SOLUTION OF NONLINEAR SISO SYSTEM THROUGH DIFFERENT PARADIGMS

Rubem Vasconcelos Pacelli[1]
[1]Federal University of Ceará, Forlateza – Ceará, Brazilian
e-mail: rubem070@alu.ufc.br

***Abstract* – This paper shows different methods for solving a nonlinear regression problem. The same dataset with 250 observations of a nonlinear SISO (single input, single output) system is used for the following methods: least-squares regressor by parts, kth order polynomial regressor, Mamdani fuzzy system, and 0-order Takagi-Sugeno fuzzy system. All models are analyzed via R^2 for different hyperparameters (polynomial order, number of intervals, etc...). The best solutions' residues and scatter plots are analyzed and discussed.**

***Keywords* – Nonlinear regression problem, Mamdani fuzzy system, Takagi-Sugeno fuzzy system, polynomial regressor, regression by parts.**

I. Least-Squares by parts

The Lest-Squares (LS) method is an approximation technique for overdetermined systems that aims to minimize the squared value of the residuals. Such systems are characterized by having more equations than variables and are easily found in practice [1].

Consider the example of a discrete-time SISO system, where $x_n, y_n \in \mathbb{R}$ are, respectively, their input and output values at the instant $n \in \{1, 2, \ldots, N\}$. At each instant, one has an equation where the input and output are related through a set of $K+1$ parameters, $\boldsymbol{\theta} \in \mathbb{R}^{K+1}$. Mathematically, one can define the output variable as

$$y_n = f(x_n; \boldsymbol{\theta}) \tag{1}$$

When $f(\cdot)$ is a linear function, the LS problem is commonly called Ordinary Least-Square (OLS). The least-squares method aims to minimize the following cost function

$$J\left(\hat{\boldsymbol{\theta}}\right) = \sum_{n=1}^{N} e_n^2, \tag{2}$$

where $\hat{y}_n$ and $\hat{\boldsymbol{\theta}}$ are the estimates of y_n and $\boldsymbol{\theta}$, respectively, and $e_n = y_n - \hat{y}_n$ is the residual signal. Although the OLS method has no optimality associated with it, various practical problems, such as regression analysis, can be solved via OLS since no probabilistic assumptions need to be made about the data [1].

The linear regressor output of the SISO system can be expressed as

$$\hat{y}_n = f(x_n; \hat{\boldsymbol{\theta}}) = \hat{a}x_n + \hat{b}, \tag{3}$$

where $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}^\mathsf{T}$. By using the Equation (3), we can rewrite te cost function as

$$J\left(\hat{\boldsymbol{\theta}}\right) = \sum_{n=1}^{N} \left(y_n - \hat{a}x - \hat{b}\right)^2. \tag{4}$$

The Equation (4) describes a convex function whose surface is a hyperparaboloid. The minimum value of the cost function corresponds to the set of coefficients sought [2]. By calculating the derivative of $J\left(\hat{\boldsymbol{\theta}}\right)$ with respect to $\hat{a}$ and $\hat{b}$, we get

$$\frac{\partial J\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{a}} = -2\sum_{n=1}^{N} x_n \left(\hat{y}_n - \hat{a}x_n - \hat{b}\right) = 0 \tag{5}$$

and

$$\frac{\partial J\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{b}} = -2\sum_{n=1}^{N} \left(\hat{y}_n - \hat{a}x_n - \hat{b}\right) = 0, \tag{6}$$

respectively. The solution of this system of equations is given by

$$\hat{a} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2} \tag{7}$$

and

$$\hat{b} = \hat{\mu}_y - \hat{a}\hat{\mu}_x, \tag{8}$$

where $\hat{\mu}_\cdot$ and $\hat{\sigma}_\cdot^2$ are the sample mean and sample variance with respect to its subscript, respectively, and $\hat{\sigma}_{xy}$ is the estimate of the covariance of x_n and y_n.

