Numerical Algorithms
Matrices
An $n \times m$ matrix

\[
\begin{bmatrix}
  a_{0,0} & a_{0,1} & \cdots & a_{0,m-2} & a_{0,m-1} \\
  a_{1,0} & a_{1,1} & \cdots & a_{1,m-2} & a_{1,m-1} \\
  \vdots & \vdots & \ddots & \vdots & \vdots \\
  a_{n-2,0} & a_{n-2,1} & \cdots & a_{n-2,m-2} & a_{n-2,m-1} \\
  a_{n-1,0} & a_{n-1,1} & \cdots & a_{n-1,m-2} & a_{n-1,m-1}
\end{bmatrix}
\]
Matrix Addition

Involves adding corresponding elements of each matrix to form the result matrix.

Given the elements of \( A \) as \( a_{i,j} \) and the elements of \( B \) as \( b_{i,j} \), each element of \( C \) is computed as

\[
c_{i,j} = a_{i,j} + b_{i,j}
\]

\((0 \leq i < n, 0 \leq j < m)\)
Matrix Multiplication

Multiplication of two matrices, \( \mathbf{A} \) and \( \mathbf{B} \), produces the matrix \( \mathbf{C} \) whose elements, \( c_{i,j} \) \((0 \leq i < n, 0 \leq j < m)\), are computed as follows:

\[
c_{i,j} = \sum_{k=0}^{l-1} a_{i,k} b_{k,j}
\]

where \( \mathbf{A} \) is an \( n \times l \) matrix and \( \mathbf{B} \) is an \( l \times m \) matrix.
Matrix multiplication, $C = A \times B$
Matrix-vector multiplication follows directly from the definition of matrix-matrix multiplication by making $B$ an $n \times 1$ matrix (vector). Result an $n \times 1$ matrix (vector).
A system of linear equations can be written in matrix form:

$$Ax = b$$

Matrix $A$ holds the $a$ constants

$x$ is a vector of the unknowns

$b$ is a vector of the $b$ constants.
Implementing Matrix Multiplication
Sequential Code

Assume throughout that the matrices are square \((n \times n)\) matrices. The sequential code to compute \(A \times B\) could simply be

```python
for (i = 0; i < n; i++)
    for (j = 0; j < n; j++) {
        c[i][j] = 0;
        for (k = 0; k < n; k++)
            c[i][j] = c[i][j] + a[i][k] * b[k][j];
    }
```

This algorithm requires \(n^3\) multiplications and \(n^3\) additions, leading to a sequential time complexity of \(O(n^3)\). Very easy to parallelize.
With n processors (and n x n matrices), can obtain:

- Time complexity of $O(n^2)$ with n processors
  Each instance of inner loop independent and can be done by a separate processor

- Time complexity of $O(n)$ with $n^2$ processors
  One element of A and B assigned to each processor.
  Cost optimal since $O(n^3) = n \times O(n^2) = n^2 \times O(n)$.

- Time complexity of $O(\log n)$ with $n^3$ processors
  By parallelizing the inner loop. Not cost-optimal since $O(n^3) \neq n^3 \times O(\log n)$.

$O(\log n)$ lower bound for parallel matrix multiplication.
Partitioning into Submatrices

Suppose matrix divided into \( s^2 \) submatrices. Each submatrix has \( n/s \times n/s \) elements. Using notation \( A_{p,q} \) as submatrix in submatrix row \( p \) and submatrix column \( q \):

\[
\begin{align*}
\text{for (} p \text{ = 0; } p < s; p++) \\
\text{for (} q \text{ = 0; } q < s; q++) \} \\
\text{C}_{p,q} = 0; \\
\text{for (} r \text{ = 0; } r < m; r++) \\
\text{C}_{p,q} = \text{C}_{p,q} + A_{p,r} \times B_{r,q}; \\
\end{align*}
\]

The line

\[
\text{C}_{p,q} = \text{C}_{p,q} + A_{p,r} \times B_{r,q};
\]

means multiply submatrix \( A_{p,r} \) and \( B_{r,q} \) using matrix multiplication and add to submatrix \( C_{p,q} \) using matrix addition. Known as block matrix multiplication.
Block Matrix Multiplication
Submatrix multiplication

