

Lattice Codes and Sphere Decoding

Pasi Pyrrö

School of Science

Bachelor's special assignment

Espoo 10.7.2017

Thesis supervisor:

Prof. Camilla Hollanti

Thesis advisors:

Prof. Marcus Greferath

D.Sc. Oliver Gnilke

Author: Pasi Pyrrö

Title: Lattice Codes and Sphere Decoding

Date: 10.7.2017

Language: English

Number of pages: 4+15

Degree programme: Mathematics and Systems Analysis

Supervisor: Prof. Camilla Hollanti

Advisors: Prof. Marcus Greferath, D.Sc. Oliver Gnilke

Keywords: sphere decoding, space-time lattice codes, wireless communications

Contents

Abstract	ii
Contents	iii
Symbols and abbreviations	iv
1 Introduction	1
2 Lattices in communications technology	2
2.1 Closest vector problem	3
2.2 Space-time lattice codes	4
3 Sphere decoder	6
3.1 Recursive sublattice decomposition	7
4 Simulations and results	8
5 Summary	9
References	10
A Program User Guide	11
A.1 Abbreviations	11
A.2 Preface	11
A.3 System requirements	11
A.4 Installing the program	12
A.5 Setup and Input	13
A.6 Running the program	14
A.7 Output and Plots	15
A.8 General troubleshooting	15

Symbols and abbreviations

Symbols

\mathbb{Z}	Set of integers
\mathbb{C}	Field of complex numbers
\mathbf{x}	Vector
\mathbf{X}	Matrix

Operators

$\ \cdot\ $	Euclidean norm
$\det(\mathbf{X})$	Determinant of matrix \mathbf{X}

Abbreviations

SNR	Signal to noise ratio
CVP	Closest vector problem
MIMO	Multiple input multiple output

1 Introduction

Wireless communication has been a crucial part of modern information technology for a couple of decades now. It is facing a lot of practical everyday problems which motivate the ongoing extensive research on the field. One of these problems is the noise and fading that occurs on wireless channels due to obstacles and radiation from the surroundings. To avoid data loss during transmission via wireless link one has to encode the data to be sent in such a robust way that it can still be recovered at the receiving end even in the presence of noise and fading of reasonable scale.

One way to tackle this problem, and the method this thesis focuses on, is the use of space-time lattice codes and sphere decoding. From a mathematical point of view the process of decoding can then be viewed as a problem of finding the closest lattice point to a given input vector, that is, the possibly noisy vector containing the data we receive from the wireless channel. This problem in its general form is known to be NP-hard but for communications applications, where the dimensionality and lattice shape are kept reasonable, there exist algorithms, like the sphere decoder, that offer polynomial expected complexity [1].

2 Lattices in communications technology

Before we go into communications applications of lattices such as lattice codes, let us start off with the definition of a lattice. Let n and m be positive integers such that $n \leq m$ and $\mathbf{b}_1, \dots, \mathbf{b}_n \in \mathbb{R}^m$ be linearly independent vectors. A subset Λ of \mathbb{R}^m is called a lattice of dimension n if it is defined as

$$\Lambda = \sum_{i=1}^n \mathbb{Z} \mathbf{b}_i = \{a_1 \mathbf{b}_1 + \dots + a_n \mathbf{b}_n \mid a_i \in \mathbb{Z}, 1 \leq i \leq n\} \quad (1)$$

where the set of vectors $\mathbf{b}_1, \dots, \mathbf{b}_n \in \mathbb{R}^m$ is called the basis of the lattice Λ . All points of the lattice can be obtained as a linear combination of the basis vectors \mathbf{b}_i and integer coefficients \mathbf{a}_i as stated in (1). The number of basis vectors n is also called the rank of the lattice and if $n = m$ the lattice is said to have full rank. The points of the lattice Λ form a group under addition which means that if $\mathbf{x} \in \Lambda$ then $-\mathbf{x} \in \Lambda$ and if $\mathbf{x}, \mathbf{y} \in \Lambda$ then $\mathbf{x} \pm \mathbf{y} \in \Lambda$. [2]

