A New Perspective on Low-Rank Optimization

Dimitris Bertsimas, Ryan Cory-Wright and Jean Pauphilet

Saturday Seminar

Li, Chen

June 12, 2021

Outline

Low-rank Optimization Problem

$$\begin{aligned}
\min_{X \in \mathcal{S}_+^n} \quad & C \bullet X + \Omega(X) + \mu \cdot \mathrm{rank}(X) \\
\text{s.t.} \quad & \mathcal{A}X = b, \ X \in \mathcal{K} \\
& \mathrm{rank}(X) \leq k
\end{aligned} \tag{1}$$

- $\mathcal{A}X = [\mathrm{tr}\left(A_1 \bullet X\right), ..., \mathrm{tr}\left(A_m \bullet X\right)]$, $A_i \in \mathcal{S}^n$, $b \in \mathbb{R}^m$
- $\mathcal{K}$ is a proper cone
- $C \in \mathcal{S}^n$, $\mu \geq 0$, $\Omega(\cdot)$ is represented by $\Omega(X) = \sum_{i=1}^n \omega(\lambda_i(X))$ where $\lambda_i(X)$ is the i-th largest eigenvalue of X, ω is closed convex

Examples

- Low-rank regression: Given $Y \in \mathbb{R}^{m\times n}$ and $X \in \mathbb{R}^{m\times p}$,

$$\min_{\beta\in\mathbb{R}^{p\times n}} \frac{1}{2m}\|Y - X\beta\|_F^2 + \frac{1}{2\gamma}\|\beta\|_F^2 + \mu\cdot\mathrm{rank}(\beta)$$

- Low-rank factor analysis: Given $\Sigma \in \mathcal{S}_+^n$,

$$\begin{array}{ll} \min\limits_{X,\Phi\in\mathcal{S}_+^n} & \|\Sigma - \Phi - X\|_q^q \\ \text{s.t.} & \mathrm{rank}(X) \le k \\ & \Phi_{i,j} = 0, \forall i,j \in [n], i \neq j \\ & \|X\|_\sigma \le M \end{array}$$

where $\|X\|_q^q = (\sum_{i=1}^n \lambda_i(X)^q)^{1/q}$ with $q \ge 1$.

- Non-negative Matrix Factorization: Given $A \in \mathcal{S}_+^n$,

$$\min_{X\in\mathcal{C}_+^n, \mathrm{rank}(X)\le k} \|X - A\|_F^2$$

where $\mathcal{C}_+^n = \left\{X : X = UU^T, U \in \mathbb{R}_+^{n\times n}\right\}$

Vector Case: Sparse Regression

- Given data $y \in \mathbb{R}^m$, $X \in \mathbb{R}^{m \times p}$, consider sparse regression

$$\min_{\beta \in \mathbb{R}^p} \; \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}\|\beta\|_2^2 + \mu\|\beta\|_0 \tag{2}$$

Vector Case: Sparse Regression

- Given data $y \in \mathbb{R}^m$, $X \in \mathbb{R}^{m \times p}$, consider sparse regression

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}\|\beta\|_2^2 + \mu\|\beta\|_0 \tag{2}$$

- $\|\cdot\|_0$ essentially involves integer constraint: by introducing $z_i = \mathbb{I}(\beta_i \neq 0)$, we have big-M reformulation

$$\begin{array}{cl} \min\limits_{\beta \in \mathbb{R}^p, z \in \{0,1\}^p} & \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}\|\beta\|_2^2 + \mu e^T z \\ \text{s.t.} & -Mz_i \leq \beta_i \leq Mz_i \end{array}$$

Vector Case: Sparse Regression

- Given data $y \in \mathbb{R}^m$, $X \in \mathbb{R}^{m\times p}$, consider sparse regression

$$\min_{\beta\in\mathbb{R}^p} \; \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}\|\beta\|_2^2 + \mu\|\beta\|_0 \tag{2}$$

- $\|\cdot\|_0$ essentially involves integer constraint: by introducing $z_i = \mathbb{I}(\beta_i \neq 0)$, we have big-M reformulation

$$\begin{array}{cl} \displaystyle\min_{\beta\in\mathbb{R}^p, z\in\{0,1\}^p} & \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}\|\beta\|_2^2 + \mu e^T z \\ \text{s.t.} & -Mz_i \leq \beta_i \leq Mz_i \end{array}$$

- A better formulation is **perspective reformulation**[1]

$$\begin{array}{cl} \displaystyle\min_{\beta,\rho\in\mathbb{R}^p, z\in\{0,1\}^p} & \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}e^T\rho + \mu e^T z \\ \text{s.t.} & z_i\rho_i \geq \beta_i^2 \end{array} \tag{3}$$

[1] Oktay Günlük and Jeff Linderoth. "Perspective reformulations of mixed integer nonlinear programs with indicator variables". In: *Mathematical programming* 124.1 (2010), pp. 183–205.

Vector Case: Sparse Regression

- Given data $y \in \mathbb{R}^m$, $X \in \mathbb{R}^{m \times p}$, consider sparse regression

$$\min_{\beta \in \mathbb{R}^p} \; \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}\|\beta\|_2^2 + \mu\|\beta\|_0 \tag{2}$$

- A better formulation is **perspective reformulation**[1]

$$\begin{array}{cl} \min\limits_{\beta,\rho \in \mathbb{R}^p, z \in \{0,1\}^p} & \frac{1}{2m}\|y - X\beta\|_2^2 + \frac{1}{2\gamma}e^T\rho + \mu e^T z \\ \text{s.t.} & z_i\rho_i \ge \beta_i^2 \end{array} \tag{3}$$

- A tighter reformulation:

$$\begin{array}{cl} \min\limits_{\beta \in \mathbb{R}^p, z \in \{0,1\}^p, \theta \in \mathcal{S}_+^p} & \frac{1}{2m}\|y\|_2^2 - \frac{1}{m}y^T X\beta + \left(\frac{1}{2m}X^T X + \frac{1}{2\gamma}I\right) \bullet \theta + \mu e^T z \\ \text{s.t.} & z_i\theta_{i,i} \ge \beta_i^2 \\ & \theta \succeq \beta\beta^T \end{array} \tag{4}$$

[1] **Günlük and Linderoth**, "Perspective reformulations of mixed integer nonlinear programs with indicator variables".