(a) Matrices

\[
\begin{bmatrix}
  a_{0,0} & a_{0,1} & a_{0,2} & a_{0,3} \\
  a_{1,0} & a_{1,1} & a_{1,2} & a_{1,3} \\
  a_{2,0} & a_{2,1} & a_{2,2} & a_{2,3} \\
  a_{3,0} & a_{3,1} & a_{3,2} & a_{3,3}
\end{bmatrix}
\times
\begin{bmatrix}
  b_{0,0} & b_{0,1} & b_{0,2} & b_{0,3} \\
  b_{1,0} & b_{1,1} & b_{1,2} & b_{1,3} \\
  b_{2,0} & b_{2,1} & b_{2,2} & b_{2,3} \\
  b_{3,0} & b_{3,1} & b_{3,2} & b_{3,3}
\end{bmatrix}
\]

(b) Multiplying $A_{0,0} \times B_{0,0}$ to obtain $C_{0,0}$

\[
A_{0,0} = \begin{bmatrix}
  a_{0,0} & a_{0,1} \\
  a_{1,0} & a_{1,1}
\end{bmatrix}, \quad B_{0,0} = \begin{bmatrix}
  b_{0,0} & b_{0,1} \\
  b_{1,0} & b_{1,1}
\end{bmatrix}
\]

\[
A_{0,0} \times B_{0,0} + A_{0,1} \times B_{1,0} = C_{0,0}
\]

\[
\begin{bmatrix}
  a_{0,0}b_{0,0} + a_{0,1}b_{1,0} & a_{0,0}b_{0,0} + a_{0,1}b_{1,1} \\
  a_{1,0}b_{0,0} + a_{1,1}b_{1,0} & a_{1,0}b_{0,0} + a_{1,1}b_{1,1}
\end{bmatrix}
\begin{bmatrix}
  a_{0,2} & a_{0,3} \\
  a_{1,2} & a_{1,3}
\end{bmatrix}
\begin{bmatrix}
  b_{2,0} & b_{2,1} \\
  b_{3,0} & b_{3,1}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
  a_{0,0}b_{0,0} + a_{0,1}b_{1,0} + a_{0,2}b_{2,0} + a_{0,3}b_{3,0} & a_{0,0}b_{0,1} + a_{0,1}b_{1,1} + a_{0,2}b_{2,0} + a_{0,3}b_{3,1} \\
  a_{1,0}b_{0,0} + a_{1,1}b_{1,0} + a_{1,2}b_{2,0} + a_{1,3}b_{3,0} & a_{1,0}b_{0,1} + a_{1,1}b_{1,1} + a_{1,2}b_{2,0} + a_{1,3}b_{3,1}
\end{bmatrix}
\]

\[
= C_{0,0}
\]
Direct Implementation

One processor to compute each element of $C$. One row of elements of $A$ and one column of elements of $B$ needed. Some elements sent to more than one processor. Can use submatrices.
Using tree construction $n$ numbers can be added in $\log n$ steps using $n$ processors:

Computational time complexity of $O(\log n)$ using $n^3$ processors.
Recursive Implementation

Apply same algorithm on each submatrix recursively.

Excellent algorithm for a shared memory systems because of locality of operations.
Recursive Algorithm

mat_mult(A, B, s)
{
    if (s == 1)  /* if submatrix has one element */
        C = A * B;  /* multiply elements */
    else {      /* continue to make recursive calls */
        s = s/2;  /* no of elements in each row/column */
        P0 = mat_mult(A_pp, B_pp, s);
        P1 = mat_mult(A_pq, B_qp, s);
        P2 = mat_mult(A_pp, B_pq, s);
        P3 = mat_mult(A_qq, B_qp, s);
        P4 = mat_mult(A_qp, B_pp, s);
        P5 = mat_mult(A_qq, B_qp, s);
        P6 = mat_mult(A_qp, B_pq, s);
        P7 = mat_mult(A_qq, B_qq, s);
        C_pp = P0 + P1;  /* add submatrix products to */
        C_pq = P2 + P3;  /* form submatrices of final matrix */
        C_qq = P4 + P5;
    }
    return (C);  /* return final matrix */
}
Mesh Implementations

- Cannon’s algorithm
- Fox’s algorithm (not in textbook but similar complexity)
- Systolic array

All involve using processor arranged a mesh and shifting elements of the arrays through the mesh. Accumulate the partial sums at each processor.
Mesh Implementations
Cannon’s Algorithm

Uses a mesh of processors with wraparound connections (a torus) to shift the A elements (or submatrices) left and the B elements (or submatrices) up.

