Matrix Rank Minimization with Applications

Maryam Fazel

Information Systems Lab
Electrical Engineering Department
Stanford University

Outline

1. **Problem statement and motivation**

2. Existing approaches

3. New approaches and results:

3.1. Trace and log-det heuristics for PSD matrices
3.2. Semidefinite embedding
3.3. Trace and log-det heuristics for general matrices

4. Applications

Rank Minimization Problem (RMP)

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, X \\ \text{subject to} & X \in \mathcal{C}, \end{array}$$

$X \in \mathbf{R}^{m\times n}$ is the optimization variable, $\mathcal{C}$ is a convex set

- RMP is difficult **non-convex** problem (NP-hard)
- RMP arises in many application areas
- usual meaning: find simplest or minimum order system; model with fewest parameters or least complexity ... (Occam's razor)

this talk: new methods to (approximately) solve RMP

Example 1: constrained factor analysis

goal: find lowest rank covariance matrix close to measured (noisy) $\hat{\Sigma}$, satisfying prior info

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, \Sigma \\ \text{subject to} & \|\Sigma - \hat{\Sigma}\| \leq \epsilon \\ & \Sigma \geq 0 \\ & \Sigma \in \mathcal{C}, \end{array}$$

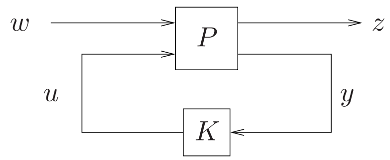
where $\Sigma \in \mathbf{R}^{n\times n}$, $\hat{\Sigma}$ is measured covariance, ϵ is tolerance, $\mathcal{C}$ denotes assumptions or prior information on Σ

$\mathbf{Rank}\, \Sigma$ corresponds to the number of factors that explain Σ

without last constraint, can solve by eigenvalue decomposition (EV); but in general, a difficult problem

Example 2: reduced order controller design

goal: given system P, design stabilizing controller K with lowest possible order to achieve specified performance



in practical implementation, lower order controller means savings in computation time and memory usage

problem can be expressed as RMP (details later)

Example 3: minimum order system realization

find minimum order system that satisfies time-domain specs (rise-time, slew-rate, overshoot, settling-time, ...)

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank} \begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & h_3 & \cdots & h_{n+1} \\ \vdots & \vdots & & \vdots \\ h_n & h_{n+1} & \cdots & h_{2n-1} \end{bmatrix} \\ \text{subject to} & F(h_1, \ldots, h_n) \preceq g \end{array}$$

variables are impulse response $h_1, \ldots, h_{2n-1}$;
linear inequality constraints involve only $h_1, \ldots, h_n$

Special case: Maximum sparsity problems

suppose $X = \mathbf{diag}(x) = \begin{bmatrix} x_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & x_n \end{bmatrix}, \quad x \in \mathbf{R}^n,$

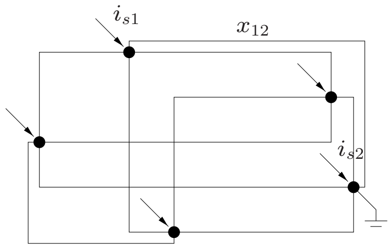
then $\mathbf{Rank}\, X = \mathbf{Card}(x)$, number of nonzero x_i

RMP reduces to finding the **sparsest** vector in convex set $\mathcal{C}$:

$$\begin{array}{ll} \text{minimize} & \mathbf{Card}(x) \\ \text{subject to} & x \in \mathcal{C}. \end{array}$$

meaning: find simplest model, design with fewest components, sparse signal representation; extract sparse signals from noise, ... (same theme as before)

Example: sparse power-ground mesh design



- n nodes, some grounded, connected by full graph of wires, i_{sk}= current injected to node k
- choose wire widths x_{ij} between nodes i & j subject to convex constraint on x ($e.g.$, $0 \leq x_{ij} \leq W_{\max}$, low power dissipation)