Cardinality vs. Low-rank Constraints

- Cardinality constraints $\|x\|_0 \leq k$

$$\Longleftrightarrow \exists z \in \{0,1\}^n \text{ s.t. } e^T z \leq k, x = z \cdot x$$

- Low-rank constraints $\operatorname{rank}(X) \leq k$

$$\Longleftrightarrow \exists Y \in \mathcal{Y}_n \text{ s.t. } I \bullet Y \leq k, X = YX$$

where $\mathcal{Y}_n = \left\{Y \in \mathcal{S}^n : Y = Y^2\right\}$ is the set of orthogonal project matrices.

Cardinality vs. Low-rank Constraints

- Cardinality constraints $\|x\|_0 \le k$

$$\Longleftrightarrow \exists z \in \{0,1\}^n \text{ s.t. } e^T z \le k, x = z \cdot x$$

- Low-rank constraints $\mathrm{rank}(X) \le k$

$$\Longleftrightarrow \exists Y \in \mathcal{Y}_n \text{ s.t. } I \bullet Y \le k, X = YX$$

where $\mathcal{Y}_n = \left\{ Y \in \mathcal{S}^n : Y = Y^2 \right\}$ is the set of orthogonal project matrices.
Proof.
  - $\Longleftarrow$: $\mathrm{rank}(X) = \mathrm{rank}(XY) \le \mathrm{rank}(Y) \le k$
  - $\Longrightarrow$: Do reduced SVD $X = U\Sigma V^T$, let $Y = UU^T$, then $Y^T Y = U(U^T U)U^T = UU^T = Y$ and $I \bullet Y = \mathrm{tr}\left(UU^T\right) \le k$, $YX = U(U^T U)\Sigma V^T = X$.

Comparison

Cardinality constraints

$$\begin{array}{ll} \min & \|\beta\|_2^2/2 + \mu\|\beta\|_0 \\ \text{s.t.} & \|\beta\|_0 \leq k \end{array}$$

$$\Longleftrightarrow \min \quad e^T\rho/2 + \mu e^T z$$
$$\begin{pmatrix} \rho_i & \beta_i \\ \beta_i & z_i \end{pmatrix} \succeq 0$$
$$z \in \{0,1\}^n, e^T z \leq k$$

- Cardinality on binaries is linear: $\|z\|_0 = e^T z$
- Binary variables are idempotent scalars which satisfy $z^2 = z$
- Model logical constraints by $x = zx$

Low-rank constraints

$$\begin{array}{ll} \min & \|\beta\|_F^2/2 + \mu \cdot \mathrm{rank}(\beta) \\ \text{s.t.} & \mathrm{rank}(\beta) \leq k \end{array}$$

$$\Longleftrightarrow \min \quad \mathrm{tr}\left(\theta\right)/2 + \mu \cdot \mathrm{tr}\left(Y\right)$$
$$\begin{pmatrix} \theta & \beta \\ \beta^T & Y \end{pmatrix} \succeq 0$$
$$Y \in \mathcal{Y}_n, I \bullet Y \leq k$$

- Rank on projection matrices is linear: $\mathrm{rank}(Y) = Y \bullet I$
- Projection matrices are idempotent matrices which satisfy $Y^2 = Y$
- Model rank constraints by $X = YX$

Outline

Perspective Functions

For a closed convex function f with $0 \in \mathrm{dom} f$

$$g_f(x,\lambda) = \begin{cases} \lambda f(x/\lambda) & \text{if } \lambda > 0, x/\lambda \in \mathrm{dom} f \\ 0 & \text{if } \lambda = 0, x = 0 \\ +\infty & \text{otherwise} \end{cases}$$

Properties:

- f is closed and convex if and only if g_f is closed and convex
- If f is closed and proper, g_f is closed if and only if $\lim_{s\to\infty} \frac{f(x+sz)-f(x)}{s} = \infty$ for any nonzero $x, z \in \mathrm{dom} f$.
- If $\lambda > 0$, then $(x, \lambda, s) \in \mathrm{epi} g_f$ iff $(x/\lambda, s/\lambda) \in \mathrm{epi} f$

Perspective Reformulation Technique

Consider a logically-constrained problem of the form

$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\subseteq\{0,1\}^n} \left\{c^T z + f(x) + \Omega(x) : x_i = 0 \text{ if } z_i = 0, \forall i \in [n]\right\} \quad (5)$$

Assumption 1 (Separability)

The function $\Omega(x) = \sum_{i=1}^n \Omega_i(x_i)$ *where* Ω_i *is closed convex and* $0 \in \operatorname{dom}(\Omega_i)$

Lemma 1

Under separability assumption, the problem (5) *is equivalent to*

$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\subseteq\{0,1\}^n} \left\{c^T z + f(x) + \sum_{i=1}^n g_{\Omega_i}(x_i, z_i) + (1 - z_i)\Omega_i(0)\right\} \quad (6)$$

Examples: Perspective Reformulation

Table 1: Convex substructures which frequently arise in MIOs and their perspective reformulations.

