1. Methodology

**2.01) Importing I-V data to MATLAB workspace**

To fulfil the goal of high throughput characterization, a MATLAB routine that imports multiple I-V data files to the workspace was paramount. The Biologic Science Instruments VMP3 16 -channel potentiostat instrument will be used to measure I-V characteristic for a solar cell under one sun illumination 1. This measurement device is expected to export raw I-V data in TXT files, and therefore this program was designed to read TXT file. The program expects the TXT file to contain V and I data in the first and second column respectively with no headers; the I and V data were then stored temporarily in separate single column arrays (vectors) and used for subsequent analysis.

In standard MATLAB coding practice, the main script, function files and other data to be imported are all grouped in the same folder. However, the program is expected to be used for datasets by multiple researchers, and therefore a likelihood that some important function files may be deleted in the process of removing old TXT files to accommodate new ones exists. Hence, separate folders were made for the function files and the data files; the folders must be labelled as ‘***Funcfiles’*** and ‘***Datafiles’*** otherwise the program will return an error.

Once the TXT files have been uploaded to ‘Datafiles’ the main script is run, and all the I-V data in the TXT files are characterized. For each TXT file, the experimental and fitted data were plotted on the same figure and and were stored in a TXT file for future reference.

**2.02) Extracting and from data**

To extract accurate and values, the raw current data is first smoothed using the “rlowess” smoothing function in MATLAB. This function fits a first order polynomial and minimizes the weighted least squares for a small set of data and rejects data six mean absolute deviations2. The data window input to this function was specified as 10. However, in some cases, the data points available may be too small for smoothing in which case the raw data was used.

The program assumes the datafiles contain I-V data for the two non-active quadrants because interpolation was used to calculate Isc and Voc. The nearest two I datapoints in first and second quadrants were interpolated to obtain , and similarly the nearest two voltage datapoints in the second and third quadrant to obtain .

**2.03) Extracting and for forward and reverse scanned data**

The program must be capable of returning - from forward and reverse scanned data; this feature is important since I-V data generated for perovskite solar cells will vary according to scan direction.

The dot product on the I and V vectors was carried out to obtain a vector containing the power generated by the solar cell for the I-V data in the active region only, a subset of the original data. The program finds the maximum power in this vector and returns - which produced the maximum power.

However, depending on the scan direction, the position of and and in the vector will be affected. In the forward scanned data, and will have higher indices in the vector, whereas lower indices in reversed scanned data. The program can identify the scan direction by comparing the index positions of and and return the correct position of the - in the original data set.

**2.04) Extracting gradients from I-V data**

Two important derivatives that must be extracted from the I-V data are defined below.

and must be calculated as they will be used for constraints and the initial guess formula which will be discussed later. Herein two second order polynomials were used to fit the data and differentiated to solve for equation 1 and 2; the figure below shows the two polynomial curves overlaid on top of simulated data.

