03/06/18. Rank regularization for $A(X) = b$, $X \in \mathbb{R}^{m \times n}$ and its convex relaxation.

Ex. Recommender systems / collaborative filtering

User \ Product	1	2	3	4	...
1		5	1		
2	3		1	5	...
3		4	?	2	
4	5		3		
⋮					

→ matrix completion.

One idea (not the only one): the vector space of all user rating vectors has small dimension → every user's rating vector is a linear combination of a few representative ones. (only so many types of users).

→ $\min \operatorname{rank}(X) : \underbrace{A(X) = y}_{\text{restriction}}$

$X = U \operatorname{diag}(\sigma) V^* \Rightarrow \operatorname{rank}(X) = \|\sigma\|_{\ell_0}$

→ combinatorially hard

Def. (Schatten p-norms).
$\|X\|_p = \|\sigma\|_p$ where $X = U \operatorname{diag}(\sigma) V^*$

Ex. $p = 2$: $\|X\|_2 = \sqrt{\sum_i \sigma_i^2} = \sqrt{\sum_{ij} X_{ij}^2}$
Frobenius norm.

induced 2-norm

$p=\infty$: $\|X\|_\infty = \sigma_1 = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|}$

Spectral/operator norm, also $\|X\|$.
⚠ often denoted $\|X\|_2$
in non-optimization contexts.

$p=1$: $\|X\|_1 = \sum \sigma_i$

Nuclear norm.
Also denoted $\|X\|_*$, because it is dual to $\|X\| = \|X\|_\infty$.

Def. (Trace inner product)

$$\langle X, Y\rangle = tr(X^T Y) = tr(X Y^T) = \sum X_{ij} Y_{ij}.$$

Thm (Von Neumann's trace inequality)

$$|\langle X, Y\rangle| \leq \sum_i \sigma_i(X)\sigma_i(Y)$$

and equality iff X, Y share U, V^*.

Def. (Dual norm)

$$\|Y\|_* = \sup\{\langle X, Y\rangle : \|X\| \leq 1\}.$$

Ex 2 dual of 2
1 dual of ∞
∞ dual of 1

Def. (Convex conjugate)

$$f^*(Y) = \sup_X \langle X, Y\rangle - f(X).$$

Thm (Fazel, 2002).

The convex envelope (conv. conj. of conv. conj.) of rank(X), over the set $\|X\| \leq 1$, is $\|X\|_1$.

→ Nuclear norm regularization

$$\min \|X\|_* \text{ s.t. } A(X) = y. \qquad (NN)$$

For matrix completion: Candès, Recht 2009

Rmk When $X = X^T$, $X \succeq 0$ (positive semidefinite), then $\|X\|_* = \mathrm{tr}(X) = \sum \lambda_i$, so better to consider

$$\min \mathrm{tr}(X) \text{ s.t. } A(X) = b, \quad X \succeq 0.$$

Rmk. Can still use semidefinite programming (SDP) in general: (NN) is equivalent to

$$\min \mathrm{tr}\, Y : \quad Y = \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$

$$A(X) = y$$

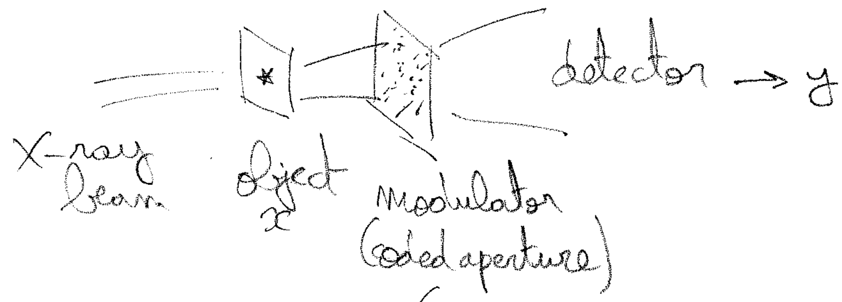
linear matrix inequality

W_1, W_2: slack variables

(minimizer has the form $\begin{bmatrix} U\Sigma U^* & U\Sigma V^* \\ V\Sigma U^* & V\Sigma V^* \end{bmatrix}$)

→ shows yet another way that $\min \|X\|_* : A(X) = b$ is convex.

Ex. Phase retrieval



$$y_{ik} = |(F D_k x)_i| \qquad i \in \text{3D grid}$$

Fourier transform

Intensity measurement, complex modulus.

$\Rightarrow \quad y_j = |a_j^T x| \qquad$ for some a_j

$j = 1, \dots, m > n = \text{length}(x)$

Conventional idea: let $c_j = a_j^T x \iff c = Ax$

Initial guess $c^{(0)}$, then

(1) $\tilde{c} = P_{\text{Ran}A}\, c^{(k)}$

(2) $c_j^{(k+1)} = \dfrac{\tilde{c}_j}{|\tilde{c}_j|}\, y_j \qquad$, repeat.

Not guaranteed to converge; may stagnate, may get stuck.

Convexification by <u>lifting</u>:

$$X = xx^T,$$

$$y_j^2 = |a_j^T x|^2 = \mathrm{tr}(a_j a_j^T x x^T)$$
$$= \mathrm{tr}(A_j X) \quad \text{with } A_j = a_j a_j^T$$
$$= \langle A_j, X \rangle$$

and $\mathrm{rank}(X) = 1$, $\quad X \succcurlyeq 0$.

$\rightarrow$ $\min \mathrm{rank}(X)$, $\quad X \succcurlyeq 0$, $\quad \langle A_j, X \rangle = y_j^2$.

Relax to a SDP (convex):

$$\min \mathrm{tr}(X), \quad X \succcurlyeq 0, \quad \langle A_j, X \rangle = y_j^2$$

(Candès et al. 2013)

$\rightarrow$ no more problem of dubious convergence, but the relaxation should be tight (no indeterminacy).

then extract x via $Xv = \lambda v$ (top),
$x = \sqrt{\lambda}\, v$.