Although the solution of the SISO OLS method is rather straightforward, many input-output relationships found in practice have nonlinearities. In these cases, one can resort to applying a transformation to the data in order to linearize the problem. Another approach is to utilize the OLS method in intervals where the scatter plot behaves approximately linear, yielding a set of linear curves with their respective parameters for each path. It is also possible to exploit other nonlinear regression methods, such as polynomial regression. The scatter plot of the dataset, shown in Figure 1, suggests that the input-output relationship is severely nonlinear. Nonetheless, there are intervals where the function can be approximated to a linear curve. A natural hyperparameter arises in this approach: the number of intervals considered. The main trade-off is that the more intervals considered, the better the performance of the coefficient of determination tends to be. However, more parameters are needed to characterize the curve. The best solution is to solve the nonlinear problem with as few parameters as possible.

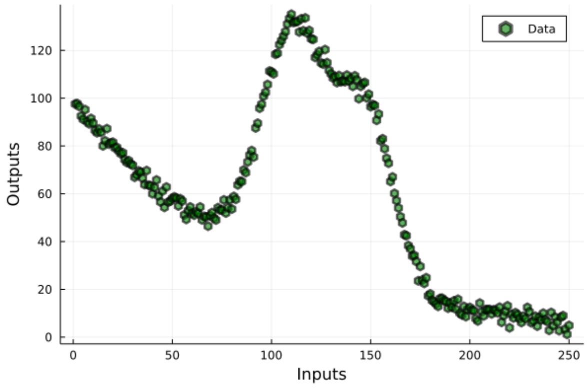


Fig. 1. Scatter plot of the dataset.

In this paper, the OLS algorithm by parts is implemented for different sets of curve intervals, $\{R_i\}_{i=1}^{I}$, where $I \in \{4,5,6\}$ and R_i is the ith interval. Since it is obtained a coefficient of determination, R^2, for each interval, the mean, μ_{R^2}, and the variance, $\sigma^2_{R^2}$, is analyzed for each configuration. The Figure 2 shows where the delimiters have been placed, in addition to the linear curves obtained by the OLS algorithm. The Algorithm 1 summarizes the behavior of the OLS algorithm by parts, and the Table I shows the performance for each value of I.

Algorithm 1: OLS algorithm by parts

Input: $\{(x_n, y_n)\}_{n=1}^{N}$

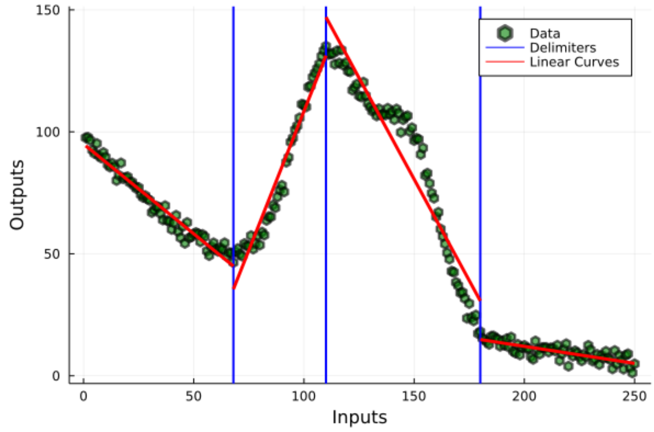
for $I \in \{4,5,6\}$ **do**
  for $i \in \{1,2,...I\}$ **do**
    for $(x_n, y_n) \in R_i$ **do**
      $N_i \leftarrow$ number of samples that belongs to R_i
      $\hat{\mu}_x \leftarrow \frac{1}{N_i}\sum x_n$
      $\hat{\mu}_y \leftarrow \frac{1}{N_i}\sum y_n$
      $\hat{\sigma}_{xy} \leftarrow \frac{1}{N_i}\sum x_n y_n - \hat{\mu}_x\hat{\mu}_y$
      $\hat{\sigma}^2_x \leftarrow \frac{1}{N_i}\sum x^2_n - \hat{\mu}^2_x$
      $\hat{a} \leftarrow \frac{\hat{\sigma}_{xy}}{\hat{\sigma}^2_x}$
      $\hat{b} \leftarrow \hat{\mu}_y - \hat{a}\hat{\mu}_x$

