WAIT: Workshop on the Advances in Information Technology, Budapest, 2017

Curve Fitting with Minimal Relative Error

Ágota Kacsó and László Szirmay-Kalos

Budapest University of Technology and Economics (e-mail: szirmay@iit.bme.hu)

Abstract

Fitting a parametric curve to a set of noisy points by searching for the minimal Euclidean distance is a standard least squares fitting problem. This approach minimizes the absolute error between the curve and the data points. However, in certain applications, the relative, rather than the absolute error expresses the quality of the fitting. This paper examines the possibilities of defining the relative error for function fitting and generalizes the Levenberg-Marquardt procedure to minimizing it. We present an application aiming at dynamic tomography.

1. Introduction

The curve fitting problem can be summarized as the following. We have a set of observations $(t_1,x_1),(t_2,x_2),\ldots,(t_N,x_N)$ and search for a member of a parametric function family $f(\mathbf{p},t)$ so that error $E(\mathbf{p})$ of the approximation is minimal. As a member of the function family is identified by parameter vector $\mathbf{p}$, the minimization takes place with respect to its elements. Error E has a local extremum if all partial derivatives forming the gradient vector is zero:

$$\frac{\partial E}{\partial \mathbf{p}_P} = 0, \quad P = 1,\ldots,N_P.$$

This extremum is a minimum, i.e. neither a maximum nor a saddle point, if Hessian matrix $\mathbf{H}$ defined by

$$\mathbf{H}_{P,Q} = \frac{\partial^2 E}{\partial \mathbf{p}_P \partial \mathbf{p}_Q}$$

is positive definite. Error function E can be chosen flexibly, but we require it to have a local minimum when the fitting is perfect, i.e. when $f(\mathbf{p},t_i) = x_i$. For now on, we use the shorthand notation of $f_i(\mathbf{p}) = f(\mathbf{p},t_i)$.

This paper investigates the problem of curve fitting with minimal relative error. First, in Section 2 we review the classical least squares fitting method and show that it minimizes the absolute error. Section 3 presents our new approach. First we discuss what relative error may mean in a curve fitting application, find an appropriate definition, and then generalize the Levenberg-Marquardt algorithm for this scenario. Finally, we show the application of the proposed method in dynamic positron emission tomography reconstruction.

2. Least squares fitting

If the sum of squared difference of f_i and x_i, called Euclidean distance or L_2 norm, is used to define E, we need to minimize

$$E(\mathbf{p}) = \frac{1}{2}\|\mathbf{f}(\mathbf{p}) - \mathbf{x}\|_2^2 = \frac{1}{2}\sum_{i=1}^{N} \left(f_i(\mathbf{p}) - x_i\right)^2.$$

where $\|.\|_2$ is the Euclidean norm, $\mathbf{f} = (f_1,\ldots,f_N)$, and $\mathbf{x} = (x_1,\ldots,x_N)$.

In least squares fitting, the equation stating that the gradient is zero becomes:

$$\frac{\partial E}{\partial \mathbf{p}_P} = \sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} \left(f_i(\mathbf{p}) - x_i\right) = 0. \tag{1}$$

In this equation, a data point x_i has a contribution that depends on its distance from the curve, thus the *absolute error* is used. From another point of view, neither the error measure nor the optimal fitting alters when we add a constant value to both the data points and to the fitted function. Instead of the L_2 norm, we can use the general L_p norm defined by

$$\|\mathbf{f}(\mathbf{p}) - \mathbf{x}\|_p = \left(\sum_{i=1}^{N} |f_i(\mathbf{p}) - x_i|^p\right)^{1/p},$$

from which the $p = 1$ case, called *total variation* or L_1 norm, is also very popular [9, 7, 5]:

$$\|\mathbf{f}(\mathbf{p}) - \mathbf{x}\|_1 = \sum_{i=1}^{N} |f_i(\mathbf{p}) - x_i|$$

These options also focus on the absolute error, but are more difficult to minimize that of using the Euclidean error.

