

D 4 Algorithms for Optimization

H. Lustfeld

Institut für Festkörperforschung

Forschungszentrum Jülich GmbH

Contents

1	Introduction	2
2	Tomographic Problems, Neural Networks and the Steepest Descent Method	4
3	The Cost Function of the Electron Microscope	6
4	Minimizing a System with the Simulated Annealing Method	8
5	Criticism, Variants and Power of the Simulated Annealing Method	10
6	Minimizing a System with the Genetic Algorithm	11
7	Criticism, Variants and Power of the Genetic Algorithm	12
8	Ground States of Molecules as an NP Complete Problem, Simulated Annealing and Molecular Dynamics	13
9	Conclusion	16
A	Modified Steepest Descent Methods and Regularization	17

1 Introduction

Optimization is one of the most important techniques in natural sciences, in engineering sciences and economic sciences. In contrast to simple one dimensional text book problems where nearly any method works, optimization problems are usually of a quite different nature. To avoid abstractness typical representatives are listed here:

- A. a) tomographic problems:
 - i) computer tomography in medicine
 - ii) exploration of soil structure [1, 2]
- b) learning in neural networks [3]
- B. yes/no decision about the success of experiments
- C. reconstruction of noisy and/or deformed patterns [4]
- D. a) traveling salesman problem [5, 6, 7]
- b) magnetic ground state of solids (spin glass) [8]
- c) formation of molecule clusters in the ground state [9, 10]
- E. optimization problems in economy

In all these examples a *cost function* or *evaluation function* is defined. It depends on parameters or variables of the system and indicates how well these parameters/variables have been chosen compared with other sets of parameters/variables. The best set minimizes the cost function. Obviously this optimization has to be done typically not in a 2 or 3 dimensional space but in a *multidimensional* space. In fact, the dimension can be of order 10^1 or more (problem of group *C* and group *Dc*), of order 10^2 (again problems of group *C* and group *Dc*) up to 10^6 in problems of group *A*, *B*, *D* and *E*.

In addition to the multidimensionality another difficulty arises in all problems (except *B*) and in particular in those of group *D*: There exists not only a global minimum but local minima as well. In group *D* the number of these local minima increases enormously, possibly even exponentially with the system size. Problems having this property are called *NP complete* [6, 5].

A word on problems of group *B* is in order. In problems of this kind the method of *support vector machines* [11, 12, 13, 14] (SVM) can be applied successfully. This success rests on the fact that within this framework the cost functions have only one single minimum¹.

The above listed problems can be treated with various methods. The simplest one is that of *steepest descent* [5]. This method and its more sophisticated variants as *conjugate gradient method* [5], *Gauss-Newton method* [15] and *Levenberg-Marquardt method* [16] including possible *regularization* [17] will be discussed in section 2. It is surprising how well these procedures

¹After a tricky transformation into high dimensions a problem of group *B* becomes linear and analogous to the task of optimizing the perceptron. (The perceptron is one of the earliest neural networks and optimization of its cost function is simple since there is only a single minimum.) A discussion of SVM is beyond the scope of this paper. But one should keep in mind that it is the only method in which a nonlinear complicated problem can be mapped onto one in which only a single minimum appears that has to be determined. This makes this method very interesting.

work in problems of group *A*, *B* and partly in problem *C* although these methods are searching for the next local minimum and are expected to get stuck there.

In problems of group *C* and even more of group *D* it is not sufficient to detect a minimum because there are many local ones well above the global minimum. And either the global minimum is of interest (e.g. when determining the ground state of a molecule) or at least a minimum that is close to the global minimum (e.g. when determining the ground state of a spin glass, in the traveling salesman problem, etc.). The most useful methods applicable here are methods derived from thermodynamics: It is an experimental fact that a system first heated and then slowly cooled down will get into its ground state. This is due to thermal fluctuations allowing the system to overcome local minima. The *Metropolis algorithm* [18] is a mathematical analogy of this. Defining the cost function as an energy, introducing an inverse temperature β and applying the Metropolis procedure leads to the *simulated annealing* [5, 19, 20] method which is most useful in dealing with NP complete problems [5, 6]. To demonstrate how this method is used it is applied to the cost function of the electron microscope [4]. This cost function is derived in section 3 and the demonstration is presented in section 4. In section 5 problems and variants of the simulated annealing method are discussed.

Closely related to simulated annealing is the *genetic algorithm* which starts with an ensemble that repeatedly generates ‘mutations’ in analogy to the simulated annealing method. The new ingredient is the generation of ‘offsprings’ to include parameters that have an exponentially small probability of being included in the simulated annealing method. The genetic algorithm is applied to the cost function of the electron microscope (section 6). Problems and variants of that method are discussed in section 7.

Both the simulated annealing method and the genetic algorithm face difficulties if the accessible states are continuous. One suggestion in this case is applying the *downhill simplex method* together with simulated annealing [5]. The downhill simplex method generates a moving structure of $N + 1$ points in an N dimensional space, takes into account $N + 1$ values at the same time and therefore can overcome a local minimum without further ingredients if the basin of attraction of that minimum is sufficiently small [21]. Therefore the combination of this method with simulated annealing is quite interesting.

But there is a different method: A cost function, say E , with parameters, say \mathbf{q} , is interpreted as a potential energy $V(\mathbf{q})$ and added to it is a (fictitious) kinetic energy term K . Thus we obtain a fictitious Hamiltonian $H = K + V$ and can follow fictitious trajectories. Along these trajectories H is a constant of motion and V is not. This means that parameter values are allowed with higher values of V in an interval width $\Delta V \propto 1/\beta$. Reducing the kinetic energy K (e.g. by adding a dissipative term) leads to a decrease of ΔV and corresponds to reducing the temperature. In this way one can hope to find a global minimum of the cost function V . This scheme has been applied to finding the ground state of molecules [22] and will be presented in section 8. The scheme demonstrates also connections between *molecular dynamics* [23] and simulated annealing. The conclusion ends the paper.

We will not elaborate on problems of group *E* in the following text. The reason they were mentioned at all is to give a quite different example for the applicability of the optimization methods described here. These methods are essential for solving problems far beyond the range for which they were originally developed.

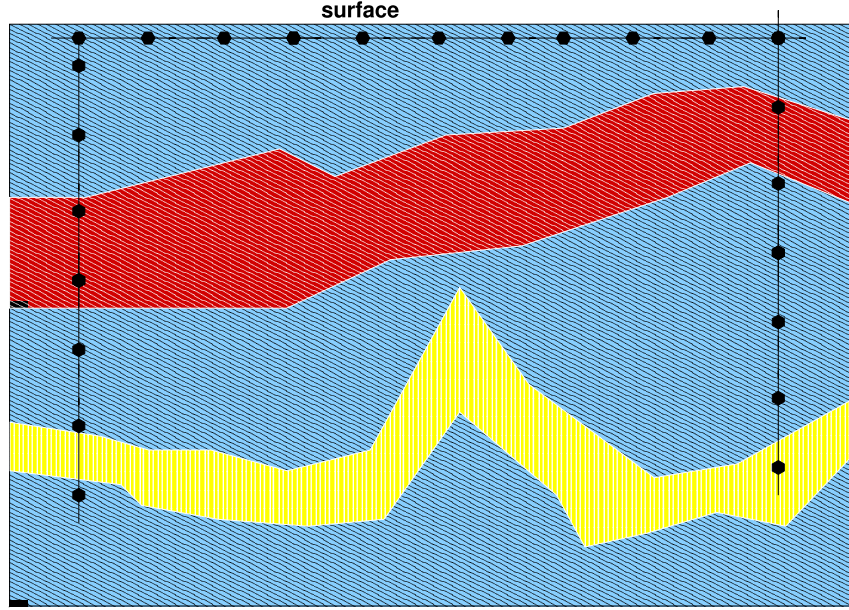


Fig. 1: Schematic view of different soil structures and electrodes which can either inject positive and negative electric current into the soil or measure the electric potential Φ .

