6. Theoretical Foundations

6.1. The rank of a matrix with infinite rows

Lemma 1 *The rank of* $\mathbf{Z} \in \mathcal{Z}_p = \{\mathbf{T} \in \mathbb{R}^{\mathbb{N}\times p}, \|\mathbf{T}\|_{\mathcal{F}} < \infty\}$ *exists, is finite and not larger than p. Moreover, such a matrix has p (non-negative) singular values.*

Proof The proof has three steps: 1) prove that both the row and column rank cannot exceed p, 2) sketch that these two ranks coincide and therefore the rank operator is well defined and 3) show that there p singular values.

1. The row rank and the column rank of $\mathbf{Z}$ cannot exceed p. On the one side, the rows are all p-dimensional vectors and therefore they cannot span a space with dimensionality higher than p. On the other side, there are only p columns, and therefore they cannot span a space with dimensionality higher than p.

2. The row and column ranks coincide, and therefore we can write rank($\mathbf{Z}$), and state $0 \leq \text{rank}(\mathbf{Z}) \leq p$. Indeed, it is straightforward to extend the result for finite matrices to any matrix in $\mathcal{Z}_p$.

3. $\mathbf{Z}$ has p singular values, that are, by definition, the square root of the eigenvalues of the (finite) matrix $\mathbf{Z}^\top\mathbf{Z}$. Since $\mathbf{Z}^\top\mathbf{Z}$ is a symmetric non-negative definite square matrix of size p, $\mathbf{Z}$ has exactly p non-negative singular values. ■

6.2. The nuclear norm is the tighest convex envelope of the rank also in $\mathcal{Z}_p$

Theorem 1 *On the set* $\mathcal{Z}_p^M = \{\mathbf{Z} \in \mathbb{R}^{\mathbb{N}\times p}, \|\mathbf{Z}\|_{\mathcal{F}} < M\}$ *the tightest convex envelope of rank*($\mathbf{Z}$) *is* $g(\mathbf{Z}) = \frac{1}{M}\|\mathbf{Z}\|_*$.

Proof The proof has two steps. First, we prove that the bi-conjugate of a real-valued function in a Hilbert space ($\mathcal{Z}_p^M$ is one) is the tighest convex envelope. After that, we prove that the bi-conjugate of the rank is the nuclear norm.

We first recall the definition of the *conjugate* of a function:

Definition 1 *Let $\mathcal{H}$ be a Hilbert space, with scalar product denoted by $\langle\cdot,\cdot\rangle$. Let $f \neq \infty$ minorized by an affine function over $\mathcal{H}$, then the conjugate of f is a function f^* defined as:*

$$f^*(s) := \sup\{\langle s, x\rangle - f(x), x \in \text{dom}f\}, s \in \mathcal{H}. \quad (19)$$

Intuitively, dom$f$ is the set of slopes of all affine functions minorizing f over $\mathcal{H}$. From this definition it is easy to see that f^* also satisfies the conjugability conditions. Therefore we can consider the following object f^{**}, referred to as the *biconjugate* of the function f. Importantly, the following lemma proves that the biconjugate of a function is the tighest convex envelope of the function. In other words:

Lemma 2 *For a given f satisfying the conjugability conditions, f^{**} is the pointwise supremum of all the affine functions on $\mathcal{H}$ majorized by f.*

Proof If $\Omega_f \subset \mathcal{H}\times\mathbb{R}$ denotes the set of pairs (y,r) defining affine functions $x \mapsto \langle y, x\rangle - r$ majorized by f, we have:

$$\begin{aligned}(y,r) \in \Omega_f &\Leftrightarrow f(x) \geq \langle y,x\rangle - r \quad \forall x \in \mathcal{H}\\ &\Leftrightarrow r \geq \sup\{\langle y,x\rangle - f(x), x \in \mathcal{H}\}\\ &\Leftrightarrow r \geq f^*(y).\end{aligned}$$

Then we obtain:

$$\begin{aligned}\sup_{(y,r)\in\Omega_f} \langle y,x\rangle - r &= \sup_{(y,r)}\{\langle y,x\rangle - r, y \in \text{dom}f^*, -r \leq -f^*(y)\}\\ &= \sup_y\{\langle y,x\rangle - f^*(y), y \in \text{dom}f^*\}\\ &= f^{**}(x),\end{aligned}$$

as we wanted to prove. ■

The previous result corresponds to part of Theorem X.1.3.5 of [2] extended to any Hilbert space and allows us to write the biconjugate of a function as:

$$f^{**}(x) = \sup_{r,s}\{\langle y,x\rangle - r, \langle y,z\rangle - r \leq f(z), \forall z \in \mathcal{H}\}. \quad (20)$$

Therefore, the biconjugate of a function f is its tighest convex envelope, and this concludes the first part of the proof. In the second part of the proof, we shall see that the biconjugate of the rank is the nuclear norm. First of all we remark that both the rank and the nuclear norm are well defined for matrices in $\mathcal{Z}_p^M$ as proven in Lemma 1 in the main manuscript.

