Outline

1. What is it?
   - The definition
   - Main results
2. Why did it?
   - Matrix optimization problems (MOPs)
   - The Moreau-Yosida regularization: Extending metric projection
   - Matrix completion
   - Beyond the Moreau-Yosida regularization
3. Main results: More details
   - The eigenvalue decomposition
   - The singular value decomposition
   - An example of detailed statement: Differentiability

Outline

What is it?

- Let $\mathcal{X} := \mathcal{S}^{m_0} \times \mathbb{R}^{m \times n}$ $(m \leq n)$ be the Cartesian product of a symmetric matrix space and an $m \times n$ real matrix space.

- The spectral operator $G : \mathcal{X} \to \mathcal{X}$ with respect to a given function $g = (g_1, g_2), g_1 : \mathbb{R}^{m_0} \times \mathbb{R}^m \to \mathbb{R}^{m_0}, g_2 : \mathbb{R}^{m_0} \times \mathbb{R}^m \to \mathbb{R}^m$, is defined by

$$G(X) = (G_1(X), G_2(X)) \in \mathcal{X}, \quad X = (Y, Z) \in \mathcal{X}\,,$$

with

$$\begin{cases} G_1(X) = P\mathrm{diag}\big(g_1(\kappa(X))\big)P^T\,, \\ G_2(X) = U\left[\mathrm{diag}\big(g_2(\kappa(X))\big) \;\; 0\right]V^T\,, \end{cases}$$

where $\kappa(X) := (\lambda(Y), \sigma(Z))$

— $\lambda_1(Y) \geq \lambda_2(Y) \geq \ldots \geq \lambda_{m_0}(Y)$ are the eigenvalues of Y
— $\sigma_1(Z) \geq \sigma_2(Z) \geq \ldots \geq \sigma_m(Z)$ are the singular values of Z

and the orthogonal matrices $P \in \mathcal{O}^{m_0}$, $U \in \mathcal{O}^m$, $V \in \mathcal{O}^n$ satisfy

$$Y = P\mathrm{diag}(\lambda(Y))P^T, \quad Z = U[\mathrm{diag}(\sigma(Z)) \;\; 0]V^T\,.$$

The symmetric function

The given function $g : \mathbb{R}^{m_0} \times \mathbb{R}^m \to \mathbb{R}^{m_0} \times \mathbb{R}^m$ is symmetric, if for any permutation matrix Q_1 and signed permutation matrix Q_2,

$$g(x) := (g_1(x), g_2(x)) = (Q_1^T g_1(Qx), Q_2^T g_2(Qx)) \quad \forall\, x = (y, z) \in \mathbb{R}^{m_0} \times \mathbb{R}^m\,,$$

in short,

$$g(x) = Q^T g(Qx) \quad \forall\, x \in \mathbb{R}^{m_0} \times \mathbb{R}^m\,.$$

Summary of results

Let $G : \mathcal{X} \to \mathcal{X}$ be the spectral operator with g.

- If g is symmetric, then the corresponding spectral operator $G : \mathcal{X} \to \mathcal{X}$ is well-defined.
- G is continuous $\Longleftrightarrow$ g is continuous.
- G is (continuously) differentiable $\Longleftrightarrow$ g is (continuously) differentiable.
- G is locally Lipschitz continuous $\Longleftrightarrow$ g is locally Lipschitz continuous.
- G is Hadamard directionally differentiable $\Longleftrightarrow$ g is Hadamard directionally differentiable.
- If g is locally Lipschitz continuous, G is directionally differentiable $\Longleftrightarrow$ g is directionally differentiable.
- G is ρ-order B(ouligand)-differentiable$\Longleftrightarrow$ g is ρ-order B(ouligand)-differentiable $(0 < \rho \leq 1)$.
- G is ρ-order G-semismooth $\Longleftrightarrow$ g is ρ-order G-semismooth $(0 < \rho \leq 1)$.
- The characterization of Clarke's generalized Jacobian of G

Outline

Matrix optimization problems (MOPs)

The primal MOP takes the form:

$$\begin{array}{lll} \text{(P)} & \min & \langle C, X\rangle + f(X) \\ & \text{s.t.} & \mathcal{A}X = b, \quad X \in \mathcal{X}\,. \end{array}$$

- $\mathcal{X} := \mathcal{S}^{m_1} \times \ldots \times \mathcal{S}^{m_{s_0}} \times \mathbb{R}^{m_{s_0+1} \times n_{s_0+1}} \times \ldots \times \mathbb{R}^{m_s \times n_s}$;
- $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ are the natural inner product and the induced norm.
- $f : \mathcal{X} \to (-\infty, +\infty]$ is a closed proper convex function;
- $\mathcal{A} : \mathcal{X} \to \mathbb{R}^p$ is a given linear operator;
- $C \in \mathcal{X}$ and $b \in \mathbb{R}^p$ are given.