$x_{ij} = 0$ means no wire between i & j, thus **sparse** mesh

Outline

1. Problem statement & motivation

2. **Existing approaches**

3. New approaches and results:

3.1. Trace and log-det heuristics for PSD matrices
3.2. Semidefinite embedding
3.3. Trace and log-det heuristics for general matrices

4. Applications

Exact solution methods

- special cases with analytical solutions ($e.g.$, via SVD, EV)

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, X \\ \text{subject to} & \|X - A\|_{2,F} \le b \end{array}$$

  - solution: sum of r leading dyads in SVD of A, where $\sigma_r > b$, $\sigma_{r+1} \le b$
  - applications: factor analysis, MDS, array signal processing, sub-space based methods, ...
- special cases that can be reduced to convex problems ($e.g.$, [Mesbahi '97])

- global optimization ($e.g.$, branch and bound)

Heuristic solution methods

- Factorization methods ($e.g.$, [Iwasaki '99])

  $\mathbf{Rank}\, X \le r$ iff there are $F \in \mathbf{R}^{m \times r}$, $G \in \mathbf{R}^{r \times n}$ s.t. $X = FG$

  solve feasibility problem with variables F, G; use bisection on r

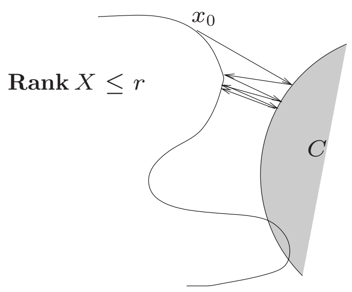
- Analytic anti-centering [David '94]

  use Newton method in reverse to *maximize* log barrier or potential function

- Alternating projections [Grigoriadis & Beran '00]

  alternate between

  - projecting onto constraint set $\mathcal{C}$ via convex optimization
  - projecting onto set of matrices of rank r via SVD



Our goal

Provide methods to (approximately) solve RMP that

- are **general:** handle any matrix, any convex constraint
- do not require user to supply an initial point
- are numerically **efficient** ($e.g.$, based on solving convex problems)
- have supporting **theory**
- are **effective** in practice

Outline

1. Problem statement and motivation
2. Existing approaches
3. **New approaches and results:**

 3.1. **Trace and log-det heuristics for PSD matrices**
 3.2. Semidefinite embedding
 3.3. Trace and log-det heuristics for general matrices

4. Applications

Trace heuristic for PSD matrices

observation: for $X = X^T \geq 0$, minimizing trace tends to give low-rank solution [Mesbahi '97, Pare '00]

RMP:	Trace heuristic:
minimize $\;\mathbf{Rank}\, X$ subject to $\;X \in \mathcal{C}$	minimize $\;\mathbf{Tr}\, X$ subject to $\;X \in \mathcal{C}$

- **convex** problem, hence efficiently solved, no initial point needed
- if feasible set bounded, provides **lower bound** on objective
- variation: weighted trace minimization

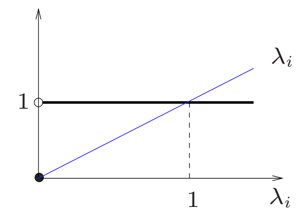
$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\, WX \\ \text{subject to} & X \in \mathcal{C}, \end{array}$$

where $W = W^T > 0$.

intuition: replace $\mathbf{Rank}\, X$ with convex function $\mathbf{Tr}\, X$

- note $\mathbf{Rank}\, X$= (# of non-zero λ_i's), $\quad \mathbf{Tr}\, X = \sum_i \lambda_i$

  where λ_is are eigenvalues of X.



