Title

A numerical optimisation method for obtaining technically smooth magnetisation characteristics

Names of authors

Hans Vande Sande [(1)], Uwe Pahner [(1)], Herbert De Gersem [(1)], Kay Hameyer [(1)] and Ludo Froyen [(2)]

Corresponding author

Hans Vande Sande

Affiliations

[(1)] K.U.Leuven, Dept. ESAT, Div. ELECTA, Kasteelpark Arenberg 10, B-3001 Heverlee-Leuven, Belgium

[(2)] K.U.Leuven, Dept. MTM, Kasteelpark Arenberg 44, B-3001 Heverlee-Leuven, Belgium

Complete address of corresponding author

Hans Vande Sande

KULeuven

Dept. ESAT – Div. ELECTA

Kasteelpark Arenberg 10

B-3001 Heverlee-Leuven

Belgium

e-mail: hans.vandesande@esat.kuleuven.ac.be

Tel.: +32 16 32 10 20

Fax.: +32 16 32 19 85

Abstract

In general, if measurements can be repeated several times assuming the same conditions, the measurement error can significantly be decreased by statistically evaluating the measurements. However, an uncertainty band always remains. Non-linear numerical simulations based on e.g. the Newton-Raphson method may establish a poor convergence if they are provided directly with measured data. Therefore, data pre-processing is required. Here, a neural network approach is employed. A two-layer perceptron is fitted on a measured magnetisation curve, thereby restricting the solution to be technically feasible while accepting the statistical nature of the data. By using a perceptron, an analytical expression of the magnetisation curve is obtained and expressions for its derivatives can easily be computed.

1. Introduction

The standard Newton iteration scheme to solve a non-linear system of equations obtained from the finite element method is based on the updating of the field dependent element reluctivity and its derivative. Usually, the manufacturer of the ferromagnetic material provides a *BH*-characteristic as diagram or in the form of a table of data samples. The influence of the material properties, in particular their accurate numerical representation, is significant for the rate of convergence during the Newton iterations. Here, a numerical optimization aiming at a technically smooth non-linear characteristic is performed to obtain a higher rate of convergence of the Newton iteration scheme. The neural network approach is adopted for representing the magnetisation curve.

2. Measurement data

Fig. 1 shows the measured magnetisation curve of iron with a carbon content of 0.55 %. The high field behaviour reveals a differential permeability which is almost zero, although in practice it cannot be smaller than the permeability of air. Obviously, this is caused by measurement errors. Here, the relative error ε_{rel} introduced by the measurement equals 0.05, both for the magnetic induction B (in T) and the magnetic field strength H (in A/m). The measured magnetisation curve is used in a particular finite element model, in which it is more appropriate to search for a characteristic in terms of the magnetic reluctivity ν (in Am/Vs) and the square of the magnetic induction B^2 (in T^2) (Fig. 2) [1]. However, by this transformation the relative error on ν and B^2 becomes $2\varepsilon_{rel}$, thus 0.10. These errors are considered below, in order to find a suitable representation of the measured data.

3. Neural network approach

It can be proven that any feed-forward neural network with two layers of adaptive weights is capable of modelling any continuous functional mapping. Perceptrons form a special class of feed-forward neural networks [2]. Here, a biased one-input one-output two-layer perceptron with sigmoidal activation functions and linear output units is chosen to approximate the measured magnetisation curve. The mathematical representation of this perceptron is

$$\begin{aligned}\nu &= \sum_{j=1}^{M} w_j^{(2)} \phi_j\left(a_j\right) + w_0^{(2)} \\ &= \sum_{j=1}^{M} w_j^{(2)} \phi_j\left(w_{j1}^{(1)} B^2 + w_{j0}^{(1)}\right) + w_0^{(2)}\end{aligned} \quad , \tag{1}$$