The dual MOP

The dual MOP takes the form:

$$\begin{aligned} \text{(D)} \quad \max \quad & \langle b, y \rangle - f^*(X^*) \\ \text{s.t.} \quad & \mathcal{A}^* y - C = X^* \,. \end{aligned}$$

- f^* is the conjugate function of f, i.e.,

$$f^*(X^*) := \sup \{ \langle X^*, X \rangle - f(X) \,|\, X \in \mathcal{X} \} \,, \quad X^* \in \mathcal{X} \,;$$

- $\mathcal{A}^* : \mathbb{R}^p \to \mathcal{X}$ is the adjoint of the linear operator $\mathcal{A}$.

The KKT condition of MOP

The KKT condition of MOP:

$$\begin{cases} C - \mathcal{A}^* y + \Gamma = 0\,, \\ \mathcal{A}X - b = 0\,, \\ 0 \in -X + \partial f^*(\Gamma)\,, \end{cases}$$

— $\partial f^*(\cdot)$ is maximal monotone.

The KKT condition of MOP:

$$\left\{ \begin{array}{l} C - \mathcal{A}^* y + \Gamma = 0\,, \\ \mathcal{A}X - b = 0\,, \\ 0 \in -X + \partial f^*(\Gamma)\,, \end{array} \right.$$

— $\partial f^*(\cdot)$ is maximal monotone.

The equivalent form:

$$F(X, y, \Gamma) = \left[\begin{array}{c} C - \mathcal{A}^* y + \Gamma \\ \mathcal{A}X - b \\ X - P_f(X + \Gamma) \end{array} \right] = 0\,,$$

— $P_f(\cdot)$ is the proximal point mapping (M-Y projection) of f, the unique optimal solution of the Moreau-Yosida regularization ψ_f of f.

The KKT condition of MOP:

$$\begin{cases} C - \mathcal{A}^* y + \Gamma = 0\,, \\ \mathcal{A}X - b = 0\,, \\ 0 \in -X + \partial f^*(\Gamma)\,, \end{cases}$$

— $\partial f^*(\cdot)$ is maximal monotone.

The equivalent form:

$$F(X, y, \Gamma) = \begin{bmatrix} C - \mathcal{A}^* y + \Gamma \\ \mathcal{A}X - b \\ X - P_f(X + \Gamma) \end{bmatrix} = 0\,,$$

— $P_f(\cdot)$ is the proximal point mapping (M-Y projection) of f, the unique optimal solution of the Moreau-Yosida regularization ψ_f of f.
— A possible question: Is P_f semismooth?

Recall: The Moreau-Yosida regularization

For the given closed proper convex function $f : \mathcal{E} \to (-\infty, +\infty]$, the Moreau-Yosida regularization ψ_f of f with respect to $\eta > 0$ is defined as

$$\psi_f(x) := \min_{z \in \mathcal{E}} \left\{ f(z) + \frac{1}{2\eta} \|z - x\|^2 \right\}, \quad x \in \mathcal{E}\,. \tag{1}$$

Denote the corresponding unique optimal solution of (1) by $P_f(x)$, the proximal point of x associated with h.

- $P_f(x), x \in \mathcal{X}$ is well defined if f is unitarily invariant. It is shown that $P_f(x)$ is a spectral operator as defined above.
- ψ_f is continuously differentiable, and it holds that

$$\nabla \psi_f(x) = \frac{1}{\eta}(x - P_f(x))\,.$$

The M-Y projection is closely related to the proximal point approach, which includes the augmented Lagrangian method.

