

Numerical Solution of Steady-State Heat Equation

Jiaqi Li

November 2018

1 Governing Equations

In this document, we attempt to solve the steady-state heat equation in one- and two- dimensions.

In 1D, we have

$$-kT''(x) = q(x), \quad x \in \Omega \quad (1)$$

In 2D, we have

$$-k\nabla^2 T(x, y) = q(x, y), \quad (x, y) \in \Omega \quad (2)$$

where k is the thermal conductivity, T is temperature, and q is a given heat source term. To simplify the problem, we make the following assumptions:

- k is constant throughout the domain.
- $\Omega = (0, l)$ in 1D and $\Omega = (0, l) \times (0, l)$ in 2D.
- Dirichlet boundary condition:

$$T|_{\partial\Omega} = g \quad (3)$$

where $\partial\Omega$ denotes the boundary of domain Ω , and g is a given function defined on $\partial\Omega$.

From PDE theory we know that the combination of equation (1)/(2) and (3) forms a well-posed problem provided that q and g are smooth.

2 Discretization Schemes

Central difference is used to discretize the second-order derivative in the heat equation.

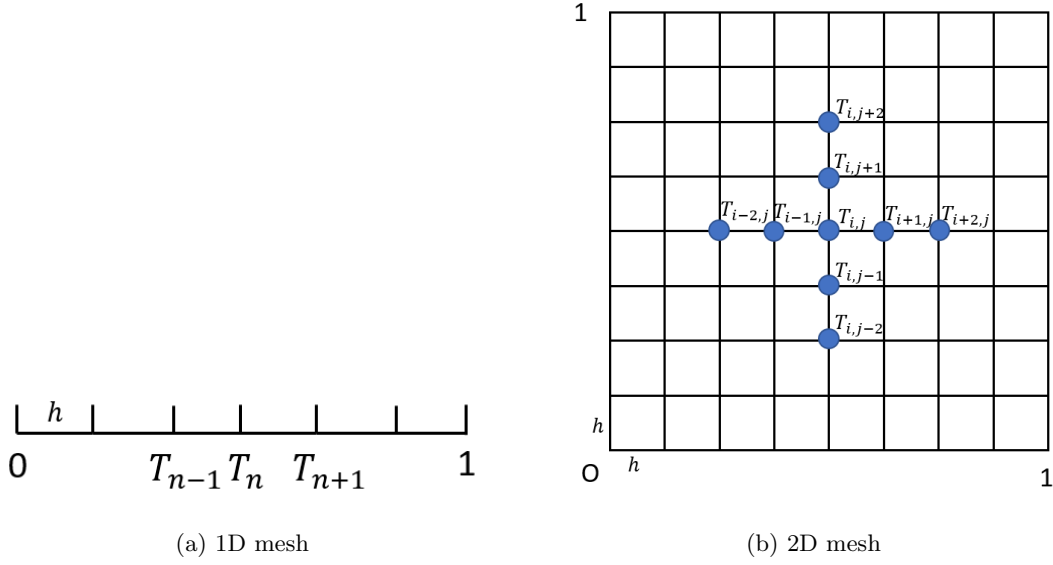


Figure 1: Representative figures of 1D and 2D mesh

2.1 1D case

In 1D case, we partition the domain into uniform intervals, as sketched in Figure 1(a). Assume $0 = x_0 < x_1 < \dots < x_m = l$, and $x_n - x_{n-1} = h$, $n = 1, 2, \dots, m$. The second-order approximation to the second-order derivative is

$$T''(x_n) = \frac{1}{h^2} (T_{n-1} - 2T_n + T_{n+1}) + O(h^2)$$

where $T_n = T(x_n)$. Then the discrete approximation to equation (1) is

$$\frac{1}{h^2} (T_{n-1} - 2T_n + T_{n+1}) + O(h^2) = -\frac{q_n}{k}$$

where $q_n = q(x_n)$. In order to get an approximate solution, we neglect the second-order term, and obtain

$$T_{n-1} - 2T_n + T_{n+1} = -\frac{q_n h^2}{k}, \quad n = 1, 2, \dots, m-1 \quad (4)$$

The boundary condition (3) implies

$$T_0 = g(0), \quad T_m = g(1) \quad (5)$$

Equations (4) and (5) form a solvable linear system.