There is also a matrix representation for the same lattice

$$\Lambda = \{\mathbf{x} \mid \mathbf{x} = \mathbf{B}\mathbf{a}\} \quad (2)$$

where $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]$ is an $m \times n$ matrix called the generator matrix of the lattice Λ and \mathbf{a} is an n -dimensional integer column vector. Note that \mathbf{B} is not uniquely determined by the lattice as in fact there exists infinitely many bases for the same lattice [2]. If \mathbf{B} is a basis for lattice Λ then \mathbf{B}' is also a basis for the same lattice if the following holds

$$\mathbf{B}' = \mathbf{W}\mathbf{B}, \quad (3)$$

$$\det(\mathbf{W}) = \pm 1 \quad (4)$$

where \mathbf{W} is an $m \times m$ matrix with integer entries [3]. Some bases are however better in some sense than others, especially in communications applications, as one could imagine. Usually reasonable orthogonality and relatively small norm for the basis vectors are desired [3]. A better basis can be obtained via a process called basis reduction, which is illustrated in figure 1. Clearly basis vectors \mathbf{u}_i obtained from the basis reduction are more pleasant to work with than the original vectors \mathbf{v}_i although they both span the same lattice.

What we considered earlier were lattices spanned over real m -dimensional space \mathbb{R}^m but in similar sense we can consider a lattice consisting of complex valued vectors from \mathbb{C}^m . The principles of such complex lattice are almost the same, however, now \mathbf{B} and \mathbf{a} take values from \mathbb{C} . Note that a complex lattice has an equivalent real representation which has double the rank of the corresponding complex lattice. Consider a lattice $\Lambda \subset \mathbb{C}^m$ with a generator matrix $\mathbf{B} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$. Now we can always express it with a real valued lattice $\Lambda_{\text{real}} \subset \mathbb{R}^{2m}$ and the corresponding $2m \times n$ generator matrix is given by

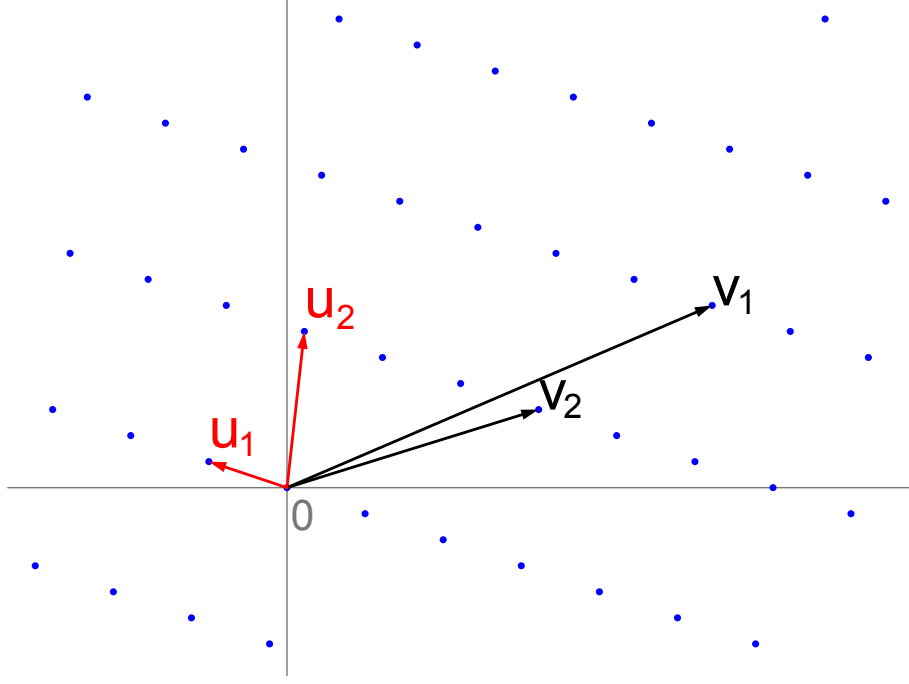


Figure 1: Two different bases for the same two dimensional lattice.

$$\mathbf{B}_{\text{real}} = \begin{bmatrix} \text{Re}(z_{11}) & \dots & \text{Re}(z_{1n}) \\ \text{Im}(z_{11}) & \dots & \text{Im}(z_{1n}) \\ \vdots & \ddots & \vdots \\ \text{Re}(z_{m1}) & \dots & \text{Re}(z_{mn}) \\ \text{Im}(z_{m1}) & \dots & \text{Im}(z_{mn}) \end{bmatrix}. \quad (5)$$

