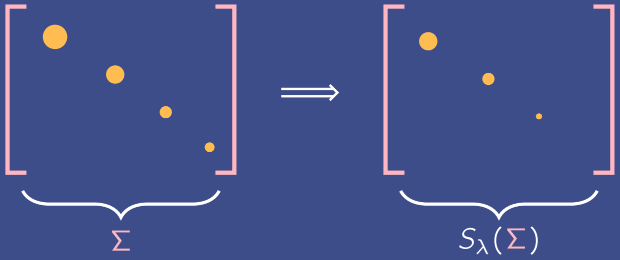
Singular Value Thresholding

the proximal of the nuclear norm



$$\operatorname{prox}_{\lambda\|\cdot\|_*}(U\Sigma V^\top) = U\, S_\lambda(\Sigma) V^\top$$

Howard Heaton

Typal Academy

Motivation

"We can't quite minimize rank, but we can get close enough." –a proximal wizard

Setting For convex f and the nuclear norm $\|\cdot\|_*$, solve

$$\min_X f(X) + \|X\|_*$$

Subproblem of Interest For $\lambda > 0$, compute nuclear norm proximal operator:

$$\text{prox}_{\lambda\|\cdot\|_*}(X) = \operatorname*{argmin}_Z \lambda\|Z\|_* + \frac{1}{2}\|Z - X\|_F^2$$

Why Subproblem matters

Prevalent Arises in matrix completion, imaging, system identification, *etc.*

Modular Can use this proximal in algorithms (*e.g.* proximal gradient, ADMM)

Why Nuclear Norm in Place of Rank

For a matrix X, the notions of rank and nuclear norm are related:

$$\text{rank}(X) = \#\{i : \sigma_i > 0\} \qquad \text{(number of nonzero singular values)}$$

$$\|X\|_* = \sum_i \sigma_i \qquad \text{(sum of singular values)}$$

Nuclear norm is surrogate for matrix rank, *i.e.* we approximate $\text{rank}(X) \approx \|X\|_*$

- Rank function is nonconvex and NP-hard to minimize
- Nuclear norm is tightest convex relaxation of rank
- Surrogate problems with this convex relaxation can be efficiently solved

Proximal Operator and Decomposition

The proximal operator for a function ϕ is

$$\mathrm{prox}_{\phi}(x) = \operatorname*{argmin}_{z} \phi(z) + \frac{1}{2}\|z - x\|^2.$$

This generalizes the projection onto a set $\mathcal{C}$:

$$\mathrm{proj}_{\mathcal{C}}(x) = \operatorname*{argmin}_{z \in \mathcal{C}} \frac{1}{2}\|z - x\|^2.$$

Theorem 1[†] For a norm $\|\cdot\|$ and $\lambda > 0$, the proximal can be decomposed via

$$\mathrm{prox}_{\lambda\|\cdot\|}(x) = x - \lambda\, \mathrm{proj}_{\mathcal{C}}\left(\frac{x}{\lambda}\right),$$

where $\mathcal{C} = \{z \ : \ \|z\|^* \leq 1\}$ and $\|\cdot\|^*$ is the norm dual to $\|\cdot\|$.

[†] See Theorem 6.46 and Example 6.47 in Beck's *First-Order Methods in Optimization*.

Projection of Matrix onto Unit Ball

Lemma 1 Let $\mathcal{C}$ be the unit ball $\{Z : \|Z\|_2 \leq 1\}$. If $X = U\Sigma V^\top$, then

$$\mathrm{proj}_{\mathcal{C}}(X) = UDV^\top, \quad \text{where} \quad D = \mathrm{diag}(d_i) \;\; \text{and} \;\; d_i = \min\{\sigma_i, 1\}.$$

Proof Given a matrix Z, set $S = U^\top ZV$ so that $Z = USV^\top$. Write $S = D + M$, where $D = \mathrm{diag}(d_i)$ and any nonzeros of M are off-diagonal. This yields

$$\|Z - X\|_F^2 = \|U(S - \Sigma)V^\top\|_F^2 = \|S - \Sigma\|_F^2 = \|D - \Sigma\|_F^2 + \|M\|_F^2 \geq \|D - \Sigma\|_F^2.$$

As the inquality is strict for nonzero M, for optimal Z we have $M = 0$. Thus,

$$\min_{\|Z\|_2 \leq 1} \|Z - X\|_F^2 = \min_{\|D\|_2 \leq 1} \|UDV^\top - X\|_F^2 = \min_{\|D\|_2 \leq 1} \|D - \Sigma\|_F^2 = \min_{|d_i| \leq 1} \sum_i (d_i - \sigma_i)^2.$$

Each diagonal entry d_i can be independently computed as $d_i = \min\{\sigma_i, 1\}$. ■

Proximal for Nuclear Norm

The soft-thresholding operator S_λ is an element-wise operator defined by

$$S_\lambda(X)_{ij} = \operatorname{sign}(X_{ij}) \cdot \max\{|X_{ij}| - \lambda,\ 0\}.$$

