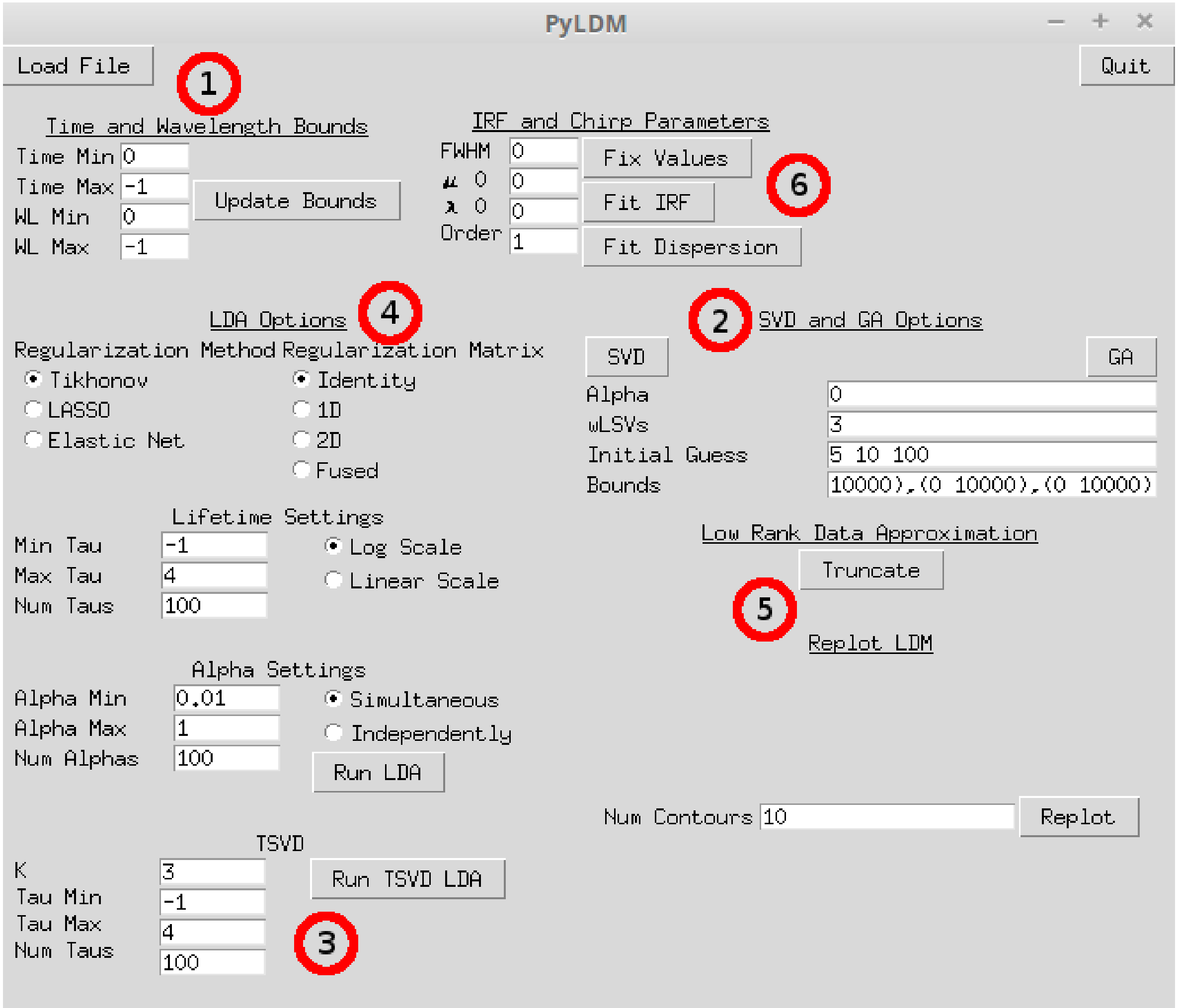
PyLDM Manual

Suggestions Before Beginning

- To get a sense of appropriate initial guesses for global analysis use of TSVD regularized maps can be helpful. K values of 3, 4 or 5 may be appropriate given your dataset.
- TSVD maps will NOT be useful to understand true kinetic distributions or dynamic processes.
- When beginning to analyze matrix regularized lifetime maps, begin with Tikhonov regularization and an identity matrix. This combination imposes the fewest assumptions on the data. Progress to the LASSO if sparser solutions are needed. This will be slower.
- Elastic net is slower than either option and should likely be the last option investigated.
- Whenever possible measure chirp and IRF experimentally and correct the data prior to modelling.

Contents

The Interface



1 Loading Data and Example Data Sets

1. All data should already be in a ΔA format (not transmission). The file should be CSV formatted as shown below, were the file to be opened in Microsoft excel, or a similar program.