The unitary invariance

Unitary invariance: for any orthogonal matrices $P \in \mathcal{O}^{m_0}$ and $U \in \mathcal{O}^m$, $V \in \mathcal{O}^n$,

$$f(X) = f(P^T Y P, U^T Z V) \quad \forall\, X = (Y, Z) \in \mathcal{S}^{m_0} \times \mathbb{R}^{m \times n}.$$

If f is unitarily invariant, then (cf. [von Neumann, 1937][1], [Davis, 1957][2])

(i) $\exists$ a convex function $g : \mathbb{R}^{m_0+m} \to (-\infty, +\infty]$ such that

$$f(X) = g(\kappa(X)).$$

(ii) g is invariant under permutations, i.e., for any permutation matrix Q_1 and signed permutation matrix Q_2,

$$g(x) = g(Q_1 y, Q_2 z) \quad \forall\, x = (y, z) \in \mathbb{R}^{m_0} \times \mathbb{R}^m.$$

---

[1] J. von Neumann, *Some matrix inequalities and metrization of metric space*, Tomsk University Review, 1 (1937), pp. 286–300.

[2] C. Davis, *All convex invariant functions of hermitian matrices*, Archiv der Mathematik, 8 (1957), pp. 276–278.

For Moreau-Yosida regularization

f is a unitarily invariant closed proper convex function and $f(X) = g(\kappa(X))$. Then,

(1) the Moreau-Yosida regularization function ψ_f of f is also unitarily invariant;

(2) the proximal mapping $P_f(X) = G(X)$ is the spectral operator with respect to $\psi_g = g(x)$ (g is symmetric).

For the given $X \in \mathcal{X}$ and $\eta > 0$, the proximal point is given by

$$P_f(X) = G(X) = (G_1(X), G_2(X)) \ ,$$

with

$$\begin{cases} G_1(X) = P\mathrm{diag}\big(g_1(\kappa(X))\big)P^T \ , \\ G_2(X) = U\left[\mathrm{diag}\big(g_2(\kappa(X))\big) \ \ 0\right]V^T \ , \end{cases}$$

where the orthogonal matrices P, U, $V \in \mathcal{O}^n$ satisfy

$$Y = P\mathrm{diag}(\lambda(Y))P^T, \quad Z = U[\mathrm{diag}(\sigma(Z)) \ \ 0]V^T \ .$$

Matrix completion

Given a matrix $M \in \mathbb{R}^{m\times n}$ with entries in the index set Ω given, find a low-rank matrix X such that $X_{ij} \approx M_{ij}$ for all $(i,j) \in \Omega$.

— Under suitable assumptions, one can recover M with high probability by solving the following nuclear norm minimization problem, see e.g., [Recht, Fazel & Parrilo, 2010][3], [Candès & Recht, 2009][4] :

$$\min\left\{\|X\|_* \mid P_\Omega(X) = P_\Omega(M)\right\}.$$

— For applications with noisy data, one may consider the following problem [Candès & Plan, 2010][5]:

$$\min\left\{\frac{1}{2}\|P_\Omega(X) - P_\Omega(M)\|^2 + \rho\|X\|_*\right\}.$$

— Useful in recommender systems, e.g. Netflix, Amazon; also in reducing “the total-variation" in image processing.

[3] B. Recht, M. Fazel, P. Parrilo, *Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization.* SIAM Review 52, pp. 471–501 (2010).

[4] E. Candès and B. Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics, 9 (2009), pp. 717–772.

[5] E. Candès and Y. Plan, *Matrix completion with noise.* Proceedings of the IEEE, 98,pp. 925?-936 (2010).

The primal MC form:

$$
\begin{aligned}
\text{(P)} \quad \min \quad & \langle 0, X\rangle + \langle 0, z\rangle + \frac{1}{2}\|z\|^2 + \rho\|X\|_* \\
\text{s.t.} \quad & \mathcal{A}X - z = b\,.
\end{aligned}
$$

— $(z, X) \in \mathcal{X} = \mathbb{R}^{|\Omega|} \times \mathbb{R}^{m\times n}$, $b = P_\Omega(M)$, and $\mathcal{A}(X) = P_\Omega(X)$.

