Optical Model Parameter Search with
Simulated Annealing and Marquardt-Levenberg Method

Young-Ouk LEE, Choong-sup GIL, and Jonghwa CHANG
Korea Atomic Energy Research Institute,
Tokio FUKAHORI
Japan Atomic Energy Research Institute
e-mail: yolee@lui.kaeri.re.kr

1. Introduction

The search for the best set of optical model parameters having energy dependent terms is a non-linear multi-dimensional minimization problem with a given set of reference measurements and physical constraints. Deterministic approaches, such as the gradient or non-linear least square methods, often lead a resulting χ^2 to local minima for their fast convergence. On the other hand, grid search approach such as the simulated annealing (SA) method [1] is known to give a pseudo-global minimum of χ^2 [2, 3]. But the SA method has two major shortcomings; the convergence speed and the absence of a covariance matrix. The convergence of the SA method becomes extremely slow as χ^2 approaches the minimum. Furthermore, the SA method does not provide a covariance matrix, which the deterministic method does, and this is useful for a parametric study of the optical model analysis. In the present work, the grid search and deterministic approaches were sequentially applied to take advantage of both methods.

2. Simulated Annealing Method

Simulated annealing method tries to find coefficients of the global optimum parameters, giving minimum χ^2 with respect to the reference measured data.

The actual procedure of SA operation in the present work is :

- Start with initial coefficients within their upper and lower bounds provided by the user and set initial sampling range VM_j as half of the bounds. Estimate χ^2 with starting set of coefficients.

$$\alpha_j \in [\text{LB}_j, \text{UB}_j]$$
$$\text{VM}_j = \frac{1}{2}(\text{UB}_j - \text{LB}_j)$$
$$(\chi^2)^0 = \chi^2 \ (\alpha_j^0) \tag{1}$$

- For next step, it randomly chooses a trial point within the step length of VM_j and χ^2 is evaluated at this point with random variable y_j

$$\alpha_j^{k+1} = \alpha_j^k + y_j \cdot r_j, \quad y_j \in [-1, 1]$$
$$(\chi^2)^{k+1} = \chi^2(\alpha_j^{k+1}) \tag{2}$$

- This χ^2 is compared to its previous one. All downhill moves are accepted and the algorithm continues from that trial point. i.e.

$$\text{accepted if} \quad (\chi^2)^{k+1} < (\chi^2)^k \tag{3}$$

- Uphill moves may be accepted; the decision is made by the Metropolis criteria with random variable z

$$\text{also accepted if} \quad \exp\left(\frac{(\chi^2)^k - (\chi^2)^{k+1}}{T}\right) > z, \quad z \in [0, 1] \tag{4}$$

It uses T(temperature) and the size of the uphill move in a probabilistic manner. The smaller T and the size of the uphill move are, the more likely that move will be accepted.

- If the trial is accepted, the algorithm moves on from that point. If it is rejected, another point is chosen instead of a trial evaluation.

- Each element of VM_j is periodically adjusted so that half of all function evaluations in that direction are accepted.

$$\beta_j = \frac{\text{number of accepted}}{\text{number of trial}} \tag{5}$$

$$\begin{aligned} VM_j &= VM_j \ (1.0 + c \ (\beta_j - 0.6)/0.6) \quad \text{if} \quad \beta_j > 0.6 \\ VM_j &= VM_j \ (1.0 + c \ (0.4 - \beta_j)/0.4) \quad \text{if} \quad \beta_j < 0.4 \end{aligned} \tag{6}$$

- A fall in T is imposed upon the system with the RT variable by

$$T_{n+1} = RT \cdot T_n \tag{7}$$

where n is the nth iteration with the same T. Thus, as T declines, uphill moves are less likely to be accepted and the percentage of rejections rise. Given the scheme for the adjusting VM, VM falls. Thus, as T declines, VM falls and SA focuses upon the most promising area for optimization.

- The termination criteria for the search is set if the last four χ^2's from the last four different T's differ from the current χ^2 by less than the user-defined tolerance(EPS) and the current χ^2 at the current T differs from the current optimal χ^2 by less than EPS.

The parameter T is crucial in using SA successfully. It influences VM, the step length over which the algorithm searches for optima. For a small initial T, the step length may be too small; thus the search may fail to find the global optima. The user should carefully examine to make sure that VM is appropriate. The relationship between the initial temperature and the resulting step length is function-dependent.

