Stanford University, Dept of Management Science and Engineering

MS&E 318 (CME 338) Large-Scale Numerical Optimization

Instructor: Michael Saunders Spring 2009

Notes 6: **LUSOL — a Basis Factorization Package**

1 Origins

LUSOL is a set of procedures for computing and updating LU factors of a general sparse matrix A. The design aims are:

- To allow A to be square or rectangular with arbitrary rank.
- To factorize $A = LU$ directly, finding suitable row and column orderings.
- To replace a column or row of A.
- To add or delete a column or row (thus altering the size of A).
- To perform a general rank-one update $A \leftarrow A + vw^T$ for sparse v and w.
- To balance stability and sparsity throughout, keeping L well-conditioned.

The primary application of LUSOL has been for square basis factorizations $B = LU$ (and column replacement) as part of the optimization packages MINOS, SNOPT, and SQOPT, and the nonlinear complementarity packages MILES and PATH. In the optimization packages it is also applied to rectangular matrices $\begin{pmatrix} B & S \end{pmatrix}$ (transposed) to find a column ordering that makes B better conditioned [4].

The original LUSOL procedures are described by Gill, Murray, Saunders and Wright [5]. The main factorization uses a traditional Markowitz strategy with Threshold Partial Pivoting similar to other early sparse LU packages—notably Y12M [12], LA05 [7], LA15 [8], MA28 [2] and MOPS [10]. The Bartels-Golub update for column-replacement follows the sparse implementation of Reid [7]. This involves a "forward sweep" of eliminations. The other updates are implemented similarly (some of them requiring a backward sweep).

LUSOL has continued to evolve, and it is now available as open source software [9]. The current version is implemented in Fortran 77, suitable for f77, f90, and f95 compilers and convertible to C by the f2c translator. Recent additions have focused on rank-revealing LU factorizations for sparse matrices, using either Threshold Rook Pivoting or Threshold Complete Pivoting. One option is intended for symmetric quasi-definite matrices [11]. In 2004, a Fortran 77 → Pascal → C translation was created by Kjell Eikland for use within the open source LP/MILP system lp_solve [6]. Since 2005 it has been lp_solve's default basis factorization package (BFP).

2 Purpose

LUSOL maintains a sparse factorization $A = LU$ and permutations P, Q such that PUQ is a sparse upper triangular (or upper trapezoidal) matrix, stored explicitly as a set of *sparse rows*. Depending on input parameters, L tends to be well-conditioned, while the condition and rank of U reflect the condition and rank of A.

The main functions of LUSOL are as follows:

Factor Determine L, U, P, Q directly from A. The initial $L = L_0$ is such that PL_0P^T is a sparse lower triangular matrix with unit diagonals and bounded off-diagonals, stored explicitly as a set of *sparse columns*. The initial U_0 is stored explicitly as a set of *sparse rows*.

Solve Given a dense vector y, use the current factors to solve one of the following systems: $Lx = y$, $L^Tx = y$, $Ux = y$, $U^Tx = y$, $Ax = y$, $A^Tx = y$.

Multiply Given a dense vector y, use the factors to form one of the following products: $x = Ly$, $x = L^Ty$, $x = Uy$, $x = U^Ty$, $x = Ay$, $x = A^Ty$.

Update Modify L, U, P, Q to reflect one of the following changes to A:

Add a column,	Replace a column,	Delete a column,
Add a row,	Replace a row,	Delete a row,
Add a rank-one matrix σvw^T.		

The initial L_0 is not altered but updates are accumulated in a product form $L = L_0M_1M_2 \dots M_\ell$ for a sequence of stabilized elementary matrices M_j. Updated U factors are maintained explicitly as sparse rows.

3 Factor

The factorization procedure *lu1fac* receives A as a list of triples (i, j, A_{ij}), where each A_{ij} is typically nonzero. (Any zero or tiny elements are deleted.) The nonzeros are sorted into a *column list* containing pointers to the start of each column and the corresponding column lengths (the number of entries). Contiguous storage is used for the entries in any given column, but if new entries arise during factorization, the whole column may be moved to another part of storage.

A similar *row list* is constructed to store the sparsity structure of A by rows. To save storage the row list does not contain the numerical values themselves (although they would improve the efficiency of Threshold Rook Pivoting as mentioned below).

Additional data structures are used to sort the columns and rows in order of increasing length.

Each stage of the LU factorization computes l and u, the next column of L and the next row of U, using a nonzero A_{ij} as "pivot element":

$$l = A_{.j}/A_{ij}, \qquad u^T = A_{i.}, \qquad L \leftarrow \begin{bmatrix} L & l \end{bmatrix}, \qquad U \leftarrow \begin{bmatrix} U \\ u^T \end{bmatrix}, \qquad A \leftarrow A - lu^T.$$

Note that the ith row and jth column of $A - lu^T$ are empty (zero). The data structure holding A has a decreasing number of rows and columns, but a certain amount of *fill* (new nonzeros) may be generated by the rank-one term lu^T.