- if $\lambda_i \leq 1$, $\mathbf{Tr}\, X \leq \mathbf{Rank}\, X$.

method applicable only when $X = X^T \geq 0$

Log-det heuristic for PSD matrices

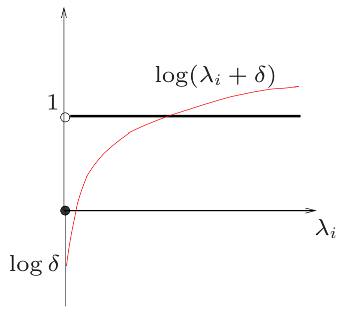
suggested heuristic for $X = X^T \geq 0$:
($\delta > 0$ is small constant for regularization)

RMP:		Log-det heuristic:	
minimize	$\mathbf{Rank}\, X$	minimize	$\log\det(X + \delta I)$
subject to	$X \in \mathcal{C}$	subject to	$X \in \mathcal{C}$

- objective is **non-convex** (in fact, concave)
- can use any local optimization method to find a local minimum

intuition: note $\mathbf{Rank}\, X$= (# of non-zero λ_i's),

$$\log\det(X + \delta I) = \log \prod_i (\lambda_i + \delta) = \sum_i \log(\lambda_i + \delta)$$



again, method applicable only when $X = X^T \geq 0$

Iterative linearization method

linearize (concave) objective at $X_k \succ 0$:

$$\log\det(X + \delta I) \approx \log\det(X_k + \delta I) \; + \; \mathbf{Tr}(X_k + \delta I)^{-1}(X - X_k)$$

minimize linearized objective (a convex problem):

$$X_{k+1} = \underset{X \in \mathcal{C}}{\operatorname{argmin}} \, \mathbf{Tr}(X_k + \delta I)^{-1} X$$

$i.e.$, iterative weighted trace minimization

- can show iterations converge to a local minimum of $\log\det(X + \delta I)$ (using concavity of $\log\det(X + \delta I)$ in X)
- with $X_0 = I$, first iteration same as trace heuristic
- typically, only a few iterations needed (about 5 or 6)

Outline

1. Problem statement & motivation

2. Existing approaches

3. **New approaches and results:**

 3.1. Trace and log-det heuristics for PSD matrices
 3.2. **Semidefinite embedding**
 3.3. Trace and log-det heuristics for general matrices

4. Applications

Semidefinite embedding

question: can we extend last two heuristics to general matrices?

Lemma 1. *Let $X \in \mathbf{R}^{m \times n}$. Then $\mathbf{Rank}\, X \leq r$ iff $\exists\ Y = Y^T \in \mathbf{R}^{m \times m}$ and $Z = Z^T \in \mathbf{R}^{n \times n}$ s.t.*

$$\mathbf{Rank} \begin{bmatrix} Y & 0 \\ 0 & Z \end{bmatrix} \leq 2r, \quad \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0.$$

conclusion: can relate the rank of a general matrix to that of a PSD (symmetric) matrix

thus, can **embed** general rank problem in a larger one involving only PSD matrices

Proof: ($\Rightarrow$) suppose $\mathbf{Rank}\, X = r_0 \leq r$. then $X = L\, R$, with $L \in \mathbf{R}^{m \times r_0}$, $R \in \mathbf{R}^{r_0 \times n}$, and $\mathbf{Rank}\, L = \mathbf{Rank}\, R = r_0$.

setting $Y = L\, L^T$ and $Z = R^T\, R$ satisfies conditions in lemma, since

$$\begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} = \begin{bmatrix} L \\ R^T \end{bmatrix} \begin{bmatrix} L^T & R \end{bmatrix} \geq 0.$$

($\Leftarrow$) w.l.o.g, suppose $\mathbf{Rank}\, Y \leq \mathbf{Rank}\, Z$. by Schur complements

$$\begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \quad \Leftrightarrow \quad \begin{cases} \text{(i)} & Y \geq 0 \\ \text{(ii)} & X^T(I - YY^\dagger) = 0 \\ \text{(iii)} & Z - X^T Y^\dagger X \geq 0 \end{cases}.$$

from (ii), and using $\mathbf{Rank}\, X = n - \dim \mathcal{N}(X) = m - \dim \mathcal{N}(X^T)$ follows

$$\begin{array}{lll} \mathcal{N}(X^T) \supseteq \mathcal{N}(Y) & \Rightarrow & \dim \mathcal{N}(X^T) \geq \dim \mathcal{N}(Y), \\ & \Rightarrow & \mathbf{Rank}\, Y \geq \mathbf{Rank}\, X^T = \mathbf{Rank}\, X, \end{array}$$

thus, $2\,\mathbf{Rank}\, X \leq \mathbf{Rank}\, Y + \mathbf{Rank}\, Z$, or $\mathbf{Rank}\, X \leq r$. $\square$