3. Nonlinear model and Marquardt-Levenberg Method

Let us briefly review the nonlinear fitting using excerpts of ref. [4]. With nonlinear dependences, the minimization must proceed iteratively for a nonlinear model with the set of M unknown parameters $a_k, k = 1, 2, \ldots, M$.

We expect the χ^2 function to be well approximated by a quadratic form, written as

$$\chi^2(\mathbf{a}) \approx \gamma - \mathbf{d} \cdot \mathbf{a} + \frac{1}{2}\mathbf{a} \cdot \mathbf{D} \cdot \mathbf{a} \tag{8}$$

where $\mathbf{d}$ is an M-vector and $\mathbf{D}$ is an $M \times M$ matrix. Let us examine how to calculate the gradient and Hessian of χ^2 merit function. The model to be fitted is

$$y = y(x; \mathbf{a}) \tag{9}$$

and the χ^2 merit function is

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{N} \left[\frac{y_i - y(x_i; \mathbf{a})}{\sigma_i} \right]^2 \tag{10}$$

The gradient of χ^2 with respect to the parameters a, which will be zero at the χ^2 minimum, has components

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^{N} \frac{[y_i - y(x_i; \mathbf{a})]}{\sigma_i^2} \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \quad K = 1, 2, \ldots, M \tag{11}$$

Taking an additional partial derivative gives

$$\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} = 2 \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \frac{\partial y(x_i; \mathbf{a})}{\partial a_l} - [y_i - y(x_i; \mathbf{a})] \frac{\partial^2 y_i(x_i; \mathbf{a})}{\partial a_l \partial a_k} \right] \tag{12}$$

It is conventional to remove the factors of 2 by defining

$$\beta_k \equiv -\frac{1}{2}\frac{\partial \chi^2}{\partial a_k} \quad \alpha_{kl} \equiv \frac{1}{2}\frac{\partial^2 \chi^2}{\partial a_k \partial a_l} \tag{13}$$

making $[\alpha] = \frac{1}{2}\mathbf{D}$ in equation (8), and rewriting the equation as a set of linear equations

$$\sum_{l=1}^{M} a_{kl}\delta a_l = \beta_k \tag{14}$$

This set is solved for the increments δ_l that, added to the current approximation, give the next approximation. In the context of least-squares, the matrix $[\alpha]$, equal to one-half times the Hessian matrix, is usually called the curvature matrix. iNote that the components α_{kl} of the Hessian matrix (12) depend both on the first derivatives and on the second derivatives of the basis functions with respect to their parameters. Second derivative term can be dismissed when it is zero (as in the linear case), or small enough to be negligible when compared to the term involving the first derivative. As the definition of the α_{kl}, we will use the formula

$$\alpha_{kl} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}\left[\frac{\partial y(x_i;\mathbf{a})}{\partial a_k}\frac{\partial y(x_i;\mathbf{a})}{\partial a_l}\right] \tag{15}$$

Marquardt has put forth a method, related to an earlier suggestion of Levenberg, for varying smoothly between the extremes of the inverse-Hessian method and the steepest descent method. The latter method is used far from the minimum, switching continuously to the former as the minimum is approached. This Marquardt-Levenberg method works well in practice and has become the standard of nonlinear least-squares routines. The method is based on two elementary insights. Marquardt's first insight is that the components of the Hessian matrix, even if they are not usable in any precise fashion, give some information about the order-of-magnitude scale of the problem. The quantity χ^2 is non-dimensional, i.e. is a pure number; this is evident from its definition(10).

On the other hand, β_k has the dimensions of $1/a_k$, which may well be dimensional, i.e. have units like cm^{-1}, or kilowatt-hours, or whatever. The constant of proportionality between β_k and δa_k must therefore have the dimensions of a_k^2. Scan the components of $[\alpha]$ and you see that there is only one obvious quantity with these dimensions, and that is $1/\alpha_{kk}$, the reciprocal of the diagonal element. So that must set the constant by some (non-dimensional) fudge factor λ, with the possibility of setting $\lambda \gg 1$ to cut down the step, In other words, replace equation (??) by

$$\delta a_l = \frac{1}{\lambda a_{ll}}\beta_l \quad \text{or} \quad \lambda a_{ll}\delta a_l = \beta_l \tag{16}$$

It is necessary that a_{ll} be positive, but this is guaranteed by definition (15).