with $w_{j0}^{(1)}$ the first layer bias weights to neuron j, $w_{j1}^{(1)}$ the first layer weights from input B^2 to neuron j, a_j the activation of neuron j, $\phi_j(a_j)$ the activation function of neuron j, $w_0^{(2)}$ the second layer bias weight to output ν, $w_{j\neq 0}^{(2)}$ the second layer weights from neuron j to output ν and M the number of neurons (Fig. 3). Each circle corresponds to a single neuron, which transforms its input into an output by means of a sigmoidal activation function (Fig. 4):

$$\phi_j\left(a_j\right) = \frac{1}{1+e^{-a_j}} \quad . \tag{2}$$

Once the weights have been determined, their values are fixed. The first and second derivative of the output with respect to the input are then given by:

$$\frac{d\nu}{dB^2} = \sum_{j=1}^{M} w_j^{(2)} \frac{d\phi_j\left(a_j\right)}{da_j} w_{j1}^{(1)} \quad , \tag{3}$$

with

$$\frac{d\phi_j\left(a_j\right)}{da_j} = \phi_j\left(a_j\right)\left(1-\phi_j\left(a_j\right)\right) \quad , \tag{4}$$

and

$$\frac{d^2\nu}{\left(dB^2\right)^2} = \sum_{j=1}^{M} w_j^{(2)} \frac{d^2\phi_j\left(a_j\right)}{da_j^2} \left(w_{j1}^{(1)}\right)^2 \quad , \tag{5}$$

with

$$\frac{d^2\phi_j\left(a_j\right)}{da_j^2} = \phi_j\left(a_j\right)\left(1-\phi_j\left(a_j\right)\right)\left(1-2\phi_j\left(a_j\right)\right) \quad . \tag{6}$$

respectively. The derivatives (3) and (5) are used during the optimisation process.

4. Sum-of-squares error

In order to optimise the weights of the perceptron, a sum-of-squares error E is defined:

$$E = \sum_{n=1}^{N} \left(\nu_p^{(n)} - \nu_m^{(n)} \right)^2 \quad , \tag{7}$$

with $\nu_p^{(n)}$ the reluctivity computed by the perceptron for the n^{th} of N measurements and $\nu_m^{(n)}$ the measured reluctivity. This sum-of-squares error only depends on the weights and is therefore denoted by $E(\mathbf{w})$. The weights are then determined by minimising this sum-of-squares error in an iterative algorithm.

To improve the convergence rate of the optimisation algorithm, the gradient $\nabla E(\mathbf{w})$ must be provided. This gradient can also be obtained analytically. The partial derivative of E with respect to the second layer weights is

$$\frac{\partial E}{\partial w_j^{(2)}} = 2 \sum_{n=1}^{N} \left(\nu_p^{(n)} - \nu_m^{(n)} \right) \phi_j^{(n)} \left(a_j^{(n)} \right) \quad . \tag{8}$$

The partial derivatives of E with respect to the first layer weights are more expensive to compute:

$$\begin{aligned} \frac{\partial E}{\partial w_{j1}^{(1)}} &= 2 \sum_{n=1}^{N} \left(\nu_p^{(n)} - \nu_m^{(n)} \right) B_m^{2\,(n)} \left(\sum_{j=1}^{M} w_j^{(2)} \frac{d\phi_j^{(n)}\left(a_j^{(n)}\right)}{da_j^{(n)}} \right) \\ \frac{\partial E}{\partial w_{j0}^{(1)}} &= 2 \sum_{n=1}^{N} \left(\nu_p^{(n)} - \nu_m^{(n)} \right) \left(\sum_{j=1}^{M} w_j^{(2)} \frac{d\phi_j^{(n)}\left(a_j^{(n)}\right)}{da_j^{(n)}} \right) \end{aligned} \quad . \tag{9}$$

5. Optimisation of the weights

The output of the optimised perceptron must be within the uncertainty band of the measurements. It is observed that this is not always the case when simply minimising the sum-of-squares error. Therefore, the optimisation problem is extended with function constraints on the output of the perceptron. However, the result may still not be suitable for use in a non-linear finite element model, because it may hamper the convergence rate of the Newton-Raphson iteration [1,5]. As a result, to make the magnetisation curve technically feasible, constraints need also to be added on the derivatives of the perceptron output [4].