TABLE I
OLS by parts performance - R^2

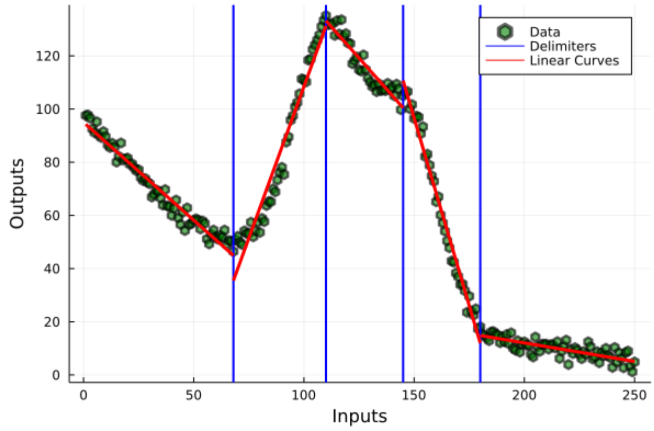
I	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	μ_{R^2}	$\sigma^2_{R^2}$
4	0.962	0.952	0.911	0.627	NaN	NaN	0.863	0.018
5	**0.962**	**0.952**	**0.895**	**0.985**	**0.627**	**NaN**	**0.884**	**0.017**
6	0.962	0.952	0.876	0.314	0.985	0.627	0.786	0.059

The best solution is found for $I = 5$, highlighted in the Table I. Although it is expected to achieve better performance as it is increased the number of intervals, it is possible to notice by the Figure 2c that the placement of the delimiters does not contribute for a good linear approximation (there are outliers for $i = 4$). This fact decreases R^2 and makes it worse when compared with the case where $I = 5$.

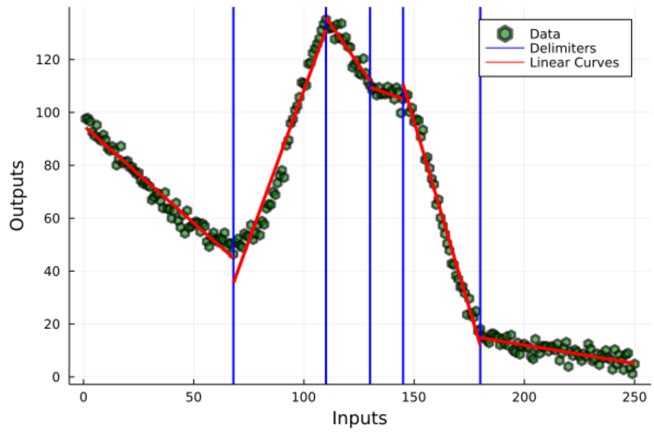
In addition to the regressor performance, one can analyze



(a) $I = 4$



(b) $I = 5$



(c) $I = 6$

Fig. 2. The OLS method by parts.

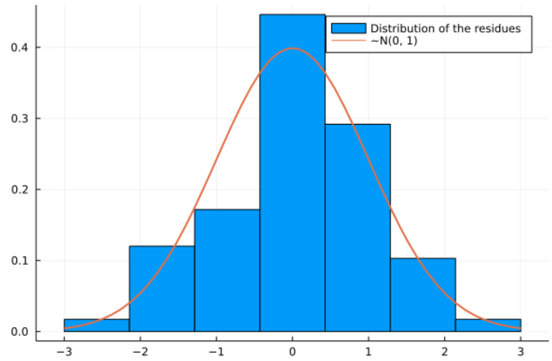


Fig. 3. Distribution of the residuals.

the distribution of the residuals. Let ξ_n be the normalized residual value, that is, $\xi_n = e_n/\sigma_e$. For a good placement of the intervals, the normalized residual distribution approximates to a zero-mean Gaussian distribution with unitary variance, i.e., $\xi_n \sim N(0,1)$. The Figure 3 shows the distribution (for $i = 1$) of the residuals along with the Gaussian distribution.