The fourth-order scheme is

$$T''(x_n) = \frac{1}{h^2} \left(-\frac{1}{12}T_{n-2} + \frac{4}{3}T_{n-1} - \frac{5}{2}T_n + \frac{4}{3}T_{n+1} - \frac{1}{12}T_{n+2} \right) + O(h^4)$$

The corresponding discrete approximation to equation (1) is

$$\frac{1}{h^2} \left(-\frac{1}{12}T_{n-2} + \frac{4}{3}T_{n-1} - \frac{5}{2}T_n + \frac{4}{3}T_{n+1} - \frac{1}{12}T_{n+2} \right) + O(h^4) = -\frac{q_n}{k}$$

For practical calculation, we neglect the fourth-order term, obtaining

$$-\frac{1}{12}T_{n-2} + \frac{4}{3}T_{n-1} - \frac{5}{2}T_n + \frac{4}{3}T_{n+1} - \frac{1}{12}T_{n+2} = -\frac{q_n h^2}{k}, \quad n = 2, 3, \dots, m-2 \quad (6)$$

The boundary condition (5) still applies. However, we need two more equations to close the system. A simple choice is to use second-order difference at the nodes adjacent to the boundary.

$$T_{n-1} - 2T_n + T_{n+1} = -\frac{q_n h^2}{k}, \quad n = 1, m-1 \quad (7)$$

Equations (6) together with (5) and (7) form a solvable linear system.

2.2 2D case

In 2D case, we partition the domain uniformly in both x and y direction, as illustrated in Figure 1(b). Let $0 = x_0 < x_1 < \dots < x_m = l$, and $0 = y_0 < y_1 < \dots < y_n = l$. Let Δx and Δy be the mesh size in x and y direction, respectively. The second-order approximation to derivatives in the heat equation (2) is

$$\begin{aligned} \nabla^2 T(x_i, y_j) &= \left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} \right) T|_{(x_i, y_j)} \\ &= \frac{1}{\Delta x^2} (T_{i-1,j} - 2T_{i,j} + T_{i+1,j}) + \frac{1}{\Delta y^2} (T_{i,j-1} - 2T_{i,j} + T_{i,j+1}) + O(\Delta x^2) + O(\Delta y^2) \end{aligned}$$

The approximation to the equation is

$$\frac{1}{\Delta x^2} (T_{i-1,j} - 2T_{i,j} + T_{i+1,j}) + \frac{1}{\Delta y^2} (T_{i,j-1} - 2T_{i,j} + T_{i,j+1}) + O(\Delta x^2) + O(\Delta y^2) = -\frac{q_{i,j}}{k}$$

where $q_{i,j} = q(x_i, y_j)$. When $m = n$, and therefore $\Delta x = \Delta y = h$, the above equation can be simplified as

$$T_{i,j-1} + T_{i-1,j} - 4T_{i,j} + T_{i+1,j} + T_{i,j+1} = -\frac{q_{i,j} h^2}{k}, \quad i, j = 1, 2, \dots, n-1 \quad (8)$$

where we have omitted the second-order error $O(h^2)$. The boundary condition (3) implies

$$T_{0,j} = g(0, y_j), \quad T_{n,j} = g(1, y_j), \quad T_{i,0} = g(x_i, 0), \quad T_{i,n} = g(x_i, 1) \quad (9)$$

for $i = 1, 2, \dots, n-1$ and $j = 0, 1, \dots, n$. Equations (8) and (9) form a closed linear system.