The optimisation process of the perceptron weights may not converge when directly starting the optimisation considering all required constraints on the perceptron output and its derivatives. Therefore, the optimisation process is performed in three successive steps:

1. An unconstrained minimisation of $E(\mathbf{w})$;
2. A constrained minimisation of $E(\mathbf{w})$, requiring the output of the perceptron to be within the uncertainty band of the measurement;
3. A constrained optimisation of $E(\mathbf{w})$ as in the previous step, but with the additional restriction that the output of the perceptron must be technically feasible.

Before starting the unconstrained minimisation of $E(\mathbf{w})$, all perceptron weights are initialised to zero. The unconstrained minimisation problem is then solved using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) Quasi-Newton Line-Search method [3]. The algorithm is terminated when the sum-of-squares error varies less than 0.5 % during five successive iterations. A sufficiently accurate initial approximation for the first constrained optimisation step is now determined.

The function constraints are then added to the minimisation process. These constraints can be computed from the relative error on the measurements. This error defines an uncertainty band in the ν - B^2 plane around each measurement. If this band is defined by the intervals $\left[B^2_{\min}\ B^2_{\max}\right]$ and $\left[\nu_{\min}\ \nu_{\max}\right]$, the constraints to be supplied are (Fig. 5):

$$\begin{aligned} \nu\left(B^2_{\min}\right) &\le \nu_{\max} \\ \nu\left(B^2_{\max}\right) &\ge \nu_{\min} \end{aligned} \quad . \tag{10}$$

This constrained minimisation problem is solved by Sequential Quadratic Programming, in which the Hessian of the Lagrangian is updated at each iteration using the BFGS formula. The quadratic subproblem at each iteration is solved by the active set method [3]. Like in the previous step, the algorithm is terminated when the sum-of-squares error varies less than 0.5 % during five successive iterations. Some constraints may still be violated and are transferred to the last optimisation step.

Now, the perceptron output is within the uncertainty band of the majority of measurements. However, directly applying this result in a finite element package can reduce the convergence rate of the Newton-Raphson algorithm significantly, due to small oscillations in the first derivative. Therefore, some extra constraints are added, such that the resulting magnetisation curve is technically feasible [1]. For the example studied here, with no data provided in the Rayleigh region of the magnetisation curve, it is at first suggested to impose monoticity to the reluctivity. It is also supposed that the low field behaviour is linear, although this is only true for a limited class of ferromagnetic materials. For that, the first derivative with respect to B^2 is set to zero if B equals zero. Moreover, monoticity is imposed to the first derivative with respect to B^2 over the whole range of the data. The latter is not true for very high field strengths, as the reluctivity cannot exceed the reluctivity of air. However, this extra constraint is justified as these high fields seldomly occur in practical applications and as this modification significantly improves the convergence rate of the Newton-Raphson algorithm. Mathematically, these constraints can be applied by imposing:

$$\left.\frac{d\nu}{dB^2}\right|_{B=0} = 0$$
$$\left.\frac{d^2\nu}{\left(dB^2\right)^2}\right|_{B^2} \geq 0 \quad \forall B^2 \qquad . \tag{11}$$

If the Rayleigh region is known, the first constraint in (11) can be omitted. As it is impossible to impose the inequality constraint over the whole domain, a number of uniformly distributed points is chosen in which this constraint must be satisfied. The same constrained optimisation method is used as in the previous case. However, the algorithm is now terminated once all second derivative constraints and most function constraints are satisfied. The latter, because it may happen that some measurements are outliers.

