**Banded triangular systems**

Banded lower triangular systems are defined by triangular matrices which have a nonzero main diagonal and \( m \) nonzero subdiagonals. Typically, \( m \ll n \).

These arise in recurrence relations which involve a constant number of terms \( m < n \). A typical construct assumes \( y(1:n) \) has been initialized and executes

```plaintext
do i = 1, n
    y(i) = y(i) + F(y(i-1), \ldots, y(i-m))
end do
```

where \( F() \) is a linear function and arguments with negative subscripts are ignored. Since \( F \) is linear we can expand this code into a doubly nested loop

```plaintext
do i = 1, n
    do j = 1,m
        y(i) = y(i) + COEF(i,i-j)*y(i-j)
    end do
end do
```

where \( COEF(1:n,1:n) \) is an array which contains the coefficients that define the linear function \( F \).
If \( f = (\phi_1, \ldots, \phi_n)^T \) is a vector that contains the initial values of the elements of the array \( y(1:n) \), \( x = (\xi_1, \ldots, \xi_n)^T \) is a vector containing the final values of the elements of the array \( y(1:n) \) and \( L \in \mathbb{R}^{n \times n} \) is a matrix of order \( n \) with elements \( \lambda_{ij} = \text{COEF}(i,j) \) for \( i = 1, \ldots, n \) and \( j = i - m, \ldots, i - 1 \), \( \lambda_{ii} = 1 \) for \( i = 1, \ldots, n \), and \( \lambda_{ij} = 0 \) for all other \((i,j)\); then we have

\[
\xi_i = \phi_i + \sum_{j=1}^{m} \lambda_{i,i-j} \xi_{i-j}
\]

for \( i = 1, \ldots, n \). This can be written as \( Lx = f \) where \( L \) is a banded unit lower triangular system.

Example: \( n = 6 \) and \( m = 2 \)

\[
\begin{pmatrix}
1 \\
\lambda_{21} & 1 \\
\lambda_{31} & \lambda_{32} & 1 \\
0 & \lambda_{42} & \lambda_{43} & 1 \\
0 & 0 & \lambda_{53} & \lambda_{54} & 1 \\
0 & 0 & 0 & \lambda_{64} & \lambda_{65} & 1
\end{pmatrix}
\begin{pmatrix}
\xi_1 \\
\xi_2 \\
\xi_3 \\
\xi_4 \\
\xi_5 \\
\xi_6
\end{pmatrix}
= 
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\phi_3 \\
\phi_4 \\
\phi_5 \\
\phi_6
\end{pmatrix}
\]
The sequential solution of a banded triangular system of order $n$ requires $2mn + n - m^2 - m$ operations (without the $\lambda = 1$ which reduces the operation count by $n$). The parallel solution without redundant operations has the same $O(n)$ critical path as a dense triangular system but requires only $m$ processors.

Dependence graph for a banded triangular system with 2 nonzero block subdiagonals
Spike algorithm

There are many ways to solve these systems in parallel. Chen, Kuck and Sameh have proposed the spike algorithm for solving banded systems. The solver is intended for a moderate number of processors and is in fact quite practical. It effectively uses a divide and conquer strategy.

Given $p$, partition $L$ as follows

$$
\begin{bmatrix}
L_1 \\
R_1 & L_2 \\
& \ddots \\
& & R_{p-1} & L_p
\end{bmatrix}
$$

where $L_i$ is a banded lower triangular matrix of order $n/p$ and $R_i$ is a matrix of order $n/p$ whose nonzeros are confined to the upper right-hand corner which is an upper triangular matrix of order $m$. Let $R_i = (0, S_i)$ where $S_i$ consists of the last $m$ columns of $R_i$. Now partition $f$ and $x$ into $k = n/m$ pieces $f = (f_1^T, \ldots, f_k^T)^T$, and $x = (x_1^T, \ldots, x_k^T)^T$. Also define a grouping of these subvectors into $p$ partitions of length $n/p$ each, i.e. $\tilde{x}_i = (x_{j_1}^T, \ldots, x_{j_u}^T)^T$ where $j = (i-1)k + 1$ and $u = (k/p) - 1$ and similarly for $\tilde{f}_i$.  