The fourth-order approximation to derivatives is

$$\begin{aligned} \nabla^2 T(x_i, y_j) &= \frac{1}{\Delta x^2} \left(-\frac{1}{12}T_{i-2,j} + \frac{4}{3}T_{i-1,j} - \frac{5}{2}T_{i,j} + \frac{4}{3}T_{i+1,j} - \frac{1}{12}T_{i+2,j} \right) + O(\Delta x^4) \\ &\quad + \frac{1}{\Delta y^2} \left(-\frac{1}{12}T_{i,j-2} + \frac{4}{3}T_{i,j-1} - \frac{5}{2}T_{i,j} + \frac{4}{3}T_{i,j+1} - \frac{1}{12}T_{i,j+2} \right) + O(\Delta y^4) \end{aligned}$$

The approximation to the heat equation (2) is

$$\begin{aligned} & \frac{1}{\Delta x^2} \left(-\frac{1}{12}T_{i-2,j} + \frac{4}{3}T_{i-1,j} - \frac{5}{2}T_{i,j} + \frac{4}{3}T_{i+1,j} - \frac{1}{12}T_{i+2,j} \right) + O(\Delta x^4) \\ & + \frac{1}{\Delta y^2} \left(-\frac{1}{12}T_{i,j-2} + \frac{4}{3}T_{i,j-1} - \frac{5}{2}T_{i,j} + \frac{4}{3}T_{i,j+1} - \frac{1}{12}T_{i,j+2} \right) + O(\Delta y^4) = -\frac{q_{i,j}}{k} \end{aligned}$$

Assume $m = n$, $\Delta x = \Delta y = h$. Then the above equation can be simplified:

$$\begin{aligned} & -\frac{1}{12}T_{i,j-2} + \frac{4}{3}T_{i,j-1} - \frac{1}{12}T_{i-2,j} + \frac{4}{3}T_{i-1,j} - 5T_{i,j} \\ & + \frac{4}{3}T_{i+1,j} - \frac{1}{12}T_{i+2,j} + \frac{4}{3}T_{i,j+1} - \frac{1}{12}T_{i,j+2} = -\frac{q_{i,j}h^2}{k} \quad (10) \end{aligned}$$

for $i, j = 2, 3, \dots, n-2$, where we have omitted the fourth-order error $O(h^4)$. The boundary condition (9) still holds, but we need $(4n - 8)$ more equations to close the system. A simple choice is to use second-order difference at the nodes adjacent to the boundary:

$$T_{i,j-1} + T_{i-1,j} - 4T_{i,j} + T_{i+1,j} + T_{i,j+1} = -\frac{q_{i,j}h^2}{k} \quad (11)$$

where $1 \leq i, j \leq n-1$ and at least one of i, j equals 1 or $n-1$. Equations (10) together with (9) and (11) form a closed linear system.

3 Implementation of Numerical Methods

3.1 Resulting linear system

3.1.1 1D case

Let

$$\mathbf{u} = \begin{bmatrix} T_0 & T_1 & \dots & T_m \end{bmatrix}^T$$

be a column vector of unknowns. The second-order finite difference scheme (4), (5) gives the following tridiagonal linear system

$$\begin{bmatrix} 1 & & & & & \\ 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & \end{bmatrix} \begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ \vdots \\ T_{m-1} \\ T_m \end{bmatrix} = \begin{bmatrix} g(0) \\ -q_1 h^2/k \\ -q_2 h^2/k \\ \vdots \\ -q_{m-1} h^2/k \\ g(1) \end{bmatrix}$$

where zeros in the matrix are omitted. There are three non-zero entries on an interior row of the matrix.

The fourth-order difference scheme (6), (5), and (7) results in the following linear system

$$\begin{bmatrix} 1 & & & & & \\ 1 & -2 & 1 & & & \\ -\frac{1}{12} & \frac{4}{3} & -\frac{5}{2} & \frac{4}{3} & -\frac{1}{12} & \\ & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & -\frac{1}{12} & \frac{4}{3} & -\frac{5}{2} & \frac{4}{3} & -\frac{1}{12} \\ & & & 1 & -2 & 1 & \\ & & & & & 1 & \\ & & & & & & 1 \end{bmatrix} \begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ \vdots \\ T_{m-2} \\ T_{m-1} \\ T_m \end{bmatrix} = \begin{bmatrix} g(0) \\ -q_1 h^2/k \\ -q_2 h^2/k \\ \vdots \\ -q_{m-2} h^2/k \\ -q_{m-1} h^2/k \\ g(1) \end{bmatrix}$$

There are five non-zero entries on an interior row of the matrix.