When X is nonnegative, this simplifies to

$$S_\lambda(X)_{ij} = \max\{X_{ij} - \lambda,\ 0\}.$$

Theorem 2 Given a matrix X with SVD $U\Sigma V^\top$ and scalar $\lambda > 0$, the nuclear norm proximal operator is given by

$$\operatorname{prox}_{\lambda\|\cdot\|_*}(X) = U S_\lambda(\Sigma) V^\top.$$

Theorem 2 Proof

By the decomposition in Theorem 1,[†]

$$\mathrm{prox}_{\lambda\|\cdot\|_*}(X) = X - \lambda\, \mathrm{proj}_{\mathcal{C}}\left(\frac{X}{\lambda}\right),$$

where $\mathcal{C}$ is as in Lemma 1. Applying Lemma 1 yields

$$\lambda\, \mathrm{proj}_{\mathcal{C}}\left(\frac{X}{\lambda}\right) = UDV^\top, \quad \text{where} \quad D = \mathrm{diag}(d_i) \quad \text{and} \quad d_i = \begin{cases} \sigma_i & \text{if } \sigma_i \leq \lambda \\ \lambda & \text{otherwise.} \end{cases}$$

Thus,

$$\mathrm{prox}_{\lambda\|\cdot\|_*}(X) = X - UDV^\top = U(\Sigma - D)V^\top = U\, S_\lambda(\Sigma)V^\top. \qquad \blacksquare$$

[†] The spectral norm is dual to the nuclear norm.

Toy Example

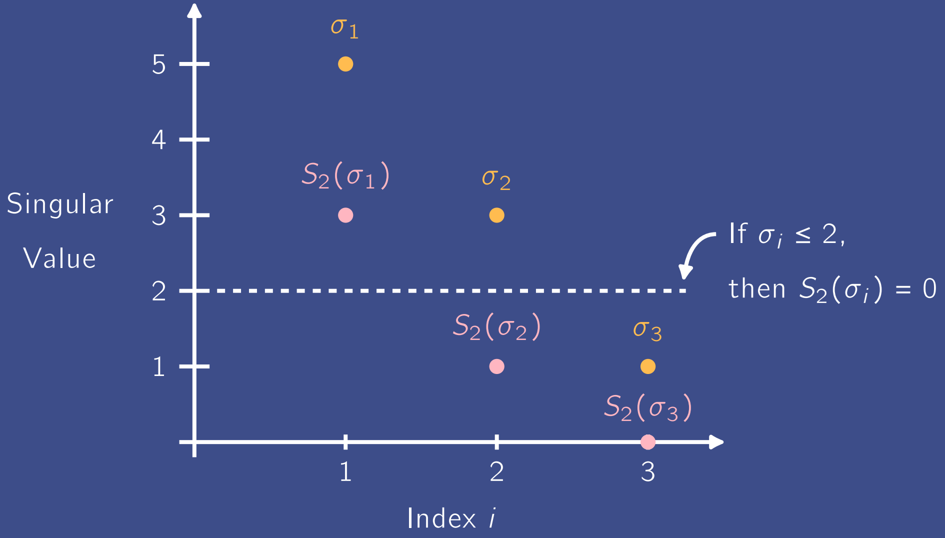
Consider the matrix

$$X = \begin{bmatrix} 0 & 3 & 0 \\ 5 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{V^\top} = U\Sigma V^\top$$

For $\lambda = 2$, the nuclear norm proximal for X is

$$\mathrm{prox}_{2\|\cdot\|_*}(X) = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S_2(\Sigma)} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{V^\top} = \begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Toy Example (Continued)



Rank Minimization and Sparsity

For a vector x, notions of sparsity and ℓ_1 norm are related:

$$\|x\|_0 = \#\{i \ : \ x_i \neq 0\} \qquad \text{(number of nonzero values)}$$

$$\|x\|_1 = \sum_i |x_i| \qquad \text{(sum of absolute values)}$$

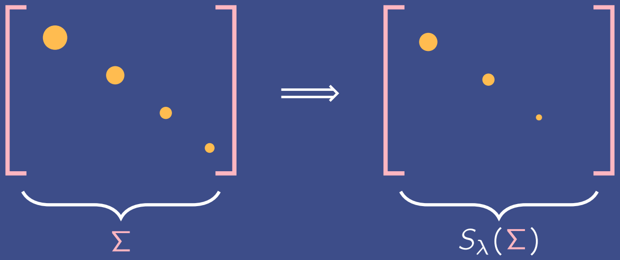
Rank minimization and sparsity promotion share analogous approximations, which have analogous proximal operators:

$$\text{prox}_{\lambda\|\cdot\|}(x) = S_\lambda(x)$$

$$\text{prox}_{\lambda\|\cdot\|_*}(X) = U\, S_\lambda(\Sigma) V^\top$$

Takeaways

- Nuclear norm is for low-rank matrices like ℓ_1 is for sparse vectors
- The nuclear norm proximal formula is $\text{prox}_{\lambda\|\cdot\|_*}(X) = U\, S_\lambda(\Sigma) V^\top$
- Proximal decomposition results can be helpful in deriving proximal formulas



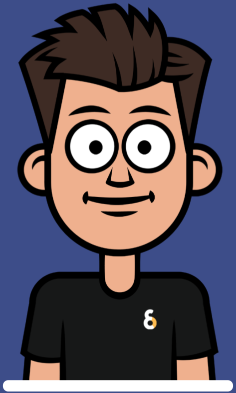
Found this useful?

➕ Follow for more

♻️ Share with friends

Building and need optimization help?

🤝 Let's connect



Howard Heaton

Typal Academy

References

- Cai, Candés, Shen. *A Singular Value Thresholding Algorithm for Matrix Completion*. 2008.
- Cai, Osher. *Fast Singular Value Thresholding without Singular Value Decomposition*. 2010.
- Beck. *First-Order Methods in Optimization*. 2017.