Bollettino Unione Matematica Italiana

Rossella Cancelliere, Mario Gai

Function approximation of Seidel aberrations by a neural network

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. **7-B** (2004), n.3, p. 687–696.

Unione Matematica Italiana

<http://www.bdim.eu/item?id=BUMI_2004_8_7B_3_687_0>

Bollettino dell'Unione Matematica Italiana, Unione Matematica Italiana, 2004.

Bollettino U. M. I.
(8) 7-B (2004), 687-696

Function Approximation of Seidel Aberrations by a Neural Network.

Rossella Cancelliere - Mario Gai

Sunto. – *In questo articolo viene studiata la possibilità di usare una rete neurale feedforward per identificare eventuali discrepanze tra un'immagine astronomica reale ed un suo modello predefinito. Questo compito viene affrontato grazie alla capacità delle reti neurali di risolvere un problema di approssimazione non lineare di funzioni attraverso la costruzione di un'ipersuperficie approssimante un insieme dato di punti sparsi. La codifica delle immagini viene effettuata associando ciascuna di esse ad alcuni momenti statistici opportunamente scelti, calcolati relativamente agli assi* $\{x, y\}$, *ottenendo in tal modo un metodo computazionalmente economico che permette un approccio realmente efficace alla diagnostica delle aberrazioni di Seidel.*

Summary. – *This paper deals with the possibility of using a feedforward neural network to test the discrepancies between a real astronomical image and a predefined template. This task can be accomplished thanks to the capability of neural networks to solve a nonlinear approximation problem, i.e. to construct an hypersurface that approximates a given set of scattered data couples. Images are encoded associating each of them with some conveniently chosen statistical moments, evaluated along the* $\{x, y\}$ *axes; in this way a parsimonious method is obtained that allows a really effective approach to Seidel aberration diagnostics.*

1. – Introduction.

It is well-known that an intimate connection between approximation theory and neural networks exists: given a compact set $D \subset \mathbb{R}^P$, a multilayer perceptron with one hidden layer can uniformly approximate any continuous function in $C(D)$ to any required accuracy. This result was established by Cybenko [3], Hornik et al. [9] e Funahashi [5], and an excellent survey of this topic is available in [4].

The first largely successful neural network model, the multilayer perceptron, was presented in 1986 by D. Rumelhart et al. [12] as an extension of the perceptron model [11].

Recently some attempts to use neural networks in astronomy have been performed, mainly in the field of adaptive optics: the reader can find details in the papers by Lloyd-Hart et al. [10] and Wizinowich et al. [13].

A possible application concerns the location of the position of a stellar image; this is possible with accuracy well below its characteristic size, when the signal to noise ratio (*SNR*) is sufficiently high. The location uncertainty is $\sigma = \alpha \cdot L/SNR$, where α is a factor keeping into account geometric factors and the centring algorithm and L is the root mean square width of the image ([7]). The best estimate of image position is obtained by a least square approach, evaluating the discrepancy between the data and the template describing the reference image. The location algorithm is therefore very sensitive to any variation of the actual image with respect to the selected template.

It is of paramount importance to check the compatibility between the real image and the reference profile; also important is the capability of extracting from the data some parameters suitable for a new definition of the template, in order to improve its similarity to the data. Self-calibration of the data, by deduction of the parameters for optimisation of the image template, is a key element in the control of the systematic effects in the position measurement.

Because of these reasons our target is the implementation of a tool for analysis of realistic images and deduction of a set of aberration parameters able to describe their discrepancy with respect to the ideal, non-aberrated image.

In Section 2 we resume the main features of sigmoidal neural networks and backpropagation algorithm, with a brief reminder of the specific definitions. In Section 3 we discuss the image characterisation problem addressed in the present work; in Section 4 we describe the generation of the data set, its processing and the current results.

2. – Sigmoidal neural networks.

In this section we just remind some of the basic definitions and characteristics; a comprehensive review on neural network properties and applications can be found in [8].