2 Tomographic Problems, Neural Networks and the Steepest Descent Method

Tomographic problems consist of a *forward problem*, a *backward problem* and a cost function connecting the two. Take the exploration of soil structure via dc currents as an example. Quite often the conductivity of the soil is sufficient to identify its structure. This fact is exploited in the **electric resistivity tomography** scheme (ERT) [2], in which the conductivity of the soil is determined: cables containing electrodes are dug in the soil partly close below its surface, partly in boreholes, cf. Fig. 1. Currents flowing to and from electrodes through the soil generate voltages at the electrodes which are measured. Given a configuration of σ values²

$$\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_{N_d}) \quad (1)$$

the electric potential Φ can be computed at any point³ \mathbf{r} in the soil for that configuration (forward problem):

$$\Phi = \Phi(\mathbf{r}, \boldsymbol{\sigma}) \quad (2)$$

The quality of the solution can be found when it is compared with the measurements at points \mathbf{r}_i , $i = 1, \dots, N_m$. To get the best $\boldsymbol{\sigma}$ values we define a cost function

$$E(\boldsymbol{\sigma}) = \frac{1}{2N_m} \sum_{i=1}^{N_m} (\Phi(\mathbf{r}_i, \boldsymbol{\sigma}) - \Phi_i)^2. \quad (3)$$

The minimum of the cost function gives the best choice of $\boldsymbol{\sigma}$. The *backward problem* consists in finding this minimum.

²We assign σ values to small volumes of soil. Thus the dimension N_d of $\boldsymbol{\sigma}$ can become quite large. E.g. choosing a relative resolution in each direction of $1/30$ means $N_d = 30^3$.

³To get Φ a partial differential equation of the Poisson type has to be solved.

Feed forward neural networks are well known [3]. They have an input layer with input \mathbf{x} , an output layer with output $G(\mathbf{x})$, G real⁴, and one or more inner layers with ‘neurons’ representing nothing but sigmoid functions⁵ $\varphi(s)$, s and φ are real. The efficiency of the networks rests on the *universal approximation theorem* [3] which states:

$G(\mathbf{x})$ real and continuous, \mathbf{x} a vector of a closed subset $S \subset R^N$ then for any $\epsilon > 0$ there exist $M < \infty$, coefficients α_i and θ_i and a real matrix W_{ij} such that

$$|G(\mathbf{x}) - G_{ap}(\mathbf{x})| < \epsilon$$

$$G_{ap}(\mathbf{x}) = \sum_{i=1}^M \alpha_i \varphi\left(\sum_{l=1}^N W_{il} x_l - \theta_i\right) \quad (4)$$

This theorem is quite important but it does not tell how to calculate the coefficients - in contrast to the Fourier theorem for example. However, that is not an disadvantage because it is impossible to compute Fourier coefficients by a Fourier transform for high dimensions, ($\mathcal{O}(100)$ is typical). The *learning process* of the network is nothing but finding out appropriate coefficients of the expansion. To achieve this a cost function is defined

$$E(\boldsymbol{\sigma}) = \frac{1}{2N_v} \sum_{i=1}^{N_v} (G_i - G_{ap}(\mathbf{x}_i))^2, \quad \boldsymbol{\sigma} = \boldsymbol{\alpha}, \mathcal{W}, \boldsymbol{\theta}. \quad (5)$$

G_i are values $G(\mathbf{x}_i)$, $i = 1, \dots, N_v$ that are known. The minimum of the cost function gives the best approximation for the coefficients $\boldsymbol{\alpha}, \mathcal{W}, \boldsymbol{\theta}$.

It is obvious that the minimum search in the backward problem of tomography and the learning phase of the feed forward neural networks are mathematically identical. The simplest method is obtained when looking at small changes⁶ of E :

$$dE = \nabla E \cdot d\mathbf{x} \quad (6)$$

Obviously the *steepest descent* occurs for

$$d\mathbf{x} = -\eta \nabla E, \quad \eta > 0, \quad \eta \ll 1. \quad (7)$$

The size of η depends on the problem. One possibility is the finding of the minimum along the direction of $-\nabla E$, in other cases η is treated as a parameter. But we can do better. The reason for this is the special structure of the cost functions. They have the form

$$E = \frac{1}{2} \mathbf{f}^+ \mathbf{f} \quad (8)$$

and its global minimum value is 0 (if no noise is present) or very small (if noise is present). This means that in the final stage of the minimum search the second derivative of E is well

⁴ G need neither be a scalar nor real, but we treat the simplest case here.

⁵a real function $\varphi(s)$ is sigmoid if it is monotonously increasing and $\varphi(s = -\infty) = 0$, $\varphi(s = \infty) = 1$.

⁶We assume for simplicity that \mathbf{x} is a real quantity in Eq. (6). Otherwise the always real function E could not be differentiated with respect to \mathbf{x} . However, if \mathbf{x} is complex, E can be written as $E(\mathbf{x}_1, \mathbf{x}_2)$ with $\mathbf{x}_1 = \mathbf{x}$ and $\mathbf{x}_2 = \mathbf{x}^*$. Applying this trick E becomes differentiable with respect to \mathbf{x}_1 and \mathbf{x}_2 . It is then sufficient to discuss the steepest descent with respect to \mathbf{x}_1 since the steepest descent with respect to \mathbf{x}_2 does not lead to a new condition. Thus the modifications for complex \mathbf{x} are trivial and will not be discussed here.