In other words we convert each column of \mathbf{B} to real vector by separating each of their complex elements into two adjacent real elements, that is the real and imaginary parts of the original complex element. This process doubles the amount of rows in \mathbf{B}_{real} but both representations ultimately describe the same lattice. [4]

2.1 Closest vector problem

One relevant mathematical problem related to sphere decoding, a communications application of interest in this paper, is the closest vector problem (CVP). Given a lattice $\Lambda \subset J$ and an input point $\mathbf{y} \in J$ the problem is to find a lattice point $\hat{\mathbf{x}} \in \Lambda$ that is closest to \mathbf{y} . More precisely $\hat{\mathbf{x}} \in \Lambda$ has to meet the following condition

$$\|\mathbf{y} - \hat{\mathbf{x}}\| \leq \|\mathbf{y} - \mathbf{x}\|, \quad \forall \mathbf{x} \in \Lambda. \quad (6)$$

J is the vector space, in which Λ belongs in. In this thesis we mostly consider $J = \mathbb{C}^n$.

For a fixed point $\mathbf{x}' \in \Lambda$ the set of vectors $\mathbf{y}_i \in J$ that satisfy the inequality (6) is called the Voronoi region (see figure 2) of lattice point \mathbf{x}' , $\mathcal{V}_{\mathbf{x}'} \subset J$. This means that if $\mathbf{y} \in \mathcal{V}_{\mathbf{x}'}$ then the solution to CVP is \mathbf{x}' as can be clearly seen from figure 2.

If $\mathbf{y} \in \partial\mathcal{V}_{\mathbf{x}'}$ the solution to CVP is a random choice between points whose Voronoi region contains that boundary. The volume of the $\mathcal{V}_{\mathbf{x}'}$, also known as the lattice constant as it is independent of the choice of basis, is given by $\det(\Lambda) = \sqrt{\det(\mathbf{B}^T\mathbf{B})}$, where \mathbf{B} is the generator matrix of Λ .



Figure 2: The voronoi cells of the two dimensional hexagonal lattice A_2 .

CVP is trivial for orthogonal lattices as one could just round every coordinate of \mathbf{y} to nearest integer and get the correct answer, but CVP has been shown to be NP-hard as the function of the lattice rank in general case when we consider skewed lattices and thus all known algorithms for CVP have exponential worst case complexity [3]. Non-trivial lattices appear in most communications applications, for example when using maximum likelihood decoding for a lattice codes sent via channel that induces fading. Such decoding process can be reduced to CVP and it's later explained in further detail.

2.2 Space-time lattice codes

Space-time lattice codes are a robust way of encoding the data on a wireless channel that uses multiple transmit and receiver antennas, i.e. a MIMO-system. Wireless channels are usually very error prone due to interference and fading caused by the surrounding electrical devices, obstacles and nature. This method helps to improve the reliability of the data transmission on such channel, meaning we get decoding errors less likely. A single codeword from certain space-time lattice code constellation can be represented as a matrix that is defined like so

$$\mathcal{L} = \left\{ \sum_{i=1}^k a_i \mathbf{X}_i \mid \mathbf{X}_i \in \mathcal{M}_{m \times l}(\mathbb{C}), a_i \in S \subset \mathbb{Z} \right\}. \quad (7)$$

The $\mathbf{X}_i, i = 1, \dots, k$ denote the constant basis matrices of the lattice code and the coefficients $a_i, i = 1, \dots, k$ are integers from a finite signal set S which represent the data we want to send.

3 Sphere decoder

In this chapter we will discuss an efficient decoding algorithm for the space-time block codes known as the sphere decoder. This thesis is particularly interested in the Schnorr-Euchner implementation of the algorithm which combines the Pohst strategy and the Babai nearest plane algorithm [3]. The problem of decoding the correct codeword from the set of all possible codewords from the received noisy codeword can be reduced to CVP that was discussed earlier. What the sphere decoder basically does is that instead of searching through all lattice points within the finite signal set boundaries it only considers lattice points within a hypersphere around the received vector of certain radius. This significantly reduces the complexity of the search algorithm.