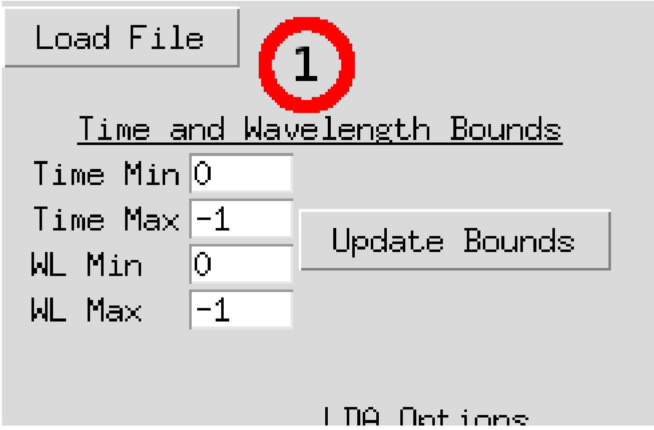
,	**Wavelength λ_1,**	**Wavelength λ_2,**
Time t_1,	$\Delta A(t_1, \lambda_1)$,	$\Delta A(t_1, \lambda_2)$,
Time t_2,	$\Delta A(t_2, \lambda_2)$,	$\Delta A(t_2, \lambda_2)$

2. When a dataset is loaded, it will be displayed on a symbolic lin-log plot. This means that a pseudo log scale will be used for negative time points. In addition, the plot will be linear between -1 and 1. Data for the three examples explored in the paper is included in the data subdirectory under the following names:

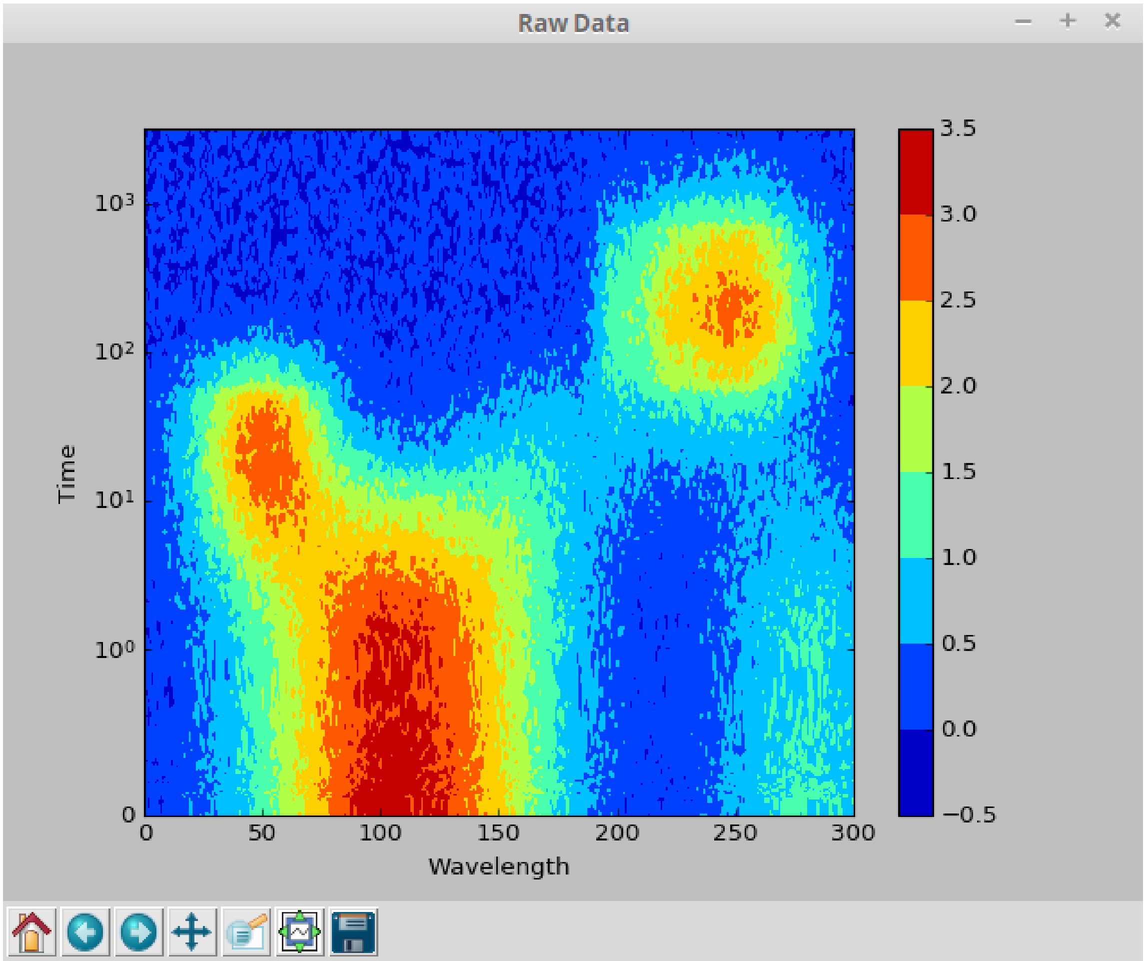
Data Set	**Noiseless File**	**File with noise**
Dynamically Moving Band	dynamic.csv	dynamic_noise10.csv
Multiple Decay Constants	hettaus.csv	hettaus_noise10.csv
Combination (Full Data Set)	fulldata.csv	fulldata_noise10.csv

3. The Time Min/Max and Wl Min/Max options displayed below will update automatically on data loading. These are the **indices** for the range of data to be used in the fitting. By default the minimum time point is set to the index of time zero. If no time zero is present then this point is the smallest positive time value. Note that including negative time points in the fitting will result in an error. The wavelength range is set to be the full range by default.