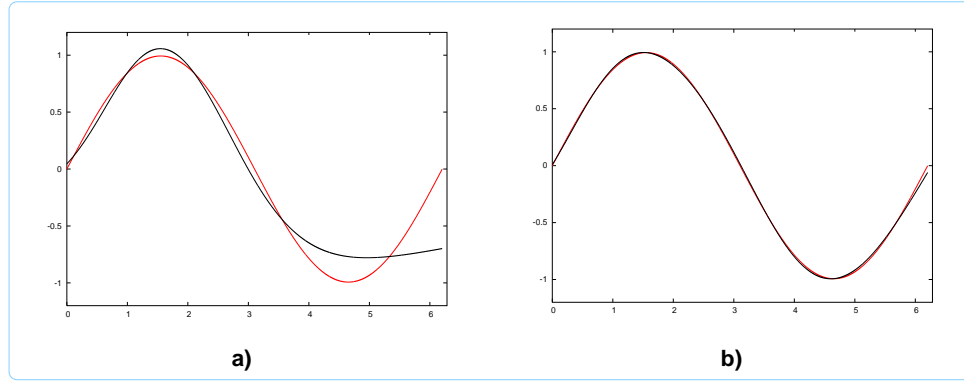


Fig. 2: Approximation of $\sin(x)$ in $[0, 2\pi]$ using the universal approximation theorem [25]. The number of neurons is 9, $\eta = 0.1$, (cf. Eq. (7)). As sigmoid function the Fermi function has been chosen: $\varphi(s) = 1/(1 + e^{-s})$. Red line: $\sin(x)$, black line: approximation. a) cost function $E < 10^{-2}$, b) cost function $E < 10^{-6}$.

approximated by first derivatives of \mathbf{f} . This fact is exploited in various iteration schemes like the *Gauss-Newton method*, *Levenberg-Marquardt method* and in connection with those: the *conjugate gradient method*. Details are given in appendix A.

Quite often the minimum of the cost function is ill conditioned, which spoils the uniqueness of the solution because there is a null space present. If a unique solution is required as in tomographic problems a regularization becomes necessary [17]. This acts as a filter enforcing the uniqueness. Details are given in appendix A.

The idea behind the steepest descent method is not very deep, the method itself is simple and, as expected, its convergence is not fast. Despite of that it is one of the most used methods in feed forward neural networks⁷ and it builds the fundamental system of equations in the *Kohonen model* [3, 24] of *neural networks*⁸.

The main objection against the steepest descent and the related refined methods comes from a quite different argument: We approach the minimum of the first basin of attraction, and this minimum does not need to be the global minimum. In the view of this fact it is really surprising how well these methods work, cf. Fig. 2. Nevertheless, as soon as there are many local minima around these methods are bound to fail. This can be detected by starting at very different points in phase space and comparing the results. If the results do not coincide, probably those methods have to be applied that will be described in sections 4, 5 and 6. Their mode of action will be demonstrated in a relatively uncomplicated case, the cost function of the electron microscope. This cost function is derived in the following section.

3 The Cost Function of the Electron Microscope

In the high resolution electron microscope (resolution 0.8\AA , magnification $1 : 10^6$) an electron beam is generated, accelerated and focused when passing through a system of magnetic lenses. It hits a target of thickness about 30 nm located in the object plane of the microscope. After having passed the object plane the beam is directed by another magnetic lens system (system 2)

⁷The method is called *back-propagation* [3] in this context.

⁸In Kohonen's network the gradient is replaced by stochastic terms that are gradients only in the average.

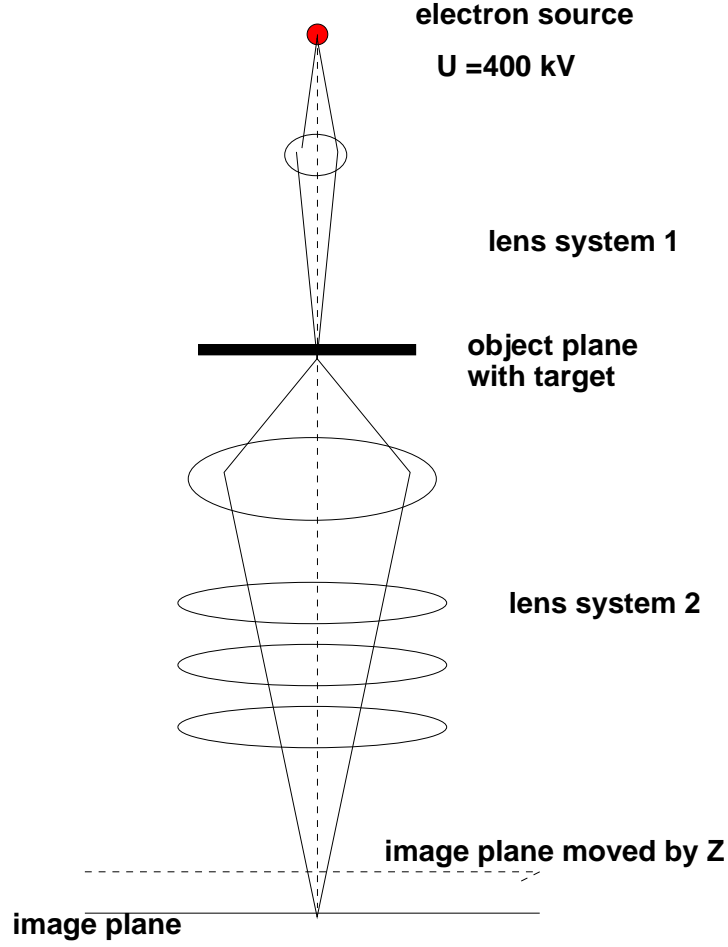


Fig. 3: Scheme of the electron microscope: the electrons are generated in the electron source, accelerated to 400 keV, focused in the object plane. After penetrating the target and passing lens system 2 they hit the image plane. The image plane can be moved up and down by changing the defocus Z to measure the intensity for different phases.

to the image plane where a high-resolution CCD camera (resolution $20 \mu\text{m}$ per pixel) measures the intensity distribution of the beam for a given focus value Z_l , cf. Fig. 3

$$I_{ms}(\mathbf{r}_i, Z_l), \quad i = 1..N \quad (9)$$

(The \mathbf{r}_i are uniformly distributed on the image plane.) Typical high-resolution intensity images for different Z values are shown in Fig. 5.

The intensity distribution in one image plane contains insufficient information. Needed is the electronic wave function Ψ after just having passed the target:

$$\Psi(\hat{\mathbf{r}}) = \int \hat{F}(\hat{\mathbf{k}}) e^{2\pi i \hat{\mathbf{k}} \hat{\mathbf{r}}}, \quad \hat{\mathbf{r}} = (x, y, z), \quad \hat{\mathbf{k}} = (k_x, k_y, k_z) \quad (10)$$

The complex coefficients $\hat{F}(\hat{\mathbf{k}})$ have to be found.

In vertical direction (i.e. perpendicular to the target) the \mathbf{k} vector component is in very good approximation determined by the kinetic energy of the beam, thus constant and we can write

$$\hat{F}(\hat{\mathbf{k}}) = \hat{F}(\mathbf{k}, k_z = \text{const}) = F(\mathbf{k}), \quad \mathbf{k} = (k_x, k_y). \quad (11)$$

The connection between intensity distribution I in the image plane and the amplitude F of Ψ is given by [4]

$$I(\mathbf{q}, Z) = \int F^*(\mathbf{k})F(\mathbf{k} + \mathbf{q})T(\mathbf{k} + \mathbf{q}, \mathbf{k}, Z)d^2k. \quad (12)$$

The so called transfer function $T(\mathbf{k}_1, \mathbf{k}_2, Z)$ is a known function of the electron microscope, reflecting its properties including lens aberrations (spherical aberration, astigmatism, coma etc.) and in particular the change due to defocusing by an amount Z .