6. Remarks

The reluctivity varies in magnitude over two decades, which hampers the training of the neural networks. The optimisation algorithm converges slowly. To avoid this, the measured data must be pre-processed. The simplest way to perform this, is normalising the data by subtracting the average value and dividing by the deviation. A more sophisticated method is called 'whitening' and takes correlations between the data into account [2].

Here, the data normalisation is combined with a preceding logarithmic transformation of the reluctivity:

$$\begin{cases} X = B^2 \\ Y = \nu \end{cases} \Rightarrow \begin{cases} X^* = X \\ Y^* = \ln Y \end{cases} \Rightarrow \begin{cases} X^{**} = \dfrac{X^* - \mu\left(X^*\right)}{\sigma\left(X^*\right)} \\ Y^{**} = \dfrac{Y^* - \mu\left(Y^*\right)}{\sigma\left(Y^*\right)} \end{cases} \tag{12}$$

with μ and σ the average and the standard deviation of the quantity between brackets respectively. The network is then trained on the normalised data X^{**} and Y^{**}. Hence, the constraints must also be transformed. This is done by applying the chain rule:

$$\nu = \exp\left(Y^{**}\sigma\left(Y^*\right) + \mu\left(Y^*\right)\right) \tag{13}$$

$$\frac{d\nu}{dB^2} = \frac{\sigma\left(Y^*\right)}{\sigma\left(X^*\right)} \exp\left(Y^{**}\sigma\left(Y^*\right) + \mu\left(Y^*\right)\right) \frac{dY^{**}}{dX^{**}} \tag{14}$$

$$\frac{d^2\nu}{\left(dB^2\right)^2} = \frac{\sigma\left(Y^*\right)}{\sigma\left(X^*\right)^2} \exp\left(Y^{**}\sigma\left(Y^*\right) + \mu\left(Y^*\right)\right).\left[\sigma\left(Y^*\right)\left(\frac{dY^{**}}{dX^{**}}\right)^2 + \frac{d^2Y^{**}}{\left(dX^{**}\right)^2}\right] \tag{15}$$

Due to this transformation, five extra floating point operations have to be performed in order to compute a reluctivity for a given squared flux density. However, this is not a disadvantage, as less neurons are required for accurately modelling the characteristic.

The input data are not uniformly distributed over the interval of interest. As a consequence, the minimisation of the sum-of-squares error concentrates on intervals in which the density of data points is large. This can be avoided by weighting the sum-of-squares error, considering the distribution of the data in a histogram. The modified sum-of-squares error then becomes:

$$E = \frac{\sum_{n=1}^{N} \frac{1}{P^{(n)}} \left(\nu_p^{(n)} - \nu_m^{(n)}\right)^2}{\sum_{n=1}^{N} \frac{1}{P^{(n)}}} \quad , \tag{16}$$

with $P^{(n)}$ the probability of the n^{th} measurement input.

The neural network approach presented here, provides an alternative to the commonly applied cubic splines for modelling magnetisation curves. A comparison of both approaches from a computational point

of view is presented in [6]. Cubic splines should also be smoothed in order to improve the convergence rate of the non-linear iterations. This can e.g. be achieved by the stochastic optimisation method proposed in [1].

7. Network training results

Fig. 6 compares the optimal solution with the measurements and the solution obtained after the unconstrained minimisation, for a perceptron with five neurons. Obviously, the unconstrained minimisation converges to the conditional average of the measured reluctivities [2], whereas the constrained minimisation also must satisfy the constraints. This is illustrated for higher values of B^2.

The derivative and the second derivative of the output with respect to the input are normalised and plotted in Fig. 7. These characteristics prove that the first derivative and the second derivative are always positive, as was imposed by the constraints. The smoothness of both curves may be improved by adding more constraints to the problem. This slows down the optimisation procedure, but this is a minor problem, since it only has to be done once.