The proof of the second part follows step-by-step the proof of Theorem 1 in [1], and therefore we just sketch the main line of reasoning. The proof starts by focusing on the set $\mathcal{Z}_p^1$, since the generalization to an arbitrary M is straightforward. Firstly we compute the conjugate of the rank:

$$\phi^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\|\leq 1}\left(\langle\mathbf{Y},\mathbf{X}\rangle - \phi(\mathbf{X})\right), \quad (21)$$

with $\phi(\mathbf{X}) = \text{rank}(\mathbf{X})$. Since the scalar product is defined as $\langle\mathbf{Y},\mathbf{X}\rangle = \text{Tr}\left(\mathbf{Y}^\top\mathbf{X}\right)$ we write:

$$\phi^*(\mathbf{Y}) = \sup_{\|\mathbf{X}\|\leq 1}\left(\text{Tr}\left(\mathbf{Y}^\top\mathbf{X}\right) - \text{Rank}(\mathbf{X})\right), \quad (22)$$

Using von Neumann's trace theorem, *i.e.*:

$$\text{Tr}\left(\mathbf{Y}^\top\mathbf{X}\right) \leq \sum_{i=1}^p \sigma_i(\mathbf{Y})\sigma_i(\mathbf{X}), \quad (23)$$

where σ_i denotes the i-th singular value, we get to the following result:

$$\phi^*(\mathbf{Y}) = \sum_{i=1}^{p} \left(\sigma_i(\mathbf{Y}) - 1\right)_+ , \qquad (24)$$

where $(y)_+$ is defined as $(y)_+ = \max\{y, 0\}$. The computation of ϕ^{**} from ϕ^* follows a similar reasoning. The main difference is that

$$\phi^{**}(\mathbf{Z}) = \sup_{\mathbf{Y}} \left(\sum_{i=1}^{p} \sigma_i(\mathbf{Z})\sigma_i(\mathbf{Y}) - \left(\sum_{i=1}^{r} \sigma_i(\mathbf{Y}) - r \right) \right),$$

diverges for $\|\mathbf{Z}\| > 1$, and therefore can only be computed for $\mathbf{Z} \in \mathcal{Z}_p^1$. In that case we obtain the desired result (see (Fazel 2002) for details):

$$\phi^{**}(\mathbf{Z}) = \|\mathbf{Z}\|_*, \qquad (25)$$

concluding the proof. ■

6.3. The derivative with respect to $\mathbf{L}_1$

Definition 2 (Fréchet derivative) *Let V and W be two Banach spaces, and $f : U \subset V \to W$, where U is an open set of V. f is Fréchet differentiable at $x \in U$ is there exists an bounded linear operator $T : V \to W$ such that.*

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Th\|_W}{\|h\|_V} = 0, \qquad (26)$$

where $\|\cdot\|_V$ and $\|\cdot\|_W$ denote the norms of V and W.

Since $\mathcal{Z}_p^M$ is a Hilbert (and therefore Banach) space, the previous definition can be applied to our case. In particular we introduce the following lemma:

Lemma 3 *Let $\mathbf{\Phi} \in \mathcal{Z}_p^M$, $\mathbf{Q} \in \mathbb{R}^{m \times r}$ and $\mathbf{L} \in \mathcal{Z}_r^M$. The Fréchet derivative of $f : \mathcal{Z}_r^M \to \mathbb{R}$ defined as $f(\mathbf{L}) = \|\mathbf{\Phi} - \mathbf{L}\mathbf{Q}^\top\|_V^2 = \left\langle \mathbf{\Phi} - \mathbf{L}\mathbf{Q}^\top, \mathbf{\Phi} - \mathbf{L}\mathbf{Q}^\top \right\rangle$ is $T = 2\left(\mathbf{L}\mathbf{Q}^\top\mathbf{Q} - \mathbf{\Phi}\mathbf{Q}\right)$ seen as a linear operator $\mathcal{Z}_r^M \to \mathbb{R}$.*

Proof The proof is straightforward from the definition. ■

References

[1] M. Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

[2] Hiriart-Urruty, J.-B., and Lemaréchal, C. 1996. *Convex Analysis and Minimization Algorithms: Part I Fundamentals and Part II Advanced Theory and Bundle Methods*, volume 305. Springer Science & Business Media.