Equation 1 can also be given an estimation theory based interpretation. The solution is the maximum likelihood estimation of the expectation of the parameter vector assuming that data points are of Gaussian distribution with the same standard deviation σ. From another point of view, in this type of fitting the noise is assumed to be Gaussian with zero mean. To prove this, let us consider the probability density of sampling x_i

$$\mathrm{pdf}(x_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - f_i)^2}{2\sigma^2}\right),$$

thus the probability density of the whole measurement is

$$\mathrm{pdf}(x_1,\ldots,x_N) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - f_i)^2}{2\sigma^2}\right),$$

The log-likelihood function is

$$\log \mathrm{pdf}(x_1,\ldots,x_N) = C(\sigma) - \sum_{i=1}^{N} \frac{(x_i - f_i)^2}{2\sigma^2}$$

where $C(\sigma)$ depends only on the variance of the measured data and is independent of the parameter vector. According to maximum-likelihood reconstruction, the unknown parameter vector should be computed to maximize the log-likelihood, which requires the gradient vector be zero:

$$\frac{\log \mathrm{pdf}(x_1,\ldots,x_N)}{\partial \mathbf{p}_P} = -\sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} (f_i(\mathbf{p}) - x_i) = 0.$$

Indeed this is the same as Equation 1.

It is worth generalizing the estimation problem to take into account that different data points x_i may be associated with different standard deviations σ_i. Repeating the previous derivation, we can obtain the following requirement for optimal fitting:

$$\frac{\log \mathrm{pdf}(x_1,\ldots,x_N)}{\partial \mathbf{p}_P} = -\sum_{i=1} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} \frac{f_i(\mathbf{p}) - x_i}{\sigma_i^2}) = 0,$$

which leads to *weighted least squares fitting* where the weight of the term is inversely proportional to the variance σ_i^2 of a data point.

Depending on the actual form of function f, the equation of zero gradient is non-linear, so its solution requires numeric techniques. A particularly popular and robust method is the *Levenberg-Marquardt algorithm* [4, 8]. Assume we have an estimate $\mathbf{p}$ of the parameter vector in iteration step n. In the next step, this estimate is refined by adding a hopefully small vector $\mathbf{d}$ to it. First, f is linearized:

$$f_i(\mathbf{p}+\mathbf{d}) \approx f_i(\mathbf{p}) + \sum_{Q=1}^{N_P} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q} \mathbf{d}_Q.$$

Substituting this into Equation 1, we get:

$$\sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} \left(f_i(\mathbf{p}) + \sum_{Q=1}^{N_P} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q} \mathbf{d}_Q - x_i \right) = 0.$$

Rearranging the terms, we obtain a system of linear equations for $\mathbf{d}$:

$$\sum_{Q=1}^{N_P} \left(\sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q} \right) \mathbf{d}_Q = \sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} (x_i - f_i(\mathbf{p})).$$

or in matrix form

$$\mathbf{A} \cdot \mathbf{d} = \mathbf{b} \tag{2}$$

where

$$\mathbf{A}_{P,Q} = \sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q}.$$

Note that this matrix is symmetric and positive definite, thus Equation 7 can be efficiently solved with *Cholesky factorization* [2].

3. Fitting with minimal relative error

In order to study the relative error, let us assume that f is strictly positive while sample points can take zero or positive values. First we should clarify what fitting with relative error means. Using the analogy of fitting with absolute error, we say that the fitting process is relative error based if neither the error measure nor the optimal fitting curve changes if both the data values and the fitted curve are multiplied by the same positive value. Similarly to the fitting with absolute error, there are many possibilities to construct relative error based fitting.

3.1. Least squares of the relative error

A simple way to introduce relative error based fitting is to replace the squared absolute difference by the squared difference of the relative error and 1 in the error function:

$$E(\mathbf{p}) = \frac{1}{2} \sum_{i=1}^{N} \left(1 - \frac{x_i}{f_i(\mathbf{p})}\right)^2.$$

Note that this option is similar to weighted least squares if the standard deviation associated with a data point is equal to the yet unknown function value. The zero gradient equation is then:

$$\sum_{i=1}^{N} \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P} \left(1 - \frac{x_i}{f_i(\mathbf{p})}\right) \frac{-x_i}{f_i^2(\mathbf{p})} = 0.$$

This equation is partially similar to equation 1 in that the terms are products of the partial derivatives and the error of a given sample. However, these terms are also weighted now with $x_i / f_i^2(\mathbf{p})$ while there is no such additional weighting in the formula of the absolute error. This weighting downscales the effect of samples that are small while they are supposed to be large, while emphasizes samples that are larger than

required. In the extreme case when $x_i = 0$, the sample has no effect on the parameter fitting process, although it is perfectly possible that noise replaces a positive value by zero. Thus, the least squares of the relative error is not appropriate for curve fitting.