Neural networks learn from examples, that is, given the training set of N multi-dimensional data pairs $\{(x_i, F(x_i))/x_i \in \mathbb{R}^P, F(x_i) \in \mathbb{R}^Q\}$, $i = 1, \ldots, N$, after the training if x_i is the input to the network, the output is close to, or coincident with, the desired answer $F(x_i)$ and the network has generalization properties too, that is it gives as output $F(x_i)$ even if the input is only «close to» x_i, for instance a noisy or distorted or incomplete version of x_i.

This can be expressed as a classical approximation problem: given a set of points x_i, $i = 1, \ldots N$, distinct and generally scattered, in a domain $D \subset \mathbb{R}^P$, and a linear space $\Phi(D)$, spanned by continuous real basis functions o_j,

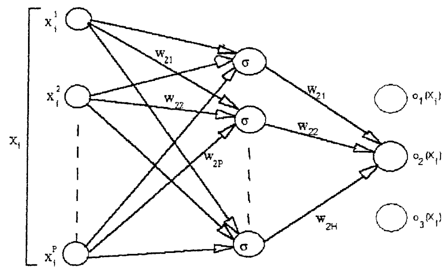

Fig. 1. – A multilayer perceptron with one hidden layer.

$j = 1, \ldots, H$, the multivariate approximation problem at scattered data pairs $(x_i, F(x_i))$ consists in finding a function $o(x) \in \mathbb{R}^Q$, whose components are a linear superposition of the basis functions o_j, minimising the error functional

$$E \equiv \sum_i \sum_{m=1}^{Q} (F_m(x_i) - o_m(x_i))^2 = \sum_i \sum_m \left(F_m(x_i) - \sum_{j=1}^{H} w_{mj}\, o_j(x_i) \right)^2 .$$

The multilayer perceptron, with sigmoidal units in the hidden layers, is one of the most known and used neural network model: it computes distances in the input space (i.e. among patterns $x_i \in \mathbb{R}^P$) using a metric based on inner products and it is usually trained by the backpropagation algorithm.

The architecture of a sigmoidal neural network is schematically shown in Fig. 1, in which we find the most common three-layers case. The network is described by the following equations:

$$a_j^{k+1} = \sum_{j'} w_{jj'}\, o_{j'}^k + bias_j$$

$$o_j^{k+1} = \sigma(a_j^{k+1}) \equiv \frac{1}{1 + e^{-a_j^{k+1}}}$$

$$o_m^{\text{out}} \equiv \sum_j w_{mj}\, o_j^{\text{out}-1}$$

Here a is the input to each unit, o is its output and w_{ij} is the weight associated to the connection between units i and j; each unit is defined by two indexes, a superscript specifying its layer (i.e. input, hidden or output layer) and a subscript labelling each unit in a layer.

The training procedure must identify the best set of weights $\{w_{ij}\}$ solving the approximation problem $o(x_i) \approx F(x_i)$ and this is usually reached by the iterative process corresponding to the standard backpropagation algorithm.

At each step, each weight is modified accordingly to the gradient descent rule (a more detailed description can be found in [12]), completed with the momentum term, $w_{ij} = w_{ij} + \Delta w_{ij}$, $\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}$ where E is the error functional defined above.

This procedure is iterated many times over the complete set of examples $\{x_i, F(x_i)\}$ (the training set), and under appropriate conditions it converges to a suitable set of weights defining the desired approximating function. Convergence is usually defined in terms of the error functional, evaluated over the whole training set; when a pre-selected threshold E_T is reached, the neural network can be tested using a different set of data $\{x_i', F(x_i')\}$, the so called test set.

3. – Telescope images.

The ideal image produced by a telescope corresponds to that of a circular aperture, derived in basic textbooks on optics; a star can be considered as a point-like source at infinity, producing a flat wavefront for an observer outside the atmosphere. We adopt the notation from [1]. The ideal monochromatic image produced by an unobstructed circular pupil of diameter D, fed by such a planar wavefront, at wavelength λ, is radial and described by the squared Airy function:

$$I(r) = k[2\, J_1(r)/r]^2. \tag{1}$$

Here J_1 is the Bessel function of the first kind, order one, k a normalisation constant, and $r = D/2$ the aperture radius. The Airy diameter, between the first two minima, is $2.44\lambda/D$ in angular units; the linear scale is defined by the magnification.