- ***Note:*** Python indexing begins at 0 instead of 1 as in Matlab
- ***Note:*** -1 is the index for the last item in the list.

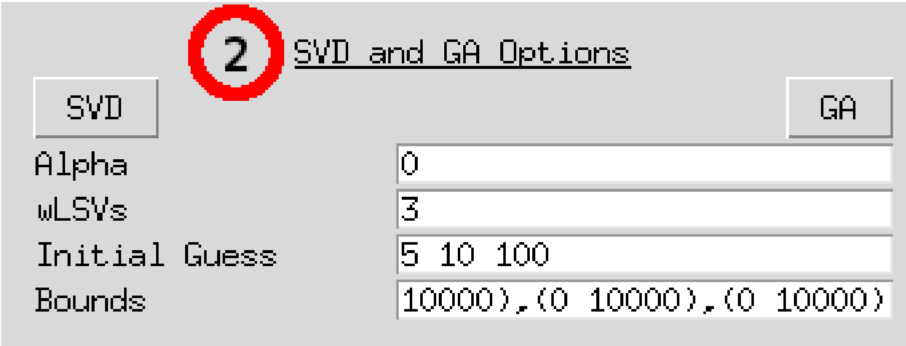


4. The hettaus_noise10.csv dataset will be used for the remainder of this manual. When loaded it should display as:

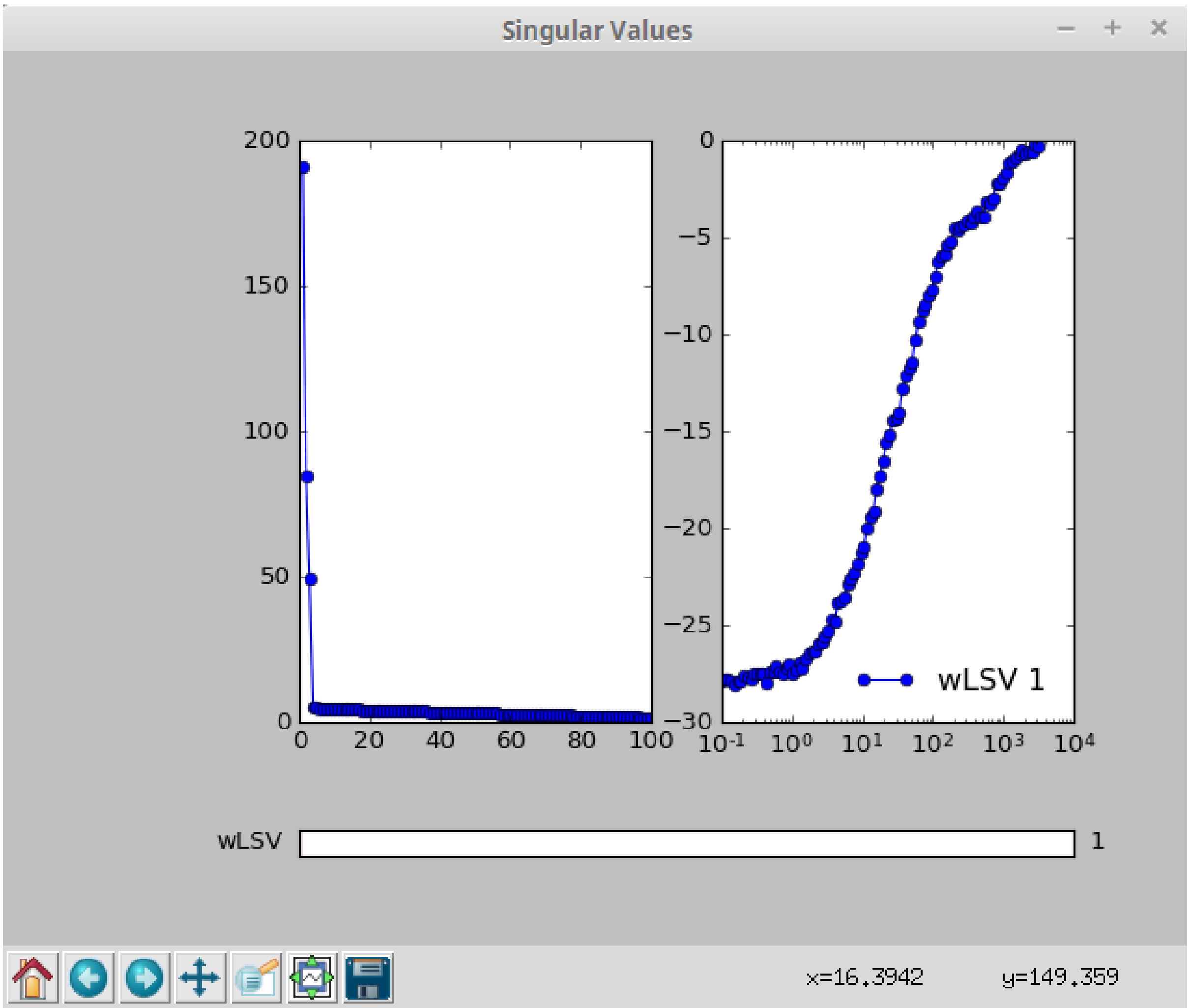


2 SVD and Global Analysis

1. The options for SVD and global analysis are shown below.



2. Once data has been loaded, the SVD button will allow you to explore the singular values and weighted left singular values.