With this knowledge the cost function

$$E = E(\{F\}) \quad (13)$$

must be constructed. Now because of Eq. (12) for each set $\{F\}$ and each defocus Z an intensity

$$I(\{F\})(\mathbf{r}_i, Z) = \sum_k I(\{F\})(\mathbf{k}, Z)e^{2\pi i\mathbf{k}\mathbf{r}_i} \quad (14)$$

can be computed and the vectors

$$\mathbf{I}(Z_l) = (I(\{F\})(\mathbf{r}_1, Z_l), \dots, I(\{F\})(\mathbf{r}_N, Z_l)) \quad (15)$$

$$\mathbf{I}_{ms}(Z_l) = (I_{ms}(\mathbf{r}_1, Z_l), \dots, I_{ms}(\mathbf{r}_N, Z_l)) \quad (16)$$

be defined, the components of which are just the computed intensity distribution. Normalizing the intensities a cost function can be defined as⁹

$$\begin{aligned} E_0(\{F\}) &= \sum_l \left(\frac{\mathbf{I}(Z_l)}{|\mathbf{I}(Z_l)|} - \frac{\mathbf{I}_{ms}(Z_l)}{|\mathbf{I}_{ms}(Z_l)|} \right)^2 \\ &= 2 - 2 \sum_l \frac{\mathbf{I}(Z_l) \cdot \mathbf{I}_{ms}(Z_l)}{|\mathbf{I}_{ms}(Z_l)| \|\mathbf{I}_{ms}(Z_l)\|}. \end{aligned}$$

The scalar product defines a $\cos \varphi$ between the measured and computed distribution. We require high sensitivity for small deviations i.e. for small angles. Since E_0 is rather insensitive in that regime we prefer to define the cost function as [4]

$$E(\{F\}) = \sum_l \arccos \varphi_l, \quad \cos \varphi_l = \frac{\mathbf{I}(Z_l) \cdot \mathbf{I}_{ms}(Z_l)}{|\mathbf{I}_{ms}(Z_l)| \|\mathbf{I}_{ms}(Z_l)\|}. \quad (17)$$

The next task is minimizing E . This will be achieved with the simulated annealing method in section 4 and with the genetic algorithm in section 5.

4 Minimizing a System with the Simulated Annealing Method

Simulated annealing is now applied to minimize the cost function E [4]: Interpreting the cost function as an energy of the corresponding configuration, we can do thermodynamics in the spirit of Monte Carlo [18]: We attribute to each configuration the Boltzmann factor $e^{-\beta E}$, β acting as a parameter, the ‘inverse temperature’. Then we go through the following steps:

⁹It is perhaps counterintuitive that out-of-focus intensities, i.e. intensities with focus values $Z \neq 0$ are used, but this is an ingenious trick yielding further important information on the complex amplitudes of Ψ , in particular on the phases.

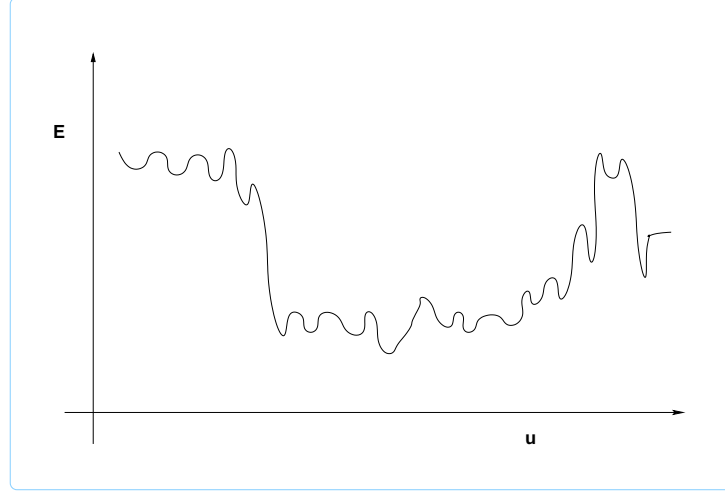


Fig. 4: Schematic plot of the cost function as function of u . There are many local minima possible far above the global minimum. This situation can occur in nearly any optimization problem and is present in the so called NP complete problems. An exception are problems to which the support vector machinery can be applied. Then there exists only one minimum, cf. the introduction.

- 1.: Initializing the variables $\{F\}$ at the beginning:

These are determined by a random number generator. The idea behind this is that at high temperature each configuration has the same weight. Consistent with this is the choice of a small β value.

- 2.: As in Monte Carlo we define a transition probability $\pi_{p \rightarrow n}$ from the present state F_p to a new one F_n by

$$\pi_{p \rightarrow n} \propto e^{-\beta(E_n - E_p)}$$

We require $\pi_{p \rightarrow n} = 1$ for $E_n - E_p < 0$. This means because of continuity

$$\pi_{p \rightarrow n} = \begin{cases} e^{-\beta(E_n - E_p)} & \text{for } E_n \geq E_p \\ 1 & \text{otherwise} \end{cases} \quad (18)$$

Eq. (18) is the *Metropolis algorithm*. This is the heart of the method: *unfavorable states are not completely excluded as in standard methods. Rather they can appear in the next step with a nonzero probability*. The idea behind it is to overcome the local minima by a simulated thermal fluctuation, cf. Fig. 4. In our example one $F(\mathbf{k})$, randomly chosen, is changed using a random number generator, with this new F the difference $E_n - E_p$ is computed and the new configuration is chosen with a probability given by Eq. (18).

- 3.: Step 2 is repeated N_s times. In analogy to thermodynamics one could argue that N_s should be so large that

$$\mathcal{O}(\Delta E) = \frac{1}{\beta}, \quad \Delta E = |E_n - E_p| \quad (19)$$

In reality N_s is an experimental parameter.

- 4.: The value of β is increased by $\Delta\beta$, which reduces the possible fluctuations of E . $\Delta\beta$ is a parameter as well. Now step 3 is repeated.

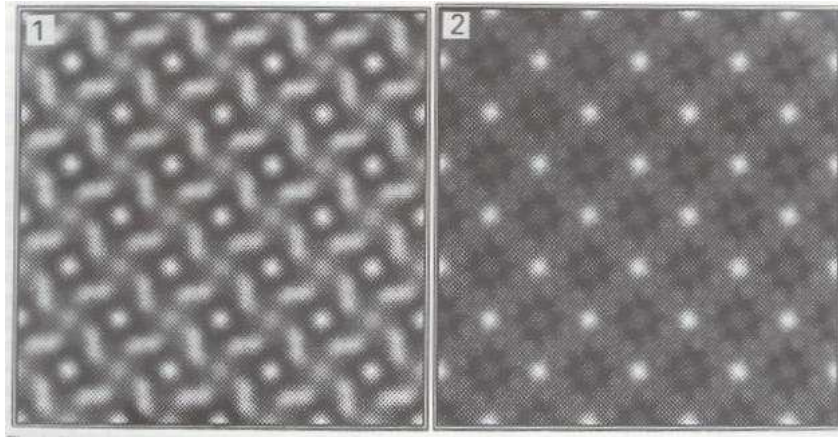


Fig. 5: Simulated high-resolution intensity images of Ni_4Mo [4]. 4×4 unit cells are shown. Image 1 is simulated for a defocus value of -35 nm, image 2 is simulated for a defocus value of -70 nm.