II. Polynomial Regression

The Polynomial regression is a nonlinear method which extends the concept of the SISO OLS algorithm. At the instant n, the input x_n is utilized to generate the tuple $(h_{n,1}, h_{n,2}, \ldots, h_{n,K})$, where $h_{n,k} = x_n^k$, $\forall\ k \in \{1, 2, \ldots, K\}$. Therefore, the SISO model is transformed into a MISO (Multiple Input, Single Output) model, where the kth input of the nth instant is x_n to the power k. Note that, albeit this model is not linear with relation to x_n, it is with relation to $h_{n,k}$, that is,

$$\begin{aligned} y_n &= f(x_n; \hat{\boldsymbol{\theta}}) = \mathbf{h}_n^\mathsf{T}\hat{\boldsymbol{\theta}} \\ &= \hat{\theta}_0 + \hat{\theta}_1 h_{n,1} + \hat{\theta}_2 h_{n,2} + \cdots + \hat{\theta}_K h_{n,K}, \end{aligned} \tag{9}$$

where $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_0 & \hat{\theta}_0 & \cdots & \hat{\theta}_K \end{bmatrix}^\mathsf{T} \in \mathbb{R}^{K+1}$ and $\mathbf{h}_n = \begin{bmatrix} 1 & h_{n,1} & h_{n,2} & \cdots & h_{n,K} \end{bmatrix}^\mathsf{T} \in \mathbb{R}^{K+1}$. The matricial notation for all N observations is given by

$$\hat{\mathbf{y}} = \mathbf{H}\hat{\boldsymbol{\theta}} \tag{10}$$

where $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}^\mathsf{T} \in \mathbb{R}^N$ and

$$\mathbf{H} = \begin{bmatrix} 1 & h_{1,1} & h_{1,2} & \cdots & h_{1,K} \\ 1 & h_{2,1} & h_{2,2} & \cdots & h_{2,K} \\ \vdots & & & \ddots & \vdots \\ 1 & h_{N,1} & h_{N,2} & \cdots & h_{N,K} \end{bmatrix} \in \mathbb{R}^{N\times(K+1)} \tag{11}$$

The cost function is given by

$$\begin{aligned} J\left(\hat{\boldsymbol{\theta}}\right) &= \left(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\theta}}\right)^\mathsf{T}\left(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\theta}}\right) \\ &= \mathbf{y}^\mathsf{T}\mathbf{y} - 2\mathbf{y}^\mathsf{T}\mathbf{H}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}^\mathsf{T}\mathbf{H}^\mathsf{T}\mathbf{H}\hat{\boldsymbol{\theta}} \end{aligned} \tag{12}$$

By differentiating the Equation (12) with respect to $\hat{\boldsymbol{\theta}}$ and setting its result to zero, we get

$$\begin{aligned} \frac{\partial J\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{\boldsymbol{\theta}}} &= -2\mathbf{H}^\mathsf{T}\mathbf{y} + 2\mathbf{H}^\mathsf{T}\mathbf{H}\hat{\boldsymbol{\theta}} = 0 \\ \therefore \hat{\boldsymbol{\theta}} &= \mathbf{H}^\dagger\mathbf{y}, \end{aligned} \tag{13}$$

where $\mathbf{H}^\dagger = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}$ is the pseudoinverse of $\mathbf{H}$, also called left inverse since $\mathbf{H}$ is a tall matrix ($N > K$). The fact that $\mathbf{H}$ is full rank ($rank(\mathbf{H}) = K+1$) guarantees the invertibility of $\mathbf{H}^T\mathbf{H}$ [3]. The procedure of the Kth-order polynomial regression is summarized in the Algorithm 2, where the hyperparameter K is analyzed for the set $\{5, 6, 7\}$.

Algorithm 2: Kth-order polynomial regressor.

```
Input: {(x_n, y_n)}_{n=1}^N
for K ∈ {5,6,7} do
    y ← generated from {y_n}_{n=1}^N
    H ← generated from {x_n}_{n=1}^N
    H† = (H^T H)^{-1} H   /* The pseudoinverse          */
    θ̂ = H† y
```

The Figure 4 shows the curve fitting for the Kth-order polynomial regressor. As shown in the Table II, the 6th-oder polynomial curve reached the best result. For this case, the model achieved the Pearson correlation coefficient between $\mathbf{y}$ and $\hat{\mathbf{y}}$ of 97.84%.

TABLE II
OLS by parts performance - R^2

$K = 5$	$K = 6$	$K = 7$
0.85	0.957	0.566

III. ORGANIZATION OF THE PAPER

This section presents the main issues for editing the manuscript.