3. After examining the singular values and their corresponding left singular vectors the GA button will fit the left singular vectors based on following input parameters. As mentioned in the suggestions the TSVD can be used to help determine initial guesses (see next section).

   (a) **Alpha:** By default this value is set to 0. If set to a value other than 0 Tikhonov regularization will be applied using this alpha value. A value of 0 corresponds to ordinary least squares.

   (b) **wLSVs:** This determines the number of wLSVs (i.e. components) to include in the analysis. A single number, e.g. 3, will include the first 3 vectors. Alternatively, a list can be entered to select specific vectors, with a space between the number of each vector, e.g.: 1 2 4 5 . ***Note:*** Vectors are NOT 0 indexed. The first vector is 1.

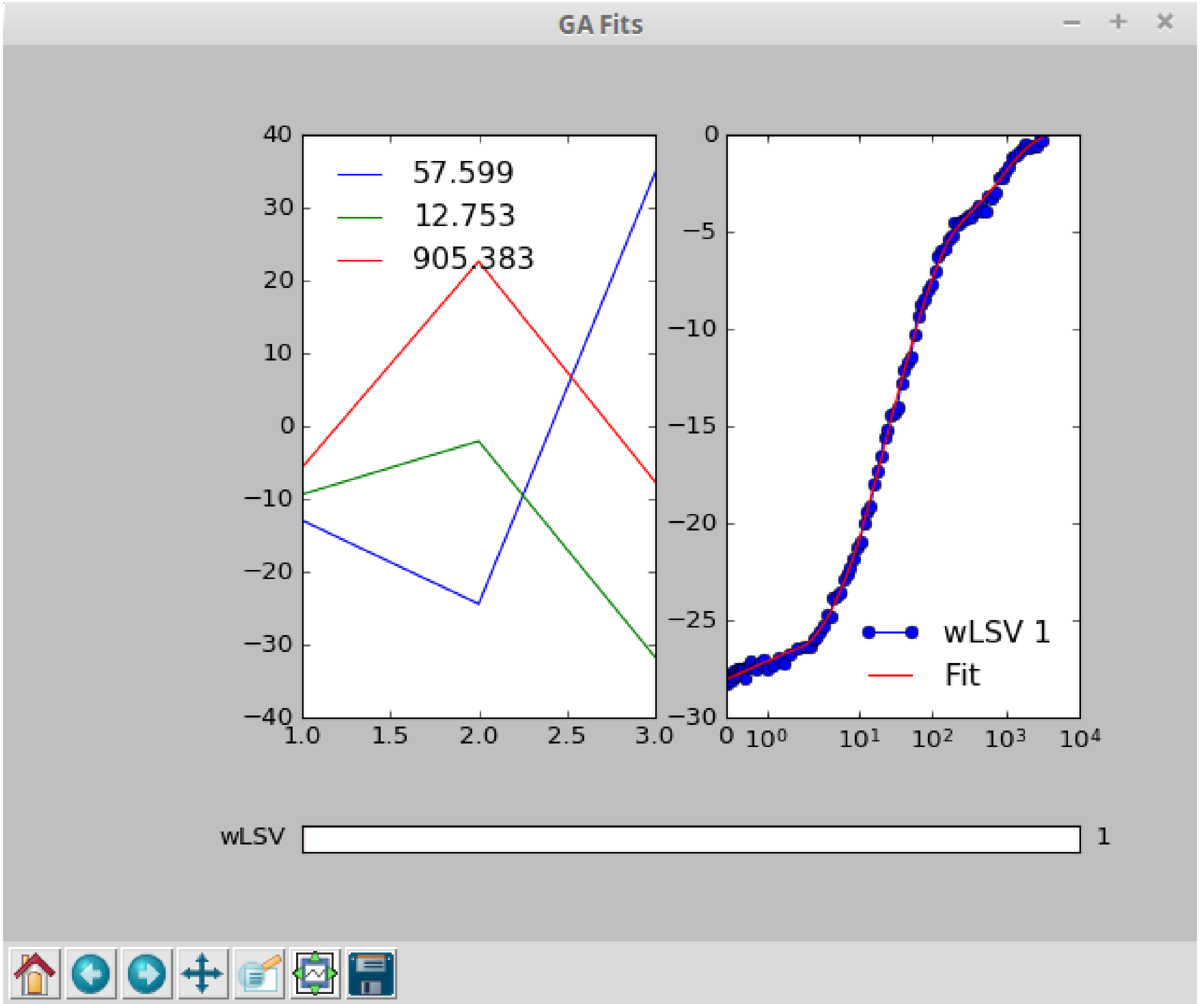
   (c) **Initial Guess:** This specifies the initial guess for the lifetime of each component. The number of guesses must match the number of wLSVs to be fit.

