The singular value decomposition of a matrix (SVD) is one of the most useful techniques employed in the statistical analysis of data matrices, and should always be used to calculate the rank when data sets appear difficult to analyze unambiguously due to singularity.

A m by n matrix A can always be factored as

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

where U and V are orthogonal, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r)$, $r = \min(m, n)$, with $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$. The σ_i are the singular values, the first r columns of V are the right singular vectors, and the first r columns of U are the left singular vectors.

Example 1: m > n

From the main SIMFIT menu choose [Statistics], [Numerical analysis], then the singular value decomposition and analyze data contained in the default test file `f08kff.tf1` to obtain the following results for a case with $m > n$.

Current matrix: SVD data for `f08kff.tf1` with rank = 4

-5.7000000E-01	-1.2800000E+00	-3.9000000E-01	2.5000000E-01
-1.9300000E+00	1.0800000E+00	-3.1000000E-01	-2.1400000E+00
2.3000000E+00	2.4000000E-01	4.0000000E-01	-3.5000000E-01
-1.9300000E+00	6.4000000E-01	-6.6000000E-01	8.0000000E-02
1.5000000E-01	3.0000000E-01	1.5000000E-01	-2.1300000E+00
-2.0000000E-02	1.0300000E+00	-1.4300000E+00	5.0000000E-01

Index	σ_i	Fraction	Cumulative	σ_i^2	Fraction	Cumulative
1	3.9987197E+00	0.4000	0.4000	1.59898E+01	0.5334	0.5334
2	3.0005164E+00	0.3002	0.7002	9.00310E+00	0.3003	0.8337
3	1.9967125E+00	0.1998	0.9000	3.98686E+00	0.1330	0.9666
4	9.9994082E-01	0.1000	1.0000	9.99882E-01	0.0334	1.0000

Right singular vectors by row (V-transpose)

8.2514556E-01	-2.7935888E-01	2.0479917E-01	4.4626307E-01
-4.5304486E-01	-2.1212912E-01	-2.6220881E-01	8.2522611E-01
-2.8285271E-01	-7.9609569E-01	4.9515860E-01	-2.0259308E-01
1.8406389E-01	-4.9314451E-01	-8.0257199E-01	-2.8072616E-01

Left singular vectors by column (U)

-2.0271367E-02	2.7939484E-01	4.6900514E-01	7.6917561E-01
-7.2841546E-01	-3.4641438E-01	-1.6941649E-02	-3.8290338E-02
4.3926966E-01	-4.9545699E-01	-2.8679802E-01	8.2222482E-02
-4.6784650E-01	3.2584052E-01	-1.5355623E-01	-1.6362605E-01
-2.2003450E-01	-6.4277549E-01	1.1245506E-01	3.5724829E-01
-9.3523390E-02	1.9268002E-01	-8.1318411E-01	4.9572409E-01

Example 2: m < n

The test file `f08kff.tf2` has $m < n$ and yields the following results.

Current matrix: SVD data for `f08kff.tf2` with rank = 4

-5.4200000E+00	3.2800000E+00	-3.6800000E+00	2.7000000E-01	2.0600000E+00	4.6000000E-01
-1.6500000E+00	-3.4000000E+00	-3.2000000E+00	-1.0300000E+00	-4.0600000E+00	-1.0000000E-02
-3.7000000E-01	2.3500000E+00	1.9000000E+00	4.3100000E+00	-1.7600000E+00	1.1300000E+00
-3.1500000E+00	-1.1000000E-01	1.9900000E+00	-2.7000000E+00	2.6000000E-01	4.5000000E+00

Index	σ_i	Fraction	Cumulative	σ_i^2	Fraction	Cumulative
1	7.9987291E+00	0.3077	0.3077	6.39797E+01	0.3677	0.3677
2	7.0059360E+00	0.2695	0.5771	4.90831E+01	0.2821	0.6498
3	5.9952457E+00	0.2306	0.8077	3.59430E+01	0.2066	0.8564
4	4.9988922E+00	0.1923	1.0000	2.49889E+01	0.1436	1.0000

Right singular vectors by row (V-transpose)

-7.9334138E-01	3.1632760E-01	-3.3417094E-01	-1.5140353E-01	2.1421955E-01	3.0010505E-01
1.0023978E-01	6.4422154E-01	4.3706509E-01	4.8903426E-01	3.7714369E-01	5.0128185E-02
1.1079521E-02	1.7244335E-01	-6.3666698E-01	4.3538482E-01	-4.3019579E-02	-6.1105243E-01
2.3606094E-01	2.1558447E-02	-1.0247828E-01	-5.2856734E-01	7.4604456E-01	-3.1199799E-01

Left singular vectors by column (U)

8.8835057E-01	1.2751497E-01	4.3306768E-01	8.3818775E-02
7.3269096E-02	-8.2640879E-01	1.9433222E-01	-5.2336903E-01
-3.6065168E-02	5.4351983E-01	7.5594729E-02	-8.3520712E-01
4.5184534E-01	-7.3311896E-02	-8.7691095E-01	-1.4658902E-01

Fractions and Cumulatives

For an explanation of the tables of fractions and cumulatives displayed and plotted by SimFIT the following expanded expression for the singular value decomposition of a m by n matrix A with $m > n$ should be noted

$$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_n u_n v_n^T,$$

where u_i are the columns of U, v_i are the columns of V and n is the rank of A. If a m by n matrix B of rank $k < n$ is required which best approximates A in the least squares sense we need to minimize the sum of squares $S = (A - B)(A - B)^T$, i.e.

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - b_{ij})^2,$$

and this is satisfied by

$$B = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_k u_k v_k^T.$$

It follows from this result that, if A is a mean-centered m by n matrix of rank n with $m > n$, then the expression

$$f_k = \frac{\sigma_1^2 + \cdots + \sigma_k^2}{\sigma_1^2 + \cdots + \sigma_n^2}$$

is the fraction of the total variance of A accounted for by approximating A by a matrix B of rank k formed by evaluating the usual expression for the SVD for matrix A but with all $\sigma_i = 0$ for $i > k$.