Real images can be described by an extension of the formalism which introduces perturbations to the planar wavefront, as in the Seidel aberration set, which includes five terms:

- A_s: Spherical aberration;
- A_c: Coma;
- A_a: Astigmatism;
- A_d: Defocus (field curvature);
- A_t: Distortion (tilt).

The five Seidel aberrations define the contributions to the local phase over the aperture (i.e. the deviation from planarity of the wavefront) by means of the lower order circular functions; the *phase aberration* Φ is

$$\Phi(\varrho, \theta) = \frac{2\pi}{\lambda}[A_s \varrho^4 + A_c \varrho^3 \cos\theta + A_a \varrho^2 \cos^2\theta + A_d \varrho^2 + A_t \varrho \cos\theta], \tag{2}$$

where ϱ, θ are the radial pupil coordinates (normalised radius, azimuth).

The generic diffraction image on the focal plane associated to a set of aberration values is described by the square modulus of the Fourier Transform of

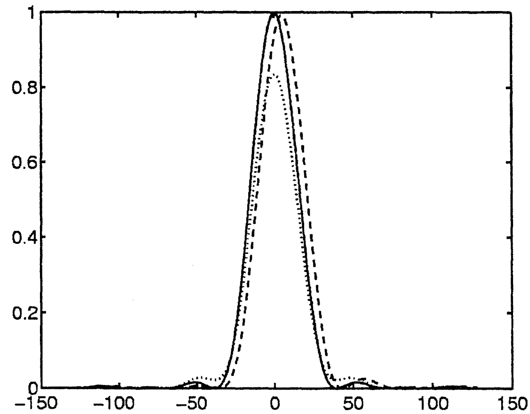

Fig. 2. – A section of the ideal image (solid line), and of two aberrated images generated with one wavelength of coma (dashed line) and one wavelength of defocus (dotted line).

the pupil function $e^{i\Phi}$:

$$I(r,\,\phi) = \frac{k}{\pi^2}\left|\int\limits_0^1 d\varrho \int\limits_0^{2\pi} d\theta\,\varrho e^{i\Phi(\varrho,\,\theta)}\,e^{-i\pi r\varrho\cos(\theta-\phi)}\right|^2 \tag{3}$$

where r and ϕ are the image coordinates. If $\Phi = 0$ (non-aberrated case, $\{A_n\} = 0$), eq. (1) is retrieved. In fig. 2, we show the ideal image (solid line), and two aberrated images corresponding to one wavelength of respectively coma (dashed line) and defocus (dotted line); the deformation induced on the image is asymmetric in the former case, corresponding to a displacement of the image center of gravity, whereas it is symmetric in the latter case, which leaves unaltered the apparent image position.

By replacement of eq. (2) in eq. (3), it is possible to put in evidence the nonlinear relation between the aberration set and the image.

Hereafter, we adopt the regime of small aberrations, corresponding to the classical Rayleigh criterion of one quarter of wavelength ($\{A_n\} \in [-0.25, 0.25]$), for the test set. We select for training a slightly larger interval ($\{A_n\} \in [-0.3, 0.3]$), to avoid potential boundary problems. In case of larger aberrations, the image quality degrades rapidly. Our interest is focused on the classification capability of the neural network in the case of small image perturbations, i.e. small aberrations.

3.1. *Image encoding.*

Typical astronomical images are sampled over a small number of pixels, to maximise the field of view, i.e. observe simultaneously a large area. The minimum sampling requirements, related to the Nyquist-Shannon criterion, are of