Marquardt's second insight is that equations (16) and (14) can be combined if we define a new matrix α' by the following prescription

$$\begin{aligned}\alpha'_{jj} &\equiv \alpha_{jj}(1+\lambda)\\\alpha'_{jk} &\equiv \alpha_{jk}(j \neq k)\end{aligned} \tag{17}$$

and then replace both (16) and (14) by

$$\sum_{l=1}^{M} \alpha'_{kl}\delta a_l = \beta_k \tag{18}$$

When λ is very large, the matrix α' is forced into being diagonally dominant, so equation goes over to be identical to (16). On the other hand, as λ approaches zero, equation (18) goes over to (14).

Given an initial guess for the set of parameters $\mathbf{a}$, the present work performs the following procedure [5] :

- Compute $\chi^2(\mathbf{a})$.

- Pick a modest value for λ.

- (†) Solve the linear equations (18) for $\delta\mathbf{a}$ and evaluate $\chi^2(a + \delta\mathbf{a})$.

- If $\chi^2(\mathbf{a} + \delta\mathbf{a}) \geq \chi^2(\mathbf{a})$, increase λ by a factor of 10 and go back to (†).

- If $\chi^2(\mathbf{a} + \delta\mathbf{a}) < \chi^2(\mathbf{a})$, decrease λ by a factor of 10, update the trial solution $\mathbf{a} \leftarrow \mathbf{a} + \delta\mathbf{a}$, and go back to (†).

Once the acceptable minimum has been found, one wants to set $\lambda = 0$ and compute the matrix

$$[C] \equiv [\alpha]^{-1} \tag{19}$$

which is the estimated covariance matrix of the standard errors in the fitted parameters $\mathbf{a}$.

4. Application and Results

A reference measurements set was constructed based on the analyses of raw experimental data and/or other sources of evaluations. The search procedure starts with the SA algorithm with the reference data set and physical constraints imposed on the functional forms of optical model parameters. The importance of the reference points were dynamically shifted in the course of the SA search, varying their corresponding errors, based on the physical constraints and eye guidance. Appropriate conditions were studied on whether the SA reaches the vicinity of the global minimum of the χ^2. Then the procedure is switched into a deterministic way using the Marquardt-Levenberg Method to accelerate the convergence to the global minimum, as well as to produce a covariance matrix. Figure 1 shows the characteristics of minimum χ^2's achievable from three cases, namely the simulated annealing alone, the Marguardt-Levenberg alone, and combination of the two methods, in the domain of adjustable parameters. In fig. 2, the change of χ^2 is plotted as the iteration proceeds for three cases. As shown, a global minimum is obtained by combining the Marquardt-Levenberg (ML) Algorithm and the simulated annealing (SA) methods. It is also noted that the SA method alone stops at the vicinity of the real minimum and the M-L method traps in the local minimum.