RMP in embedded PSD form

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, X \\ \text{subject to} & X \in \mathcal{C}, \end{array}$$

equivalent to PSD RMP

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank} \begin{bmatrix} Y & 0 \\ 0 & Z \end{bmatrix} \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \\ & X \in \mathcal{C}, \end{array}$$

with variables $X \in \mathbf{R}^{m \times n}$, $Y \in \mathbf{R}^{m \times m}$ and $Z \in \mathbf{R}^{n \times n}$.

- X is **general**; Y and Z are symmetric PSD
- any method for minimizing rank of PSD matrices applies
- can exploit structure of X ($e.g.$, block-diagonal, symmetric) to reduce number of variables

Outline

1. Problem statement & motivation
2. Existing approaches
3. **New approaches and results:**

   3.1. Trace and log-det heuristics for PSD matrices
   3.2. Semidefinite embedding
   3.3. **Trace and log-det heuristics for general matrices**

4. Applications

Trace heuristic for general matrices

applying trace heuristic to PSD form of RMP yields

$$\begin{array}{ll} \text{minimize} & \mathbf{Tr}\begin{bmatrix} Y & 0 \\ 0 & Z \end{bmatrix} \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \geq 0 \\ & X \in \mathcal{C}, \end{array}$$

can show this is equivalent to

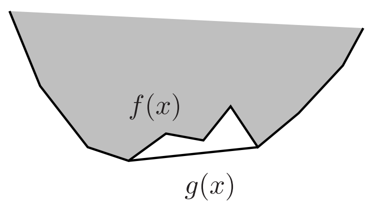
$$\begin{array}{ll} \text{minimize} & \|X\|_* \\ \text{subject to} & X \in \mathcal{C}, \end{array}$$

where $\|X\|_* = \sum_i \sigma_i(X)$, called **nuclear norm** of X, is dual of the spectral (or maximum singular value) norm.

question: how is $\|X\|_*$ related to $\mathbf{Rank}\, X$?

Convex envelope

definition: convex envelope of $f : \mathcal{C} \to \mathbf{R}$, $\mathcal{C} \subseteq \mathbf{R}^n$, is largest convex function g s.t. $g(x) \leq f(x)$ for all $x \in \mathcal{C}$.



- 'best' convex lower approximation
- epigraph of g is convex hull of epigraph of f

Convex envelope of rank

Lemma 2. *$\|X\|_*$ is the convex envelope of $\mathbf{Rank}\, X$ on the set $\{X \in \mathbf{R}^{m\times n} \mid \|X\| \leq 1\}$.*

conclusions:

- trace heuristic minimizes the **convex envelope** of rank ($i.e.$, the *best* convex approximation to rank) over unit ball
- hence, heuristic provides lower bound on objective

lemma provides theoretical support for use of trace/nuclear norm heuristic

Special case: Maximum sparsity problem

suppose $X = \mathbf{diag}(x)$, $x \in \mathbf{R}^n$, applying trace heuristic yields

$$\begin{array}{ll} \text{minimize} & \|x\|_1 \\ \text{subject to} & x \in \mathcal{C}. \end{array}$$

- well-known ℓ_1 heuristic for finding sparse solutions
- used in LASSO methods in statistics [Tibshirani '94], signal decomposition by basis pursuit [Donoho '96], ...
- $\|x\|_1$ is convex envelope of $\mathbf{Card}(x)$ over $\{x \mid \|x\|_\infty \leq 1\}$
- thus, trace/nuclear norm heuristic can be considered an extension of ℓ_1 heuristic to matrix case

Log-det heuristic for general matrices

applying log-det heuristic to PSD form of RMP yields

$$\begin{array}{ll} \text{minimize} & \log\det\left(\begin{bmatrix} Y & 0 \\ 0 & Z \end{bmatrix} + \delta I\right) \\ \text{subject to} & \begin{bmatrix} Y & X \\ X^T & Z \end{bmatrix} \succeq 0 \\ & X \in \mathcal{C}. \end{array}$$

can show this is equivalent to

$$\begin{array}{ll} \text{minimize} & \sum_i \log(\sigma_i(X) + \delta) \\ \text{subject to} & X \in \mathcal{C}, \end{array}$$

can iteratively linearize as before to obtain iterations in X, Y, Z.