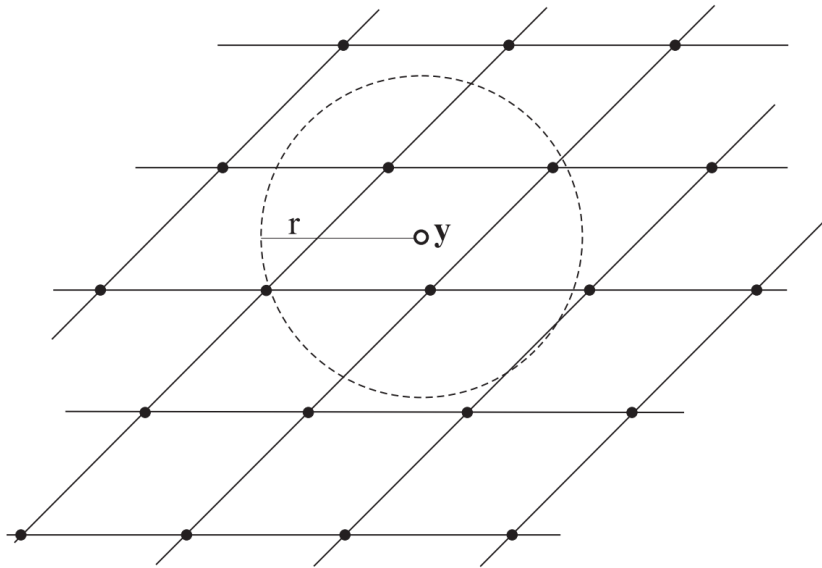


Figure 3: The CVP with sphere decoding in two dimensional lattice. The search complexity can be reduced by giving an upper bound for the closest point distance [1].

One question remains though, so how do we choose the radius? If it's too small, no lattice points lie within the sphere or if it's too large, the complexity will increase and slow down the algorithm. The answer to that question is, with our Schnorr-Euchner implementation of the algorithm, we can simply set it to infinity. It uses by definition the Babai nearest plane algorithm, which requires no upper bound for the distance to the closest point, to find a nearby point which we can use to dynamically update the upper bound of the search algorithm to the distance to that point as clearly the closest point is either that point we just found, also known as the Babai point, or one within that radius.

3.1 Recursive sublattice decomposition

For understanding how the sphere decoder works we introduce the concept of sublattice decomposition from n -dimensional lattice to a stack of $(n - 1)$ -dimensional sublattices. As one could guess this can also be done recursively and ultimately we end up with 0-dimensional lattices also known as the lattice points. The motivation to this recursive representation of a lattice is that one only needs to consider intervals of sublattice layer indices when doing the distance comparison in each dimension from n to 0.

Let us consider a $m \times n$ lattice generator matrix \mathbf{G} for a lattice Λ and $m \times (n - 1)$ matrix \mathbf{G}' so that

$$\mathbf{G} = [\mathbf{G}' \quad \mathbf{v}_n] \quad (8)$$

where the last column vector can be written as $\mathbf{v}_n = \mathbf{v}_{\parallel} + \mathbf{v}_{\perp}$, with \mathbf{v}_{\parallel} in the column space of \mathbf{G}' and \mathbf{v}_{\perp} in the null space of \mathbf{G}' . With this notation we can decompose any n -dimensional lattice as follows

$$\Lambda(\mathbf{G}) = \bigcup_{u_n=-\infty}^{+\infty} \left\{ \mathbf{c} + \mathbf{u}_n \mathbf{v}_{\parallel} + u_n \mathbf{v}_{\perp} \mid \mathbf{c} \in \Lambda(\mathbf{G}') \right\} \quad (9)$$

4 Simulations and results

5 Summary

References

- [1] Mäki, M. *Space-time block codes and the complexity of sphere decoding*. Doria, Referenced 10.7.2017. Available: <https://www.doria.fi/bitstream/handle/10024/54404/gradu2008maki-miia.pdf>
- [2] Cassels, J.W.S. *An introduction to the Geometry of Numbers*, New York, Springer–Verlag, 1971.
- [3] Agrell, E., Eriksson, T. and Zeger, K. *Closest point search in lattices* IEEE Transactions on Information Theory, Vol. 48, pp.2201-2214, August 2002.
- [4] Conway, J.H. and Sloane, N.J.A. *Sphere packings, lattices and groups*. Third edition, New York, Springer–Verlag, 1998.