Penalty	$\Omega(x)$	$g_\Omega(x,z)$	Formulation
Big-M	$\begin{cases} 0 & \text{if } \lvert x\rvert \le M, \\ +\infty & \text{otherwise} \end{cases}$	$\begin{cases} 0 & \text{if } \lvert x\rvert \le Mz \\ +\infty & \text{otherwise} \end{cases}$	$\lvert x\rvert \le Mz$
Ridge	$\frac{1}{2\gamma}x^2$	$\begin{cases} x^2/2\gamma z & \text{if } z>0 \\ 0 & \text{if } x=z=0 \\ +\infty & \text{otherwise} \end{cases}$	$\min \theta$ s.t. $\theta z \ge \frac{1}{2\gamma}x^2$
Ridge + Big-M	$\frac{1}{2\gamma}x^2 + \begin{cases} 0 & \text{if } \lvert x\rvert \le M \\ +\infty & \text{otherwise} \end{cases}$	$\begin{cases} x^2/2\gamma z & \text{if } z>0, \lvert x\rvert \le Mz \\ 0 & \text{if } x=z=0 \\ +\infty & \text{otherwise} \end{cases}$	$\min \theta$ s.t. $\theta z \ge \frac{1}{2\gamma}x^2,\ \lvert x\rvert \le Mz$
Power	$\lvert x\rvert^p,\ p \ge 1$	$\begin{cases} \lvert x\rvert^p z^{1-p} & \text{if } z>0, \\ 0 & \text{if } x=z=0 \\ +\infty & \text{otherwise} \end{cases}$	$\min \theta$ s.t. $(\theta, z, x) \in \mathcal{K}_{\text{pow}}^{1/p}$
Logarithm	$-\log(x+\epsilon) : \epsilon > 0$	$\begin{cases} -z\log(x/z+\epsilon) & \text{if } x,z>0 \\ 0 & \text{if } x=z=0 \\ +\infty & \text{otherwise} \end{cases}$	$\min \theta$ s.t. $(x+z\epsilon, z, -\theta) \in \mathcal{K}_{\text{exp}}$
Entropy	$x\log x$	$\begin{cases} x\log(x/z) & \text{if } x,z>0 \\ 0 & \text{if } x=z=0 \\ +\infty & \text{otherwise} \end{cases}$	$\min \theta$ s.t. $(z, x, -\theta) \in \mathcal{K}_{\text{exp}}$
Softplus	$\log(1+\exp(x))$	$\begin{cases} z\log(1+\exp(x/z)) & \text{if } z>0 \\ 0 & \text{if } x=z=0 \\ +\infty & \text{otherwise} \end{cases}$	$\begin{aligned} \min\ & \theta \\ \text{s.t.}\ & z \ge u+v, \\ & (u, z, -\theta) \in \mathcal{K}_{\text{exp}}, \\ & (v, z, x-\theta) \in \mathcal{K}_{\text{exp}} \end{aligned}$

Matrix Perspective Functions

Definition 1

The matrix function $f : \mathcal{X} \to \mathcal{S}^n$ is convex if

$$f(tX + (1-t)Y) \preceq tf(X) + (1-t)f(Y),\ \forall X, Y \in \mathcal{X}, t \in [0,1]$$

where $\mathcal{X} \subseteq \mathcal{S}^n$ is convex. The generalized matrix perspective function is

$$g_f(X,Y) = \begin{cases} Y^{\frac{1}{2}} f(Y^\dagger X) Y^{\frac{1}{2}} & \text{if } Y \succeq 0, Y^\dagger X \in \mathcal{X} \\ +\infty & \text{otherwise} \end{cases}$$

- Typically one requires $Y \succ 0$ in the def of matrix perspective function.
- Note $Y^\dagger X = XY^\dagger = Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}}$ since it is symmetric by def.
- Note $X \in \mathrm{span}(Y)$ to avoid blowing up to $+\infty$

Examples

- Spectral constraint $f(X) = \begin{cases} 0 & \text{if } \|X\|_\sigma \preceq M \\ \infty & \text{otherwise} \end{cases}$

$$g_f(X,Y) = \begin{cases} 0 & \text{if } -MY \preceq X \preceq MY \\ \infty & \text{otherwise} \end{cases}$$

- Convex quadratic $f(X) = X^2$

$$g_f(X,Y) = \begin{cases} XY^\dagger X & \text{if } Y \succeq 0 \\ \infty & \text{otherwise} \end{cases}$$

  Note that minimizing $\operatorname{tr}(g_f(X,Y)) =$ is a SDP:

$$\min_{\theta, Y} \left\{ \operatorname{tr}(\theta) : \begin{pmatrix} \theta & X \\ X & Y \end{pmatrix} \succeq 0 \right\}$$

  by Generalized Schur Complement Lemma

- The convex quadratic $f(X) = X^T X$ extends to rectangular case.

Examples

- Logarithm: $f(X) = -\log\det(X)$

$$g_f(X,Y) = \begin{cases} Y^{\frac{1}{2}}(\log Y - \log X)Y^{\frac{1}{2}} & \text{if } X, Y \succ 0 \\ \infty & \text{otherwise} \end{cases}$$

- Von Neumann entropy $f(X) = X^{\frac{1}{2}} \log X X^{\frac{1}{2}}$

$$g_f(X,Y) = \begin{cases} X^{\frac{1}{2}} \log(Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}})X^{\frac{1}{2}} & \text{if } X, Y \succ 0 \\ \infty & \text{otherwise} \end{cases}$$

The epigraph of g_f is the quantum relative entropy cone

$$\mathcal{K}_{mat}^{op,rel} := \left\{ \begin{array}{l} (X_1, X_2, X_3) \in \mathcal{S}^n \times \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} : \\ X_1 \succeq -X_2^{\frac{1}{2}} \log(X_2^{-\frac{1}{2}} X_3 X_2^{-\frac{1}{2}}) X_2^{\frac{1}{2}} \end{array} \right\}$$

Examples

Table 2: Analogy between perspectives of scalars and perspectives of matrix convex functions.