   (d) **Bounds:** These specify the bounds for the non-linear solver. The number of sets of bounds must match the number of vectors (and guesses). The bounds are entered in the following format:

   $$(\text{lower\_bound}_1\ \text{upper\_bound}_1),(\text{lower\_bound}_2\ \text{upper\_bound}_2)...$$

   There are spaces separating the respective pairs of lower and upper bounds for a single lifetime. The sets of bounds are separated by **only** a **comma**, and no space.

4. Using the settings displayed under item 1. above for the hettaus_noise10.csv dataset, the following results screen will be displayed.



The slider on the bottom will allow you to click through each wLSV and fit. You can rerun the fitting with new initial parameters if you are unhappy with the agreement of the fit and vectors. The left panel contains the decay associated spectra (i.e. the pre-exponential functions) for the corresponding lifetimes across the fitted vectors.

3 TSVD Lifetime Maps

1. The following settings are used to run a truncated SVD regularization to produce a lifetime density map.

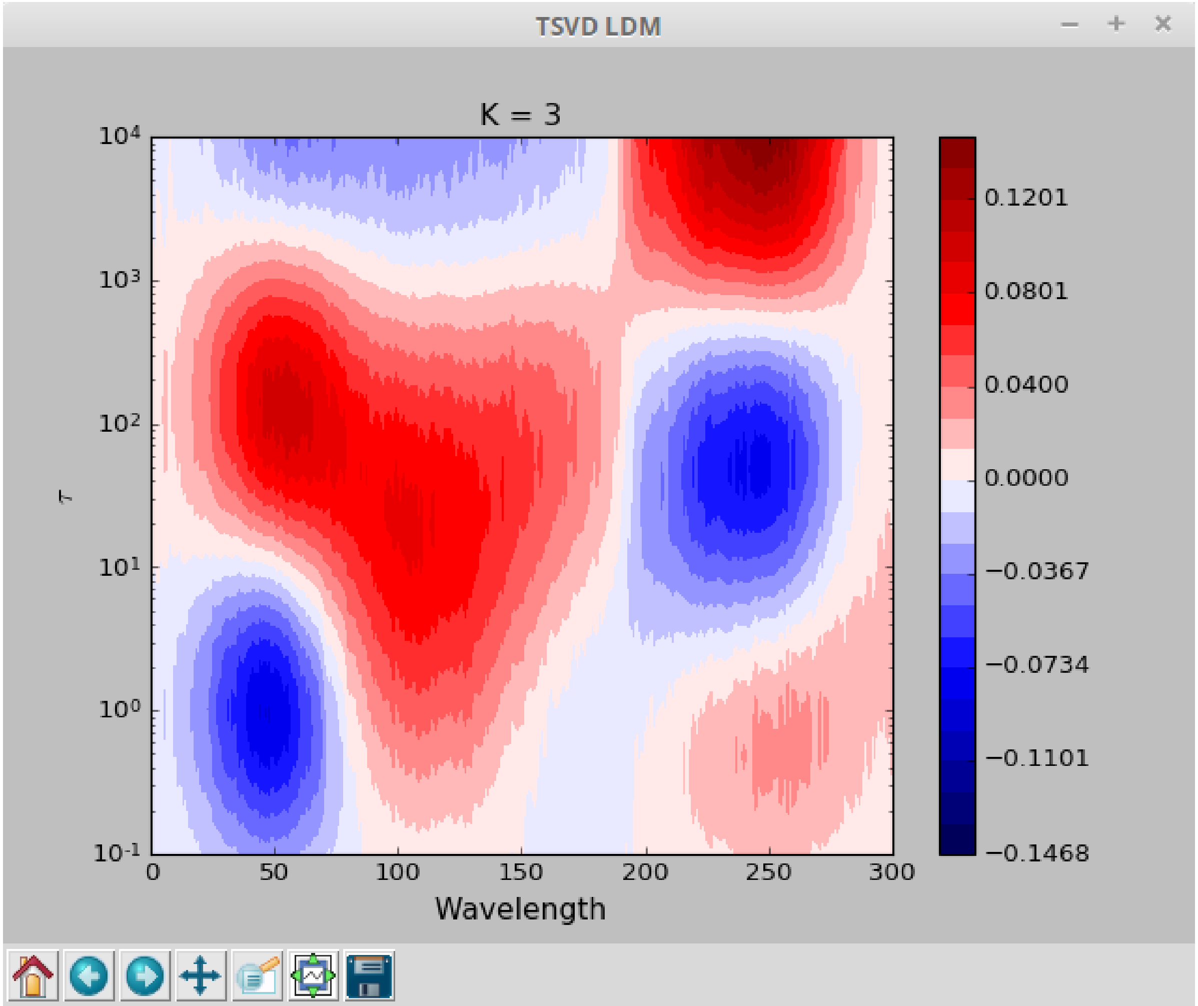
2. These maps should be used only to help a user examine a dataset initially to orient themselves, and perhaps provide some indication as to appropriate initial guesses for fitting of wLSVs.