Iterative ℓ_1 heuristic for maximum sparsity problem

suppose $X = \mathbf{diag}(x)$, $x \in \mathbf{R}^n$, applying log-det heuristic yields

$$\begin{array}{ll} \text{minimize} & \sum_i \log(|x_i| + \delta) \\ \text{subject to} & x \in \mathcal{C}. \end{array}$$

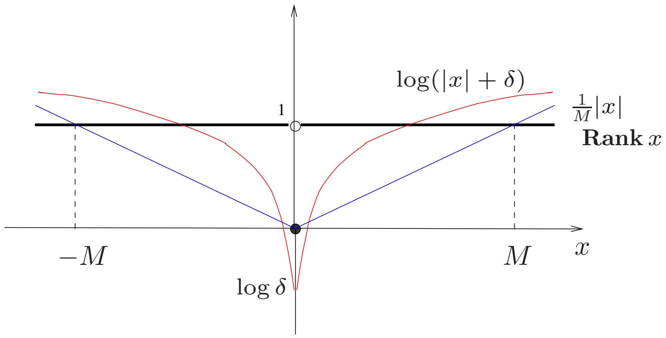
iterative linearization and minimization yields

$$\begin{aligned} x^{(k+1)} &= \underset{x \in \mathcal{C}}{\operatorname{argmin}} \sum_i w_i^{(k)} |x_i| \\ w_i^{(k)} &= \frac{1}{|x_i^{(k)}| + \delta}. \end{aligned}$$

- each step is **weighted ℓ_1 norm** minimization
- **interpretation:** if $x_i^{(k)}$ is small, its weight in next minimization step is large, so small entries in x are generally pushed towards zero as far as constraints allow

An interpretation

objectives for the two heuristics plotted for scalar case, $i.e.$, $X = x \in \mathbf{R}$.



- in this case $\sigma(X) = |x|$
- note $\frac{1}{M}|x|$ is convex envelope of $\mathbf{Rank}\, x$ on $\{x \mid |x| \leq M\}$.

Outline

1. Problem statement and motivation
2. Existing approaches
3. New approaches and results:

   3.1. Trace and log-det heuristics for PSD matrices
   3.2. Semidefinite embedding
   3.3. Trace and log-det heuristics for general matrices

4. **Applications**
   - System realization with time-domain constraints
   - Reduced order controller design
   - Euclidean distance matrix problems
   - Portfolio optimization with fixed transaction costs

System realization with time-domain constraints

goal: find minimum order system whose step response up to nth sample satisfies specs, $e.g.$, certain rise-time, slew-rate, overshoot, settling characteristics and given sample delay ...

