Why is minimizing the nuclear norm of a matrix a good surrogate for minimizing the rank?

A method called "Robust PCA" solves the matrix decomposition problem

$$L^*, S^* = \arg\min_{L,S} \|L\|_* + \|S\|_1 \quad \text{s.t. } L + S = X$$

as a surrogate for the actual problem

$$L^*, S^* = \arg\min_{L,S} rank(L) + \|S\|_0 \quad \text{s.t. } L + S = X,$$

i.e. the actual goal is to decompose the data matrix X into a low-rank signal matrix L and a sparse noise matrix S. In this context: **why is the nuclear norm a good approximation for the rank of a matrix?** I can think of matrices with low nuclear norm but high rank and vice-versa. Is there any intuition one can appeal to?

(linear-algebra) (matrices) (norm)

edited Jun 29 '12 at 7:26

asked Jun 28 '12 at 14:44

blubb
181 1 8

2 Consider this question: Why is the ℓ_1 norm a surrogate for ℓ_0? – p.s. Jun 28 '12 at 16:33

You write "low-rank" but the formula has $\arg\max$... – user31373 Jun 28 '12 at 23:01

@LeonidKovalev: Oops. Of course I meant $\min$. – blubb Jun 29 '12 at 7:26

4 Answers

Why does compressed sensing work? Because the ℓ_1 ball in high dimensions is extremely "pointy" -- the extreme values of a linear function on this ball are very likely to be attained on the faces of low dimensions, those that consist of sparse vectors. When applied to matrices, the sparseness of the set of eigenvalues means low rank, as @mrig wrote before me.

answered Jun 28 '12 at 23:05

user31373

Thank you, the reduction to sparse coding of eigenvalues was indeed the link I was looking for! – blubb Jun 29 '12 at 7:39

The nuclear norm can be thought of as a convex relaxation of the number of non-zero eigenvalues (i.e. the rank).

answered Jun 28 '12 at 14:48

mrig
161 2

Thanks for asking this question. I have problems understanding the same thing. Particularly as follows.

Although the nuclear norm is a relaxation for the rank, I don't understand how the nuclear norm can work as surrogate for minimising rank unless it is an upper bound for the rank.

As was pointed out in the original post, the nuclear norm does not appear to provide a bound (lower or upper) to the matrix rank. For e.g. consider the following.

```
1)
A: #<BASIC-MATRIX (2 2)      2.00000      0.00000 \\
                             0.00000      2.00000 >
 (RANK A): 2
 (NUCLEAR-NORM A): 4.0d0


2)
 A: #<BASIC-MATRIX (2 2)      0.10000      0.00000 \\
                              0.00000      0.10000 >
 (RANK A): 2
 (NUCLEAR-NORM A): 0.20000000298023224d0
```

In 1) the nuclear norm is larger than the rank while in 2) it is lower than the rank.

answered Nov 28 '14 at 0:48

dingo
11 3

It's not that you want something that's actually a lower or upper bound for your quantity of interest, it's that you want something that grows bigger/smaller together with it. For instance, `sqrt(x)` can be either smaller or bigger than `x`, but if you want to minimize `sqrt(x)`, you can perhaps more easily minimize `x`. – Ken Williams Aug 13 '15 at 19:06

Welcome to Math.SE. I think you should ask a separate question and link to this question. That way, you'll get more attention and people are not forced to answer in a comment. :) – blubb Feb 19 at 6:17

A nuclear norm of a matrix is equivalent to the L1-norm of the vector of its eigenvalues. Thus, you are injecting sparsity to the vector of eigenvalues. Essentially, this sparsity means you are reducing the rank of the original matrix.

answered Feb 19 at 4:56

DebangaRaj
11 1