The dual MC form:

$$
\begin{aligned}
\text{(D)} \quad \max \quad & \langle b, y\rangle - \frac{1}{2}\|z^*\|^2 - \delta_{B_2^\rho}(X^*) \\
\text{s.t.} \quad & \mathcal{A}^*y - X^* = 0, \quad y + z^* = 0\,,
\end{aligned}
$$

— $B_2^\rho := \{Z \in \mathbb{R}^{m\times n} \mid \|Z\|_2 \le \rho\}$.
$\|\cdot\|_*$ is the nuclear norm of matrices, i.e., the sum of singular values. $\|\cdot\|_2$ is the spectral norm of matrices, i.e., the largest singular value.

More applications

The MOP is a broad framework including many optimization problems:

- SDP: $\mathcal{X} = \mathcal{S}^n$, $f = \delta_{\mathcal{S}^n_+}$ and $f^* = \delta_{\mathcal{S}^n_-}$.
- Matrix norm approximation

$$\min \left\{ \|B_0 + \sum_{k=1}^{p} y_k B_k\|_2 \,|\, y \in \mathbb{R}^p \right\}$$

- Robust matrix completion/Robust PCA [6]:

$$\min \left\{ \|X\|_* + \rho\|Y\|_1 \,|\, P_\Omega(X) + P_\Omega(Y) = P_\Omega(M) \right\}$$

- Fastest Mixing Markov Chain (FMMC) [7]:

$$\min \left\{ \|\mathcal{P}(p)\|_{(2)} \,|\, p \geq 0,\ Bp \leq e \right\}$$

— $\|\cdot\|_{(k)}$ is Ky Fan k-norm of matrices, i.e., the sum of the k largest singular values.

---

[6] E. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, Journal of the ACM (JACM), 58 (2011), p. 11.

[7] S. Boyd, P. Diaconis, and L. Xiao, *Fastest mixing Markov chain on a graph*, SIAM review, 46 (2004), pp. 667–689.

For MOPs, if the Moreau-Yosida regularization of f to be “tractable", then

— admits a closed form solution or can be computed efficiently

— second order information

Some “tractable" cases:

- $f = \delta_{\mathcal{S}_+}$, SDP
- $f = \|\cdot\|_*$, the nuclear norm of matrices
- $f = \|\cdot\|_2$, the spectral norm of matrices
- $f = \|\cdot\|_{(k)}$, the Ky Fan k-norm of matrices
- $f = \delta_{\mathcal{K}}$, $\mathcal{K}$ is the epigraph cone of $\|\cdot\|_{(k)}$
- ...

Beyond the Moreau-Yosida regularization

The spectral operator may not necessarily be the gradient of a certain function. For example, define $F : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ by

$$F(Z) = U\left[\mathrm{diag}\big(f(\sigma(Z))\big) \;\; 0\right] V^T, \quad Z \in \mathbb{V}^{m\times n}$$

associated with the function $f : \mathbb{R}^m \to \mathbb{R}^m$

$$f_i(z) = \begin{cases} \phi\left(\dfrac{z_i}{\|z\|_\infty}\right) & \text{if } z \in \mathbb{R}^m \setminus \{0\}, \\ 0 & \text{otherwise,} \end{cases} \qquad z \in \mathbb{R}^m,$$

where $(U, V) \in \mathcal{O}^{m,n}(Z)$ and the scalar function $\phi : \mathbb{R} \to \mathbb{R}$ takes the form

$$\phi(t) = \mathrm{sgn}(t)(1+\varepsilon^\tau)\frac{|t|^\tau}{|t|^\tau + \varepsilon^\tau}, \quad t \in \mathbb{R},$$

for some $\tau > 0$ and $\varepsilon > 0$.

— f is symmetric

— the spectral operator $F(\cdot)$ is used in [Miao et al. 2012][8]

---

[8] W.M. Miao, D.F. Sun and S.H. Pan, *A Rank-Corrected Procedure for Matrix Completion with Fixed Basis Coefficients*, Preprint available at http://arxiv.org/abs/1210.3709.