The error evolution of the (normalised) network is plotted in Fig. 8. The constraints increase the minimum error slightly. When switching to the second constrained minimisation step, the error almost remains the same and the weights are only slightly altered.

8. Influence of network size

The sum-of-squares error obtained after the unconstrained optimisation phase can be decreased by increasing the number of neurons, because more degrees of freedom are involved. However, the better a perceptron fits to the exact data, the more curvature it features. One of the easiest ways to avoid those overfitted solutions is the permanent monitoring of the sum-of-squares error during training on a completely independent data set, called the validation set [2]. The training process is stopped when the error on the validation set starts to increase. For perceptrons with less than six neurons, it is observed that other stopping criteria determine the end of the unconstrained optimisation phase.

The influence of the number of neurons on the sum-of-squares error after the constrained optimisation phases is plotted in Fig. 9. Overfitted solutions cannot be obtained here, because curvature constraints are

applied. For a fixed network size, four networks have been separately optimised, each network having a different random initial set of weights. A trendline is drawn through the averaged value. It is obvious that the error decreases for an increasing number of neurons. Further increasing the network size may lower the error slightly. A compromise must be made between this error and the time to compute a network output, as eventually the network must be implemented in a numerical simulation package.

9. Computation of an electromagnet

As an example, the computation of an electromagnet using a perceptron approximation for the *BH*-characteristic of the armature is given. The armature is driven into saturation by applying a sufficiently high current to the coil. The static non-linear problem is solved by the finite element method.

Fig. 10a shows the expected solution, for the fully optimised five-neuron perceptron of Fig. 6. In Fig. 11, the convergence behaviour of the Newton method is plotted for the optimised perceptron and for the partially optimised perceptron obtained by replacing the second derivative constraint in (11) by a first derivative constraint which must be positive. Obviously, the latter converges slower, because the smoothness of the resulting characteristic is smaller in that case. If no constraints are applied, it is possible that the Newton method gets trapped in a local minimum, yielding erroneous solutions.

11. Conclusions

A method is described which optimises the weights of a perceptron in order to obtain a technically feasible magnetisation curve which approximates the measured data. It is shown that a perceptron with five neurons is sufficient to reach this aim. An advantage of using a perceptron is the fact that all its derivatives can be calculated analytically. The data are normalised by a logarithmic transformation of the reluctivity. Although this implies a few extra floating point operations, it yields a decrease of the required number of neurons and it improves the convergence rate of the optimisation algorithm. When the number of neurons is small, it is not necessary to monitor the model on an independent validation set. The sum-of-squares error of the optimised perceptron decreases for an increasing network size. A compromise has to be made

between the network size and the time to compute the network output. An example is given to illustrate the feasibility of the neural network approximation.

Acknowledgements

The authors are grateful to the Belgian “Fonds voor Wetenschappelijk Onderzoek Vlaanderen” for its financial support of this work and the Belgian Ministry of Scientific Research for granting the IUAP No. P5/34 on Coupled Problems in Electromagnetic Systems. The research Council of the K.U.Leuven supports the basic numerical research.

References

[1] U. Pahner, R. Belmans, K. Hameyer, COMPEL **18/4** (1999), pp. 629-637.

[2] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, U.K., 1995.

[3] J. Nocedal, S.J. Wright, Numerical Optimization, Springer-Verlag, New York, U.S.A, 1999.

[4] R. Rosenberger, K.D. Dettmann, K. Heuck, R. Kegel, etzArchiv, VDE-Verlag, Germany **6/3** (1984), pp. 115-120.

[5] L. Jänicke, A. Kost, IEEE Trans. on Magnetics **34/5** (1998), pp. 2505-2508.

[6] H. Vande Sande, K. Hameyer, IEE Proc.-Science, Measurement and Technology, **149** (Sept. 2002), accepted.