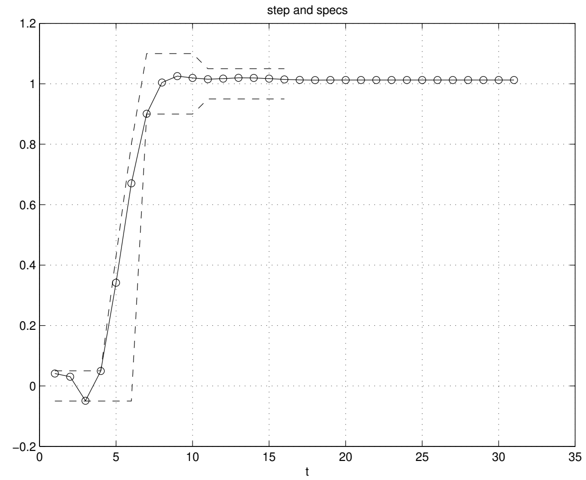
can express as

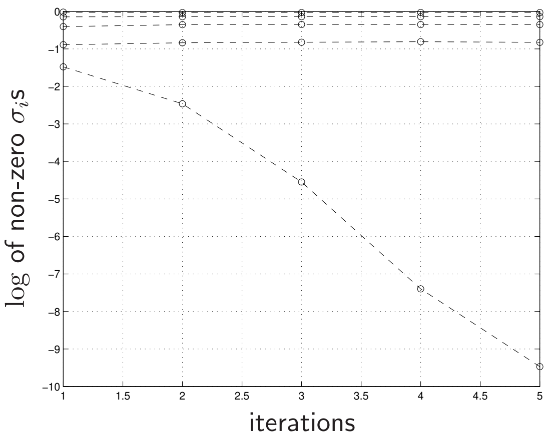
$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, H_n \\ \text{subject to} & l_i \leq s_i \leq u_i, \quad k = 1, \ldots, n \\ & h_{n+1}, \ldots, h_{2n-1} \in \mathbf{R}, \end{array}$$

where h_is are impulse response, $s_k = \sum_{i=1}^{k} h_i$ are step response, l_i, u_i are samples of lower and upper time domain specs, and H_n is hankel matrix

$$H_n = \begin{bmatrix} h_1 & h_2 & \ldots & h_n \\ h_2 & h_3 & \ldots & h_{n+1} \\ \vdots & \vdots & & \vdots \\ h_n & h_{n+1} & \ldots & h_{2n-1} \end{bmatrix}.$$

example: $n = 16$, specs denoted by dashed region

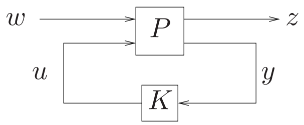




- conclude that specs can be met by an LTI system of order four

Reduced order controller design

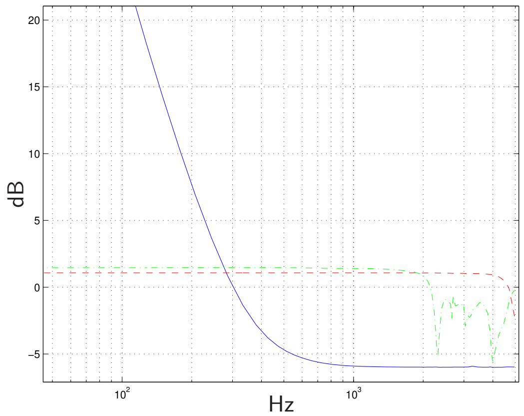
goal: given system P, design stabilizing controller K with lowest possible order to achieve specified performance ($e.g.$, closed loop l_2-gain from w to z less than γ)



- $x \in \mathbf{R}^n$ state, w and u disturbance and control inputs, z and y performance and measured outputs
- known: there exists stabilizing controller K of order $n_K \leq n$, achieving γ iff there exist PSD matrices $R, S \in \mathbf{R}^{n\times n}$ satisfying certain LMIs.

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank} \begin{bmatrix} R & I \\ I & S \end{bmatrix} \\ \text{subject to} & \begin{bmatrix} R & I \\ I & S \end{bmatrix} \geq 0 \\ & \text{other LMIs in } R \text{ and } S. \end{array}$$

example: MIMO flexible positioning mechanism P of order 29, 5 inputs, 3 outputs, with γ within 5% of its minimum achievable value



Maximum singular value plots: open-loop system (blue) closed-loop system with 29th order controller corresponding to $\gamma_{\rm opt}$ (red), closed loop system with 20th order controller with 5% relaxation of $\gamma_{\rm opt}$ (green)

- 20th order controller obtained using log-det heuristic, converged in 5 iterations
- controller achieves same performance as full-order one within less than 0.5dB, with 30% reduction in number of states
- leads to savings in computation time and memory usage

Euclidean distance matrix (EDM) with smallest embedding dimension

construct configurations of points from information about interpoint distances

- D is EDM iff $\exists\ x_1, \ldots, x_n \in \mathbf{R}^r$ s.t. $D_{ij} = \|x_i - x_j\|^2$
- minimum possible r called **embedding dimension**
- [Schoenberg '35] D is EDM with embedding dimension r iff
  - $D_{ii} = 0$,
  - $VDV \le 0$, where $V = I - \frac{1}{n}\mathbf{1}\mathbf{1}^T$,
  - $\mathbf{Rank}\, VDV \le r$.