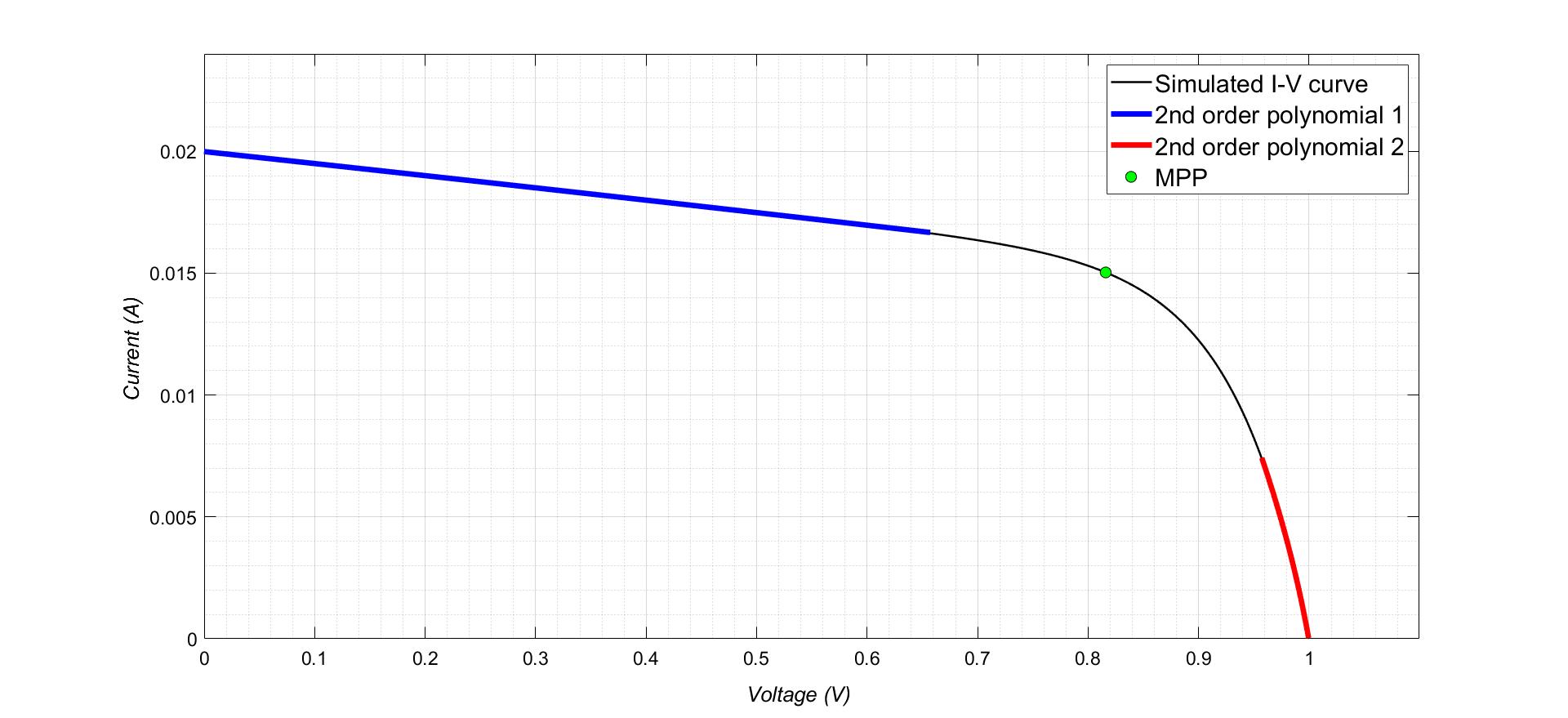


Figure 1: Two separate polynomial curves fitted on simulated I-V data to describe the regions of the curve for voltages above and below

The polynomial curve 1 fits the data for V between V = 0 and V = 0.8 whereas polynomial curve 2 fits data between I = 0 and I = . The two polynomials have the equation of the form

where are constants of the polynomial equation and the index specifies which polynomial. If the polynomial fits data with then , otherwise

This equation is differentiated to obtain

Substituting Eq (4) into Eq (1) & Eq (2) gives,

**2.05) Initial parameter guessing methods**

The performance of numerical methods for solving least squares problems depend on the initial guesses for the parameters3 4. If the initial guess is close to the actual solution, then locating the global minimum for the error function is more likely to be found3 5 6.

Many different initial guess models are published in literature, but in this study only one method was considered. The initial guess method from Zhang *et al* 4, a highly cited study, was modified because it required linear fitting of data for voltages near , and there was no description on how to choose the number of data points to be considered.

In this study, the initial guess method calculates near the maximum power point and The method is further discussed below.

The derivative of V against I written in terms of and is shown below.

Zhang *et al* approach assumes and the exponential terms is negligible to obtain to obtain Eq (8) 4.

Since is generally large, it is assumed that

for the initial guess of this parameter.

Two more equations are required to solve ,and they can be derived by using and the gradient at maximum power point. At the maximum power point the following equation is true regardless of the model used.

Eq (6) and Eq (10) are substituted to Eq (8) to obtain

Eq (11) and Eq (12) are subtracted and rearranged to obtain to as function of known parameters.

can be recovered by substituting the calculated into Eq (10) and rearranged to make as the subject of the formula, the equation is shown below.

In summary, equations 9, 13 and 14 were calculated and inputted into the solver as initial guesses.

**2.06) Deriving constraints for and**

1. constraint

As resistances cannot be negative, the lowest minimum possible should be equal to 0. The maximum for this parameter can be found by assuming for equation (7) which can be reduced to

If equation 15 is evaluated at and the equation can be rewritten as

Since all terms are positive on the right-hand side of equation 16,

Therefore

1. constraint

A similar analysis can be carried by `assuming and evaluating equation equation 7 at V = 0 and I = Isc to find the maximum possible . However, is usually very large for many cases and accurate estimates cannot be obtained. Additionally, more noise is also observed in the I-V curve near V = 0. Hence, the maximum was chosen to equal to . It is very unlikely that a experimental solar cell will have shunt resistance this high.

