Singular Value Decomposition

Let A be a general real m-by-n matrix. The **singular value decomposition (SVD)** of A is the factorization $A = U\Sigma V^T$, where U and V are orthogonal, and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ $r = \min(m, n)$, with $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$. If A is complex, then its SVD is $A = U\Sigma V^H$ where U and V are unitary, and Σ is as before with real diagonal elements. The σ_i are called the **singular values**, the first r columns of V the **right singular vectors** and the first r columns of U the **left singular vectors**.

The routines described in this section, and listed in Table 2.12, are used to compute this decomposition. The computation proceeds in the following stages:

1.
    The matrix A is reduced to bidiagonal form: $A = U_1 B V_1^T$ if A is real ($A = U_1 B V_1^H$ if A is complex), where U_1 and V_1 are orthogonal (unitary if A is complex), and B is real and upper-bidiagonal when $m \geq n$ and lower bidiagonal when $m < n$, so that B is nonzero only on the main diagonal and either on the first superdiagonal (if $m \geq n$) or the first subdiagonal (if $m < n$).
2.
    The SVD of the bidiagonal matrix B is computed: $B = U_2 \Sigma V_2^T$, where U_2 and V_2 are orthogonal and Σ is diagonal as described above. The singular vectors of A are then $U = U_1 U_2$ and $V = V_1 V_2$.

The reduction to bidiagonal form is performed by the subroutine xGEBRD, or by xGBBRD for a band matrix.

The routine xGEBRD represents U_1 and V_1 in factored form as products of elementary reflectors, as described in section 5.4. If A is real, the matrices U_1 and V_1 may be computed explicitly using routine xORGBR, or multiplied by other matrices without forming U_1 and V_1 using routine xORMBR. If A is complex, one instead uses xUNGBR and xUNMBR, respectively.

If A is banded and xGBBRD is used to reduce it to bidiagonal form, U_1 and V_1 are determined as products of Givens rotations , rather than as products of elementary reflectors. If U_1 or V_1 is required, it must be formed explicitly by xGBBRD. xGBBRD uses a vectorizable algorithm, similar to that used by xSBTRD (see Kaufman [77]). xGBBRD may be much faster than xGEBRD when the bandwidth is narrow.

The SVD of the bidiagonal matrix is computed either by subroutine xBDSQR or by subroutine xBDSDC. xBDSQR uses QR iteration when singular vectors are desired [32,23], and otherwise uses the dqds algorithm [51]. xBDSQR is more accurate than its counterparts in LINPACK and EISPACK: barring underflow and overflow, it computes all the singular values of B to nearly full relative precision, independent of their magnitudes. It also computes the singular vectors much more accurately. See section 4.9 and [32,23,51] for details.

xBDSDC uses a variation of Cuppen's divide and conquer algorithm to find singular values and singular vectors [69,58]. It is much faster than xBDSQR if singular vectors of large matrices are desired. When singular values only are desired, it uses the same dqds algorithm as xBDSQR [51]. Divide-and-conquer is not guaranteed to compute singular values to nearly full relative precision, but in practice xBDSDC is often at least as accurate as xBDSQR. xBDSDC represents the singular vector matrices U_2 and V_2 in a

compressed format requiring only $O(n \log n)$ space instead of n^2. U_2 and V_2 may subsequently be generated explicitly using routine xLASDQ, or multiplied by vectors without forming them explicitly using routine xLASD0.

If $m \gg n$, it may be more efficient to first perform a QR factorization of A, using the routine xGEQRF, and then to compute the SVD of the n-by-n matrix R, since if $A = QR$ and $R = U\Sigma V^T$, then the SVD of A is given by $A = (QU)\Sigma V^T$. Similarly, if $m \ll n$, it may be more efficient to first perform an LQ factorization of A, using xGELQF. These preliminary QR and LQ factorizations are performed by the drivers xGESVD and xGESDD.

The SVD may be used to find a minimum norm solution to a (possibly) rank-deficient linear least squares problem (2.1). The effective rank, k, of A can be determined as the number of singular values which exceed a suitable threshold. Let $\hat{\Sigma}$ be the leading k-by-k submatrix of Σ, and $\hat{V}$ be the matrix consisting of the first k columns of V. Then the solution is given by:

$$x = \hat{V}\hat{\Sigma}^{-1}\hat{c}_1$$

where $\hat{c}_1$ consists of the first k elements of $c = U^T b = U_2^T U_1^T b$. $U_1^T b$ can be computed using xORMBR, and xBDSQR has an option to multiply a vector by U_2^T.

Table 2.12: Computational routines for the singular value decomposition

Type of matrix and storage scheme	Operation	Single precision		Double precision	
		real	complex	real	complex
general	bidiagonal reduction	SGEBRD	CGEBRD	DGEBRD	ZGEBRD
general band	bidiagonal reduction	SGBBRD	CGBBRD	DGBBRD	ZGBBRD
orthogonal/unitary	generate matrix after bidiagonal reduction	SORGBR	CUNGBR	DORGBR	ZUNGBR
	multiply matrix after bidiagonal reduction	SORMBR	CUNMBR	DORMBR	ZUNMBR
bidiagonal	SVD using QR or dqds	SBDSQR	CBDSQR	DBDSQR	ZBDSQR
	SVD using divide-and-conquer	SBDSDC		DBDSDC	

Singular Value Decomposition and Voting

5-4 or 9-0

The nine justices of the United States Supreme Court do not always issue unanimous decisions, nor do their votes match the pattern that would be shown by nine totally independent thinkers.