3. K represents the singular value of truncation. This value is **NOT** zero-indexed. Appropriate initial guesses will likely be between 3 and 6 depending on the dataset.

4. Tau Min/Max are the bounds for a logarithmic scale for producing the lifetimes to be fit. E.g. the settings in item 1. above define the bounds to be between 10^{-1} and 10^{-4}.

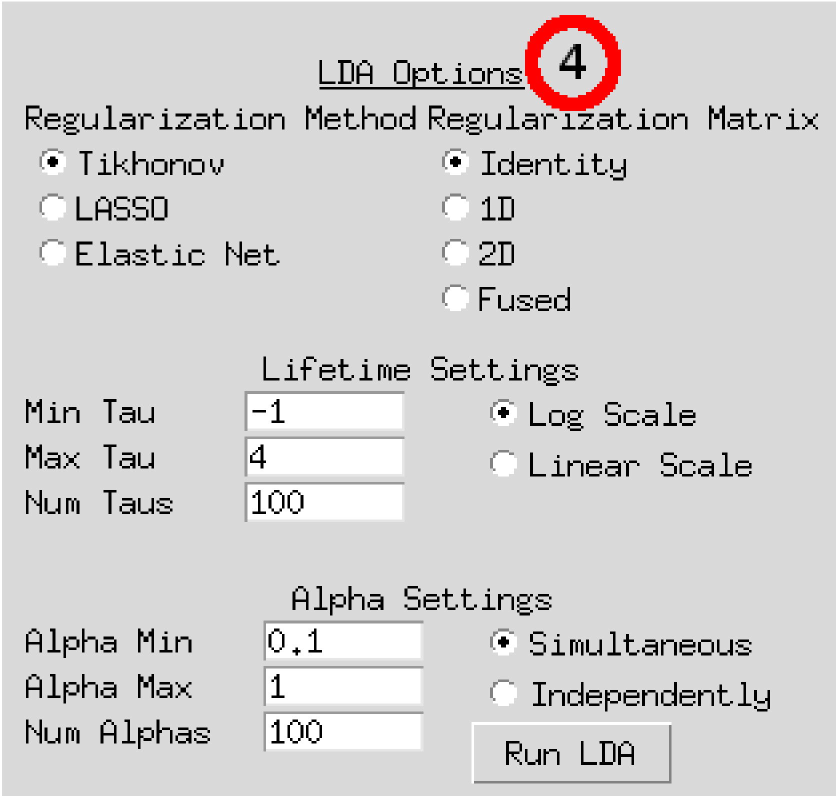
5. Num Taus specifies the number of lifetimes to use.

6. Running the settings in item 1. on the hettaus_noise10.csv dataset produces the following output.


TSVD LDM
K = 3
10^4
10^3
10^2
10^1
10^0
10^{-1}
τ
0
50
100
150
200
250
300
Wavelength
0.1201
0.0801
0.0400
0.0000
−0.0367
−0.0734
−0.1101
−0.1468

4 Matrix Regularization Options

1. This first group of settings allow you to choose the type of matrix regularization and which sort of penalty matrix to use. In general it is a good idea to begin with Tikhonov regularization and the Identity matrix option.



2. The matrix choices are as follows (see e.g. [1] for some discussion):

   (a) **Identity** is an identity matrix

   (b) **1D** is a first derivative approximation

$$\frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 & \cdots \\ 0 & 1 & -1 & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}$$

(c) **2D** is a second derivative approximation

$$\frac{1}{4}\begin{bmatrix} -2 & 1 & 0 & 0 & \cdots \\ 1 & -2 & 1 & 0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \cdots & 1 & -2 & 1 \end{bmatrix}$$

(d) **Fused** is a combination of the identity and first derivative matrices:

$$\begin{bmatrix} 1 & 0 & 0 & \cdots \\ 0 & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & 1 \\ 1 & -1 & 0 & \cdots \\ 0 & \ddots & \ddots & \ddots \\ 0 & \cdots & 1 & -1 \end{bmatrix}$$