1. Initially processor Pi,j has elements ai,j and bi,j (0 <= i < n, 0 <= k < n).
2. Elements are moved from their initial position to an “aligned” position. The complete ith row of A is shifted i places left and the complete jth column of B is shifted j places upward. This has the effect of placing the element ai,j+i and the element bi+j,j in processor Pi,j,. These elements are a pair of those required in the accumulation of ci,j.
3. Each processor, Pi,j, multiplies its elements.
4. The ith row of A is shifted one place right, and the jth column of B is shifted one place upward. This has the effect of bringing together the adjacent elements of A and B, which will also be required in the accumulation.
5. Each processor, Pi,j, multiplies the elements brought to it and adds the result to the accumulating sum.
6. Step 4 and 5 are repeated until the final result is obtained (n - 1 shifts with n rows and n columns of elements).
Movement of $A$ and $B$ elements
Step 2 — Alignment of elements of $A$ and $B$
Step 4 - One-place shift of elements of $A$ and $B$
Matrix-Vector Multiplication

Pumping action

\[ a_{0,3} \ a_{0,2} \ a_{0,1} \ a_{0,0} \]
\[ a_{1,3} \ a_{1,2} \ a_{1,1} \ a_{1,0} \]
\[ a_{2,3} \ a_{2,2} \ a_{2,1} \ a_{2,0} \]
\[ a_{3,3} \ a_{3,2} \ a_{3,1} \ a_{3,0} \]
Solving a System of Linear Equations

\[ a_{n-1,0}x_0 + a_{n-1,1}x_1 + a_{n-1,2}x_2 \quad \ldots \quad + a_{n-1,n-1}x_{n-1} = b_{n-1} \]

\[ \vdots \]

\[ a_{2,0}x_0 + a_{2,1}x_1 + a_{2,2}x_2 \quad \ldots \quad + a_{2,n-1}x_{n-1} = b_2 \]

\[ a_{1,0}x_0 + a_{1,1}x_1 + a_{1,2}x_2 \quad \ldots \quad + a_{1,n-1}x_{n-1} = b_1 \]

\[ a_{0,0}x_0 + a_{0,1}x_1 + a_{0,2}x_2 \quad \ldots \quad + a_{0,n-1}x_{n-1} = b_0 \]

which, in matrix form, is

\[ Ax = b \]

Objective is to find values for the unknowns, \( x_0, x_1, \ldots, x_{n-1} \), given values for \( a_{0,0}, a_{0,1}, \ldots, a_{n-1,n-1} \), and \( b_0, \ldots, b_n \).
Solving a System of Linear Equations

Dense matrices

Gaussian Elimination - parallel time complexity $O(n^2)$

Sparse matrices

By iteration - depends upon iteration method and number of iterations but typically $O(\log n)$

- Jacobi iteration
- Gauss-Seidel relaxation (not good for parallelization)
- Red-Black ordering
- Multigrid
Gaussian Elimination

Convert general system of linear equations into triangular system of equations to be solved by Back Substitution. Uses characteristic of linear equations that any row can be replaced by that row added to another row multiplied by a constant. Starts at the first row and works toward the bottom row. At the $i$th row, each row $j$ below the $i$th row is replaced by row $j + (\text{row } i) \left(-\frac{a_{j,i}}{a_{i,i}}\right)$. The constant used for row $j$ is $-\frac{a_{j,i}}{a_{i,i}}$. Has the effect of making all the elements in the $i$th column below the $i$th row zero because

$$ a_{j,i} = a_{j,i} + a_{i,i} \left( -\frac{a_{j,i}}{a_{i,i}} \right) = 0 $$
Gaussian elimination

Column

Row

Row $i$

Row $j$

Already cleared to zero

Cleared to zero

$\text{Column } i$

$\text{Step through}$

$\text{Column}$

$a_{ji}$
Partial Pivoting

If $a_{i,i}$ is zero or close to zero, we will not be able to compute the quantity $-a_{j,i}/a_{i,i}$.

Procedure must be modified into so-called partial pivoting by swapping the $i$th row with the row below it that has the largest absolute element in the $i$th column of any of the rows below the $i$th row if there is one. (Reordering equations will not affect the system.)