Stage 0

\[ \begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4 \\
x_5 \\
x_6 \\
x_7 \\
x_8 \\
x_9 \\
\end{bmatrix} = \begin{bmatrix}
L_1 \\
\vdots \\
L_2 \\
\end{bmatrix}\begin{bmatrix}
\hat{x}_1 \\
\hat{x}_2 \\
\hat{x}_3 \\
\end{bmatrix} = \begin{bmatrix}
f_1 \\
f_2 \\
f_3 \\
\end{bmatrix}
\]

\[ n = 9m \quad p = 3 \]

= m by m upper triangular matrix

= m by m lower triangular matrix
Stage 1: Solve

\[ L_1 \hat{x}_1 = \hat{f}_1 \]

and for \( i = 2, \ldots, p \)

\[ L_i(G_i, \hat{h}_i) = (S_i, \hat{f}_i) \]
Stage 1

\[ \begin{bmatrix} I & I \\ & \ddots & \ddots & \ddots \\ & & I & I \\ & & & \ddots & \ddots & \ddots \\ & & & & I & I \\ & & & & & \ddots & \ddots & \ddots \\ & & & & & & I & I \\ & & & & & & & \ddots & \ddots & \ddots \\ & & & & & & & & I & I \\ & & & & & & & & & I \\ & & & & & & & & & & I \\ & & & & & & & & & & & I \\ & & & & & & & & & & & \end{bmatrix} \]

\[ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} \]

- \( I \) = m by m identity
- \( \) = dense matrix with m columns
- \( \) = m by m dense matrix

Equivalent to transformation by block diagonal matrix:

\[ \begin{bmatrix} L_1^{-1} & & & & \\ & L_2^{-1} & & & \\ & & L_3^{-1} & & \\ & & & R_1 & L_2 \\ & & & & R_2 L_3 \end{bmatrix} \]

also applied to \( \hat{\mathbf{f}} \) to give \( \hat{\mathbf{h}} \).
Now let

\[ G_i = \begin{pmatrix} M_i \\ N_i \end{pmatrix} \]

and \( \tilde{h}_i = (h_{j}^T, \ldots, h_{j+u}^T)^T \) where \( N_i \) is a square matrix of order \( m \), and \( j = (i - 1)k + 1 \) and \( u = (k/p) - 1 \).

Notice the unknowns above each of the partitioning lines of the matrix form an independent system (see the circled matrix and vector elements in the diagram of Stage 1). This system is called the reduced system and is solved in Stage 2. It is of the form

\[
\begin{bmatrix}
I & & \\
N_2 & I & \\
& \ddots & \ddots \\
& & N_p & I
\end{bmatrix}
\begin{bmatrix}
x_{k/p} \\
x_{2(k/p)} \\
\vdots \\
x_k
\end{bmatrix}
= 
\begin{bmatrix}
h_{k/p} \\
h_{2(k/p)} \\
\vdots \\
h_k
\end{bmatrix}
\]
Stage 2

\[
\begin{bmatrix}
I & I \\
\end{bmatrix}
\begin{bmatrix}
x_3 \\
x_6 \\
x_9 \\
\end{bmatrix} = \begin{bmatrix}
h_3 \\
h_6 \\
h_9 \\
\end{bmatrix}
\]

Reduced System of order \( pm \)
The remaining elements of the solution can now be recovered in parallel by simple matrix vector multiplication. If we let \( \hat{h}_i = (b_i^T, h_{(ik)/p})^T \) and \( \hat{x}_i = (y_i^T, x_{(ik)/p})^T \). Therefore, \( b_i \) and \( y_i \) are vectors of length \( (n/p) - m \). Stage 3 performs the \( p - 1 \) independent matrix vector products

\[
y_i = b_i - M_i x_{(i-1)k/p}
\]

for \( i = 2, \ldots, p \).
### Stage 3

Circled quantities known at start of this stage

Boxed equations can be solved in parallel to recover remaining unknowns
• Operations $\approx 4nm^2$ assuming unit lower triangular matrix

• Redundancy $\approx 2m$

• Time given $\hat{p} = pm$ processors is
  \[
  \frac{2}{\hat{p}} nm^2 + \frac{3}{\hat{p}} nm + O(m^2)
  \]

• If only the last component of the $x$ is required this reduces to
  \[
  \frac{2}{\hat{p}} nm^2 + \frac{1}{\hat{p}} nm + O(m^2)
  \]

• Speedup is bounded by $p$ in terms of levels of scalar computations in the computation graphs.

• Data locality often mitigates this pessimistic result