Figures

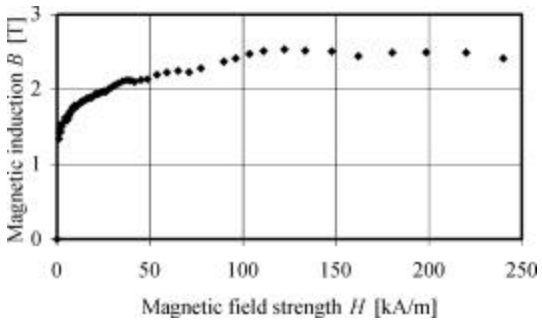

Fig.1:

Measured magnetisation curve of a ferromagnetic material with a carbon content of 0.55 %, in terms of the flux density *B* and the field strength *H.*

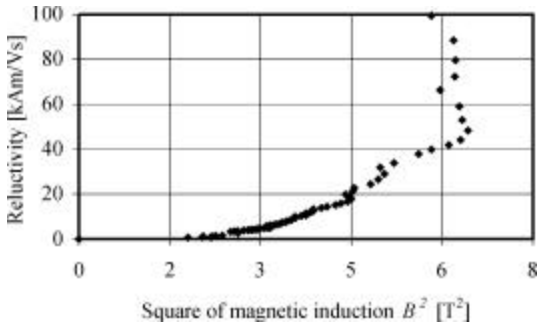

Fig. 2:

Measured magnetisation curve of the ferromagnetic material in Fig. 1, in terms of the reluctivity ν and the square of the flux density B^2.

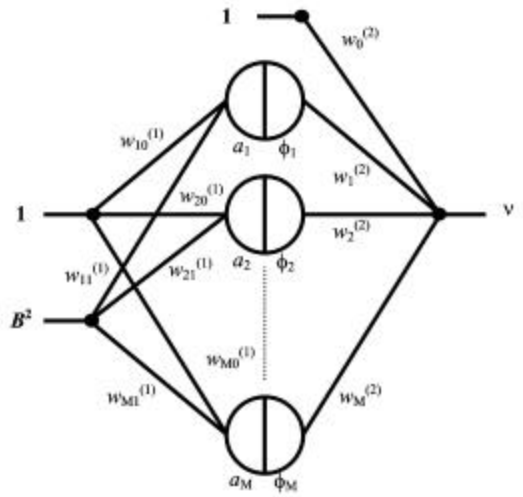

Fig. 3:

Schematic view of a biased one-input one-output two-layer perceptron.

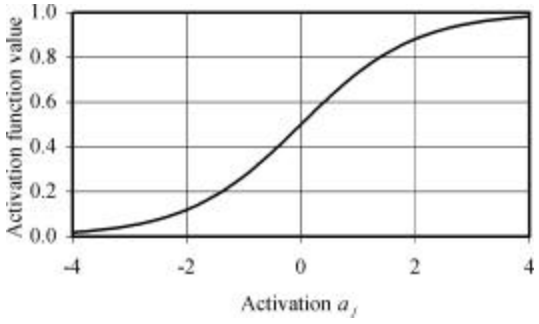

Fig. 4:

The sigmoidal activation function.

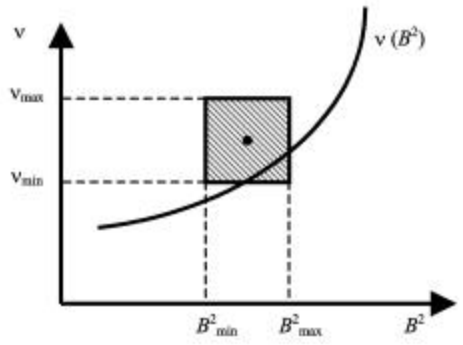

Fig. 5:

Interpretation of the function constraints.