In the following, we will not consider partial pivoting.
Sequential Code

Without partial pivoting:

```c
for (i = 0; i < n-1; i++) /* for each row, except last */
    for (j = i+1; j < n; j++) {
        /* step through subsequent rows */
        m = a[j][i]/a[i][i]; /* Compute multiplier */
        for (k = i; k < n; k++) /* last n-i-1 elements of row j*/
            a[j][k] = a[j][k] - a[i][k] * m;
        b[j] = b[j] - b[i] * m; /* modify right side */
    }
```

The time complexity is $O(n^3)$. 
Parallel Implementation

Row

Column

Row i

n – i +1 elements (including b[i])

Already cleared to zero

Broadcast ith row
Analysis

Communication

$n - 1$ broadcasts performed sequentially. $i$th broadcast contains $n - i + 1$ elements.

Time complexity of $O(n^2)$ (see textbook)

Computation

After row broadcast, each processor $P_j$ beyond broadcast processor $P_i$ will compute its multiplier, and operate upon $n - j + 2$ elements of its row. Ignoring the computation of the multiplier, there are $n - j + 2$ multiplications and $n - j + 2$ subtractions.

Time complexity of $O(n^2)$ (see textbook).

Efficiency will be relatively low because all the processors before the processor holding row $i$ do not participate in the computation again.
Pipeline implementation of Gaussian elimination
Strip Partitioning

Poor processor allocation! Processors do not participate in computation after their last row is processed.
Cyclic-Striped Partitioning

An alternative which equalizes the processor workload:
Iterative Methods

Time complexity of direct method at $O(N^2)$ with $N$ processors, is significant.

Time complexity of iteration method depends upon:

- the type of iteration,
- number of iterations
- number of unknowns, and
- required accuracy

but can be less than the direct method especially for a few unknowns i.e a sparse system of linear equations.
Jacobi Iteration

Iteration formula - $i$th equation rearranged to have $i$th unknown on left side:

$$x_i^k = \frac{1}{a_{i,i}} \left[ b_i - \sum_{j \neq i} a_{i,j} x_j^{k-1} \right]$$

Superscript indicates iteration:

$x_i^k$ is $k$th iteration of $x_i$, $x_j^{k-1}$ is $(k-1)$th iteration of $x_j$. 

Example of a Sparse System of Linear Equations

Laplace’s Equation

\[ \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0 \]

Solve for \( f \) over the two-dimensional x-y space.

For a computer solution, *finite difference* methods are appropriate.

Two-dimensional solution space is “discretized” into a large number of solution points.
Finite Difference Method
If distance between points, $\Delta$, made small enough:

\[
\frac{\partial^2 f}{\partial x^2} \approx \frac{1}{\Delta^2} \left[ f(x + \Delta, y) - 2f(x, y) + f(x - \Delta, y) \right]
\]

\[
\frac{\partial^2 f}{\partial y^2} \approx \frac{1}{\Delta^2} \left[ f(x, y + \Delta) - 2f(x, y) + f(x, y - \Delta) \right]
\]

Substituting into Laplace’s equation, we get

\[
\frac{1}{\Delta^2} \left[ f(x + \Delta, y) + f(x - \Delta, y) + f(x, y + \Delta) + f(x, y - \Delta) - 4f(x, y) \right] = 0
\]

Rearranging, we get

\[
f(x, y) = \frac{[f(x - \Delta, y) + f(x, y - \Delta) + f(x + \Delta, y) + f(x, y + \Delta)]}{4}
\]

Rewritten as an iterative formula:

\[
f^k(x, y) = \frac{[f^{k-1}(x - \Delta, y) + f^{k-1}(x, y - \Delta) + f^{k-1}(x + \Delta, y) + f^{k-1}(x, y + \Delta)]}{4}
\]

$f^k(x, y)$ - $k$th iteration, $f^{k-1}(x, y)$ - $(k - 1)$th iteration.
Boundary points (see text)