- 5.: Step 4 is repeated reducing the possible fluctuation further and further till the cost function does not decrease any more. The obtained $\{F\}$ is taken as solution.

The result can be observed in Fig. 6. The simulated annealing method has been able to reconstruct nearly completely the coefficients of the electronic wave function, compare Fig. 6b with Fig. 6a.

5 Criticism, Variants and Power of the Simulated Annealing Method

There are some objections against the method: At first sight it may seem that the method is an application of statistical theorems without making sure that their presumptions are fulfilled when interpreting the cost function as an energy. However, even if the cost function has few properties in common with a genuine energy, the method will still work as a recipe to get away from high-valued local minima of the cost function. In this sense the $\Delta\beta$ and N_s are parameters depending on the problem at hand.

The transition of the configuration $p \rightarrow n$ is much more accessible to criticism. In our example a new configuration was obtained by just changing one $F(\mathbf{k}_l)$. This is rather arbitrary. Why not something else? There is a controversy in the literature about the best replacement of one configuration by another one, in particular if the possible values are not discrete. The downhill simplex method is recommended by some authors [5]. We think that the choice depends on the system, and in our example a primitive procedure is successful. Really advantageous is the fact that no deep knowledge about the true solution is required. This feature remains true for more complicated systems.

Although the simulated annealing method is not caught in local minima with a depth of about $\Delta E < 1/\beta$ there is no guarantee that the system ends in the global minimum. There are different variants to circumvent this problem.

- i) The simplest variant consists in doing nothing - a very good procedure if the system has
 - apart from the global minimum - many other minima, which are representative as well

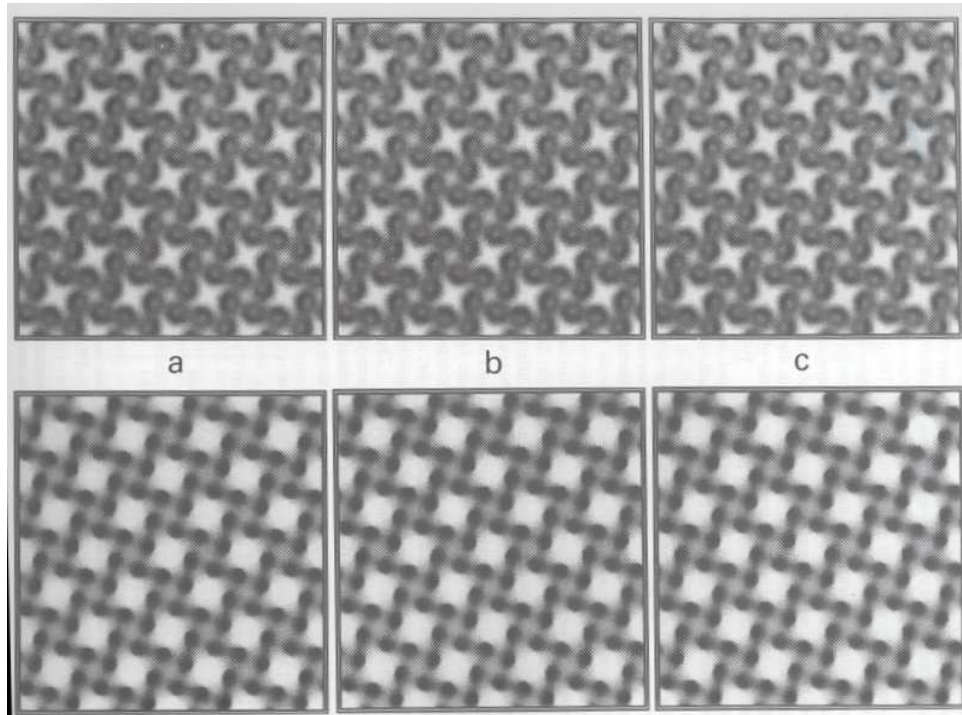


Fig. 6: Exit plane wave function of Ni_4Mo [4]. Each image comprises 4×4 unit cells. The images in the top row show the amplitude, those in the bottom row the phase of the wave function. a) Generating wave function which was used to simulate the model images of Fig. 5. b) Wave function reconstructed from the model images by means of the simulated annealing algorithm. c) Wave function reconstructed by means of the genetic algorithm.

and whose cost function is not much higher. This is the case for the problem of designing complex integrated circuits, for which this method was first applied [26], the spin glass problem [8] and the problem of the traveling salesman [5].

- ii) In the second variant the final configuration is taken as the initial configuration and the procedure is restarted (again with small β). This variant had been successful in as different topics as optimizing time-tables and computing protein configurations¹⁰ [27, 28, 30].
- iii) In the third variant one starts with an ensemble of different configurations instead of just one and selects the best one at the end.

Variant iii) is a precursor of an algorithm that will be explained next.

6 Minimizing a System with the Genetic Algorithm

We explain the method of the genetic algorithm by applying it to the cost function of the electron microscope [4] as well:

¹⁰Even more generalized procedures are discussed in [29].

- 1.: We generate an ensemble which is called *population* in this context. The configuration $\{F(\mathbf{k}_i)\}$ of all its N_{qA} members $m(i)$ is arranged as a vector, e.g. for $m(i_1)$ and $m(i_2)$:

$$\begin{aligned} m(i_1) &= (F^{(i_1)}(\mathbf{k}_1), F^{(i_1)}(\mathbf{k}_2), \dots, F^{(i_1)}(\mathbf{k}_{l_c}), \dots, F^{(i_1)}(\mathbf{k}_N)) \\ m(i_2) &= (F^{(i_2)}(\mathbf{k}_1), F^{(i_2)}(\mathbf{k}_2), \dots, F^{(i_2)}(\mathbf{k}_{l_c}), \dots, F^{(i_2)}(\mathbf{k}_N)) \end{aligned} \quad (20)$$

In the folklore of the genetic algorithm [31] the vectors are called *chromosomes* and their components are called *genes*.

- 2.: Each member of the population is rated using the cost function Eq. (17).
- 3.: *mutation of genes*: One randomly selected member is copied and one randomly selected gen (component) of the copy is changed randomly. The modified copy is rated too.
- 4.: *crossover*: The best rated member, say $m(i_1)$ and another member, say $m(i_2)$, chosen randomly, are selected. After determining an integer $1 \leq l_c < N$ randomly two *offsprings* are generated choosing for the first the genes of $m(i_1)$ up to l_c and from l_c on the genes of $m(i_2)$. It is done vice versa for the second offspring. The offsprings are rated.
- 5.: Three members are removed from the ensemble except for the latest offsprings. These members are either those with the lowest ratings or determined by a random number generator from those with bad ratings.
- 6.: Step 3 to 5 are repeated till the cost function is low enough.

The result can be observed in Fig. 6. The genetic algorithm method has been able to reconstruct the coefficients of the electronic wave function nearly completely, cf. Fig. 6c with Fig. 6a.

7 Criticism, Variants and Power of the Genetic Algorithm

This method claims to reflect some important ingredients of evolution theory. And since this led to such an admirable outcome as homo sapiens [32] the genetic algorithm with its analogies to evolution can - that is the hope - be applied to optimization problems with analogous admirable results.