A Program User Guide

A.1 Abbreviations

SNR	Signal to noise ratio
CPU	Central processing unit
RAM	Random access memory
GPU	Graphical processing unit
C++	C++ programming language
OS	Operating system

A.2 Preface

The purpose of this program is to simulate the performance of different space–time lattice codes under different levels of SNR. It is a command line application that uses simple text based files for input and output. It also has a optional plotting utility which provides graphical interface for the simulation results. The program is primarily intended to be used on Linux based operating systems but should be portable to other platforms as well (there are some slight differences in the required compiler instructions and used libraries though). This guide assumes Linux based operating system, the steps described here might differ slightly when the program is used on other operating systems. The details regarding to those differences are not discussed here. In the following chapters we will go through the installation and usage of the program in detail.

A.3 System requirements

Regarding the hardware as the rule of the thumb, the better the performance your computer has the faster the simulations, although nearly all modern computers with a keyboard and a monitor should suffice for small simulations. The program makes use of parallel CPU cores so one does not need to worry too much about single thread performance of the CPU. The program should not be too memory consuming except in some special use cases so no excessive amount of RAM is needed for fast simulations. GPU support might also be implemented in the future but for now the choice of GPU is irrelevant regarding the simulations.

Before trying to install and run the program itself, you need to make sure you have the all the required libraries and a compatible C++ compiler installed. The recommended GNU C++ compiler (g++) should be preinstalled in most Linux distributions but in case it is not, please refer to installation notes here: <https://gcc.gnu.org/wiki/InstallingGCC>.

The next thing is to gather the required library packages:

1. OpenBLAS library: <https://github.com/xianyi/OpenBLAS/wiki/Installation-Guide>
2. LAPACK – Linear Algebra PACKage: <http://www.netlib.org/lapack>
3. ARPACK library: <http://www.caam.rice.edu/software/ARPACK>

4. Armadillo C++ Linear algebra library: <http://arma.sourceforge.net/download.html>
5. Boost C++ library (optional, for plotting only): <http://www.boost.org>

You can follow the links for detailed instructions on how install those packages if you're facing problems (e.g. you're not installing the program on Linux OS) but the most straightforward way of doing this is by opening a terminal in Linux (CTRL+ALT+T) and using the package manager utility (e.g. apt-get in Ubuntu) to install those packages. The first three packages are required by the Armadillo library so you should install the packages in the given order. Also note that not Linux based operating systems might require different libraries for Armadillo to work, in this case please refer to the Armadillo documentation by following the given link.

In the Linux package manager the libraries listed should go by names: *libopenblas-dev*, *liblapack-dev*, *libarpack-dev*, *libarmadillo-dev* and *libboost-all-dev*. Note that there might be some version number before the first dash in those names. The Boost library can safely be omitted for compability reasons, but requires a few modifications in the source code and Makefile for the compilation process. If all the install processes succeed you are now all set up to install the program itself.

A.4 Installing the program

Open terminal (CTRL+ALT+T) and create a folder for the program

```
$ mkdir sphere-decoder
$ cd sphere-decoder
```

Once you are in that folder copy the contents of this Gitlab repository there: <https://version.aalto.fi/gitlab/pasi.pyrro/sphere-decoder>. This can be done by either downloading the repository as a compressed folder (e.g. zip) and extracting it into the folder created earlier or by using the following command in the terminal:

```
$ git clone git@version.aalto.fi:pasi.pyrro/sphere-decoder.git
```

After copying the files check the contents of that folder by typing this command in the terminal:

```
$ ll
```

Check if the contents of that folder look similar to those of the gitlab repository. If everything looks all right and you're using Linux with Boost library, just type the following in terminal

```
$ make
```

to compile the program. For other operating systems you likely need to modify the Makefile located in the `/src/` directory. If you didn't install the Boost C++ library you need to do the following in order to compile the program:


```
$ nano src/misc.hpp
```

In the text editor locate a line of code saying

```
#define PLOTTING // enable plotting (requires boost C++ library)
```

and either remove it or comment it out (type `'//'` in front of the line). After that save and exit by pressing CTRL+O, ENTER and CTRL+X. After that type the command:

```
$ nano src/Makefile
```

and remove all the compiler flags that have the word `'lboost'` in it. Save changes and exit the text editor as before. Now you should be able to compile the program by typing:

```
$ make
```

If the compiling succeeded, now check if the program runs correctly by typing:

```
$ make run
```

If there are no errors regarding missing libraries or such then you have just correctly installed the program! If there are such errors during the compilation or program runtime, go back to previous section and double check that you have installed the correct libraries.