3.2. Logarithmic plot

Another approach for relative error based fitting is to "log" the problem and instead of fitting $f(\mathbf{p},t)$ to $(t_1,x_1),(t_2,x_2),\ldots,(t_N,x_N)$, we fit $\log(f(\mathbf{p},t))$ to $(t_1,\log(x_1)),(t_2,\log(x_2)),\ldots,(t_N,\log(x_N))$ with minimizing the squared error:

$$E(\mathbf{p}) = \frac{1}{2}\sum_{i=1}^{N}\left(\log(f_i(\mathbf{p})) - \log(x_i)\right)^2.$$

The equation for zero gradient is then

$$\sum_{i=1}^{N}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}\frac{\log(f_i(\mathbf{p})) - \log(x_i)}{f_i(\mathbf{p})} = 0.$$

This option has also problems with zero data points, namely, for these cases f_i cannot be positive.

3.3. Poisson-like model

To find a better alternative, we take a reverse direction and first establish the zero gradient equation, which has the same structure as equation 1, but the error of a sample is modified from absolute to relative difference:

$$\sum_{i=1}^{N}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}\left(1 - \frac{x_i}{f_i(\mathbf{p})}\right) = 0. \tag{3}$$

The error function can be constructed from this by integration:

$$E(\mathbf{p}) = \sum_{i=1}^{N} f_i(\mathbf{p}) - x_i\log(f_i(\mathbf{p})) + C \tag{4}$$

where C is an arbitrary constant that is independent of the parameter vector. It is easy to see that the gradient of this error function is indeed the left side of equation 3.

This equation can also be given estimation theory interpretation if x_i values are integers, when this is the maximum likelihood estimation of the expectation of random variables of Poisson distribution. If x_i is of Poission distribution, then its probability is

$$P(x_i) = \frac{f_i^{x_i}}{x_i!}\exp(-f_i).$$

The probability of all data values assuming statistical independence is

$$P(x_1,\ldots,x_N) = \prod_{i=1}^{N}\frac{f_i^{x_i}}{x_i!}\exp(-f_i).$$

The log-likelihood function is

$$\log P(x_1,\ldots,x_N) = C - \sum_{i=1}^{N} f_i - x_i\log(f_i)$$

where $C = -\sum_{i=1}^{N}\log(x_i!)$ is a constant that is independent of the parameter vector. Note that the maximization of this is equivalent to the minimization of Equation 4.

The equivalence with the maximum likelihood estimation assuming Poisson data means that this type of relative error should be used when data points are integers and follow a Poisson distribution. However, we can use this fitting criterion also for non-integer data points, which obviously cannot have Poisson distribution.

To solve this Equation 4 with a Levenberg-Marquardt like method, we need to linearize, but unlike in the least squares method where f_i is linearized, now the same procedure is executed for $1/f_i$:

$$\begin{aligned}\frac{1}{f_i(\mathbf{p}+\mathbf{d})} &\approx \frac{1}{f_i(\mathbf{p})} + \sum_{Q=1}^{N_P}\frac{\partial 1/f_i(\mathbf{p})}{\partial \mathbf{p}_Q}\mathbf{d}_Q \\ &= \frac{1}{f_i(\mathbf{p})} - \frac{1}{f_i^2(\mathbf{p})}\sum_{Q=1}^{N_P}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q}\mathbf{d}_Q. \end{aligned}\tag{5}$$

After the substitution into Equation 3 and rearrangement, we obtain:

$$\sum_{Q=1}^{N_P}\left(\sum_{i=1}^{N}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}\frac{x_i}{f_i^2(\mathbf{p})}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q}\right)\mathbf{d}_Q = \sum_{i=1}^{N}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}\left(\frac{x_i}{f_i(\mathbf{p})} - 1\right). \tag{6}$$

or in matrix form

$$\mathbf{A}\cdot\mathbf{d} = \mathbf{b} \tag{7}$$

where the matrix of this linear system of equations

$$\mathbf{A}_{P,Q} = \sum_{i=1}^{N}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}\frac{x_i}{f_i^2(\mathbf{p})}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q}$$

is symmetric and positive-definite if data points x_i are positive and positive-semidefinite if data points may be zero, which can be seen by proving that the following quadratic form positive with arbitrary vector $\mathbf{y}$:

$$\begin{aligned}\mathbf{y}^T\cdot\mathbf{A}\cdot\mathbf{y} &= \sum_{P=1}^{N_P}\sum_{Q=1}^{N_P}\mathbf{y}_P\mathbf{A}_{P,Q}\mathbf{y}_Q \\ &= \sum_{i=1}^{N}\frac{x_i}{f_i^2(\mathbf{p})}\sum_{P=1}^{N_P}\mathbf{y}_P\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}\sum_{Q=1}^{N_P}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q}\mathbf{y}_Q. \end{aligned}\tag{8}$$