For those who have ever wondered where the court's decisions might fall on the spectrum between monolithic unity and total randomness, an answer is now in.

The voting pattern of the Rehnquist court over the last nine years "shows that the court acts as if composed of 4.68 ideal justices," says Dr. Lawrence Sirovich, a mathematician at the Mount Sinai School of Medicine in Manhattan whose day job is figuring out how the visual system works.

By another measure, "the decision space of the Rehnquist court requires only two dimensions for its description," he writes in the issue of The Proceedings of the National Academy of Sciences being published today.

Nine independent thinkers who focus solely on the merits of cases might be expected to vote in all possible combinations over a long enough period. Dr. Sirovich's analyses indicate that the Supreme Court voting falls a long way from that pattern.

His first measure entails considering that if two members were twins who always voted the same way, the court would effectively have eight members, not nine. On a court with nine members, there are 512 possible voting patterns, or half that number if a vote is marked as being in the majority or not.

But the actual number of voting patterns is very much less, as if generated by a smaller number of wholly independent individuals.

Analyzing nearly 500 opinions issued since 1995 — the court membership has not changed since Justice Stephen G. Breyer joined it in 1994 — Dr. Sirovich calculates, based on information theory, that 4.68 ideal justices would have produced the same diversity of decision making.

By ideal, Dr. Sirovich means a justice whose voting is uncorrelated with any other's. His measure, thus, points up the high degree of correlation in the court's voting pattern.

Looking only at final votes, of course, ignores many nuances about how the court operates in reality, Dr. Sirovich acknowledged. Justices often file concurring opinions or dissents, allowing them to cooperate in the creation of a majority while preserving independence in outlining the reasons behind a vote.

Considering the decisions with another technique known as singular value decomposition, Dr. Sirovich has also found considerably less diversity than might be expected.

It would take nine dimensions for a mathematician to describe the voting patterns of the nine uncorrelated justices. But Dr. Sirovich has found that just two dimensions are needed to describe almost all the decisions of the Rehnquist court.

Although his refusal to draw any political implications from his analysis may disappoint some people, the neutrality of the approach is what makes it appealing to political and legal scholars.

"People typically come up with a single dimension, readily interpretable as a left-right point of view," said Jeffrey Segal, a political scientist at the State University of New York at Stony Brook. "He comes up with two dimensions but doesn't label them."

Dr. Yochai Benkler of the New York University Law School said the analysis was important, because instead of starting from some theory about the politics of the court, Dr. Sirovich had used a purely mathematical analysis, yet one whose results fit with common sense observation.

"What you see here is not someone trying to prove a point, but someone who has said, `Beyond the stories, this is the math of how people behave and how they ally,' " Professor Benkler said. "And the interesting

thing here is the fit between the mathematical observation and the widely held intuition about the politics of adjudication."

Dr. Sirovich's specialty is in extracting patterns from information. "He is the master of S.V.D.'s," Dr. Mitchell Feigenbaum, a mathematician at the Rockefeller University, said of singular value decomposition.

Dr. Sirovich introduced the technique for use in machine recognition of human faces, and it is now the basis for most face recognition systems. It depends on the fact that the pixels in a photograph are not independent, but that the shading of each pixel is highly correlated with that of its neighbor. Because of that correlation, the 100,000 dimensions that would in principle be needed to describe the 100,000 pixels in a photo of a face can be collapsed to about a hundred dimensions.

The vectors — mathematical representations of information — in those 100 dimensions represent basic faces from which any other face can be constructed.

The same method applied to the Supreme Court's voting patterns enables vectors in just two dimensions to describe almost all its decisions. One vector is very close to the vector that represents a unanimous decision. The other lies near the vector representing the most common of the court's 5-to-4 voting patterns, including the one that decided the 2000 presidential election.

Each justice's vote can be regarded as fixed mixture of those two voting patterns, Dr. Sirovich writes. Only three decisions out of 468 are not fully captured by his two vectors. One, Rogers v. Tennessee, split both sides of the usual 5-to-4 vote, with Chief Justice William H. Rehnquist separated from his usual allies Justices Antonin Scalia and Clarence Thomas, and Justices John Paul Stevens and Breyer sundered from Justices David H. Souter and Ruth Bader Ginsburg.

The case concerned a technical issue of Tennessee law, which might account for the absence of the usual lineup.

"I think this was a case of pure law and no `politics,' and they acted independently," Dr. Sirovich said.

Dr. Sirovich applied his two yardsticks — information theory and single value decomposition — to the Warren court for two periods when its composition was unchanged, from 1959 to 1961 and from 1967 to 1969. The first Warren court was somewhat more diverse than the Rehnquist court, operating as if with 5.16 ideal justices.

But its dynamics were quite similar, with two dominant voting patterns, William O. Douglas being the regular dissenter, as Justice Stevens is in the Rehnquist court, and Tom C. Clark and Potter Stewart playing the role of the swing votes, similar to that of Justices Anthony M. Kennedy and Sandra Day O'Connor in the present court.

The second Warren court was much more liberal than the first one, and some players switched roles. Justices Douglas, Earl Warren and William J. Brennan Jr., often in minority in the first Warren court, became part of the 5-to-4 majority in the second. Justice Thurgood Marshall, in the second court, dissented just once.

The surprising similarity in the voting dynamics of the Warren and Rehnquist courts, despite their ideological differences, suggests that some deeper factor may be imposing a pattern.

One possibility, Dr. Sirovich suggests, could be "a sameness in the overall quality for cases percolating up to the court through the judicial substructure." Another possibility could be the dynamic of the size of the court, which dictates a 5-to-4 voting pattern as the winning combination.