A.5 Setup and Input

Now that the program runs it is time to take a look at the simulation parameters and the program input to make reasonable simulations. In the `/bases/` folder one can save text files containing the lattice code basis matrices for the simulation which the program takes as input. The basis file for the Alamouti code should already be there and is used as the default. When creating a new basis file, one should stick to either Mathematica or Matlab syntax for complex matrices to ensure correct interpretation of the input.

For simulation parameters and other options one should navigate to `/settings/` folder and open the default settings file called *settings.ini*. The contents of the file are displayed in figure A1. After `'//'` begins a comment sequence which is ignored by the program when reading the settings file. Make sure not to remove any of the 15 option lines or the program will give you an error. In case you end up messing up the settings file in some way, easy way to reset it is to just delete it and run the program normally. This will generate a new default settings file for you, like the one in figure A1.

When editing the settings, only modify the right hand side of the equality sign. If you change the variable names, the program will work incorrectly. You can always check the correct variable names from figure A1 or just generate a new default settings file if you happen to change them accidentally. Another rule of the thumb is that if

```
// configuration settings and simulation parameters for the sphere decoder program //

basis_file=alamouti.txt           // Text file containing the basis matrices
output_file=auto                  // Optionally specify the output filename (auto = automatic)
x-PAM=4                           // The size of the PAM signaling set
energy_estimation_samples=-1      // Number of samples to make the code energy estimation (-1 = sample all)
no_of_matrices=4                  // Number of basis matrices
no_of_transmit_antennas=2         // Number of broadcast antennas
no_of_receiver_antennas=2        // Number of receiver antennas
snr_min=6                         // Minimum value for signal-to-noise ratio
snr_max=20                       // Maximum value for signal-to-noise ratio
snr_step=2                       // Increase SNR by this value per each iteration
simulation_rounds=10000          // Number of simulation rounds to run
required_errors=-1               // Demand at minimum this many errors before the simulation ends
time_slots=2                     // Number of time slots used in the code
plot_results=1                   // Draw plots? (1 = yes, -1 = no)
spherical_shaping_max_power=75   // Defines the maximum distance from origin for codebook elements (-1 = unbounded)
```

Figure A1: The default settings file for the sphere decoder program.

you want to use default behaviour for some feature, use -1 value for that variable. The order of the setting variables does not matter, just make sure they're all on their own separate line. Remember to save the changes you make to that file. The program doesn't need to be recompiled for these changes to take effect, just run it again.

One can also create multiple settings file in that folder. To use a different settings file one just needs to give its name as a command line argument for the program.

A.6 Running the program

Once the program is set up correctly, running the program is a relatively easy task. Just make sure you're in the main folder of the local repository and type the following in terminal:

```
$ make run
```

or

```
$ ./sphdec [settings_file.ini*]
```

The first command uses the default settings file and in the second one can optionally specify the name of the settings file to be used (do not give the whole file path, just the name of a file in /settings/ folder). If no command line argument is given the two commands are equal.

Once the program has started it prints useful data related to the current actions it is taking to the terminal as well as in the log.txt file located in the /logs/ folder. Visiting the log.txt file can be useful for reviewing previous program behaviour in the future. In the log file each program runtime should be separated by an empty line.

If the program gets stuck somewhere or takes too long to finish a large simulation, pressing CTRL+C in the terminal will terminate the execution of the program instantly. Although doing this may result in a loss of data calculated so far. For large simulations one needs to be patient as the simulation complexity is NP-hard as the function of lattice code dimension.

A.7 Output and Plots

If the program exits without errors its output should be found from /output/ folder in csv format. The filenames are by default named so that they are in chronological order so the latest simulation output should be the last one in the file listing. The csv format is easy to import in other programs like Matlab.

If the program was compiled successfully with the PLOTTING flag enabled and the following settings variable is set to

```
plot_results=1
```

then a couple of graphical interfaces for plots generated from the simulation data should pop up. If there are errors check that you also have GNUPlot installed on your system. Plotting gives you a quick review on how the simulation went. For more rigorous data analysis it is recommended to import the output csv file to some numerical scientific calculation software like Matlab.

A.8 General troubleshooting

Some general miscellaneous tips are listed here to help you troubleshoot errors while using the program:

1. Do not include the dollar sign in terminal commands. It's just there to indicate terminal input.