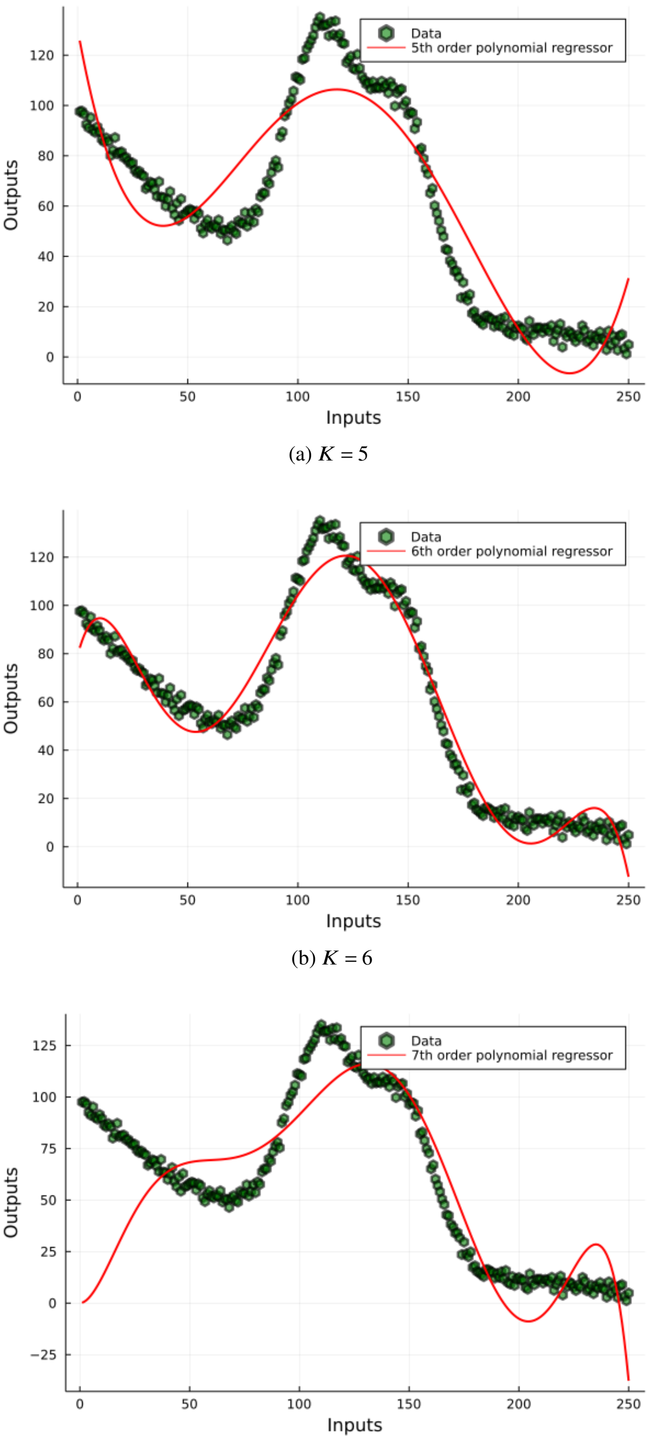
A. General Organization

The papers that shall be published in the Brazilian Power Electronics Journal must contain the following main sections: 1) Title; 2) Authors and Affiliations; 3) Abstract and Keywords; 4) Introduction; 5) Body Text; 6) Conclusions; 7) References; 8) Biographies. This order must be respected, unless the authors add some items, such as: Nomenclature; Appendices and Acknowledgements.

Some comments regarding the main items of the manuscripts are presented below.

1) Title: The title of the paper should be as succinct as possible, stating the subject of the paper in a very clear manner. It should be centered at the top of the first page, in bold, type size 14 points, with the whole title in capital letters.

2) Authors and affiliations: Below the title (leaving one blank line), also centered, the name(s) of the author(s) must be included. The middle names may be abbreviated, but the first and last names must be written in their complete



(a) $K = 5$

(b) $K = 6$

(c) $K = 7$

Fig. 4. The Kth-order polynomial regressor.

forms (type size 12 points). Immediately below the authors' names, their affiliations, with city, state and country, must be informed (type size 10 points). The electronic addresses must be informed just below the affiliations (type size 10 points).