The minimum possible value for can be calculated by considering the minimum fill factor possible, 25%. At this condition , and if

Therefore,

1. constraint

The minimum is equal to 1, this is because this is the ideal case. The minimum can be obtained by assuming and and the exponential term is negligible for equation 16. It can be rearranged to obtain the maximum n possible

However, in some cases may me too large that is not plausible. In general n greater than 3 is usually not seen but since research solar cells will be tested this ideality factor may be inclusive. However, n > 5 is not expected and so is allowed for a maximum of 5.

The table below summarizes the constraints that were applied

**2.07) Nonlinear least squares solver**

The MATLAB provides two algorithms which changes the parameters it has been supplied with to minimize the least squares between the fitted curve and the experimental data. The levenberg-marquardt most commonly used algorithm. However, this method cannot be constraint and negative parameters are possible if the starting guesses are not near the actual solution. There the trust region method was implemented as containsts are permitted in this algorithm. The lower bound is maintained as

An initial guess vector was created in MATLAB which contained the . The value was maintained as , but the value was calculated from the Zhang method and the value from the Phang method. The algorithm was set as trust region reflective and the tolerance step size and was set as and the function tolerance was set as with the maximum number of iterations set to 1000.

**2.08) Simulating noisy I-V data generation for fixed Voc and Isc**

To determine the accuracy of the estimated parameters, simulated data with known parameters were fitted. The single diode model photovoltaic model, which is implicit in nature, can be expressed as an explicit form using the lambert W function.

The program is expected to be robust for many different types of I-V curves, but currently it is expected that most I-V curves that will be tested will be for perovskite solar cells. Good perovskite solar cells have a fill factor [] and with an Isc and Voc of 20 mA/cm^2 and 1V. Data was simulated for a 1 perovskite solar cell that produces an Isc of 20mA and a Voc of 1V for different parameters. For this report a 8 combinations of parameters were used to generate the simulated data, which are shown in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Case** | **Rs** | **Rsh** | **n** |
| 1 (low Rs, High Rsh, low n) |  |  | 1 |
| 2(high Rs, High Rsh, low n) |  |  | 1 |
| 3(low Rs, Low Rsh, low n) |  |  | 1 |
| 4(low Rs, High Rsh, high n) |  |  | 5 |
| 5(high Rs, Low Rsh, high n) |  |  | 5 |

Table 1: The table summarises the parameters used to generate eight I-V simulated data for Isc = 20mA and Voc = 1 V.

These parameters were chosen because of the shape of the I-V plot differs greatly, and the algorithm is expected to fit the data robustly.

**2.09) Comparison of this model with literature model**

Many different models are available In literature, however, same data set is required to test the model with literature. Data presented in …. Was used and the root measure squared error was used. The equation is shown below. There are many error quanitifying methods available, however, this study was limited to only one as we predict that no new significant discussion can be obtained by testing results for this other error methods. The model is machine learning model.

1. VMP3 Multi-channel potentiostat. *Bio-Logic Science Instruments*

2. Smooth response data - MATLAB smooth - MathWorks Australia. Available at: https://au.mathworks.com/help/curvefit/smooth.html. (Accessed: 24th May 2018)

3. Gottschalg, R., Rommel, M., Infield, D. G. & Kearney, M. J. The influence of the measurement environment on the accuracy of the extraction of the physical parameters of solar cells. *Measurement Science and Technology* **10,** 796–804 (1999).

4. Zhang, C., Zhang, J., Hao, Y., Lin, Z. & Zhu, C. A simple and efficient solar cell parameter extraction method from a single current-voltage curve. *Journal of Applied Physics* **110,** 064504 (2011).

5. Enebish, N., Agchbayar, D., Dorjkhand, S., Baatar, D. & Ylemj, I. Numerical analysis of solar cell current-voltage characteristics. *Solar Energy Materials and Solar Cells* **29,** 201–208 (1993).

6. Eriksson, J. Optimization and Regularization of Nonlinear Least Squares Problems. (1996).