Note on Approximation by Nonlinear Optimization

Jaroslav Sklenar
Department of Statistics and Operations Research
University of Malta

Introduction

The purpose of this note is to discuss the use of nonlinear optimization techniques to solve approximation problems typical for example in signal identification. Different techniques based on classical and modern approaches to time series are available. The presented idea considers cases when signals are composed of a finite number of certain nonlinear functions distinct in their parameter sets, and realization of an additive random error. The focus is given to the sums of parameterized trigonometric functions. As the random error probability distribution is assumed unknown, the common LSQ criterion is replaced with its parameterized generalization. The obtained unconstrained non-smooth minimization problem can be solved either directly or after a smooth reformulation to the constrained problem. The initial values for computational procedures are estimated using heuristics and suitable statistical techniques, e.g., periodograms. The ideas are illustrated by simple explanatory examples accompanied by figures. Test results are shown for MS Excel Solver, MATLAB is used for visualization.

Problem formulation

One of important signal processing tasks is signal identification. We denote $\mathbf{x} = (x_1, ..., x_n)^T$ a vector of time points (both equidistant and non-equidistant cases are acceptable) and $\mathbf{y} = (y_1, ..., y_n)^T$a vector of related measurements. Various techniques developed for processing of this type of data may be found. There are approaches based on: analysis of trend and cyclic behaviour, followed by the estimate of random error distribution; analysis of autocorrelation structure that leads to advanced autoregressive techniques; and harmonic analysis based on estimates of frequencies. In our case, we assume that n measurements contained in $\mathbf{y}$were obtained as realizations of the following random vector:

$$\eta = (\eta_1, ..., \eta_n)^T = \left(\sum_{k=1}^{K} f_k(x_j, \beta_k) + \varepsilon_j \right)^T_{j=1,...,n},$$

where $\varepsilon_j, j = 1, ..., n$ denote random errors and functions $f_k, k = 1, ..., K$ are of known types, but unknown vectors of parameters $\beta_k, k = 1, ..., K$ have to be found. Note that the functions f_k can be very complex, so the above additive model is in fact no limitation. More complex combinations of individual functions (like product or composed functions) can always be expressed as one complex function.

Harmonic approximation

One of traditional tasks utilizing Fourier series is to identify the signal under the assumption that $f_k, k = 1, ..., K$ are trigonometric functions and proportions among unknown frequencies are rational numbers. In this case, the parameters of the following trigonometric function are estimated:

$$y = \sum_{k=1}^{K} \beta_{k1} \sin(\beta_{k2} x + \beta_{k3}).$$

As usually, we search for coefficients $\beta_k, k = 1, ..., K$ in such a way to minimize a distance between measurements $\mathbf{y} = (y_1, ..., y_n)^T$ and predicted values $\hat{\mathbf{y}} = (\hat{y}_1, ..., \hat{y}_n)^T$ where

$$\hat{y}_j = \sum_{k=1}^{K} b_{k1} \sin(b_{k2} x_j + b_{k3})$$

and $\mathbf{b}_k, k = 1, ..., K$ are estimates of unknown $\beta_k, k = 1, ..., K$. A criterion to be minimized can be defined in different ways. We utilize a parametrized distance

$$d_p(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|_p = \sqrt[p]{\sum_{j=1}^{n} |y_j - \hat{y}_j|^p},$$

where $0 < p < \infty$. Therefore, we use $(d_p(\mathbf{y}, \hat{\mathbf{y}}))^p$as a criterion to get an unconstrained optimization problem with unknown variables $\mathbf{b}_k, k = 1, ..., K$:

$$\min\{\sum_{j=1}^{n} \left| y_j - \sum_{k=1}^{K} b_{k1} \sin(b_{k2} x_j + b_{k3}) \right|^p \}.$$

Classical approach to solve this problem is based on the assumption of normal distribution of stochastically independent homoskedastic random errors, and hence $p = 2$ is used to get the common LSQ (least square) criterion used in regression methods in statistics. An alternative is the direct use of nonlinear optimization algorithms to solve the above minimization problem. The main problem is how to find an optimal solution when the objective function is in general non-differentiable (because of absolute values) and non-convex (because of sin function).