	Perspective of function			Matrix perspective of function		
Type	$f(x) : \mathbb{R} \to \mathbb{R}$	$g_f(\boldsymbol{x}, t)$	Ref.	f	g_f	Ref.
Quadratic	x^2	x^2/t	[3]	$\boldsymbol{X}^\top \boldsymbol{X}$	$\boldsymbol{X}^\top \boldsymbol{Y}^\dagger \boldsymbol{X}$	[8]
Power	$-x^\alpha : 0 < \alpha < 1$	$-x^\alpha t^{1-\alpha}$	[11]	$-\boldsymbol{X}^\alpha$	$-\boldsymbol{Y}^{\frac{1-\alpha}{2}} \boldsymbol{X}^\alpha \boldsymbol{Y}^{\frac{1-\alpha}{2}}$	Prop. 2
Log	$-\log(x)$	$-t \log(\frac{x}{t})$	[11]	$\log(\boldsymbol{X})$	$-\boldsymbol{Y}^{\frac{1}{2}} \log\left(\boldsymbol{X}^{-\frac{1}{2}} \boldsymbol{Y} \boldsymbol{X}^{-\frac{1}{2}}\right) \boldsymbol{Y}^{\frac{1}{2}}$	[24]
Entropy	$x \log(x)$	$x \log(\frac{x}{t})$	[11]	$\boldsymbol{X}^{\frac{1}{2}} \log(\boldsymbol{X}) \boldsymbol{X}^{\frac{1}{2}}$	$\boldsymbol{X}^{\frac{1}{2}} \log(\boldsymbol{Y}^{-\frac{1}{2}} \boldsymbol{X} \boldsymbol{Y}^{-\frac{1}{2}}) \boldsymbol{X}^{\frac{1}{2}}$	[43, 21]

Properties of the Matrix Perspective

Proposition 1

Let f be a matrix-valued function, and g_f its perspective function.

- *f is matrix convex iff g_f is matrix convex*[2]
- *g_f is positive homogeneous: $g_f(\mu X, \mu Y) = \mu g_f(X, Y)$*
- *Let $Y \succ 0$, $\mathrm{epi}(f) = \{(X, \theta) : \theta \succeq f(X)\}$, then $(X, Y, \theta) \in \mathrm{epi}(g_f)$ iff $(Y^{-\frac{1}{2}} X Y^{-\frac{1}{2}}, Y^{-\frac{1}{2}} \theta Y^{-\frac{1}{2}}) \in \mathrm{epi}(f)$*
- *Let $f(X)$ be a spectral function, then*

$$g_f(X, Y) = U \mathrm{Diag}(g_f(\lambda_1(X), \lambda_1(Y)), ..., g_f(\lambda_n(X), \lambda_n(Y)))U^T$$

where X, Y are simultaneously diagonalizable under U.

[2] Ali Ebadian, Ismail Nikoufar, and Madjid Eshaghi Gordji. "Perspectives of matrix convex functions". In: *Proceedings of the National Academy of Sciences* 108.18 (2011), pp. 7313–7314.

Matrix Perspective Reformulation Technique

Assumption 2

$\Omega(X) = \sum_{i=1}^{n} \omega(\lambda_i(X))$ *where* ω *is closed convex and* $0 \in \mathrm{dom}(\omega)$

Theorem 1

The low-rank problem (1) is equivalent to

$$\begin{aligned} \min_{X \in \mathcal{S}_+^n, Y \in \mathcal{Y}_n^k} \quad & C \bullet X + g_\Omega(X, Y) + \mathrm{tr}\,(I - Y)\,\omega(0) + \mu \cdot \mathrm{tr}\,(Y) \\ \text{s.t.} \quad & \mathcal{A}X = b,\ X \in \mathcal{K} \end{aligned} \tag{7}$$

where $\mathcal{Y}_n^k = \left\{ Y \in \mathcal{S}_+^n : \mathrm{tr}\,(Y) \leq k, Y^2 = Y \right\}$

Matrix Perspective Reformulation Technique (MPRT): Relax $\mathcal{Y}_n^k$ to

$$\mathrm{conv}\left(\mathcal{Y}_n^k\right) = \left\{ Y \in \mathcal{S}_+^n : \mathrm{tr}\,(Y) \leq k, Y \preceq I \right\}$$

Proof of Theorem 1

Based on projection matrix modeling of rank constraints, We only need to show equivalence of problem (7)

$$
\begin{array}{ll}
\min\limits_{X\in\mathcal{S}_+^n, Y\in\mathcal{Y}_n^k} & C\bullet X + g_\Omega(X,Y) + \mathrm{tr}\,(I-Y)\,\omega(0) + \mu\cdot\mathrm{tr}\,(Y) \\
\text{s.t.} & \mathcal{A}X = b,\ X\in\mathcal{K}
\end{array}
$$

and

$$
\begin{array}{ll}
\min\limits_{X\in\mathcal{S}_+^n, Y\in\mathcal{Y}_n^k} & C\bullet X + \Omega(X) + \mu\cdot\mathrm{tr}\,(Y) \\
\text{s.t.} & \mathcal{A}X = b,\ X\in\mathcal{K},\ X = YX
\end{array}
\tag{8}
$$

Proof of Theorem 1

Based on projection matrix modeling of rank constraints, We only need to show equivalence of problem (7)

$$\begin{array}{ll} \min\limits_{X\in\mathcal{S}_+^n, Y\in\mathcal{Y}_n^k} & C\bullet X + g_\Omega(X,Y) + \mathrm{tr}\,(I-Y)\,\omega(0) + \mu\cdot\mathrm{tr}\,(Y) \\ \text{s.t.} & \mathcal{A}X = b,\ X\in\mathcal{K} \end{array}$$

and

$$\begin{array}{ll} \min\limits_{X\in\mathcal{S}_+^n, Y\in\mathcal{Y}_n^k} & C\bullet X + \Omega(X) + \mu\cdot\mathrm{tr}\,(Y) \\ \text{s.t.} & \mathcal{A}X = b,\ X\in\mathcal{K},\ X = YX \end{array} \tag{8}$$

Let X, Y be feasible to (8), let $X = U\mathrm{Diag}(\lambda_1(X), ..., \lambda_n(X))U^T$, $Y = UU^T$, then

$$\begin{aligned} g_\Omega(X,Y) &= \mathrm{tr}\left(U\mathrm{Diag}(\mathbb{I}(\lambda_1(Y)>0)\omega(\lambda_1(X)), ..., \mathbb{I}(\lambda_n(Y)>0)\omega(\lambda_n(X)))U^T\right) \\ &= \textstyle\sum_{i=1}^n \mathbb{I}(\lambda_i(Y)>0)\omega(\lambda_i(X)) \end{aligned}$$

Hence $g_\Omega(X,Y) + (n - \mathrm{tr}\,(Y))\omega(0) = \Omega(X)$.