Using the

$$z_i = \sum_{Q=1}^{N_P}\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_Q}\mathbf{y}_Q = \sum_{P=1}^{N_P}\mathbf{y}_P\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}$$

shorthand notation, we get

$$\mathbf{y}^T\cdot\mathbf{A}\cdot\mathbf{y} = \sum_{i=1}^{N}\frac{x_i}{f_i^2(\mathbf{p})}\sum_{P=1}^{N_P}\mathbf{y}_P\frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}z_i$$

$$= \sum_{i=1}^{N} \frac{x_i}{f_i^2(\mathbf{p})} z_i \sum_{P=1}^{N_P} \mathbf{y}_P \frac{\partial f_i(\mathbf{p})}{\partial \mathbf{p}_P}$$
$$= \sum_{i=1}^{N} \frac{x_i}{f_i^2(\mathbf{p})} z_i^2 > 0. \quad (9)$$

Thus, the efficient Cholesky factorization can also be used to solve equation 6.

4. Application to dynamic emission tomography

In dynamic Positron Emission Tomography, we analyse the dynamic nature of biological processes, like accumulation and emptying drugs in certain organs. To make the measurement, radioactive material, called radiotracer, is injected into the body, which spreads into the tissues by diffusion described by a compartment model [3, 15, 14, 13]. For example, the solution of the popular *two-tissue compartment model* leads to the following time functions that describe the radiotracer density in a given compartment:

$$C(\mathbf{p},t) = (1-f_v) \sum_{i=1}^{2} a_i \alpha_i \exp(-\alpha_i t) * C_p(t) + f_v C_w(t), \quad (10)$$

where dynamic properties are defined by parameters $\mathbf{p} = (f_v, a_1, a_2, \alpha_1, \alpha_2)$. In this model

$$F(a_1, a_2, \alpha_1, \alpha_2) = a_1 \alpha_1 \exp(-\alpha_1 t) + a_2 \alpha_2 \exp(-\alpha_2 t)$$

is the impulse response of the compartment system, f_v is the fractional volume of blood, $*$ stands for convolution, $C_p(t)$ and $C_w(t)$ are the known *blood activity function* and the *whole blood concentration function*, respectively.

Radiotracer molecules are instable isotopes and randomly decay generating particles that can be detected by the detectors of the tomograph. For example, in Positron Emission Tomography (PET) [12, 11, 1, 10], decays generate positrons that annihilate with the electrons of the tissue where two opposite gamma photons are born that conserve energy and momentum. In Single Photon Emission Tomography (SPECT), on the other hand, photons are emitted by the primary decay. Detecting photon hits in small time intervals called *frames*, the reconstruction process can estimate the number of particle emissions in certain regions of the body. Since the probability of the decay of a given radiotracer isotope molecule is small but there are many of them in a single voxel and their decay is statistically independent, the number of decays in a voxel and frame is a random variable of Poisson distribution. The number of decays can obviously be only non-negative integers, but the result of the static reconstruction in a frame may result in an arbitrary non-negative number even fractions. The explanation is partly the numerical inaccuracy and partly the fact that the reconstruction provides blurred data. For example, if one voxel has a single decay and its neighbor has no decay, the reconstruction may indicate that the voxel has 0.8 decays and is neighbor 0.2 decays. Thus, the inherent noise is Poisson but the fitting process should be able to handle real values not only integers. This observation justifies that we need to use a relative error criterion during fitting.

5. Results

To demonstrate the proposed fitting process, we simulated the measurement process of a positron emission tomograph, executed a reconstruction of the simulated measurement, and fitted the two-tissue compartment model on the reconstructed activities in different frames. We have considered three different measurements generating 184, 1620, and 15981 hits in total in all frames and in all detector pairs, called LORs. Note that the number of LORs is 2115, thus the first measurement is extremely noisy, the second is moderately noisy, and we can say only for the third measurement that every detector pair gets at least a few hits.