Illustration of difficulties

Example 1: The influence of non-convexity, and hence, the importance of suitable choice of the initial solution can be seen from the following example. Figure 1 shows the negative penalty of a harmonic function approximation. The "measurements" were computed by MATLAB [1] by the formula: $y =$ 0.5sin(x) + 0.2rand - 0.1 (additive uniform noise) in 30 equidistant points in the interval [0,14.5] with the step 0.5 - see the solid line in the Figure 2. The approximation function is$\hat{y} = b_1 \sin(b_2 x)$. The mesh is based on a matrix with dimension 100x100 where the parameter b_1 changed in the interval [-1,1] with the step 0.02 and the parameter b_2 changed in the interval [-1,2] with the step 0.03. The penalty was computed for p=1 as the sum of absolute errors. Figure 1 shows the negative penalty, so each peak of the "mountain range" represents one local minimum where the nonlinear programming search algorithm can end. The global minima close to the point b_1=0.5, b_2=1 are the two highest peaks. The exact coordinates of the global minimum found by the Excel solver are b_1=0.5157, b_2=0.9894. The other global minimum is at the same but negative coordinates: sin(x) = -sin(-x). The penalty for this approximation is smaller than the penalty of the original 'ideal' function without noise. Figure 10 shows the original data and two approximations. The correct one corresponds to the global minimum, the wrong one is the left-most local minimum in Figure 9 (exact coordinates b_1=0.2699, b_2=-0.7081).

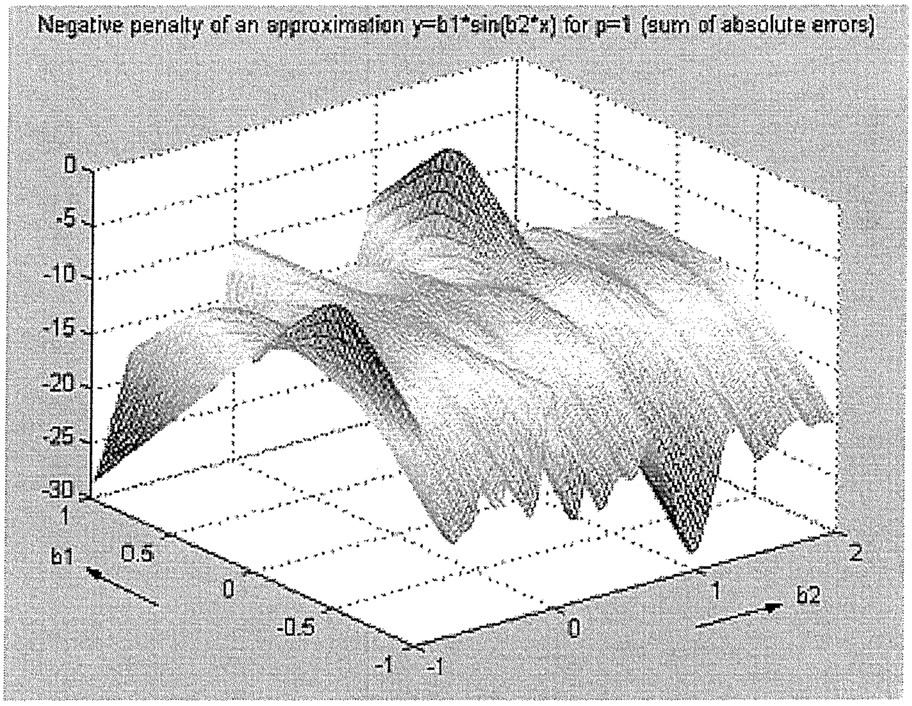

Figure 9: Penalty function of harmonic approximation

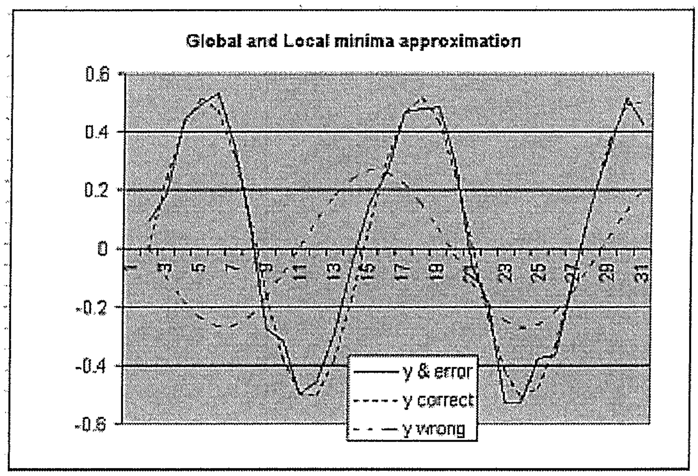

Figure 10: Harmonic approximation

Approximation methods

The introduced problem is closely related to the well known area of regression coefficients point estimates in statistics [4]. This widely used technique has computational advantages but in certain cases, it is not sufficient. For example it can fail in these cases: unknown random error probability distribution, deviations vary with changing x, and periodogram analysis generally fails for signals that contain close frequencies. In these cases, modern search techniques as genetic algorithms and neural networks are often introduced. These techniques might be slow in finding improved values and their global and local convergence is not guaranteed.