References

[1] Lee, Y.: *ECISPLOT: An interactive optical model parameter searcher with simulated annealing algorithm.* Tech. Rep. NDL-9/99, KAERI (1999).
[2] Lee, Y., *et al.*: *J. of Nucl. Sci. and Tech.*, **36**(12), 1125 (1999).
[3] Lee, Y., *et al.*: In *Proc. of PHYSOR 2000: ANS International Topical Meeting on Advances in Reactor Physics and Mathmatics and Computation into the Next Millenium.* Pittsburgh, Pennsylvania, U.S.A, May 7-12, 2000 (2000).
[4] Press, W. H., *et al.*: *Numerical Recipes.* Cambridge Univ. Press, The Pitt Building, Trumpington Street, Cambridge CB2 IRP (1986).
[5] Lee, Y.: *MLFIT : A portable fitting tool using the nonlinear least-squares (NLLS) marquardt-levenberg algorithm.* Tech. Rep. NDL-13/00, KAERI (2000).
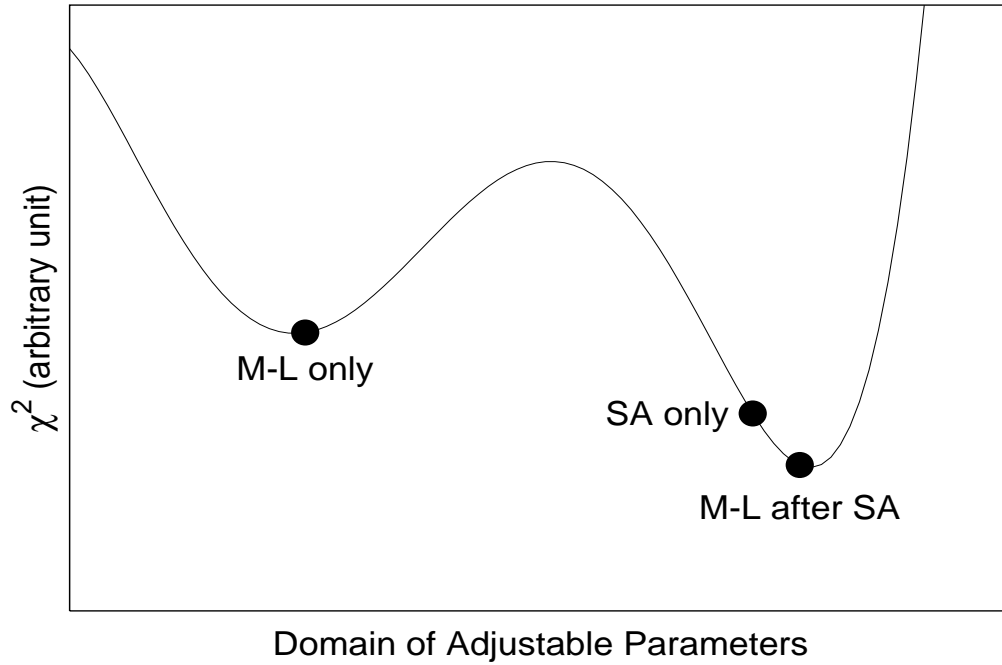
Figure 1: Characteristics of χ^2 for 3 cases: the simulated annealing (SA) alone, the Marguardt-Levenberg(ML) alone, and combination of the two methods
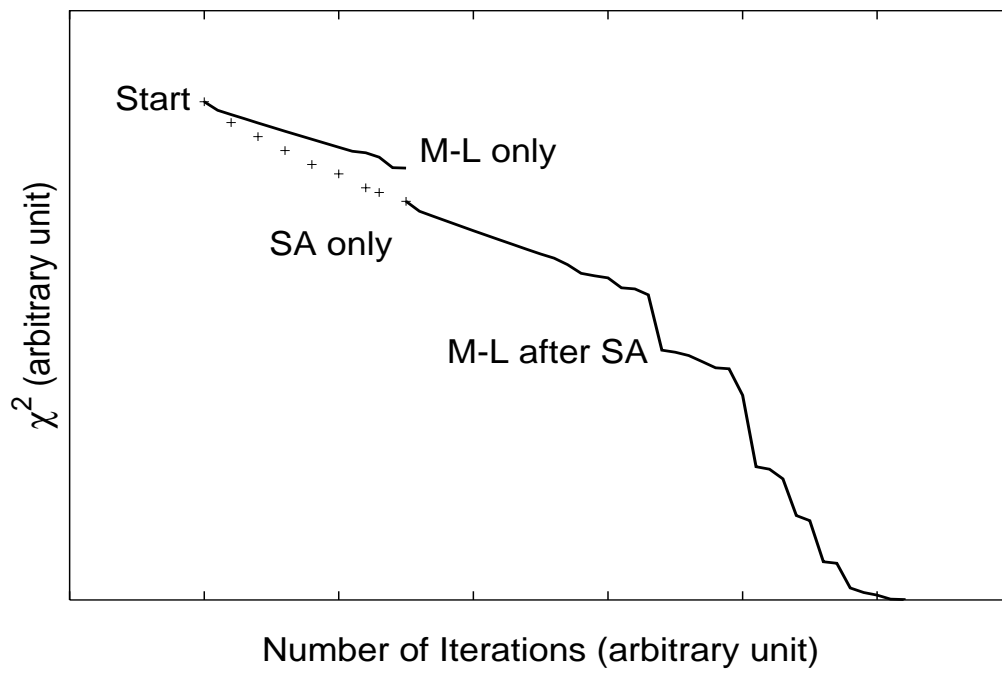

Figure 2: χ^2 changes as a function of iterations as the iteration proceeds for three cases