Taking a closer look on the method one can state the following: Without the crossover (step 4) the method is nothing but a variant of simulated annealing: One starts with an ensemble and tries new configurations (mutations) as in the simulated annealing method. Instead of applying the Metropolis criterion the new configurations are rated within the samples and need not be discarded as long as they are within the energy spread ΔE of the ensemble. In this way the ensemble can pass local minima. When approaching the minimum of the cost function ΔE of the ensemble shrinks. This corresponds to an increase of β in the simulated annealing formalism. Therefore it is no surprise that the reconstruction of the wave function has the same quality as that obtained from simulated annealing, cf. Fig. 6b and Fig. 6c.

What is really different is the crossover. Its advantage is i) getting to radically different states of the system in so far as these can be obtained by mutations with an exponentially small probability only, ii) the possibility to preserve cluster characteristics of the parent members. Its disadvantage is i) the recipe for generating these crossovers is rather arbitrary. In a higher-dimensional system the genuine clusters are not one dimensional and therefore the chromosomes should be higher dimensional as well. In this way the method would get properties of

cluster methods. ii) The recipes are in general not simple. This can be seen already in the traveling salesman problem. A simple recipe for the crossover as was presented here leads at once to routes touching some cities twice, others not at all.

In conclusion one can state that the genetic algorithm is a variant of the simulated annealing method plus allowing further configurations that keep some properties of the so far best solutions and at the same time are not easily obtained by repeatedly changing configurations in the simulated annealing method.

Really advantageous is the fact that - leaving aside the crossover procedure - no idea about the true solution is required. The efficiency of crossovers can be enhanced if the recipe for the crossover is adapted to properties of the system.

8 Ground States of Molecules as an NP Complete Problem, Simulated Annealing and Molecular Dynamics

There can be no doubt that the ground state energy problem of molecules is an NP complete problem [6]: If for simplicity the interaction between atoms is approximated by a *Lennard-Jones potential*

$$V_{LJ}(r) = \frac{1}{r^{12}} - \frac{2}{r^6}, \quad (21)$$

then it can be shown [33, 34] that the number of local minima, N_{min} , of a system with N_a atoms grows exponentially fast and is roughly estimated by

$$N_{min} \approx \exp(0.36N_a + 0.03N_a^2) \text{ (this is an extrapolation)}. \quad (22)$$

That means

$$\begin{aligned} N_a = 13, & \quad : \quad N_{min} \approx 1000, \\ N_a = 100, & \quad : \quad N_{min} \approx 10^{140}. \end{aligned}$$

The actual complicated form of the ground state energy, computed in the density functional formalism, is shown in Fig. 7 for the O_3 molecule.

Because of the many local minima steepest descent and related methods will fail, and simulated annealing and genetic algorithm methods have their day. However, there is a difficulty mentioned in the introduction: The positions of the ions change continuously, suggesting a variant of simulated annealing. In this situation a promising variant of simulated annealing has been suggested which applies a kind of *molecular dynamics* [23]. The idea is the following: The ground state energy is a function of the ionic positions \mathbf{R}_i ,

$$E_g = E_g(\mathbf{R}_1 \dots \mathbf{R}_{N_a}). \quad (23)$$

This is the cost function to be used now as a potential energy of the ions. Added to it is an (artificial) kinetic energy term

$$K = \sum_{i=1}^{N_a} \frac{1}{2m} \mathbf{P}_i^2, \quad (24)$$

where m is a fictitious mass and \mathbf{P}_i the (artificial) momentum of ion i . Thus we obtain the fictitious Hamiltonian

$$H = \sum_{i=1}^{N_a} \frac{1}{2m} \mathbf{P}_i^2 + E_g(\mathbf{R}_1, \dots, \mathbf{R}_{N_a}). \quad (25)$$

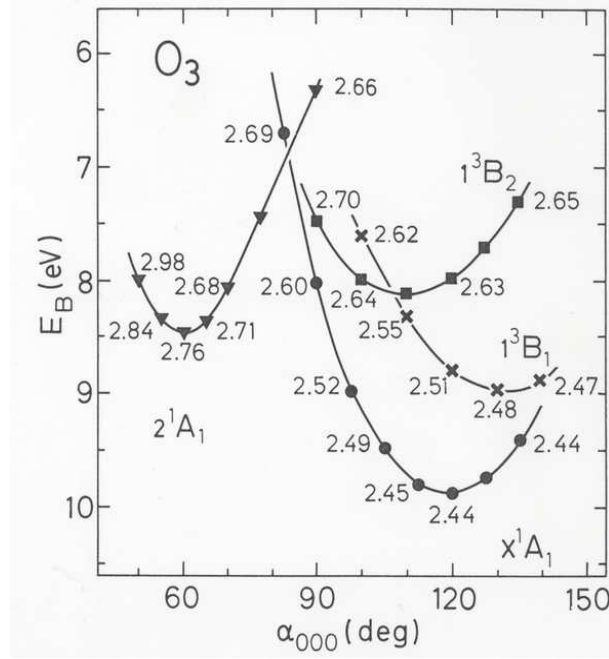


Fig. 7: Energy surfaces of states of the ozone molecule, O_3 [22]

Now we can follow fictitious trajectories of this artificial system. We have for the averaged kinetic energy, $\langle K \rangle$, the relation:

$$\langle K \rangle \propto \frac{1}{\beta}. \quad (26)$$

This gives a connection between inverse temperature β and mean kinetic energy. Furthermore the artificial hamiltonian is a constant of motion and thus E_g is *not*. Instead E_g fluctuates by an amount ΔE_g

$$\langle \Delta E_g \rangle \propto \langle K \rangle \propto \frac{1}{\beta} \quad (27)$$

along a trajectory in the phase space of the ions. Reducing the kinetic energy (e.g. by adding a dissipative term) leads to a decrease of ΔE_g and corresponds to reducing the temperature. Beginning with a promising configuration of the ions and setting their velocities to values in accordance with high temperature one can hope to detect the global minimum of the cost function E_g . Note the important point here: Only the ground state energy appears in this procedure. Therefore *density functional theory* is applicable in the standard way to obtain a good approximation for E_g . Taking excited states into account would be much harder.

This scheme connects *molecular dynamics* [23] with simulated annealing.

In spite of the similarity of this scheme with a real life annealing technique in which molecules are heated and then slowly cooled down the simulated annealing procedure differs in two ways, i) for numerical reasons the decrease of the temperature $T = \frac{1}{\beta}$ is 5 – 6 orders of magnitude faster than in reality [35], ii) whereas in an experimental setup not only the ground state but the excited states are important, in the calculation here only the ground state in its various ionic configurations is taken into account. Therefore even in this situation the annealing is ‘simulated’. Let us discuss two examples that show the power and the limitations of this variant of simulated annealing.