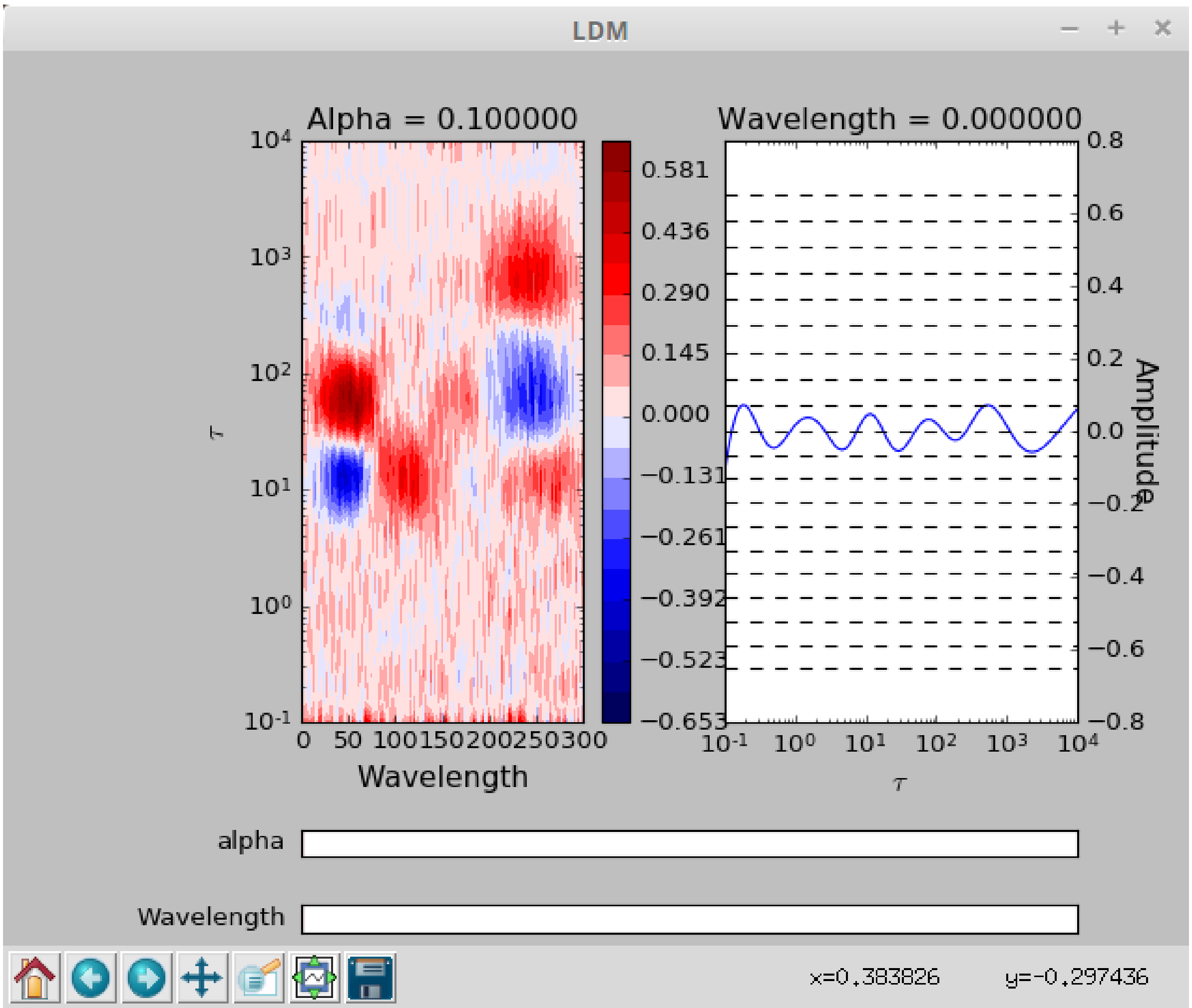
3. After selecting the regularization method, the rest of the options manage the fit parameters.

4. Min/Max Tau and Num Taus determine the min, max and number of lifetimes as previously with the TSVD regularization. These lifetimes can be distributed on a linear or log scale.

5. The alpha settings determine the number of alphas to test for model comparison. These are distributed along a linear scale. Use at least 3 alpha values to avoid errors in calculating statistics. If you really only need 1 alpha value then set the min and max alpha to be the same number. In the case that you are using Elastic Net, the ρ values are simply distributed on a scale from 0 to 1 in tenths. The relevant fit statistics can be calculated for all wavelengths taken together or for each wavelength independently. In general there is no need to calculate statistics for wavelengths independently, unless it is likely that there is structure to the noise in the dataset that affects wavelengths differently.

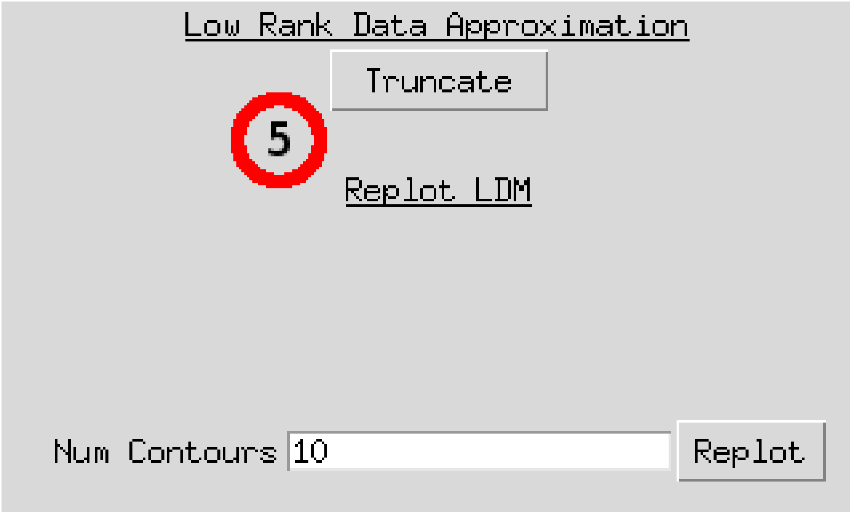
6. Running the settings used above on hettaus_noise10.csv will produce the following output LDM. There will be horizontal dashed lines on the LDM corresponding to the lifetimes from the fitting of singular vectors if this routine was run prior to the LDA. The right plot