Proof of Theorem 1

Based on projection matrix modeling of rank constraints, We only need to show equivalence of problem (7)

$$\begin{aligned} \min_{X \in \mathcal{S}_+^n, Y \in \mathcal{Y}_n^k} \quad & C \bullet X + g_\Omega(X, Y) + \operatorname{tr}(I - Y)\,\omega(0) + \mu \cdot \operatorname{tr}(Y) \\ \text{s.t.} \quad & \mathcal{A}X = b,\ X \in \mathcal{K} \end{aligned}$$

and

$$\begin{aligned} \min_{X \in \mathcal{S}_+^n, Y \in \mathcal{Y}_n^k} \quad & C \bullet X + \Omega(X) + \mu \cdot \operatorname{tr}(Y) \\ \text{s.t.} \quad & \mathcal{A}X = b,\ X \in \mathcal{K},\ X = YX \end{aligned} \tag{8}$$

Let X, Y be feasible to (7), then $X \in \operatorname{span}(Y)$, which implies $X = YX$ as $Y \in \mathcal{Y}_n$. $g_\Omega(X, Y) + (n - \operatorname{tr}(Y))\omega(0) = \Omega(X)$ still holds.

Convex Hulls of Low-Rank Sets

Convex hull of Low-rank sets is equivalent to the MPRT relaxation.

Theorem 2

Let

$$\mathcal{T} = \{X \in \mathcal{S}_+^n : \operatorname{tr}(f(X)) + \mu \cdot \operatorname{rank}(X) \le t, \operatorname{rank}(X) \le k\}$$

where $f(X) = U\operatorname{Diag}(f(\lambda_1(X)), ..., f(\lambda_n(X))U^T)$, *then its convex hull*

$$\mathcal{T}^c = \left\{X : \exists Y \in \operatorname{conv}\left(\mathcal{Y}_n^k\right) \text{ s.t. } \operatorname{tr}(g_f(X,Y)) + \operatorname{tr}(I - Y) f(0) + \mu \cdot \operatorname{tr}(Y) \le t\right\}$$

where $\operatorname{conv}\left(\mathcal{Y}_n^k\right) = \{Y \in \mathcal{S}_+^n : Y \preceq I, \operatorname{tr}(Y) \le k\}$.

Proof of Theorem 2

$\operatorname{conv}(\mathcal{T}) \subseteq \mathcal{T}^c$:

If $\operatorname{rank}(X) \leq k$, then $\exists Y \in \mathcal{Y}_n^k$ s.t. $X = YX$ and $\operatorname{rank}(X) = \operatorname{tr}(Y)$. Then $\operatorname{tr}(g_f(X,Y)) + (n - \operatorname{tr}(Y))f(0) = \operatorname{tr}(f(X)) \leq t$.

Proof of Theorem 2

$\mathcal{T}^c \subseteq \operatorname{conv}(\mathcal{T})$:
let $X \in \mathcal{T}^c$, then by def of g_f, we have

$$\mathcal{T}^c = \left\{ \begin{array}{ll} X: & \exists Y \in \operatorname{conv}\left(\mathcal{Y}_n^k\right) \text{ s.t.} \\ & \sum_{i=1}^n g_f(\lambda_i(X), \lambda_i(Y)) + (n - \operatorname{tr}(Y))f(0) + \mu \cdot \operatorname{tr}(Y) \le t \end{array} \right\}$$

Hence $(\lambda(X), \lambda(Y))$ is in

$$\left\{ (x, y) \in \mathbb{R}_+^n \times [0,1]^n : \sum_{i=1}^n y_i \le k, \sum_{i=1}^n (g_f(x_i, y_i) + (1 - y_i)f(0) + \mu \cdot y_i) \le t \right\},$$

which is the convex hull of

$$\mathcal{U} \triangleq \left\{ (x, y) \in \mathbb{R}_+^n \times \{0,1\}^n : \sum_{i=1}^n y_i \le k, \sum_{i=1}^n (f(x_i) + \mu \cdot y_i) \le t, x_i = 0 \text{ if } y_i = 0 \right\}.$$

Let $(\lambda(X), \lambda(Y)) = \sum_k \alpha_k (x_k, y_k)$ with $(x_k, y_k) \in \mathcal{U}$, then

$$X = \sum_k \alpha_k U \operatorname{Diag}(x_k) U^T \in \operatorname{conv}(\mathcal{T})$$

.

Relaxations: From Cardinality to Low-rank

Relax Logical constraint (5) as follows:

$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\subseteq\{0,1\}^n} \left\{c^T z + f(x) + \Omega(x) : x_i = 0 \text{ if } z_i = 0, \forall i \in [n]\right\}$$
$$\Updownarrow$$
$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\subseteq\{0,1\}^n} \left\{c^T z + f(x) + \Omega(x) : x_i = x_i z_i, \forall i \in [n]\right\}$$

By analog, relax rank constraint as follows:

$$\begin{aligned} \min_{X\in\mathcal{S}^n_+} \quad & C \bullet X + \Omega(X) + \mu \cdot \operatorname{rank}(X) \\ & \mathcal{A}X = b, \operatorname{rank}(X) \le k \end{aligned}$$
$$\Updownarrow$$
$$\begin{aligned} \min_{X, Y\in\mathcal{S}^n_+} \quad & C \bullet X + \Omega(X) + \mu \cdot \operatorname{tr}(Y) \\ & \mathcal{A}X = b, X = XY, Y^2 = Y, \operatorname{tr}(Y) \le k \end{aligned}$$

Relaxations: From Cardinality to Low-rank

Relax Logical constraint (5) as follows:

$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\subseteq\{0,1\}^n} \left\{c^T z + f(x) + \Omega(x) : x_i = x_i z_i, \forall i \in [n]\right\}$$
$$\Updownarrow$$
$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\cap\{0,1\}^n} \left\{c^T z + f(x) + \sum_{i=1}^n g_{\Omega_i}(x_i, z_i) + (1 - z_i)\Omega_i(0)\right\}$$

By analog, relax rank constraint as follows:

$$\begin{aligned} \min_{X,Y\in\mathcal{S}^n_+} \quad & C \bullet X + \Omega(X) + \mu \cdot \operatorname{tr}(Y) \\ & \mathcal{A}X = b, X = XY, Y^2 = Y, \operatorname{tr}(Y) \le k \end{aligned}$$
$$\Updownarrow$$
$$\begin{aligned} \min_{X,Y\in\mathcal{S}^n_+} \quad & C \bullet X + g_\Omega(X, Y) + \operatorname{tr}(I - Y)\,\omega(0) + \mu \cdot \operatorname{tr}(Y) \\ & \mathcal{A}X = b, Y^2 = Y, \operatorname{tr}(Y) \le k \end{aligned}$$

Relaxations: From Cardinality to Low-rank

Relax Logical constraint (5) as follows:

$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\cap\{0,1\}^n} \left\{c^T z + f(x) + \textstyle\sum_{i=1}^n g_{\Omega_i}(x_i, z_i) + (1-z_i)\Omega_i(0)\right\}$$
$$\Downarrow$$
$$\min_{x\in\mathbb{R}^n, z\in\mathcal{Z}\cap[0,1]^n} \left\{c^T z + f(x) + \textstyle\sum_{i=1}^n g_{\Omega_i}(x_i, z_i) + (1-z_i)\Omega_i(0)\right\}$$

By analog, relax rank constraint as follows:

$$\begin{aligned}\min_{X,Y\in\mathcal{S}^n_+} \quad & C\bullet X + g_\Omega(X,Y) + \mathrm{tr}\,(I-Y)\,\omega(0) + \mu\cdot\mathrm{tr}\,(Y)\\ & \mathcal{A}X = b, Y^2 = Y, \mathrm{tr}\,(Y) \le k\end{aligned}$$
$$\Downarrow$$
$$\begin{aligned}\min_{X,Y\in\mathcal{S}^n_+} \quad & C\bullet X + g_\Omega(X,Y) + \mathrm{tr}\,(I-Y)\,\omega(0) + \mu\cdot\mathrm{tr}\,(Y)\\ & \mathcal{A}X = b, Y \preceq I, \mathrm{tr}\,(Y) \le k\end{aligned}$$

Application: Matrix Completion

$$\min_{X \in \mathcal{S}^n_+} \sum_{(i,j) \in \mathcal{I}} (X_{ij} - A_{ij})^2 + \frac{1}{2\gamma} \|X\|_F^2 + \mu \cdot \mathrm{rank}(X)$$

has MPRT relaxation

$$\begin{array}{ll} \min\limits_{X, \theta, Y \in \mathcal{S}^n_+} & \sum_{(i,j) \in \mathcal{I}} (X_{ij} - A_{ij})^2 + \frac{1}{2\gamma} \mathrm{tr}\,(\theta) + \mu \cdot \mathrm{tr}\,(Y) \\ \text{s.t.} & Y \preceq I \\ & \begin{pmatrix} \theta & X \\ X & Y \end{pmatrix} \succeq 0 \end{array}$$

Application: Low-Rank Factor Analysis

$$
\begin{aligned}
\min_{X,\Phi\in\mathcal{S}^n_+} \quad & \|\Sigma-\Phi-X\|_q^q \\
\text{s.t.} \quad & \operatorname{rank}(X)\le k \\
& \Phi_{i,j}=0, \forall i,j\in[n], i\neq j \\
& \|X\|_\sigma \le M
\end{aligned}
$$

- Problem: $\operatorname{tr}\left((\Sigma-\Phi-X)^q\right)$

Application: Low-Rank Factor Analysis

$$\begin{aligned}
\min_{X,\Phi\in\mathcal{S}^n_+} \quad & \|\Sigma - \Phi - X\|_q^q \\
\text{s.t.} \quad & \text{rank}(X) \le k \\
& \Phi_{i,j} = 0, \forall i,j \in [n], i \ne j \\
& \|X\|_\sigma \le M
\end{aligned}$$

- Scalar analog:

$$\mathcal{T} = \{(x,y,z,t) : t \ge |x+y-d|^q, |x| \le M, z \in \{0,1\}, x = 0 \text{ if } z = 0\}$$

Calculate its convex hull:

$$\mathcal{T}^c = \left\{ \begin{array}{ll} (x,y,z,t): & \exists \beta \in \mathbb{R} \text{ s.t. } t \ge \frac{|y-\beta-(1-z)d|^q}{(1-z)^{q-1}} + \frac{|x+\beta-dz|^q}{z^{q-1}} \\ & |x| \le Mz, z \in [0,1] \end{array} \right\}$$

Application: Low-Rank Factor Analysis

$$\begin{aligned}
\min_{X,\Phi\in\mathcal{S}^n_+} \quad & \|\Sigma-\Phi-X\|_q^q \\
\text{s.t.} \quad & \operatorname{rank}(X)\le k \\
& \Phi_{i,j}=0, \forall i,j\in[n], i\neq j \\
& \|X\|_\sigma \le M
\end{aligned}$$

- Scalar analog:

$$\mathcal{T}=\{(x,y,z,t): t\ge |x+y-d|^q, |x|\le M, z\in\{0,1\}, x=0 \text{ if } z=0\}$$

Calculate its convex hull:

$$\mathcal{T}^c=\left\{ \begin{array}{ll} (x,y,z,t): & \exists\beta\in\mathbb{R} \text{ s.t. } t\ge \frac{|y-\beta-(1-z)d|^q}{(1-z)^{q-1}}+\frac{|x+\beta-dz|^q}{z^{q-1}} \\ & |x|\le Mz, z\in[0,1] \end{array} \right\}$$