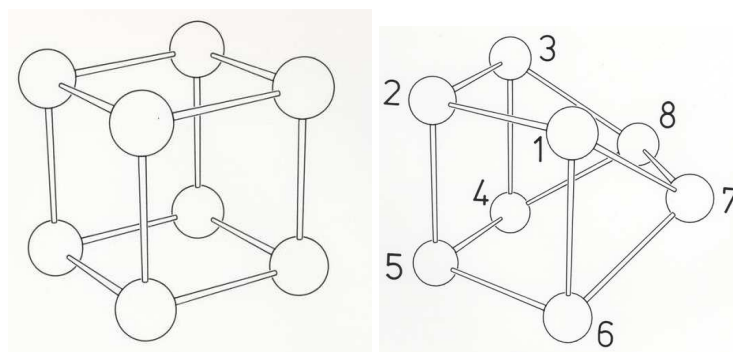


Fig. 8: *Left: The cubic structure of the P_8 molecule was believed to be the ground state configuration although its energy is higher than that of two P_4 molecules. Right: The true ground state obtained by a simulated annealing variant described in the text. The cubic structure corresponds to a local minimum.*

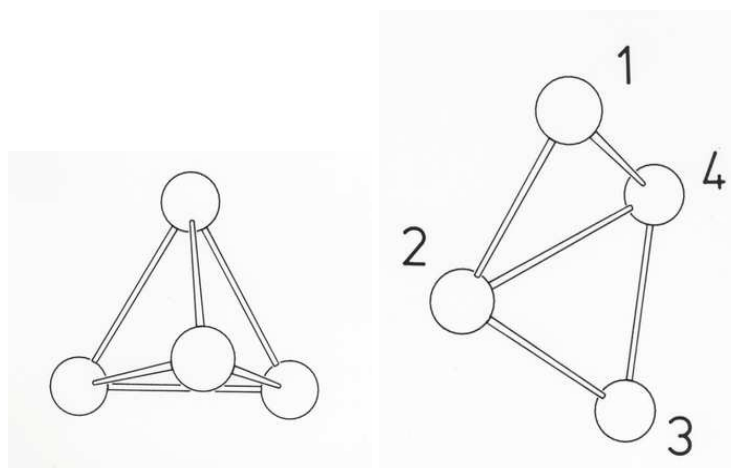


Fig. 9: *Left: Ground state structure of the P_4 molecule. Right: Cooling down using the simulated annealing variant, described in the text, usually leads to a roof structure with higher energy. The reason is the small basin of attraction of the ground state.*

- For a long time the cubic structure of the phosphorous P_8 was assumed as the ground state configuration because of its symmetry, cf. Fig. 8, in spite of the fact that the energy was higher than twice the energy of the P_4 ground state energy. Under these circumstance P_8 should have been metastable. Applying simulated annealing with an initial cubic configuration the true ground state - having *lower symmetry* - emerged with an energy less than that of two P_4 molecules [22].
- Phosphorous P_4 has a *symmetric* tetrahedral ground state. Starting at high temperature with an ionic configuration not extremely close to the ground state configuration the simulated annealing method leads to a configuration having higher energy, cf. Fig. 9. This example should caution the belief that simulated annealing will work automatically under *any circumstances* [22].

9 Conclusion

In probably any discipline of science problems occur involving a cost function E . Typically the cost function arises if

- ground state energies or best configurations or minimum efforts have to be determined,
- the fitting of a non trivial function has to be achieved,
- an ill posed problem has to be transformed into one finding a minimum (a typical procedure in tomography).

The best parameters or variables are the ones minimizing this cost function and one gets an optimization problem of the kind described in this paper.

Typically two different difficulties arise in the optimizations:

- a) the argument space of E is multidimensional
- b) a number - even a huge number - of local minima can be present.

There are optimization problems in which local minima are unimportant, either because their *depth* or the *number of local minima* or both are very small. This can occur in tomographic problems [1], in optimization problems of neural networks [3] and partly in pattern reconstruction [4, 36] - examples that are presented in this paper. In all these cases a steepest descent method or related ones as conjugate gradient, Gauss-Newton or Levenberg-Marquardt methods etc. are appropriate for solving the optimization problem.

The situation changes drastically if in addition to difficulty a) we have a system where difficulty b) is present. This happens in particular in systems which are NP complete [6] (i.e. the number of local minima increases strongly, possibly even exponentially with the system size). Examples of this type are the traveling salesman problem, and the ground state of molecules [9], of spin glasses [26], best protein configurations [28, 30] etc.. In all these systems application of steepest descent and related methods will lead to a complete failure. Quite different methods are required. In this paper the simulated annealing [19, 5] and genetic algorithm [31] methods and some of their variants have been discussed taking as examples the reconstruction problem of the electron microscope [4] and ground state configurations of molecules [22].

In spite of the very often great success of these methods - successful if the appropriate variant has been found - there are cases in which even these methods fail. This occurs if the basin of attraction of the global minimum is very small. An example is the ground state of P_4 . Its tetrahedral structure has very low energy but a very small basin of attraction. It is very difficult in such a situation to detect the global minimum just by applying simulated annealing or genetic algorithm methods [9, 22].

Acknowledgments

I would like to thank M. Lentzen for explaining the electron microscope to me, M. Lentzen, R.O. Jones and K. Steffen for stimulating discussions and M. Lentzen, R.O. Jones, and A. Bringer for a critical reading of the manuscript.

Appendix

A Modified Steepest Descent Methods and Regularization

Quite often the cost function has the special structure

$$E(\mathbf{x}) = \frac{1}{2} \mathbf{f}^+ \mathbf{f} \quad (28)$$

with

$$\mathbf{f}^+ = \frac{1}{\sqrt{N}} (U_1^* - U_1^*(\mathbf{x}), \dots, (U_N^* - U_N^*(\mathbf{x}))). \quad (29)$$

In particular in tomography there is the problem of either fewer equations than unknown parameters x_i or the problem of ill conditioning. In both cases the solution becomes unstable. This can already be seen from a simple example:

$$x_1 + x_2 = 1, \quad \epsilon x_2 = \eta, \quad |\epsilon| + |\eta| \ll 1. \quad (30)$$

If ϵ and η vanish *exactly*, then there is only one equation and two unknown quantities. Then the solution is by no means unique. The much more dangerous situation is given if $\epsilon \neq 0$ and $\eta \neq 0$ - this situation occurs e.g. in the presence of noise. Then there is a unique solution, but x_2 may be anything - and may be huge in spite of the fact that ϵ and η are small. Such an instability of the solution can be removed by adding a quadratic expression to E :

$$E \rightarrow \hat{E} = E + \frac{1}{2} \alpha^2 \mathbf{x}^+ \mathcal{B}^+ \mathcal{B} \mathbf{x}. \quad (31)$$

If α and \mathcal{B} are chosen appropriately the solution becomes stable and unique. The additional term is called *regularization* [17] term, it acts as a filter and depending on the problem the filter has to be chosen such that the interesting solution remains untouched. Setting

$$\hat{\mathbf{f}}^+ = (\mathbf{f}^+, \mathbf{x}^+ \mathcal{B}^+), \quad (32)$$

we get the structure of Eq. (28) back:

$$\hat{E} = \frac{1}{2} \hat{\mathbf{f}}^+ \hat{\mathbf{f}}. \quad (33)$$

Therefore we will assume that the cost function is given by Eq. (28) and regularization - if necessary - has already been built in.