Figures 1 and 2 show the reconstruction results obtained by setting 10 and 100 frames, respectively, and using the proposed modified Levenberg-Marquardt scheme for the solution of the non-linear equation. Note that when setting 10 or 100 frames not only the number of data points but also the noise level is affected since we distribute the same number of events. Thus, the 10 frame case is less noisy than the 100 frame case.

For comparison, we have also made the same experiment with the classical Levenberg-Marquardt algorithm minimizing the absolute error. The difference between the curves obtained with minimal absolute and relative errors is quite small. If absolute error is used, a few larger values may attract the curve significantly, which is not the case when relative error is minimized.

6. Conclusions

This paper investigated the problem of curve fitting when relative rather than absolute error is considered. We reviewed several possibilities to define the curve fitting with relative error and modified the Levenberg-Marquardt algorithm accordingly. The fitting process can also be given a probabilistic interpretation, namely, our method is proposed when Poisson noise corrupts the true data.

Acknowledgement

This work has been supported by OTKA K–104476 and VKSZ-14 PET/MRI 7T projects.

References

1. B. Csébfalvi and G. Rácz. Retailoring box splines to lattices for highly isotropic volume representations. *Computer Graphics Forum*, 35(3):411–420, 2016.
2. Dariusz Dereniowski and Marek Kubale. *Cholesky Factorization of Matrices in Parallel and Ranking of Graphs*, pages 985–992. Springer Berlin Heidelberg, 2004.

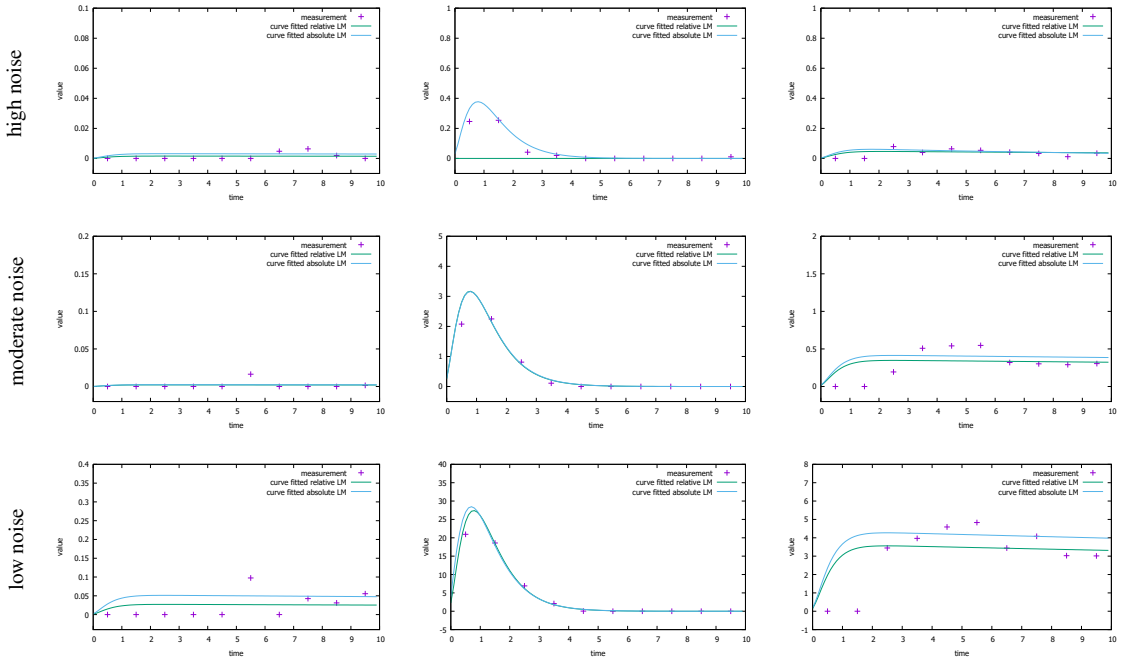

Figure 1: *Fitting of for very noisy (upper row), moderately noisy (middle row), and weakly noisy data of 10 points.*