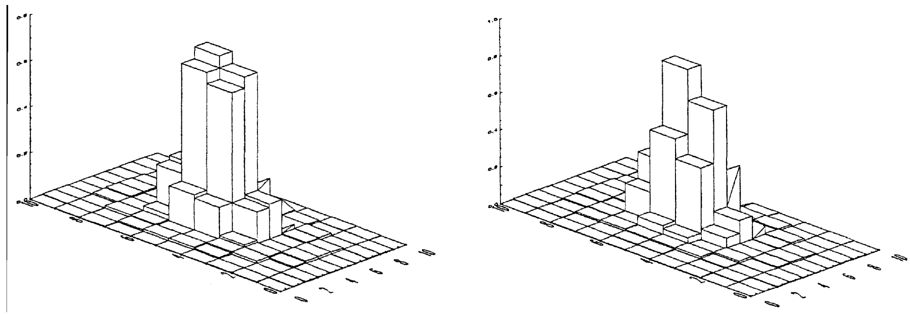

Fig. 3. – Airy image sampled over a low-resolution pixel array. Small displacements of the image vs. the detector provide large variations of the detected signal.

order of two pixels over the full width at half maximum, or about four-five pixels within the Airy diameter. The signal detected in each pixel is then affected by strong variations depending on the initial phase (or relative position) of the parent intensity distribution (the continuous image) with respect to the pixel array, as shown in Fig. 3. The pixel intensity distribution of the measured images, thus, is not convenient for evaluating the discrepancy of the underlying image (3) with respect to the nominal Airy image (1).

It may be possible to add a magnifying device, providing good sampling for the images in a small region: in this case, the resolution is adequate to minimise the effects of the finite pixel size ([6]). However, assuming a sampling of 20 pixels per Airy diameter, and reading up to the third Airy ring, the image size is $60 \times 60 = 3600$ pixels. Direct usage of such images as input data to the neural network is impractical, because of the large computational load involved, and identification of a more compact encoding, possibly removing the need for additional custom hardware, appears as appealing.

The encoding scheme we adopt for the images allows extraction of the desired information for classification: each input image is described by the centre of gravity and the first central moments, up to the fourth order:

$$(4) \qquad \mu_x = \frac{\iint dx\, dy\, x \cdot I(x,y)}{\iint dx\, dy\, I(x,y)} \qquad \mu_y = \frac{\iint dx\, dy\, y \cdot I(x,y)}{\iint dx\, dy\, I(x,y)}$$

$$\sigma_x^2 = \frac{\iint dx\, dy\, (x-\mu_x)^2 \cdot I(x,y)}{\iint dx\, dy\, I(x,y)} \qquad \sigma_y^2 = \frac{\iint dx\, dy\, (y-\mu_y)^2 \cdot I(x,y)}{\iint dx\, dy\, I(x,y)}$$

$$M(i,j) = \frac{\iint dx\, dy \left(\frac{x-\mu_x}{\sigma_x}\right)^i \left(\frac{y-\mu_y}{\sigma_y}\right)^j \cdot I(x,y)}{\iint dx\, dy\, I(x,y)}$$

The central moments are much less sensitive than the image itself to the effects related to the finite pixel size; therefore, they can be deduced also from the low resolution images mentioned above, without the need for high resolution detectors. Moreover, the central moments have an immediate physical meaning:

- the first order moment provides the centre of gravity of the image
- the second order central moment is the mean square width
- the third order central moment (skewness) is an index of the image asymmetry
- the fourth order central moment (kurtosis) is an index of how much peaked is the distribution; for a Gaussian, its value is 3.

It is important to verify the sensitivity of these quantities to aberrations; for instance the x square width is shown in fig. 4, where the values obtained for independent variation of each aberration, are plotted side by side for ease of comparison over the selected range $[-0.3\lambda, 0.3\lambda]$. It appears to be sensitive mostly to defocus, much less to spherical aberration and coma, and insensitive to astigmatism and distortion. We introduced this encoding tecnique in [2], where it is possible to find more details on this argument. Here we select as suitable input variables for our neural network ten moments i.e. σ_x^2, σ_y^2, $M(0, 3)$, $M(0, 4)$, $M(1, 1)$, $M(2, 1)$, $M(1, 2)$, $M(3, 1)$, $M(1, 3)$ and $M(2, 2)$; because aberrations differently influence different sections of each image, we measured the moments, according to eq. (4), in two different subareas of the global image, each treated as a point, therefore the input vector to the neural network has 20 components.

The Seidel aberrations are the targets of our network, therefore the output vector of the neural network is five-dimensional.