3.1.2 2D case

Assume $m = n$. Let

$$\boldsymbol{\alpha}_j = [T_{0,j} \quad T_{1,j} \quad \dots \quad T_{n,j}]^T, \quad j = 0, 1, \dots, n$$

The second-order difference scheme (8), (9) gives

$$\mathbf{C}\boldsymbol{\alpha}_{j-1} + \mathbf{B}\boldsymbol{\alpha}_j + \mathbf{C}\boldsymbol{\alpha}_{j+1} = \mathbf{f}_j, \quad j = 1, 2, \dots, n-1 \quad (12)$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & & & & \\ 1 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 1 \\ & & & & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 0 \end{bmatrix}$$

are $(n+1) \times (n+1)$ matrices, and

$$\mathbf{f}_j = [T_{0,j} \quad -q_{1,j} h^2/k \quad \dots \quad -q_{n-1,j} h^2/k \quad T_{n,j}]^T$$

is column vector with length $(n+1)$. Rewriting equation (12) using block matrices, we obtain

$$\begin{bmatrix} \mathbf{I} & & & & \\ \mathbf{C} & \mathbf{B} & \mathbf{C} & & \\ & \ddots & \ddots & \ddots & \\ & & \mathbf{C} & \mathbf{B} & \mathbf{C} \\ & & & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_{n-1} \\ \boldsymbol{\alpha}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_0 \\ \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_{n-1} \\ \boldsymbol{\alpha}_n \end{bmatrix}$$

Note that $\boldsymbol{\alpha}_0$ and $\boldsymbol{\alpha}_n$ on the right hand side is given by boundary condition (9). There are five non-zero entries on an interior row of the matrix.

The second-order scheme (10), (9), (11) results in the following equations for α_j :

$$\mathbf{F}\alpha_{j-2} + \mathbf{E}\alpha_{j-1} + \mathbf{D}\alpha_j + \mathbf{E}\alpha_{j+1} + \mathbf{F}\alpha_{j+2} = \mathbf{f}_j, \quad j = 2, 3, \dots, n-2$$

where

$$\mathbf{D} = \begin{bmatrix} 1 & & & & & & \\ & 1 & -4 & 1 & & & \\ & -\frac{1}{12} & \frac{4}{3} & -5 & \frac{4}{3} & -\frac{1}{12} & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & -\frac{1}{12} & \frac{4}{3} & -5 & \frac{4}{3} & -\frac{1}{12} \\ & & & & 1 & -4 & 1 & \\ & & & & & & 1 & \\ & & & & & & & 1 \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} 0 & & & & & & \\ & 1 & & & & & \\ & & \frac{4}{3} & & & & \\ & & & \ddots & & & \\ & & & & \frac{4}{3} & & \\ & & & & & 1 & \\ & & & & & & 0 \end{bmatrix}, \mathbf{F} = \begin{bmatrix} 0 & & & & & & \\ & 0 & & & & & \\ & & -\frac{1}{12} & & & & \\ & & & \ddots & & & \\ & & & & -\frac{1}{12} & & \\ & & & & & 0 & \\ & & & & & & 0 \end{bmatrix},$$

are $(n+1) \times (n+1)$ matrices. Then the global system is

$$\begin{bmatrix} \mathbf{I} & & & & & & \\ \mathbf{C} & \mathbf{B} & \mathbf{C} & & & & \\ \mathbf{F} & \mathbf{E} & \mathbf{D} & \mathbf{E} & \mathbf{F} & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & \mathbf{F} & \mathbf{E} & \mathbf{D} & \mathbf{E} & \mathbf{F} \\ & & & \mathbf{C} & \mathbf{B} & \mathbf{C} & \\ & & & & & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{n-2} \\ \alpha_{n-1} \\ \alpha_n \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_{n-2} \\ \mathbf{f}_{n-1} \\ \alpha_n \end{bmatrix}$$

There are nine non-zero entries on an interior row of the matrix.