- Relaxation:

$$\begin{aligned}
\min_{\substack{X,\Phi,Y_1,\\ Y_2,\theta,\beta}} \quad & \operatorname{tr}(\theta) \\
\text{s.t.} \quad & Y_1+Y_2=I, \operatorname{tr}(Y_2)\le k, -MY_2\preceq X\preceq MY_2 \\
& \theta\succeq Y_1^{\frac{1-q}{2}}(\Phi-\beta-Y_1^{\frac{1}{2}}\Sigma Y_1^{\frac{1}{2}})Y_1^{\frac{1-q}{2}}+Y_2^{\frac{1-q}{2}}(X+\beta-Y_2^{\frac{1}{2}}\Sigma Y_2^{\frac{1}{2}})Y_2^{\frac{1-q}{2}} \\
& \Phi_{i,j}=0, \forall i,j\in[n], i\neq j \\
& X,\Phi,Y_1,Y_2,\theta\in\mathcal{S}^n_+, \beta\in\mathcal{S}^n
\end{aligned}$$

Application: rank-k SVD

$$\min_{X\in\mathbb{R}^{n\times m}} \left\{ \frac{1}{2}\|X-A\|_F^2 : \mathrm{rank}(X) \le k \right\}$$

is equivalent to

$$\begin{array}{cl} \displaystyle\min_{X\in\mathbb{R}^{n\times m}, Y\in\mathcal{S}_+^m, \theta\in\mathcal{S}_+^n} & \frac{1}{2}\|A\|_F^2 - X\bullet A + \frac{1}{2}\mathrm{tr}\,(\theta) \\ \text{s.t.} & \mathrm{tr}\,(Y) \le k \\ & Y \preceq I \\ & \begin{pmatrix} \theta & X \\ X^T & Y \end{pmatrix} \succeq 0 \end{array}$$

Application: Optimal Experimental Design

Given $A = (a_1, a_2, ..., a_n) \in \mathbb{R}^{m \times n}$, solve

$$\max_{z \in \{0,1\}^n, e^T z \leq k} \log\det\left(\epsilon I + \sum_{i=1}^{n} z_i a_i a_i^T\right)$$

or its reformulation

$$\begin{array}{ll} \max\limits_{z \in \{0,1\}^n, \theta \in \mathcal{S}_+^m} & \operatorname{tr}(\theta) \\ \text{s.t.} & \theta \preceq \log\left(\epsilon I + A\mathrm{Diag}(z)A^T\right) \\ & e^T z \leq k \end{array}$$

Application: Optimal Experimental Design

$$
\begin{array}{ll}
\max\limits_{z\in\{0,1\}^n,\theta\in\mathcal{S}_+^m} & \operatorname{tr}(\theta) \\
\text{s.t.} & \theta \preceq \log\left(\epsilon I + A\mathrm{Diag}(z)A^T\right) \\
& e^T z \leq k
\end{array}
$$

Let $X = A\mathrm{Diag}(z)A^T$, then

- $(-\theta, I, \epsilon I + X) \in \mathcal{K}_{mat}^{op,rel}$.
- $\mathrm{rank}(X) \leq k$, so $\exists\, Y$ such that $Y \in \mathcal{Y}_m^k$ and $X = YX$.

Apply MPRT

$$
\begin{array}{ll}
\max\limits_{\theta,Y\in\mathcal{S}_+^m,z} & \operatorname{tr}(\theta) + \operatorname{tr}(I-Y)\log\epsilon \\
\text{s.t.} & \theta \preceq Y^{\frac{1}{2}}\log\left(\epsilon I + Y^{-\frac{1}{2}}A\mathrm{Diag}(z)A^T Y^{-\frac{1}{2}}\right)Y^{\frac{1}{2}} \\
& z\in[0,1]^n, e^T z \leq k \\
& Y \preceq I, \operatorname{tr}(Y) \leq k
\end{array}
$$

The first constraint can be written as $(-\theta, Y, \epsilon I + X) \in \mathcal{K}_{mat}^{op,rel}$.
It makes sense when $m > k$, otherwise $\operatorname{tr}(Y) \leq k$ always holds.

Outline

Reduced Rank Regression

Compare relaxation "Persp"

$$\begin{array}{ll} \min\limits_{\beta \in \mathbb{R}^{p\times n}, W\in \mathcal{S}_+^n, \theta \in \mathcal{S}_+^p} & \frac{1}{2m}\|Y - X\beta\|_F^2 + \frac{1}{2\gamma}\mathrm{tr}\left(\theta\right) + \mu \cdot \mathrm{tr}\left(W\right) \\ \text{s.t.} & W \preceq I \\ & \begin{pmatrix} \theta & \beta \\ \beta^T & W \end{pmatrix} \succeq 0 \end{array}$$

and "DCL"[3]

$$\begin{array}{ll} \min\limits_{\beta \in \mathbb{R}^{p\times n}, W\in \mathcal{S}_+^n, \theta \in \mathcal{S}_+^p} & \frac{1}{2m}\|Y\|_F^2 - \frac{1}{m} Y \bullet X\beta + \left(\frac{1}{2m}X^TX + \frac{1}{2\gamma}I\right) \bullet \theta + \mu \cdot \mathrm{tr}\left(W\right) \\ \text{s.t.} & W \preceq I \\ & \begin{pmatrix} \theta & \beta \\ \beta^T & W \end{pmatrix} \succeq 0 \end{array}$$

and "NN" (nuclear norm relaxation)