3.1 Pivot Strategies

To preserve sparsity, a *Markowitz strategy* is used to select potential pivots A_{ij}. Pivots should have a low merit function $M_{ij} \equiv (r_i - 1)(c_j - 1)$, where r_i and c_j are the lengths of row i and column j of A, because M_{ij} bounds the fill that could be created by lu^T. The sparsest columns and rows are searched in turn (columns of length 1, rows of length 1, then columns of length 2, rows of length 2, and so on). The lowest M_{ij} is recorded for pivots that satisfy a specified stability test. Typically only 5 or 10 of the shortest columns and rows are searched, but some of the stability tests may require more extensive searching.

To preserve stability, one of the following *threshold pivoting strategies* is used. Let `FacTol` be a given number, typically 10 or 5 or perhaps nearer 1. (It controls the size of the off-diagonals of L and possibly U.) Further, let $A_{\max}$ be the largest remaining nonzero and $D_{\max}$ the largest remaining diagonal (both absolute values).

Strategy	Name	Stability Test
Threshold Partial Pivoting	TPP	$\|l\|_\infty \le$ `FacTol`
Threshold Rook Pivoting	TRP	$\|l\|_\infty$ *and* $\|u/A_{ij}\|_\infty \le$ `FacTol`
Threshold Complete Pivoting	TCP	$A_{\max} \le$ `FacTol` $\times \lvert A_{ij}\rvert$
Threshold Symmetric Pivoting	TSP	TRP for symmetric A
Threshold Diagonal Pivoting	TDP	$D_{\max} \le$ `FacTol` $\times \lvert A_{ii}\rvert$

In practice, TPP is used most often. It preserves sparsity well and is usually sufficiently stable with `FacTol` = 100 or 10. Smaller values are needed if A is singular or ill-conditioned and *rank-revealing* properties are desired. Ideally, any rank-deficiency should be identified by the correct number of small diagonals in U. The other options are needed for greater reliability in determining rank—for example, in the Simplex Method when the first rather arbitrary B is factorized.

TPP requires A_{ij} to be sufficiently large compared to other elements in its own column. It is efficient to implement, and every column contains at least one entry that satisfies the test, so relatively few columns need be searched. It is convenient to store the largest element in each column in a known place (e.g., as the first entry in the column).

TRP is symmetric in that A_{ij} must be sufficiently large compared to other elements in its own column *and* row. It costs more than TPP (may require more searching and produce less sparse factors), but it has more definite rank-revealing properties and seems acceptably efficient with `FacTol` as low as 2 or even 1.1.

TCP requires $A_{\max}$, the largest element in the current A. A *heap* structure is used to store the largest element in each column, and then $A_{\max}$ is always at the top of the heap. TCP satisfies the TRP test but in a more restrictive way. If `FacTol` is too close to 1, the TCP factors are likely to be unacceptably dense.

(Note that pathological examples are known for which TRP or even TCP fail to reveal rank correctly.)

TSP and TDP are intended for symmetric matrices that are either definite or *quasi-definite*. (In exact arithmetic, symmetric factorizations $PAP^T = LDL^T$ exist for such matrices for any permutation P.) TSP is implemented by pivoting on diagonals only and requiring $\|l\|_\infty \le$ `FacTol` as for TPP.

TDP requires $D_{\max}$, the largest remaining diagonal. It has not been implemented yet (and may not be needed). Another heap structure will be needed to store the diagonals $|A_{ii}|$, so that $D_{\max}$ will be available at the top of the heap.

The main benefit of TCP and TDP is that they concentrate singularities at the end of P and Q, so that PUQ will be *upper trapezoidal.* The other strategies may give small diagonals in any part of U (if A is singular). This often indicates singularity correctly, but the the following matrix illustrates that TPP may be misleading:

$$A = \begin{pmatrix} \delta & 1 & 1 & 1 \\ & \delta & 1 & 1 \\ & & \delta & 1 \\ & & & \delta \end{pmatrix}, \qquad \delta \text{ very small.}$$

TPP would accept all diagonals and return $L = I$, $U = A$. LUSOL's singularity check would then report that four diagonals of U were very small compared to other entries in their own column and that $\operatorname{rank}(A) = 0$ (!).

