Rank Deficiency and Ill-Conditioned Least Square Problems

Edward Kang

Abstract

In least squares problems, ill-conditioning happens when explanatory factors are correlated for regression problems. When there is a lot of correlation, we may have a rank deficiency problem. There are generally three approaches to solving rank deficiency problem: regularization, parameter subset selection, and principal components analysis. We are going to focus on the topic of the singular value decomposition.

Introduction

We always have to deal with rank-deficient matrices that are corrupted by noise. When our matrix comes from some kind of measurement that has uncertainty associated with it, it will be quite far from rank-deficient even if the underlying "true" matrix is rank-deficient.

SVD is a good tool for dealing with numerical rank issues, and we want to show that if a singular value is small then the norm could be large.

Definitions

Least squares problems are a special sort of minimization. Suppose $A \in R^{m\times n}$ and $m > n$. In general, we will not be able to exactly solve overdetermined equations $Ax = b$; the best we can do is to minimize the residual $r = b - Ax$. In least squares problems, we minimize the two-norm of the residual:

$$\text{minimize } \|Ax - b\|_2^2$$

However, in regression problems, the columns of A correspond to explanatory factors and ill-conditioning happens when the explanatory factors are correlated. When there is a lot of correlation and the columns of A are truly linearly dependent, or when A is contaminated by enough noise that a moderate correlation seems dangerous, then we may declare that we have a rank-deficient problem.

The singular value decomposition (SVD) is important for solving least squares problems and for a variety of other approximation tasks in linear algebra. For $A \in R^{m\times n2}$, we write

$$A = U\Sigma V^T$$

where $U \in R^{m\times m}$ and $V \in R^{n\times n}$are orthogonal matrices and $\Sigma \in R^{m\times n}$ is diagonal. The diagonal matrix Σ has non-negative diagonal entries $\sigma 1 \geq \sigma 2 \geq \ldots \geq \sigma n \geq 0$. The σi are called the singular values of A.

If we substitute $A = U\Sigma V^T$ in the least squares residual norm formula, we can "factor out" U just as we pulled out the Q factor in QR decomposition

$$\|Ax - b\| = \|U\Sigma V^T x - b\| = \|\Sigma\tilde{x} - \tilde{b}\|, \text{ where } \tilde{x} = V^T x \text{ and } \tilde{b} = U^T b.$$

Where $\tilde{x} = V^T x$ and $\tilde{b} = U^T b$. Note that $\|\tilde{x}\| = \|x\|$ and $\|\tilde{b}\| = \|b\|$. If A has rank r, then singular values $\sigma r + 1, \ldots, \sigma n$ are all zero. In this case, there are many different solutions that minimize the residual — changing the values of $\tilde{x}$r+1 through $\tilde{x}$n does not change the residual at all. One standard way to pick a unique solution is to choose the minimal norm solution to the problem, which corresponds to setting $\tilde{x}$r+1 = . . . = $\tilde{x}$n = 0. In this case, the Moore-Penrose pseudoinverse is defined as

$$A^+ = V\Sigma^+ U^T$$

where $\Sigma^+ = \mathrm{diag}(\sigma 1, \sigma 2, \ldots, \sigma r)$ and U and V consist of the first r left and right singular vectors. We solve the problem $Ax \approx b$ by writing $x = A^+ b$.

Algorithm for Pseudoinverse

1. Perform a singular value decomposition $A = U\Sigma V^T$
2. Check for small singular values in ΣΣ. A common criterion is that anything less than ‖ A ‖ϵ is small. Set those singular values to zero.
3. Form $\Sigma^+ = \mathrm{diag}(\sigma 1, \sigma 2, \ldots, \sigma r)$. Reciprocating the nonzero singular values, and leave the zero ones untouched.
4. $A^+ = V\Sigma^+ U^T$

Code for Checking Singular Values

```
[U,S,V] = svd(A); sigma = diag(S);
% compute exact data
xex = ones(n,1); bex = A*xex;
for i = 1:n-1
% data perturbation
deltad = (n-1)^(-i)*(0.5-rand(size(dex))).*dex;
d = dex+deltad;
% solution of perturbed linear least squares problem
w = U'*d;
x = V * (w(1:n) ./ sigma);
errx(i+1) = norm(x - xex); errd(i+1) = norm(deltad);
end
loglog(errd,errx,'*');
ylabel('||x^{ex} - x||_2'); xlabel('||\delta d||_2')
```

Numerical Examples

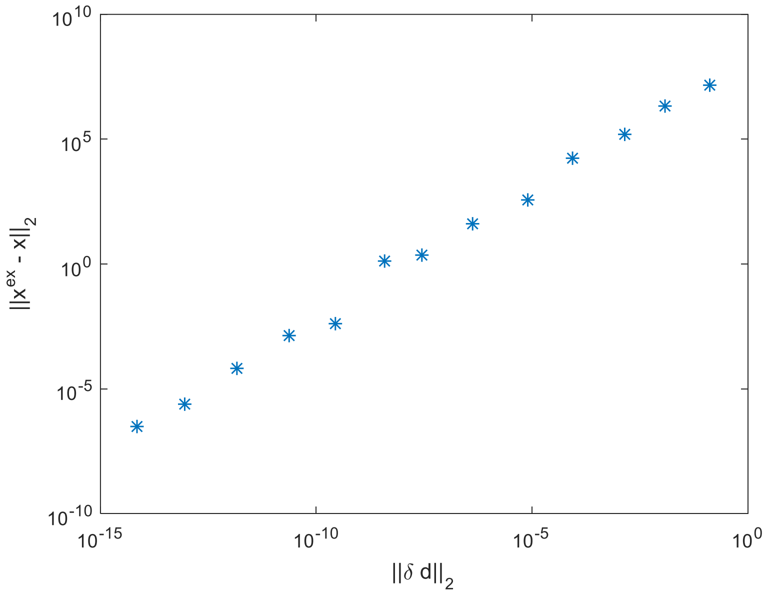
Suppose that the data d are d = dex + δd . Where δd represents the measurement error. Using the function [shaw.m]() to generate the matrix and vector in the least squares problem, then compute the σ to see whether if a singular value small, the norm could be large.