Unconstrained optimization techniques

If the criterion is differentiable (p is even) then the gradient can be computed. However the obtained system of equations is nonlinear and can only exceptionally be solved analytically. Therefore, an iteration procedure based on nonlinear optimization has to be built. If the periodogram identifies one dominating frequency, we may decompose the problem complexity starting with $K = 1$ and the function$y = b_1 \sin(b_2 x + b_3)$ with three unknown coefficients. In such simplest case, the frequency identified by the periodogram is used to initialize b_2. Then the initial value for b_3 is often set to zero and b_1 is initially estimated by the LSQ algorithm. Such estimates are often "close enough" to the global optimal solutions. Hence, the optimum can be found by efficient locally convergent algorithms.

When p is distinct from 2 (e.g. $p = 1$), the specialized nonlinear optimization algorithms for LSQ problems, such as the Marquardt - Levenberg algorithm combining the advantages of gradient method robustness and Newton's method speed, cannot be used. Still, fast conjugate directions algorithms may replace them. For example, MS Excel Solver [2] used in the above-mentioned example implements both quasi-Newton and conjugate gradient methods using quadratic approximation for line search. See [3] for details of these methods. Because two choices of approximating gradients numerically are available, the solver is robust enough to successfully deal with non-smooth functions in mid-size problems.

The MS Excel Solver is easy to use, especially for novices in optimization and it supports connection of user-friendly data inputs and graphical outputs. A more experienced users may use similar sophisticated unconstrained optimization procedures using functions from MATLAB Optimization Toolbox [1].

If the decomposition is not feasible, the previous ideas can still be used for $K > 1$. However, the problem size and complexity increases and the initial estimates may lead to local minima. In addition to the use of a pure random generation and expert's knowledge, we suggest to use exploratory analysis based on own heuristics for the initial solution choice. We may give another examples based on Excel Solver computations to show that the discussed technique may be easily implemented.

Example 2: Figure 11 shows the results of an approximation by two harmonic functions whose frequencies have irrational ratio. The 'measurements' were computed by Excel by the formula $y = \sin(0.1x) + \sin(0.1\sqrt{2}x)$+rand()$0.4 - 0.2b$ (additive uniform noise in [-0.2, 0.2]) in 200 equidistant points in the interval [1,200] with the step 1 - see the solid line in the Figure 11. The approximation function is:

$$\hat{y} = b_{11}\sin(b_{12}x + b_{13}) + b_{21}\sin(b_{22}x + b_{23}) + b_{31}$$

The Excel formula shows the 'ideal' values $(1, 0.1, 0, 1, 0.1\sqrt{2}, 0, 0)$ of the 7 approximation parameters. The values of the approximation parameters were found by the Excel solver with criterion computed for p=1 (sum of absolute errors). Figure 11 shows one good and one wrong approximations. The good one corresponds to a minimum whose exact coordinates rounded to 4 decimal places are: $b_{11} = 1.0132, b_{12} = 0.0996, b_{13} = 0.0312, b_{21} = 0.9907, b_{22} = 0.1414, b_{23} = -0.0177, b_{31} = -0.0064$. Criterion of this minimum is better (less) than the criterion of the original 'ideal' function without noise. Still it is not guaranteed that it is the global minimum. In fact by changing solver parameters the results slightly change. This may be caused by several very close minima or (probably) the global minimum is flat and the changes are caused by the solver settings. The graph obviously shows no difference. The wrong approximation in Figure 11 corresponds to one of many local minima.