shows a vertical slice through the LDM at a particular wavelength with the dashed lines corresponding to the contour levels. The alpha slider will allow you to slide through the maps produced for various alpha values, and the wavelength slider will change the slice produced on the right plot.



7. In addition, GCV, Cp and/or L-curve statistics plots will be produced if using Tikhonov or LASSO regularization.

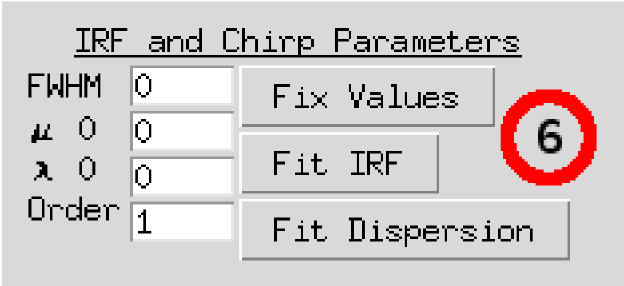
5 Replotting



1. The replot LDM will reproduce a previously fit LDM with a new number of contour levels.

2. The truncate button provides a low-rank approximation of the data matrix. The point of truncation is the value entered in the wLSVs box in the SVD and GA options section. This can suppress some noise in the data; however, is not usually necessary.

6 Chirp and IRF



1. ***Note:*** It is preferable to experimentally determine the chirp and IRF and correct data prior to modelling. The dispersion fitting procedure is most accurate for reference measurements taken for this purpose. The dispersion is calculated by fitting a polynomial of a specified order to the maximum gradient of the cross-correlation between a step function and the signal, as described in [2]. This procedure can produce erratic results in cases where there is low signal or it changes sign, which is problematic for the actual sample data of interest. The coefficients from a reference measurement can be fit and then fixed for use with the sample data.

2. The equation of the polynomial is given by:

$$\mu = \mu_0 + \sum_{i}^{order} \left(\mu_i + \frac{\lambda - \lambda_0}{100} \right)^i$$

3. λ_0 is the wavelength representing the center of the IRF dominated regime with μ_0 being the chirp or location of the IRF at λ_0.

4. Regardless of bounds set on the data for LDA and GA fitting routines, all wavelengths and time points are used for chirp fitting. The cross-correlation and fitted curve are plotted together. Inspect this curve, as the cross-correlation can produce erratic results in cases where there is low signal or it changes sign.

5. Given this fit, the data is adjusted to resynchronize time 0 to the nominal delay. This allows all wavelengths to be fit simultaneously.

6. To view the dataset after this transformation, press the Update Bounds button in the data options section. As with inspection of the cross-correlation, care should be taken in

examining the adjusted data.

7. The IRF is calculated using an iterative reconvolution during a global analysis using the parameters specified in the SVD and Global Analysis section. This routine assumes that the time zero has been synchronized across all wavelengths, i.e. chirp needs to be corrected for before using it.

8. ***Note:*** Maintaining the default values of chirp parameters and FWHM assumes that dispersion and IRF can be ignored, and so they will not be accounted for in any modelling.

7 Other Files

A number of other files are included with this package. These are enumerated below. Additionally, the github wiki for the package contains more specific information on functions.

1. **SynthData.ipynb :** This file is a Jupyter notebook (python 2.7) that contains the code used to create the example datasets.

2. Unit test files. All files can be run independently typing python <filename> in the terminal.

   (a) **test_lda.py :** Unit tests for code in lda.py .

   (b) **test_svd_ga.py :** Unit tests for code in svd_ga.py .

   (c) **test_data.py :** Unit tests for code in data.py .

3. **data/test_mat.csv :** This is a minimal matrix example used for certain unit tests contained in test_lda.py and test_svd_ga.py

4. Mathematica Notebooks. These contain Mathematica verifications for certain calculations used in the unit tests.

   (a) **test_lda.nb :** For test_lda.py unit tests.

   (b) **test_svd_ga.nb :** For test_svd_ga.py unit tests.

References

[1] Noschese S, Reichel L. Inverse problems for regularization matrices. Numerical Algorithms. 2012;60(4):531–544. doi:10.1007/s11075-012-9576-8.

[2] van Wilderen LJGW, Lincoln CN, van Thor JJ. Modelling multi-pulse population dynamics from ultrafast spectroscopy. PLoS ONE. 2011;6(3). doi:10.1371/journal.pone.0017373.