  first two conditions imply distance properties:
  $D_{ii} = 0, \quad D_{ij} = D_{ji} \ge 0, \quad D_{ik} + D_{ki} \ge D_{ij}.$

goal: given upper & lower bounds on distances D_{ij}, find EDM D with smallest embedding dimension

$$\begin{array}{ll} \text{minimize} & \mathbf{Rank}\, VDV \\ \text{subject to} & D_{ii} = 0 \qquad i = 1, \ldots, n \\ & -VDV \ge 0 \\ & L_{ij} \le D_{ij} \le U_{ij} \qquad i,j = 1, \ldots, n. \end{array}$$

applications: statistics/psychometrics (multidimensional scaling), chemistry (molecular conformation)

example: given pairwise distances of 40 points in $\mathbf{R}^6$, perturbed by noise up to 20%, find smallest embedding dimension

EDM with embedding dimension 6 found using trace heuristic ($i.e.$, 1 iteration of log-det)

Portfolio optimization with fixed transaction costs

portfolio with n assets is to be adjusted by performing transactions, after which it will be held over a fixed period

goal: maximize expected final wealth

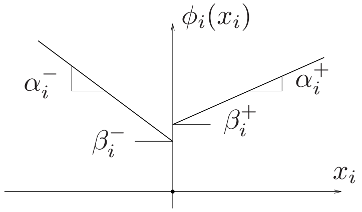
$$\begin{array}{ll} \text{maximize} & \bar{a}^T(w+x) \\ \text{subject to} & \mathbf{1}^T x + \phi(x) \leq 0 \\ & w + x \in \mathcal{S} \end{array}$$

where $w \in \mathbf{R}^n$ denotes current holdings, $x \in \mathbf{R}^n$ amounts transacted, $\bar{a} \in \mathbf{R}^n$ expected return, $\phi : \mathbf{R}^n \to \mathbf{R}_+$ transaction costs, $W = a^T(w+x)$ is end of period wealth

- $\mathcal{S}$ denotes other convex constraints on portfolio, *e.g.*,

  bound on risk $\mathbf{E}\,(W - \mathbf{E}W)^2 = (w+x)^T\Sigma\,(w+x) \leq \sigma_{\max}^2$

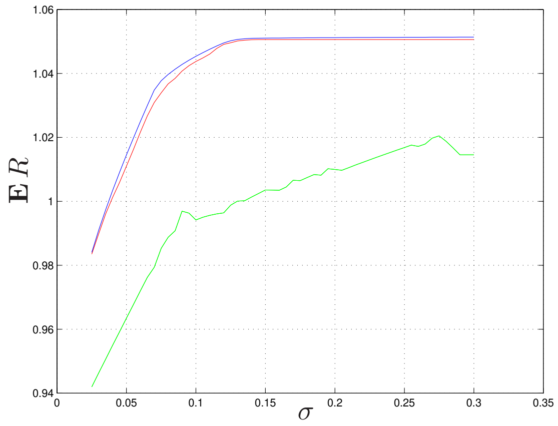
- transaction costs: $\phi(x) = \sum_{i=1}^n \phi_i(x_i)$, with



  (fixed plus linear cost model)

- seek **sparse** vector of transactions due to fixed transaction costs

example: 100 stocks plus a riskless asset with constraints on shorting



return vs risk tradeoff curve: global upper bound from ℓ_1 heuristic (top), lower bound from solution obtained by iterative ℓ_1 (middle), and solution computed without regard for fixed costs (bottom)

Conclusions

- RMP is difficult nonconvex problem, with many applications
- maximum sparsity problem is special case
- general RMP can be reduced to a PSD one, via semidefinite embedding
- trace and log-det heuristics for PSD RMP can be extended to general case
- generalization of **trace** heuristic to general matrices is **nuclear norm**, which is convex envelope of rank