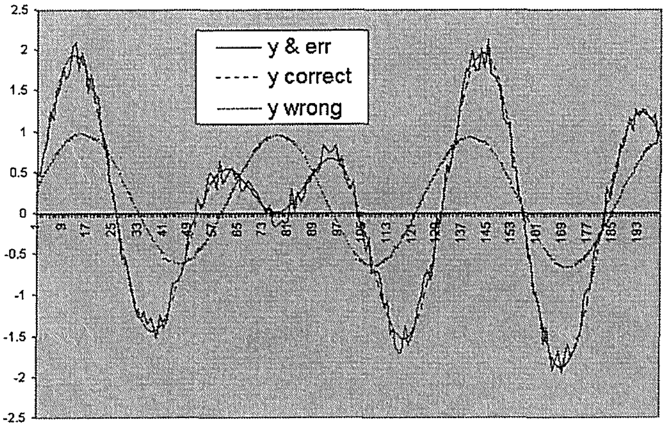

Figure 11: Approximation by two harmonic functions with irrational ratio of frequencies

Example 3: Figure 12 shows the results of an approximation by a sum of a harmonic and a linear functions:

$$\hat{y} = b_{11}\sin(b_{12}x + b_{13}) + b_{21} + b_{22}x$$

The 'measurements' were originally computed by Excel by the formula: $y = 100\sin(0.5x+3)+1+20x$ in 15 equidistant points in the interval [1,15] with the step 1. Then the values were manually modified without any clear pattern to simulate for example not accurate measurements - see the solid line in the Figure 4. Note that the reading for x=13 is very far from the original value that may be caused for example by a human error. The Excel formula shows the 'ideal' values (100, 0.5, 3, 1, 20) of the 5 approximation parameters. The values of the approximation parameters were found by the Excel solver. Figure 12 shows two good and one wrong approximations. Note the difference between the criterion computed for p=1 (sum of absolute errors) and the criterion computed for p=2 (sum of squared errors). The sum of squares is much more sensitive to excess fluctuations than the sum of absolute errors. Again, the wrong approximation corresponds to one of many local minima.

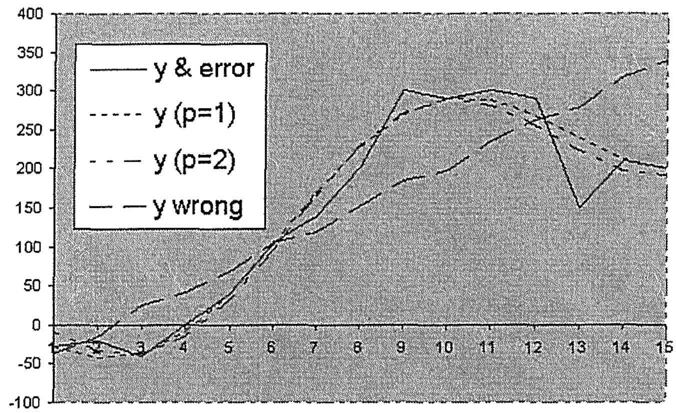

Figure 12: Approximation by a sum of a harmonic and a linear functions

Unfortunately, unconstrained optimization solvers using approximations of derivatives cannot be used in general, as their theoretical properties (guaranteed convergence) and numerical behaviour (error influence) can be questionable. Therefore, something more reliable has to be implemented for large-scale problems.

Constrained optimization technique

The above unconstrained minimization problem can be reformulated as a constrained one with 2 variables associated with each measurement:

$$0 < p < \infty : \min\{\sum_{j=1}^{n} \left(d_j^+ + d_j^-\right)^p \mid d_j^+ - d_j^- =$$
$$y_j - \sum_{k=1}^{K} f_k(x_j, \mathbf{b}_k), d_j^+ \geq 0, d_j^- \geq 0, j = 1, ..., n\}$$

Using a suitable algebraic modeling language, the problem can be described using the summation-indexed-based notation.

Conclusion

There are non-linear programming software tools based on robust search algorithms, often with the possibility to select the one that best suits the problem solved. These tools are general enough, and hence, very flexible in comparison with specialized packages. They can be easily used and require only modest programming abilities if any. Especially, they may serve well in the preparation step when the user identifies the problem features before choosing specialized algorithms and software.

References

[1] *MATLAB version 7: User's Guide,* The MathWorks Inc., 2004.

[2] *MS Excel: Solver's help and documentation*, Microsoft, Inc., 2004.

[3] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty, *Nonlinear programming: Theory and algorithms*, John Wiley & Sons, 1993.

[4] R.E. Walpole and R.H. Myers, *Probability and Statistics for Engineers and Scientists*, Prentice Hall, 1993.

[5] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, 1990.