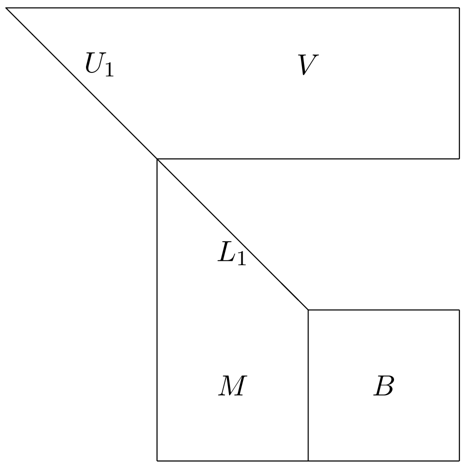
In contrast, the rook pivoting strategy would choose a δ as diagonal only if *all* elements in the same row and column were small. Instead, TRP is likely to choose the first superdiagonal as pivots (effectively permuting the first column to the end). The result is

$$PUQ \approx \begin{pmatrix} 1 & 1 & 1 & \delta \\ & 1 & 1 & -\delta^2 \\ & & 1 & \delta^3 \\ & & & -\delta^4 \end{pmatrix}, \qquad \operatorname{rank}(A) \approx 3.$$

3.2 Backward and Forward Triangles

The strategy of searching for columns of length 1 and then rows of length 1 automatically reveals the following structure in a typical sparse A. For some preliminary permutations P_1 and Q_1,

$P_1 A Q_1 =$



,

where $[U_1 \;\; V]$ is called the backward triangle (it forms the top rows of U) and $\begin{bmatrix} L_1 \\ M \end{bmatrix}$ is the forward triangle (the first columns of L). No elimination has occurred yet. It remains to compute LU factors of the bottom block B.

3.3 Sparse Elimination

A typical step in the LU factorization of block B is illustrated here. The Markowitz strategy has selected a pivot ① in a reasonably sparse column and row, and this is regarded as large enough under the TPP stability test when `FacTol` $= 3.0$. The merit function M_{11} predicts that at most $2 \times 3 = 6$ new entries will be created in rows 2–3 and columns 2–4 (an over-estimate because 16, 1 and 5 are already nonzero). Note that all other rows and columns will be unaltered.

TPP can pivot on	①
TRP can pivot on	4
TCP must pivot on	16

①	4	2	1						
2		16		×		×			×
1	1		5		×	×		×	
		×					×		
			×			×		×	
				×		×	×		×
	×				×				
		×							×
				×			×		
			×					×	

Most of the complexity of *lu1fac* arises from updating the lists that hold the nonzeros. These are implemented as arrays [*locc*, *lenc*, *indc*, *a*] for the column list and [*locr*, *lenr*, *indr*] for the row list (sparsity structure only).

When the pivot row is deleted from the column list (to become the first row of U), there is room in the modified columns for *one* new entry. In this example, *one* is sufficient and columns 2–4 can be updated *in place*.

Similarly, when the pivot column is deleted from the row list, row 3 can be updated in place. However, row 2 has *two* new entries. All of row 2 must be moved to the beginning of the row list's free storage, and its previous space marked as unused. Thus, columns and rows can migrate around memory within their own lists, leaving gaps that are periodically compressed ("garbage collection").

After most of the LU factors have been computed, the remaining rows and columns of the updated A will be sufficiently dense (say 50%) to warrant a switch to dense processing.

4 Solve

The current procedures for solving $Lx = y$, $Ux = y$, ... treat y as a dense vector. Solves with L_0 and L_0^T are the most efficient because the columns are processed in natural order (forwards or backwards) and never altered. The product-form updates to L are handled reasonably efficiently (but each involves two arbitrary indices).

Ideally, new options should be implemented in LUSOL to take advantage of sparse right-hand sides y. For example, each iteration of Primal Simplex requires solution of $Bv = a_s$, where a_s is a column of A (with an average of only 5 or 10 entries regardless of the size of A). A significant cost in some simplex implementations lies in the most trivial operation: setting a dense vector to zero before unpacking a_s. This is true even for $m = 10,000$, and certainly for $m = 1$ million.

Gilbert and Peierls [3] show how to solve a triangular system $Lx = y$ in $O(p)$ operations, where p is the number of *nonzero multiplications* needed to form the product Lx. When y is sparse, p may be very small *regardless of problem size.* The true benefits of this technique have been realized in CPLEX; see Bixby [1].

The GP algorithm requires L to be stored by columns. In LUSOL, this means $L_0x = y$, $U_0^T x = y$ and updated $U^T x = y$ could be solved efficiently. LA05 maintains both the row and the column structure of U. (What about the row structure of L_0?)