\[
\begin{array}{cccccccccccc}
\bullet & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} \\
\bullet & x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} & x_{17} & x_{18} & x_{19} & x_{20} \\
\bullet & x_{21} & x_{22} & x_{23} & x_{24} & x_{25} & x_{26} & x_{27} & x_{28} & x_{29} & x_{30} \\
\bullet & x_{31} & x_{32} & x_{33} & x_{34} & x_{35} & x_{36} & x_{37} & x_{38} & x_{39} & x_{40} \\
\bullet & x_{41} & x_{42} & x_{43} & x_{44} & x_{45} & x_{46} & x_{47} & x_{48} & x_{49} & x_{50} \\
\bullet & x_{51} & x_{52} & x_{53} & x_{54} & x_{55} & x_{56} & x_{57} & x_{58} & x_{59} & x_{60} \\
\bullet & x_{61} & x_{62} & x_{63} & x_{64} & x_{65} & x_{66} & x_{67} & x_{68} & x_{69} & x_{70} \\
\bullet & x_{71} & x_{72} & x_{73} & x_{74} & x_{75} & x_{76} & x_{77} & x_{78} & x_{79} & x_{80} \\
\bullet & x_{81} & x_{82} & x_{83} & x_{84} & x_{85} & x_{86} & x_{87} & x_{88} & x_{89} & x_{90} \\
\bullet & x_{91} & x_{92} & x_{93} & x_{94} & x_{95} & x_{96} & x_{97} & x_{98} & x_{99} & x_{100} \\
\end{array}
\]
Relationship with a General System of Linear Equations

Using natural ordering, $i$th point computed from $i$th equation:

$$x_i = \frac{x_{i-n} + x_{i-1} + x_{i+1} + x_{i+n}}{4}$$

or

$$x_{i-n} + x_{i-1} - 4x_i + x_{i+1} + x_{i+n} = 0$$

which is a linear equation with five unknowns (except those with boundary points).

In general form, the $i$th equation becomes:

$$a_{i,i-n}x_{i-n} + a_{i,i-1}x_{i-1} + a_{i,i}x_i + a_{i,i+1}x_{i+1} + a_{i,i+n}x_{i+n} = 0$$

where $a_{i,i} = -4$, and $a_{i,i-n} = a_{i,i-1} = a_{i,i+1} = a_{i,i+n} = 1$. 
Those equations with a boundary point on diagonal unnecessary for solution

To include boundary values and some zero entries (see text)

\[
\begin{bmatrix}
1 & 1 & -4 & 1 & 1 \\
1 & 1 & -4 & 1 & 1 \\
a_{i,i-n} & a_{i,i-1} & a_{i,i} & a_{i,i+1} & a_{i,i+n} \\
1 & 1 & -4 & 1 & 1 \\
1 & 1 & -4 & 1 & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_{N-1} \\
x_N \\
0 \\
0
\end{bmatrix}
= 0
\]
Gauss-Seidel Relaxation

Uses some newly computed values to compute other values in that iteration.
Gauss-Seidel Iteration Formula

\[ x_i^k = \frac{1}{a_{i,i}} \left[ b_i - \sum_{j=1}^{i-1} a_{i,j} x_j^k - \sum_{j=i+1}^{N} a_{i,j} x_j^{k-1} \right] \]

where the superscript indicates the iteration.

With natural ordering of unknowns, formula reduces to

\[ x_i^k = (-1/a_{i,i})[a_{i,i-n} x_{i-n}^k + a_{i,i-1} x_{i-1}^k + a_{i,i+1} x_{i+1}^{k-1} + a_{i,i+n} x_{i+n}^{k-1}] \]

At the \( k \)th iteration, two of the four values (before the \( i \)th element) taken from the \( k \)th iteration and two values (after the \( i \)th element) taken from the \((k-1)\)th iteration. We have:

\[ f^k(x, y) = \frac{[f^k(x - \Delta, y) + f^k(x, y - \Delta) + f^{k-1}(x + \Delta, y) + f^{k-1}(x, y + \Delta)]}{4} \]
Red-Black Ordering

First, black points computed. Next, red points computed. Black points computed simultaneously, and red points computed simultaneously.
forall (i = 1; i < n; i++)
    forall (j = 1; j < n; j++)
        if ((i + j) % 2 == 0) /* compute red points */
            f[i][j] = 0.25*(f[i-1][j] + f[i][j-1] + f[i+1][j] + f[i][j+1]);
forall (i = 1; i < n; i++)
    forall (j = 1; j < n; j++)
        if ((i + j) % 2 != 0) /* compute black points */
            f[i][j] = 0.25*(f[i-1][j] + f[i][j-1] + f[i+1][j] + f[i][j+1]);
First, a coarse grid of points used. With these points, iteration process will start to converge quickly.

At some stage, number of points increased to include points of the coarse grid and extra points between the points of the coarse grid. Initial values of extra points found by interpolation. Computation continues with this finer grid.

Grid can be made finer and finer as computation proceeds, or computation can alternate between fine and coarse grids.

Coarser grids take into account distant effects more quickly and provide a good starting point for the next finer grid.
Multigrid processor allocation