3.2 Iterative solutions for the resulting linear system

Once the linear system is formed, we utilize iterative methods to solve the system. Suppose the system is

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where \mathbf{A} is an $n \times n$ matrix.

3.2.1 Jacobi method

Algorithm 1 Jacobi method

```

 $\mathbf{x}^{(0)} \leftarrow 0$ 
for  $k = 0, 1, \dots$  do

    
$$x_i^{(k+1)} \leftarrow \frac{1}{a_{ii}} \left( b_i - \sum_{j \neq i} a_{ij} x_j^{(k)} \right), \quad i = 1, \dots, n$$


    if  $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < tol$  then
        exit
    end if
end for

```

3.2.2 Gauss-Seidel method

Algorithm 2 Gauss-Seidel method

```

 $\mathbf{x}^{(0)} \leftarrow 0$ 
for  $k = 0, 1, \dots$  do

    
$$x_i^{(k+1)} \leftarrow \frac{1}{a_{ii}} \left( b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right), \quad i = 1, \dots, n$$


    if  $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < tol$  then
        exit
    end if
end for

```

3.3 Required memory

Denote the total degrees of freedom by N , and let n be the number of grid points in one direction. Then $N = n$ in 1D, and $N = n^2$ in 2D. We need to store two discrete temperature fields (to calculate the difference between two iterations) and one source term field, which take up $3 \times 8N = 24N$ Bytes using double precision.

Besides, we need to store the non-zero entries in the matrix A and their positions. In our code, we store a sparse matrix A as its N rows and each row stores the position of non-zero entries (an integer array) and their values (array of double precision real numbers). Suppose the number of non-zero entries on each row is r , then each row would take up $r \times (4N + 8N) = 12rN$ Bytes.

From Section 3.1 we know that in 1D, $r = 3$ for second-order difference and $r = 5$ for fourth-order difference; in 2D, $r = 5$ for second-order difference and $r = 9$ for fourth-order difference.

Summing up the required memory for vectors and the matrix, we obtain Table 1.

	2nd order	4th order
1D	$60n$	$84n$
2D	$84n^2$	$132n^2$

Table 1: Required memory in different cases

4 Program Usage

This section describes how to build the program, input options, and verification procedures.

4.1 Build procedures

We use an autotools-based build system. With the configure script generated by autotools, we can generate Makefiles with the following commands (on Stampede2):

```
$ module load intel boost
$ export PKGPATH=/work/00161/karl/stampede2/public/
$ ./configure FC=ifort --with-masa=$PKGPATH/masa-intel-0.50 \
    --with-grvy=$PKGPATH/grvy-intel-0.34
```

Alternatively, the user can run the shell script “configure.sh”, which does exactly the same thing. After successful invocation of “configure”, we can proceed with the command “make” and build the code. Then a binary executable named “heateq” should be found in src/ directory.

4.2 Input options

The name of input file can be specified as a command line argument; if no argument is provided, the default name “input.dat” is assumed. We utilize the GRVY library for input parsing. An example of input file is provided below.

```
# input.dat: an input file example
dimen = 2                # dimension of the problem (1 for 1D, 2 for 2D)
side_Length = 1.0        # side length of domain
num_Mesh = 32            # number of meshes in one direction

order = 4                # order of discretization (2 or 4)
```



```

solver_Flag = 2          # solver type (1 - Jacobi, 2 - Gauss-Seidel)

verification_Flag = 1    # flag for verification mode (1 - on, 0 - off)
debug_Flag = 1           # flag for debug mode (0 - off, 1 - standard, 2 - verbose)

k_0 = 1.0                # thermal conductivity

eps = 1.0e-12            # iterative solver tolerance
max_Iter = 250000         # max solver iterations
print_Iter = 1000        # print error every print_iter iterations
output_File = 'sol.dat'  # name of output file

```

With those comments marked with “#”, the meaning of various input options is clear.