Outline

1. What is it?
   - The definition
   - Main results
2. Why did it?
   - Matrix optimization problems (MOPs)
   - The Moreau-Yosida regularization: Extending metric projection
   - Matrix completion
   - Beyond the Moreau-Yosida regularization
3. Main results: More details
   - The eigenvalue decomposition
   - The singular value decomposition
   - An example of detailed statement: Differentiability

The eigenvalue decomposition

Let $Y \in \mathcal{S}^{m_0}$ be given. Denote

$$\Lambda(Y) = \mathrm{diag}(\lambda_1(Y), \lambda_2(Y), \ldots, \lambda_{m_0}(Y))\,.$$

Let $\mu_1 > \mu_2 > \cdots > \mu_r$ be the distinct eigenvalues of Y. Define the index set

$$a_k := \{\, i \,|\, \lambda_i(Y) = \mu_k \,\}\,, \quad k = 1, \ldots, r\,.$$

Proposition 1 (D. Sun & J.S. 02, 03)

For any $H \in \mathcal{S}^{m_0}$, let P be an orthogonal matrix such that

$$P^T(\Lambda(Y) + H)P = \mathrm{diag}(\lambda(\Lambda(Y) + H))\,.$$

Then, for any $H \to 0$, we have

$$\begin{cases} P_{a_k a_l} = O(\|H\|)\,, & k, l = 1, \ldots, r,\ k \neq l\,, \\ P_{a_k a_k} P_{a_k a_k}^T = I_{|a_k|} + O(\|H\|^2)\,, & k = 1, \ldots, r\,, \\ \mathrm{dist}(P_{a_k a_k},\ \mathcal{O}^{|a_k|}) = O(\|H\|^2)\,, & k = 1, \ldots, r\,, \end{cases}$$

where for each k, $\mathcal{O}^{|a_k|}$ is the set of all $|a_k| \times |a_k|$ orthogonal matrices.

The singular value decomposition

Let $Z \in \mathbb{R}^{m\times n}$ be given. Let $\overline{\mu}_1 > \overline{\mu}_2 > \ldots > \overline{\mu}_r$ be the nonzero distinct singular values of Z. Define

$$a_k := \{i \mid \sigma_i(Z) = \overline{\mu}_k,\ 1 \le i \le m\}, \quad k = 1, \ldots, r\,.$$

Proposition 2

For any $\mathbb{R}^{m\times n} \ni H \to 0$, let $Y := [\Sigma(Z)\ \ 0] + H$. Let U and V be two orthogonal matrices satisfying $[\Sigma(Z)\ \ 0] + H = U\,[\Sigma(Y)\ \ 0]\,V^T$. Then, there exist $Q \in \mathcal{O}^{|a|}$, $Q' \in \mathcal{O}^{|b|}$ and $Q'' \in \mathcal{O}^{n-|a|}$ such that

$$U = \begin{bmatrix} Q & 0 \\ 0 & Q' \end{bmatrix} + O(\|H\|) \quad \text{and} \quad V = \begin{bmatrix} Q & 0 \\ 0 & Q'' \end{bmatrix} + O(\|H\|)\,,$$

where $Q = \mathrm{diag}(Q_1, Q_2, \ldots, Q_r)$ is a block diagonal orthogonal matrix with the k-th diagonal block given by $Q_k \in \mathcal{O}^{|a_k|}$, $k = 1, \ldots, r$. Furthermore, we have

$$S(H_{a_k a_k}) = Q_k\left(\Sigma(Y)_{a_k a_k} - \Sigma(Z)_{a_k a_k}\right)Q_k^T + O(\|H\|^2), \quad k = 1, \ldots, r$$

and

$$[H_{bb}\ \ H_{bc}] = Q'\,[\Sigma(Y)_{bb} - \Sigma(Z)_{bb}\ \ 0]\,Q''^T + O(\|H\|^2)\,.$$

where $S(H_{a_k a_k}) = (H_{a_k a_k} + H_{a_k a_k}^T)/2$, $k = 1, \ldots, r$.

The Fréchet differentiability

Theorem 1

The spectral operator G is (continuously) differentiable at $\overline{X}$ if and only if the symmetric function g is (continuously) differentiable at $\overline{\kappa} = \kappa(\overline{X})$. In this case, the derivative of G at $\overline{X}$ is given by for any $H = (A, B) \in \mathcal{X}$,

$$G'(\overline{X})H = \left(\overline{P}[L_1(\overline{\kappa}, \widetilde{H}) + \overline{\mathcal{A}}^D \circ \widetilde{A}]\overline{P}^T, \overline{U}[L_2(\overline{\kappa}, \widetilde{H}) + \mathcal{T}(\overline{\kappa}, \widetilde{B})]\overline{V}^T\right) ,$$

where $\widetilde{H} = (\widetilde{A}, \widetilde{B}) = (\overline{P}^T A\overline{P}, \overline{U}^T B\overline{V})$.