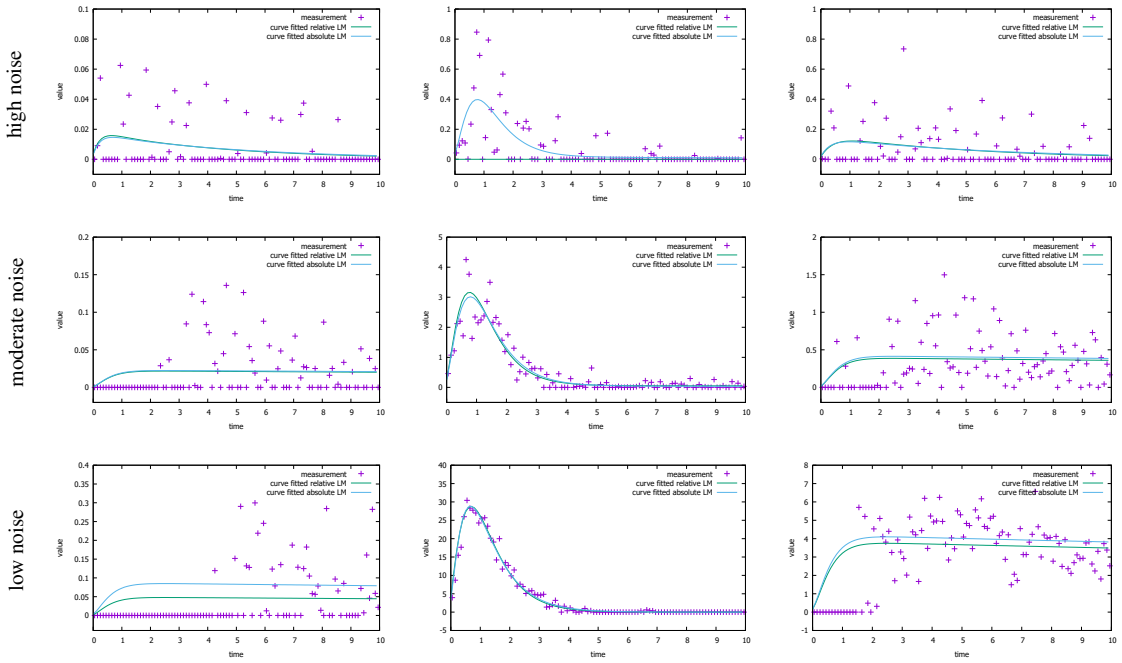

Figure 2: *Fitting of for very noisy (upper row), moderately noisy (middle row), and weakly noisy data of 100 points.*

3. M.E. Kamasak, C.A. Bouman, E.D. Morris, and K. Sauer. Direct reconstruction of kinetic parameter images from dynamic pet data. *IEEE Trans Med Imaging*, 24(5):636–650, 2005.

4. K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944.

5. Milán Magdics and László Szirmay-Kalos. *Optimization in Computer Engineering - Theory and Applications*, chapter Total Variation Regularization in Maximum Likelihood Estimation. Scientific Research Publishing, 2011.

6. Milán Magdics, Balázs Tóth, Balázs Kovács, and László Szirmay-Kalos. Total variation regularization in PET reconstruction. In *KÉPAF*, pages 40–53, 2011.

7. D.W. Marquardt. An algorithm for the least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 11(2):431–441, 1963.

8. M. Persson, D. Bone, and H. Elmqvist. Total variation norm for threedimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine and Biology*, 46(3):853–866, 2001.

9. G. Rácz and B. Csébfalvi. Tomographic reconstruction on the body-centered cubic lattice. In *Spring Conference on Computer Graphics (SCCG)*, 2015.

10. L. Szirmay-Kalos, M. Magdics, and B. Tóth. Multiple importance sampling for PET. *IEEE Trans Med Imaging*, 33(4):970–978, 2014.

11. L. Szirmay-Kalos, M. Magdics, B. Tóth, and T. Bükki. Averaging and Metropolis iterations for positron emission tomography. *IEEE Trans Med Imaging*, 32(3):589–600, 2013.

12. László Szirmay-Kalos and Ágota Kacsó. Regularizing direct parametric reconstruction for dynamic pet with the method of sieves. In *Molecular Imaging Conference*, MIC '16, pages M16D–1, 2016.

13. Guobao Wang and Jinyi Qi. An optimization transfer algorithm for nonlinear parametric image reconstruction from dynamic pet data. *IEEE Trans Med Imaging*, 31(10):1977–1988, 2012.

14. Jianhua Yan, Beata Planeta-Wilson, and R.E. Carson. Direct 4d list mode parametric reconstruction for pet with a novel em algorithm. In *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, pages 3625–3628, Oct 2008.