A =

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	4.5029e-07	4.4945e-05	6.4288e-04	0.0034	0.0089	0.0105	0.0024	0.0047	0.0437	0.0966	0.1140	0.0866	0.0435	0.0113
2	4.4945e-05	3.3199e-04	0.0020	0.0071	0.0144	0.0131	0.0011	0.0146	0.0861	0.1718	0.2001	0.1617	0.0979	0.0435
3	6.4288e-04	0.0020	0.0060	0.0140	0.0196	0.0100	9.6456e-04	0.0537	0.1852	0.3045	0.3230	0.2541	0.1617	0.0866
4	0.0034	0.0071	0.0140	0.0211	0.0167	7.9076e-04	0.0284	0.1696	0.3743	0.4893	0.4488	0.3230	0.2001	0.1140
5	0.0089	0.0144	0.0196	0.0167	0.0023	0.0172	0.1502	0.4095	0.6282	0.6435	0.4893	0.3045	0.1718	0.0966
6	0.0105	0.0131	0.0100	7.9076e-04	0.0172	0.1423	0.4222	0.7184	0.7997	0.6282	0.3743	0.1852	0.0861	0.0437
7	0.0024	0.0011	9.6456e-04	0.0284	0.1502	0.4222	0.7494	0.8863	0.7184	0.4095	0.1696	0.0537	0.0146	0.0047
8	0.0047	0.0146	0.0537	0.1696	0.4095	0.7184	0.8863	0.7494	0.4222	0.1502	0.0284	9.6456e-04	0.0011	0.0024
9	0.0437	0.0861	0.1852	0.3743	0.6282	0.7997	0.7184	0.4222	0.1423	0.0172	7.9076e-04	0.0100	0.0131	0.0105
10	0.0966	0.1718	0.3045	0.4893	0.6435	0.6282	0.4095	0.1502	0.0172	0.0023	0.0167	0.0196	0.0144	0.0089
11	0.1140	0.2001	0.3230	0.4488	0.4893	0.3743	0.1696	0.0284	7.9076e-04	0.0167	0.0211	0.0140	0.0071	0.0034
12	0.0866	0.1617	0.2541	0.3230	0.3045	0.1852	0.0537	9.6456e-04	0.0100	0.0196	0.0140	0.0060	0.0020	6.4288e-04
13	0.0435	0.0979	0.1617	0.2001	0.1718	0.0861	0.0146	0.0011	0.0131	0.0144	0.0071	0.0020	3.3199e-04	4.4945e-05
14	0.0113	0.0435	0.0866	0.1140	0.0966	0.0437	0.0047	0.0024	0.0105	0.0089	0.0034	6.4288e-04	4.4945e-05	4.5029e-07

b =	x =	For n = 14, σ =
0.6047	0.1592	2.993428170119102e+00
1.1012	0.3402	1.857061621960602e+00
1.8573	0.5946	1.034629824533204e+00
2.7718	0.8497	3.939756969825725e-01
3.4799	0.9926	5.887209995808784e-02
3.6127	0.9489	3.473468774152014e-02
3.2987	0.7538	2.467635507378432e-02
3.0062	0.5896	3.678806230766344e-03
2.8041	0.7980	9.088096029912202e-04
2.3948	1.5249	3.910056998180689e-05
1.7409	2.0342	1.682334949410932e-06
1.0905	1.5463	6.353708268792253e-07
0.6204	0.6478	9.547130253058336e-08
0.3373	0.1486	2.196890154043734e-09

The error for different value of different values of $\|\delta d\|$:



We see that small perturbations δd in the measurements can lead to large errors in the solution x of the linear least squares problem if the singular values of A are small.

Conclusion Remarks

SVD is a good tool for dealing with numerical rank issues. One standard way to pick a unique solution is to choose the minimal norm solution by Moore-Penrose pseudoinverse. In addition, we show that a small singular value could lead to large norm. A common criterion is that anything less than $\| A \|\epsilon$ is small. Set those singular values to zero. However, pseudoinverse is a computationally expensive operation. For future study, we want to find a cheaper but also stable operation.

References

"3. Methods for Rank-Deficient Problems." *Rank-Deficient and Discrete Ill-Posed Problems* (1998): 45-67. Web. 6 Mar.2016

"Least Squares Problems." *SpringerReference* (n.d.): n. pag. Web. 6 Mar. 2017.

"Linear Least-Squares Estimation: Fundamentals." *Advanced Kalman Filtering, Least-Squares and Modeling* (2011): 101-38. Web. 6 Mar. 2017.

"QR Factorization, Singular Value Decomposition, and Projections." *Numerical Linear Algebra and Applications* (n.d.): 181-236. Web. 6 Mar. 2017