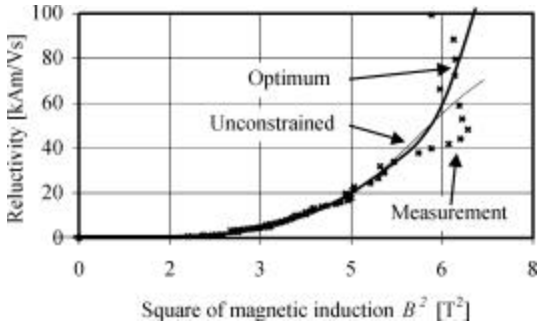

Fig. 6:

Comparison of the optimal solution with the measurements and the unconstrained solution.

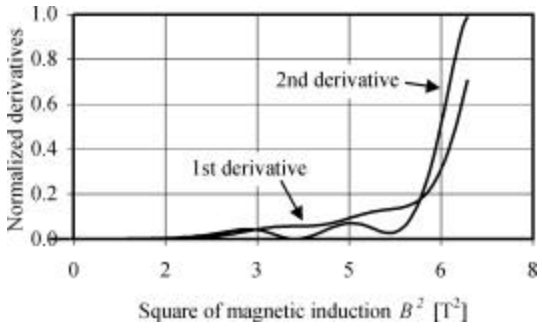

Fig. 7:

Normalized value of the first and second derivative of the solution.

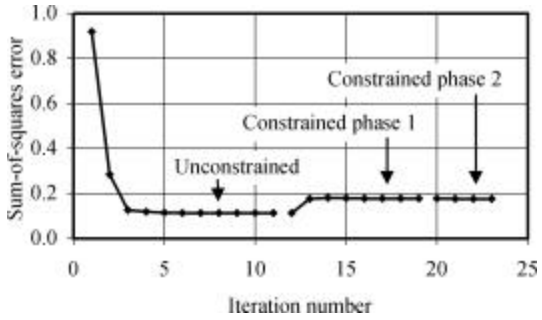

Fig. 8:

Evolution of the sum-of-squares error.

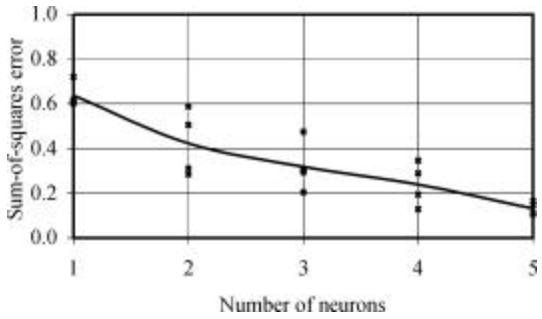

Fig. 9:

Influence of the number of neurons on the sum-of-squares error, for some networks with random initial guesses for the weights.

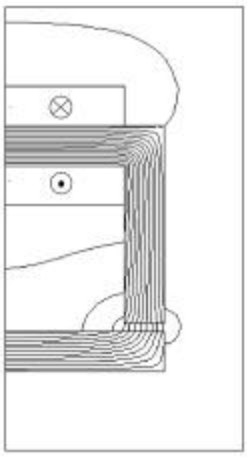

Fig. 10:

Magnetic field distribution in an electro-magnet, obtained by using a five-neuron (a) optimised and (b) non-optimised perceptron for the *BH*-characteristic of the armature.

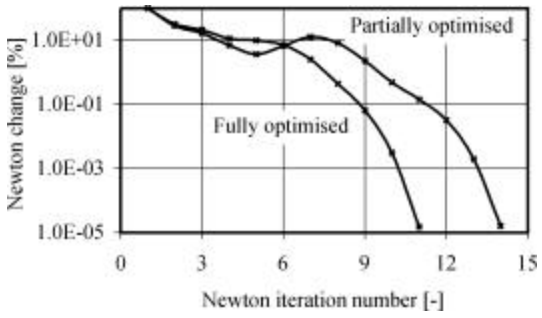

Fig. 11:

Convergence of the Newton-method, obtained by using a five-neuron partially-optimised fully-optimised perceptron for the *BH*-characteristic of the armature.