4.3 Verification procedures

In our program, we use a manufactured solution for the steady-state heat equation available within the MASA library. Simply set `verification_Flag` to 1 in the input file to enable the verification mode. The program will call MASA to get the source term $q(x, y)$, and then solve the heat equation as described before. Finally, we compare the difference between our numerical solution and the exact solution from MASA, and output their normalized L^2 difference. An example of the stdout is shown below:

```

L2 error between numerical solution and MASA:
  7.767339199626751E-006

```

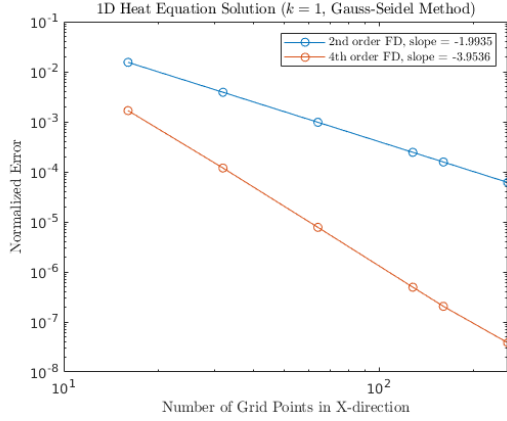
5 Program Verification and Performance

5.1 Verification results

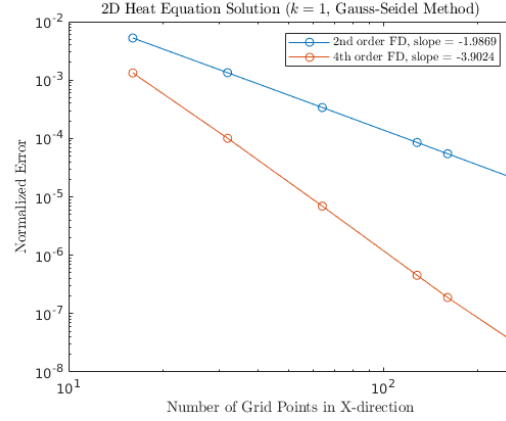
In both 1D and 2D, we perform a uniform mesh refinement study: $n = 16, 32, 64, 128, 160, 256$. In Figure 2 we present the results for Gauss-Seidel method. From Figure 2(a), we see that in 1D, the slope for second-order scheme is -1.9935, and the slope for fourth-order scheme is -3.9536. From Figure 2(b), we see that in 2D, the slope for second-order scheme is -1.9869, and the slope for fourth-order scheme is -3.9024. They are all very close to the theoretical value of -2 and -4, confirming the correctness of our program.

5.2 Runtime performance

GRVY library is utilized to measure runtime performance. In Figure 3, we present the result for 1D problem, where `num_Mesh` = 256, and we use second-order finite difference with Gauss-Seidel



(a) 1D problem



(b) 2D problem

Figure 2: Convergence plot for 1D and 2D problem

method to solve the linear system. From the result we see that the iterative solver (which updates the solution vector) takes up most of the time (89% here), and other part of Solve_System takes a large part (10%) too. In each iteration, besides calling functions to update the solution vector, the subroutine Solve_System computes the difference between two iterations, and decides whether to exit loop based on the difference.

```
Heat Equation Solver - Performance Timings:
--> Solve_System/Gauss-Seidel Iteration : 4.11932e+00 secs ( 89.3818 %) | [2.85210e-05 9.07937e-11 144431]
--> Solve_System                        : 4.67736e-01 secs ( 10.1490 %) | [4.67736e-01 0.00000e+00 1]
--> Initialize                          : 3.81494e-03 secs ( 0.0828 %) | [3.81494e-03 0.00000e+00 1]
--> Postprocess                         : 2.90084e-03 secs ( 0.0629 %) | [2.90084e-03 0.00000e+00 1]
--> Read Input                          : 2.33889e-03 secs ( 0.0507 %) | [2.33889e-03 0.00000e+00 1]
--> GRVY_Unassigned                     : 1.25675e-02 secs ( 0.2727 %)
Total Measured Time = 4.60868e+00 secs (100.0000 %)
```

Figure 3: Runtime performance measurements