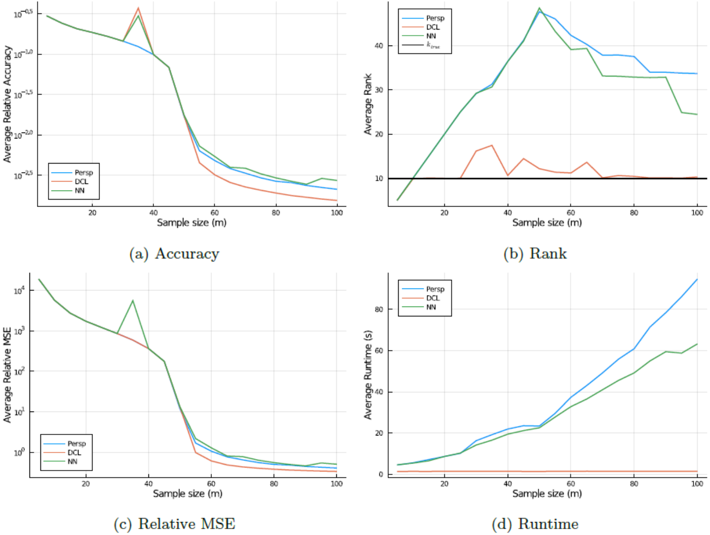
$$\min_{\beta \in \mathbb{R}^{p\times n}} \frac{1}{2m}\|Y - X\beta\|_F^2 + \frac{1}{2\gamma}\|\beta\|_F^2 + \mu \cdot \|\beta\|_*$$

[3] Hongbo Dong, Kun Chen, and Jeff Linderoth. "Regularization vs. relaxation: A conic optimization perspective of statistical variable selection". In: *arXiv preprint arXiv:1510.06083* (2015).

Reduced Rank Regression

- Goal: recover $\mathrm{rank}\text{-}k_{true}$ solution $\beta_{true} = UV^T$
- Data generation:
  - $U \in \mathbb{R}^{p \times k_{true}}, V \in \mathbb{R}^{k_{true} \times n}, X \in \mathbb{R}^{m \times p} \sim$ i.i.d. $\mathcal{N}(0, 1)$.
  - $Y = X\beta_{true} + E$ where $E_{i,j} \sim \mathcal{N}(0, \sigma)$
  - $n = p = 50$; $k = 10$, $\gamma = 10^{-6}$, $\sigma = 0.05$ and vary m
  - Cross validation to select $\mu \in [10^{-4}, 10^4]$
- Performance measure
  - Relative accuracy $\|\beta_{true} - \beta_{est}\|_F / \|\beta_{est}\|_F$
  - The rank: number of singular values $\sigma_i(\beta_{est}) > 10^{-4}$
  - Out-of-sample MSE: $\|X_{new}\beta_{est} - Y_{new}\|_F^2$ (normalized by $\|X_{new}\beta_{true} - Y_{new}\|_F^2$)
- Results are averaged over 100 random instances
- `Julia 1.5` using `JuMP.jl 0.21.6` and `Mosek 9.1`
- Code: github.com/ryancorywright/MatrixPerspectiveSoftware.

Reduced Rank Regression



(a) Accuracy (b) Rank

(c) Relative MSE (d) Runtime

Fig. 1: Comparative performance, as the number of samples m increases, of formulations (6) (Persp, in blue), (7) (DCL, in orange) and (28) (NN, in green), averaged over 100 synthetic reduced rank regression instances where $n = p = 50$, $k_{true} = 10$. The hyperparameter μ was first

Optimal Experimental Design

Compare three approaches:

- Low-rank relaxation + rounding the largest k z_i's

$$\begin{array}{ll} \max\limits_{\theta, Y \in \mathcal{S}^m_+, z} & \mathrm{tr}\,(\theta) + \mathrm{tr}\,(I - Y)\log\epsilon \\ \text{s.t.} & (-\theta, Y, \epsilon I + A\mathrm{Diag}(z)A^T) \in \mathcal{K}^{op,rel}_{mat} \\ & z \in [0,1]^n, e^T z \leq k \\ & Y \preceq I, \mathrm{tr}\,(Y) \leq k \end{array}$$

- Cardinality relaxation + rounding the largest k z_i's

$$\max_{z \in [0,1]^n, e^T z \leq k} \log\det\left(\epsilon I + \sum_{i=1}^{n} z_i a_i a_i^T\right)$$

- Greedy heuristic

Algorithm 1: Greedy submodular maximization

Initialize $\mathcal{I} = \emptyset$;
for $i = 1, 2, 3, ..., k$ **do**
  Let $j^* \leftarrow \arg\max_{j \in [n] \setminus \mathcal{I}} \log\det\left(\epsilon I + \sum_{i \in \mathcal{I} \cup \{j\}} z_i a_i a_i^T\right)$;
  $\mathcal{I} \leftarrow \mathcal{I} \cup \{j^*\}$.

Optimal Experimental Design

Compare three approaches:

- Data generation:
  - $m = 10$, $n = 20$, $A \in \mathbb{R}^{m \times n} \sim$ i.i.d. $\mathcal{N}(0, 1/\sqrt{n})$.
  - $\epsilon = 10^{-6}$
  - Vary $k < m$ over 20 random instances.
- Software: `CVX 1.22`, `Matlab R2021a`, `Mosek 9.1`, `CVXQuad`.

Optimal Experimental Design

Compare three approaches:
Performance measure: Average run time and Relative optimality gap

	Problem (23)+round		Submodular		Problem (24)+round	
k	Time(s)	Gap (%)	Time(s)	Gap (%)	Time(s)	Gap (%)
1	0.52	88.8	0.00	88.9	347.0	0.00
2	0.63	93.7	0.00	93.7	338.5	0.01
3	0.59	97.1	0.00	97.0	320.8	0.06
4	0.63	100.2	0.00	100.2	338.7	0.18
5	0.53	103.8	0.00	103.9	331.1	0.37
6	0.53	109.0	0.00	109.0	287.5	1.40
7	0.55	117.7	0.00	117.7	255.1	2.39
8	0.60	136.9	0.00	138.5	236.1	5.25
9	0.54	260.9	0.00	287.5	235.9	28.43

Conclusion

- MPRT for obtaining good relaxation of low-rank problems
- Characterization of the convex hull of epigraphs of various matrix function under low-rank constraints.
- New problems arise such as non-symmetric cone optimization involving von Neumann entropy