The global minimum of E in Eq. (28) has the property

$$|\mathbf{f}| = 0 \text{ or } |\mathbf{f}| \ll 1. \quad (34)$$

Then we obtain close to the global minimum¹¹

$$E(\mathbf{x}_0 + d\mathbf{x}) = (\mathbf{f}^+(\mathbf{x}_0) + d\mathbf{x}^+ \mathcal{A}^+)(\mathbf{f}(\mathbf{x}_0) + \mathcal{A}d\mathbf{x}) + \mathcal{O}(|d\mathbf{x}|^3), \quad (35)$$

¹¹Throughout this appendix we assume that \mathbf{x} is a real quantity. Otherwise the always real function E could not be differentiated with respect to \mathbf{x} . However, if \mathbf{x} is complex, E can be written as $E(\mathbf{x}_1, \mathbf{x}_2)$ with $\mathbf{x}_1 = \mathbf{x}$ and $\mathbf{x}_2 = \mathbf{x}^*$. Applying this trick E becomes differentiable with respect to \mathbf{x}_1 and \mathbf{x}_2 . We can then expand and set $d\mathbf{x}_2 = d\mathbf{x}_1^*$ at the end. Thus the modifications for complex \mathbf{x} are trivial and will not be discussed here.

with

$$\mathcal{A} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}.$$

This result suggests a minimization of $|\mathbf{f}(\mathbf{x}_0) + \mathcal{A}d\mathbf{x}|^2$ leading to the equation

$$\mathcal{A}^+ \mathcal{A} d\mathbf{x} = -\mathcal{A}^+ \mathbf{f}(\mathbf{x}_0). \quad (36)$$

We know that $\mathcal{A}^+ \mathcal{A}$ is positive definite¹². Therefore this equation has always a solution. And it can be obtained by applying the conjugate gradient method because this converges with certainty for positive definite matrices. In the *Gauss-Newton method* $\mathbf{x}_0 + d\mathbf{x}$ is taken as the next step of the iteration. This is not too risky since $\mathcal{A}^+ \mathcal{A}$ is the *second derivative of E* at $E = 0$. In the *Levenberg-Marquardt method* Eq. (36) is replaced by

$$(\mathcal{A}^+ \mathcal{A} + \mu \mathcal{I}) d\mathbf{x} = -\mathcal{A}^+ \mathbf{f}(\mathbf{x}_0), \quad \mu > 0. \quad (37)$$

The justification for this is that the right hand side of the equation is just $-\nabla E$ and therefore this procedure mixes steepest descent with Gauss-Newton. There are various recipes how μ , and, in the next iteration step, $\mathbf{x}_0 \rightarrow \mathbf{x}_0 + \xi d\mathbf{x}$, ξ should be chosen. But this is beyond the scope of this paper and the interested reader is referred to the literature.

¹²Any matrix $\mathcal{A}^+ \mathcal{A}$ has no negative eigenvalues, because of regularization this matrix is positive definite.

References

- [1] A. Kemna, *Tomographic Inversion of Complex Resistivity* (Der Andere Verlag, Osnabrück, 2000)
- [2] A. Kemna, J. Vanderborght, B. Kulessa, H. Vereecken, J. Hydrol **267**, 125 (2002)
- [3] S. Haykin, *Neural Networks*, MacMillan College Publishing Company, New York, 1994
- [4] A. Thust, M. Lentzen, K. Urban, Ultramicroscopy **53**, 101 (1994)
- [5] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P Flannery, *Numerical Recipes in C++*, Cambridge University Press, Cambridge, 2002
- [6] M.R. Garey and D.S. Johnson, *Computers and Intractability*, W.H. Freeman and company, San Francisco, 1979
- [7] D.P. Landau, K. Binder, *Monte Carlo Simulations in Statistical Physics*, Cambridge University Press, Cambridge, 2000
- [8] K.H. Fischer, J.A. Hertz, *Spin Glasses*, Cambridge University Press, Cambridge, 1991
- [9] R.O Jones and D. Hohl, J.Chem.Phys. **92**, 6710 (1990)
- [10] R.O. Jones and G. Seifert, J.Chem.Phys. **96**, 7564 (1992)
- [11] T. Eitrich, *Support-Vektor-Maschinen und ihre Anwendung auf Datensätze aus der Forschung*, Berichte des Forschungszentrums Jülich, Jül 4096, 2003
- [12] V. Kecman, *Learning and Soft Computing*, The MIT Press, Cambridge, 2001
- [13] S. Abe, *Support Vector Machines for Pattern Classification*, Springer Verlag, London, 2005
- [14] N. Chen, W. Lu, J. Yang, G. Li, *Support Vector Machine in Chemistry*, World Scientific, Singapore, 2004
- [15] The Gauss-Newton method is standard and can be found in the internet.
- [16] The Levenberg-Marquardt method is standard and can be found in the internet.
- [17] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill Posed Problems*, Halsted Press, Washington, 1977
- [18] K. Binder and D.W. Heermann, *Monte Carlo Simulations in Statistical Physics*, Springer Verlag, Berlin, 1997
- [19] P.J. Vanlaarhoven and E.H. Aarts, *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht, 1987
- [20] L. Davis, *Genetic Algorithms and Simulated Annealing*, Pitman Publishing, London, 1987
- [21] K. Selber, F. Nellen, B. Steffen, J. Thömmes, M.-R. Kula, J. Chromatogr. **B 743**, 21 (2000)

- [22] R.O. Jones, *Density Functional, Molecular Dynamics and More*, in: *Computational Approaches to Novel Condensed Matter Systems*, edited by D. Neilson and M.P. Das, Plenum Press, New York, 1995
- [23] D.C. Rapaport, *The Art of Molecular Dynamics Simulation*, Cambridge Univ. Press, Cambridge, 2005
- [24] H. Ritter, T. Martinetz, K. Schulten, *Neuronale Netze: eine Einführung in die Neuroinformatik selbstorganisierender Netzwerke*, Addison Wesley, Bonn, 1991
- [25] Exercise in one of my lectures, C++ code written by M. Zilly, 2003
- [26] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, *Science*, **220**, 671 (1983)
- [27] H. Müller-Krumbhaar, *Europhys.Lett.* **7**, 479 (1988)
- [28] A. Verma, A. Schug, K. H. Lee and W. Wenzel, *J. Chem. Phys.*, (in press, 2005)
- [29] A. Schug, T. Herges, A. Verma, K.H. Lee and W. Wenzel, *Chem. Phys. Chem.*, (in press, 2005)
- [30] W. Wenzel, submitted to *Phys. Rev. Lett.*, 2005
- [31] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer Verlag, Berlin, 1996
- [32] C. Darwin, *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*, London, 1859
- [33] M.R. Hoare, *Adv. Chem. Phys.* **87**, 6166 (1987)
- [34] D.M. Deaven, N. Tit, J.R. Morris, K.M. Ho, *Chem.Phys.Lett.* **256**, 195 (1996)
- [35] R.O. Jones, private communication
- [36] A. Thust, *Grundlagen und fortgeschrittene Methoden der hochauflösenden Elektronenmikroskopie*, Band 1 des 29. Ferienkurs des IFF, Forschungszentrum Jülich, ISBN 3-89336-217-7, 1998