3) Abstract and keywords: This part is considered one of the most important in the whole paper. It is based on information in Abstract and Keywords that technical papers are indexed and stored in databases.

The Abstract should have no more than 200 words, indicating the main ideas contained in the paper, as well as procedures and obtained results. The Abstract should not be confused with the Introduction and should not have any abbreviations, references, figures, etc. For writing the Abstract, as well as the whole manuscript, you should use passive voice, e. g., "... the experimental results show that..." instead of "... the results we obtained show that...". The word Abstract must be written both in italic and in bold. The Abstract text should be in bold.

Keywords are index terms that identify the main topics of the paper. The term Keywords must be both in italic and bold. The Keywords themselves should be in bold.

4) Introduction: The Introduction must prepare the reader for the paper he/she will read, including a historical overview of the subject and also presenting the main contributions of the paper. The Introduction must not be similar to the Abstract and it is the first section of the paper to be numbered as a section.

5) Body text: The authors must organize the body text in various sections, which should contain important information about the proposal of the paper, facilitating its comprehension for readers.

6) Conclusions: The conclusions should be as clear as possible, highlighting the importance of the paper in the respective research area. The advantages and disadvantages of the proposed subject should be clearly emphasized, as well as the obtained results and possible applications.

7) References: The citation of references throughout the text should appear between square brackets, just before the punctuation mark at the end of the sentence in which the reference is inserted. Only the number of the references should be used, avoiding citations such as "... according to the reference [2]...".

Papers that were accepted for publication, but were not published yet, should also be in the references along with the citation "in Press".

Papers from journals and conferences must begin with the name of the authors (initials followed by the last name), followed by the title, journal or conference name (in italic), number of volume, pages, month and year of publication.

Regarding books, following the name of the authors (initials followed by last name), the title should be in italic, and then should come the publisher, number of edition and place and year of publication.

At the end of these guidelines, there is an example of how the references should be inserted **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**.

8) Biographies: The biographies of the authors should appear in the same order as in the beginning of the paper and should basically contain the following items:

- Full name (in bold and underlined);
- Place and year of undergraduation and graduation conclusions;
- Professional experience (Institutions and companies in which they have worked, number of patents obtained, areas of expertise, relevant scientific activities, scientific societies in which they are members, etc.).

In case additional items are used, such as Nomenclature, Appendices and Acknowledgements, the following instructions should be considered:

9) Nomenclature: The nomenclature consists of the definition of quantities and symbols used throughout the paper. Its inclusion is not mandatory and this item must not be numbered. If this item is included, it should precede the Introduction. In case the authors do not include this item, the definition of quantities and symbols must occur during the text, right after they appear. In the beginning of these guidelines there is an example of this optional item.

10) Acknowledgements and appendices: The acknowledgements to any collaborators, as well as appendices, do not receive any numeration and should be at the end of the text, before the references. At the end of this text there is an example of this optional item.

On the last page of the paper, the authors should distribute the contents evenly, using both columns, in a way that both end in a parallel manner.

B. Organization of the Sections of the Paper

The organization of the manuscript in titles and subtitles is important to divide it in sections, which help the reader to find subjects of interest in the paper. They also help the authors to develop their paper in an orderly form. The paper can be organized in primary, secondary and tertiary sections.

The primary sections are the titles of the actual sections. They are written in capital letters in the center of the column separated by a blank line above and another one below them, and sequential Roman numerals should be used.

The secondary sections are the subtitles of the sections. Just the first letter of each word of the section should be written with a capital letter. It should be located at the left part of the column being separated by a blank line above from the rest of the text. The designation of the secondary sections is done with letters in uppercase form, followed by a dot. They should be in italic.

The tertiary sections are subdivisions of the secondary sections. Only the first letter of the first word of the section should be a capital letter. The designation of the tertiary sections should be done with Arabic numerals, followed by parentheses. They should be in italic.