5 Multiply

The Multiply procedures may be needed to recover parts of A from L and U if the original data has been over-written. Less trivially, products with L and L^T are needed if an iterative solver (such as LSQR) is applied to a least-squares problem in the following way. Suppose A is rectangular ($m > n$) with $\mathrm{rank}(A) = n$. A sparse factorization $A = LU$ is typically cheaper than a QR factorization. To solve

$$\min \|Ax - b\|_2^2$$

we may solve the equivalent problem

$$\min \|Ly - b\|_2^2, \qquad Ux = y$$

using the same iterative solver. Although L is typically less sparse than A, it tends to be well-conditioned even if A is not. Hence the solver may converge sufficiently quickly.

Note that U is being used as a right-preconditioner, and we assumed $A = LU$ exactly. If U is exact or just *approximate*, it may be more efficient to apply the iterative solver to the equivalent problem

$$\min \|AU^{-1}y - b\|, \qquad Ux = y.$$

6 Update

The main work for all updates comes from a forward or backward sweep of stabilized eliminations.

Forward sweeps tend to alter only a few rows of U. The most common update (Bartels-Golub-Reid column replacement) requires a sparse column to be *inserted* (possibly altering several rows) and then a forward sweep to eliminate the row spike that appears, as mentioned in Notes 5. Several rows may be added to the spike, with occasional row interchanges.

Backward sweeps may alter many rows of U. They eliminate a *column spike* and generate a row spike that grows in length and must be added to several subsequent rows, with occasional row interchanges. The final row spike must then be eliminated by a forward sweep.

The more general updates are needed for *symmetric* matrices (such as KKT systems in active-set methods).

7 Basis Repair

A vital use of LUSOL within MINOS, SQOPT, and SNOPT is to ensure that the current basis B is not too ill-conditioned. If necessary, certain columns of B can be replaced by unit vectors (associated with slack variables). The row and column permutations P and Q define which slack variables should be introduced.

Often the TPP strategy warns of singularity reliably. If not, when the basic variables are recomputed, the size of the residual or basic variables ($\|b - Bx_B - Nx_N\|$ or $\|x_B\|$) may provide a clue. The value of `FacTol` can be reduced toward 1.0, and if singularity remains, TRP and finally TCP can be activated. At the start of a major iteration (for problems with nonlinear constraints), TRP is used directly. We call this a "BR factorization" to indicate the Rank-Revealing requirement.

Another form of basis repair is sometimes invoked to select a better B from the current basic and superbasic variables $\begin{pmatrix} B & S \end{pmatrix}$. Since the condition of L is always controlled, a factorization of the matrix

$$\begin{pmatrix} B^T \\ S^T \end{pmatrix} \tag{1}$$

is relevant. LUSOL has an option to compute the row and column permutations P and Q without storing L and U. Here the first m rows of P point to the rows of (1), i.e., the columns of $\begin{pmatrix} B & S \end{pmatrix}$, that will define a well-conditioned B, to the extent that one exists. We call this a "BS factorization".

References

[1] R. E. Bixby. Solving real-world linear programs: a decade and more of progress. *Operations Research*, 50(1):3–15, 2002.

[2] I. S. Duff. MA28: A set of Fortran subroutines for sparse unsymmetric linear equations. Report R8730, AERE Harwell, Oxfordshire, England, 1977.

[3] J. R. Gilbert and T. Peierls. Sparse partial pivoting in time proportional to arithmetic operations. *SIAM J. Sci. and Statist. Comput.*, 9(5):862–874, 1988.

[4] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Review*, 47(1):99–131, 2005. SIGEST article.

[5] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Maintaining LU factors of a general sparse matrix. *Linear Algebra and its Applications*, 88/89:239–270, 1987.

[6] lp_solve open source LP and MILP solver. `http://groups.yahoo.com/group/lp_solve/`.

[7] J. K. Reid. A sparsity-exploiting variant of the Bartels-Golub decomposition for linear programming bases. *Math. Program.*, 24:55–69, 1982.

[8] J. K. Reid. LA15 basis factorization package (thread-safe version of LA05). HSL 2004 Catalogue, `http://www.aspentech.com/hsl/hslnav/hsl2004.htm`, 2004.

[9] SOL downloadable software. `http://www.stanford.edu/group/SOL/software.html`.

[10] U. Suhl and L. Suhl. Computing sparse LU-factorizations for large-scale linear programming bases. *ORSA J. of Computing*, 2:325–335, 1990.

[11] R. J. Vanderbei. Symmetric quasi-definite matrices. *SIAM J. Optim.*, 5:100–113, 1995.

[12] Z. Zlatev, J. Wasniewski, and K. Schaumburg. *Y12M: Solution of Large and Sparse Systems of Linear Algebraic Equations.* Lecture Notes in Computer Science 121. Springer Verlag, Berlin, Heidelberg, New York, 1981.