Main results: More details

$$(\overline{\mathcal{A}}^D)_{ij} := \begin{cases} \dfrac{(g_1(\overline{\kappa}))_i - (g_1(\overline{\kappa}))_j}{\lambda_i(\overline{Y}) - \lambda_j(\overline{Y})} & \text{if } \lambda_i(\overline{Y}) \neq \lambda_j(\overline{Y})\,, \\ (g'(\overline{\kappa}))_{ii} - (g'(\overline{\kappa}))_{ij} & \text{otherwise}\,, \end{cases} \qquad i,j \in \{1,\ldots,m_0\}\,,$$

$$(\overline{\mathcal{E}}_1^D)_{ij} := \begin{cases} \dfrac{(g_2(\overline{\kappa}))_i - (g_2(\overline{\kappa}))_j}{\sigma_i(\overline{Z}) - \sigma_j(\overline{Z})} & \text{if } \sigma_i(\overline{Z}) \neq \sigma_j(\overline{Z})\,, \\ (g'(\overline{\kappa}))_{ii} - (g'(\overline{\kappa}))_{ij} & \text{otherwise}\,, \end{cases} \qquad i,j \in \{1,\ldots,m\}\,,$$

$$(\overline{\mathcal{E}}_2^D)_{ij} := \begin{cases} \dfrac{(g_2(\overline{\kappa}))_i + (g_2(\overline{\kappa}))_j}{\sigma_i(\overline{Z}) + \sigma_j(\overline{Z})} & \text{if } \sigma_i(\overline{Z}) + \sigma_j(\overline{Z}) \neq 0\,, \\ (g'(\overline{\kappa}))_{ii} - (g'(\overline{\kappa}))_{ij} & \text{otherwise}\,, \end{cases} \qquad i,j \in \{1,\ldots,m\}\,,$$

and

$$(\overline{\mathcal{F}}^D)_{ij} := \begin{cases} \dfrac{(g_2(\overline{\kappa}))_i}{\sigma_i(\overline{Z})} & \text{if } \sigma_i(\overline{Z}) \neq 0\,, \\ (g'(\overline{\kappa}))_{ii} - (g'(\overline{\kappa}))_{ij} & \text{otherwise.} \end{cases} \qquad i \in \{1,\ldots,m\},\ j \in \{1,\ldots,n-m\}\,.$$

Main results: More details

For any $B = [B_1 \;\; B_2] \in \mathbb{R}^{m \times n}$, let

$$\mathcal{T}(\overline{\kappa}, B) := \left[\overline{\mathcal{E}}_1^D \circ S(B_1) + \overline{\mathcal{E}}_2^D \circ T(B_1) \;\; \overline{\mathcal{F}}^D \circ B_2\right] \in \mathbb{R}^{m \times n}.$$

Define a linear operator $L(\overline{\kappa}, \cdot) : \mathcal{X} \to \mathcal{X}$ by for any $Z = (A, B) \in \mathcal{X}$,

$$L(\overline{\kappa}, Z) = (L_1(\overline{\kappa}, Z), L_2(\overline{\kappa}, Z))$$

with

$$L_1(\overline{\kappa}, Z) := \begin{bmatrix} \theta_1(\overline{\kappa}, Z) I_{|\alpha_1|} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \theta_{r_0}(\overline{\kappa}, Z) I_{|\alpha_{r_0}|} \end{bmatrix} \in \mathcal{S}^{m_0}$$

and

$$L_2(\overline{\kappa}, Z) := \begin{bmatrix} \theta_{r_0+1}(\overline{\kappa}, Z) I_{|a_1|} & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \theta_{r_0+r}(\overline{\kappa}, Z) I_{|a_r|} & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n},$$

where $\theta_k(\overline{\kappa}, \cdot) : \mathcal{X} \to \mathbb{R}$, $k = 1, \ldots, r_0 + r$ are given by

$$\theta_k(\overline{\kappa}, Z) := \sum_{k'=1}^{r_0} \overline{c}_{kk'} \mathrm{tr}(A_{\alpha_{k'} \alpha_{k'}}) + \sum_{k'=r_0+l=r_0+1}^{r_0+r} \overline{c}_{kk'} \mathrm{tr}(S(B_{a_l a_l})) .$$

Thank you!