IV. OTHER INSTRUCTIONS

A. Editorial Rules

For papers with multiple authors, it is necessary to inform the order of presentation of the authors and filling out the Copyright form at the https://mc04.manuscriptcentral.com/revistaep, authorizing the publication of the paper.

The Brazilian Power Electronics Journal should be considered source of original publication. It reserves its right to make normative, spelling and grammatical modifications in the original files, but respecting the style of the authors. The final versions cannot be sent to the authors.

The published papers will become property of the Brazilian Power Electronics Journal, and its total or partial reprinting must be authorized by SOBRAEP.

Figures, tables and equations should follow the following guidelines.

B. Figures and Tables

Tables and figures (drawings or pictures) should be inserted in the text right after they are mentioned for the first time, as long as they fit the size of the columns; if necessary, use the whole page. Figures resolution should be at least 300 dpi and vector files should be preferably used for better print quality. Table captions should be above the tables and figure captions should be below the figures. The tables should have titles and they are designated by the word Table, being numbered in sequence by Roman numerals. Table captions must be centered and in bold.

Figures also need captions and they are designated by Figure in the text (Fig. in the caption itself), numbered with Arabic numerals in a sequenced manner, left- and right-justified, as shown in the example. The designation of the parts of a figure is done by adding lowercase letters to the numbers of the figures starting with the letter a, e.g. Figure 1(a).

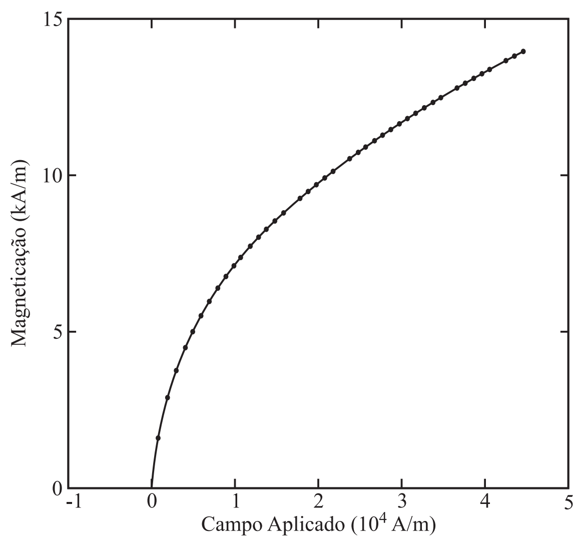


Fig. 5. Magnetization as a function of applied field. (Note that "Fig." is abbreviated and there is a period after the figure number followed by two spaces.)

To better understand graphs, the definition of their axes should be done with words not letters, except when referring to waveforms and phase planes. The units should be between parentheses. For example, use the denomination “Magnetization (A/m)”, instead of “M (A/m)”.

Figures and tables should be positioned preferably in the beginning or the end of the column, avoiding putting them in the middle. Avoid tables and figures whose sizes exceed the size of the columns. The figures should preferentially be in black, with a white background, since the printed version of the journal is in black and white. Their lines should be thick, so the impression is readable.

C. Abbreviations and Acronyms

Abbreviations and acronyms must be defined the first time they are used in the text, e.g. “... Pulse-Width Modulation (PWM)...”.

D. Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (1). The equations should be written in a compact form, centered in the column. If a nomenclature section is not included in the beginning of the text, the quantities should be defined right after the equation, such as:

$$\Delta I_L = I_o + \frac{\sqrt{3}}{2}\frac{V_i}{Z} \tag{14}$$

where:

ΔI_L - resonant inductor peak current;
I_o - load current;
V_i - source voltage;
Z - characteristic impedance.

V. CONCLUSIONS

This paper was fully written in accordance with the guidelines for submissions of papers in English.

ACKNOWLEDGEMENTS

The authors thank John Doe for the collaboration of preparing this paper. This Project was financed by the CNPq (xxyyzz process).

REFERENCES

[1] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc., 1993.

[2] P. S. Diniz, *et al.*, *Adaptive filtering*, vol. 4, Springer, 1997.

[3] G. Strang, G. Strang, G. Strang, G. Strang, *Introduction to linear algebra*, vol. 3, Wellesley-Cambridge Press Wellesley, MA, 1993.