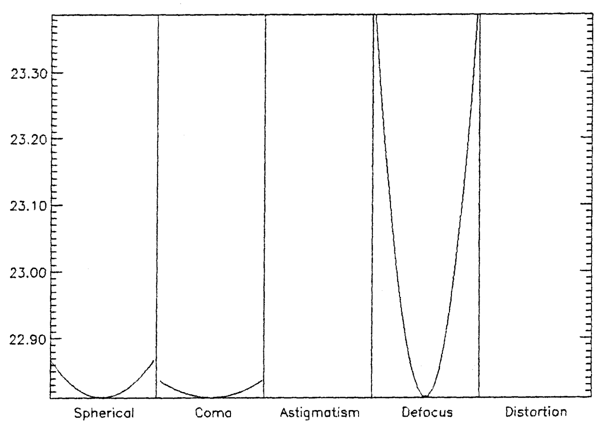

Fig. 4. – Variation of the x variance vs. each aberration.

4. – Data processing and results.

In this section we describe the generation of the training and test sets and the results related to the neural network performances.

The training set consists of $T = 2500$ instances; we placed 100 samples on each axis (variation of a single aberration at a time) while the remaining 2000 points are randomly distributed, i.e. each variable is independently chosen.

From a given set of five aberration values, the image is built accordingly to eq. (3) and (2); then the moments corresponding to eq. (4) are computed and the values inserted in the data file.

The test set consists of $T = 2500$ instances generated in a similar way. A total of 2500 test instances has been generated.

We optimized a sigmoidal neural network with an hidden layer made by 100 units on the training set and verified it on the test set; the training required 8000 iterations. Because the desired behaviour of the neural network is the computation, over the test set, of output values coincident with the pre-defined target values, the plot of output vs. target for each aberration should be ideally the bisector of the first quadrant; these plots are shown in fig. 5.

We also computed the best fit lines together with the fit errors; the results are shown in tab. 1.

The outputs computed by the neural network are quite consistent with the desired test targets; coma and distortion are the best recognized aberrations,

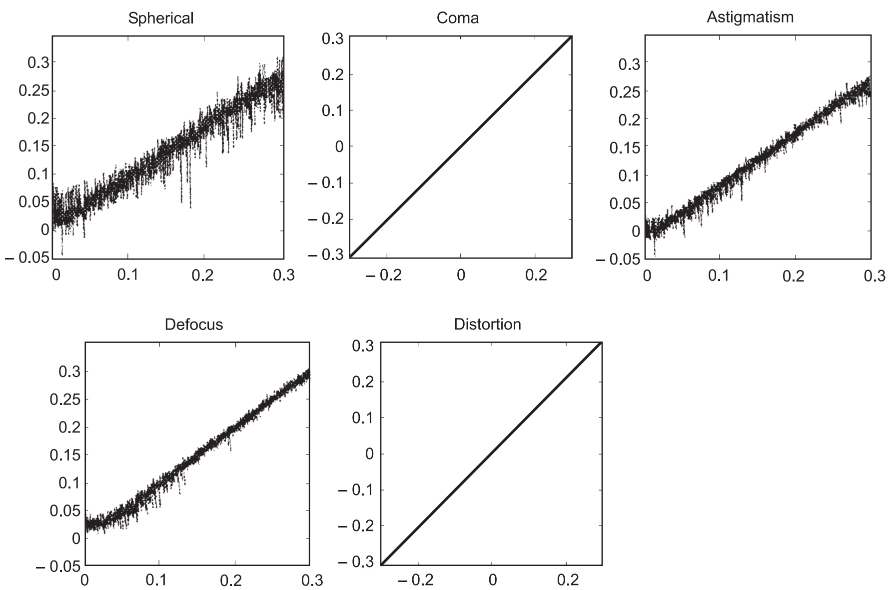

Fig. 5. – Aberration recognition.

TABLE 1. – Results of best fit between the network output and the test set targets, for each output variable.

Spherical	$a = 0.0126$ $b = 0.977$	$\sigma_a = 0.0006$ $\sigma_b = 0.003$
Coma	$a = 0.00069$ $b = 0.9999$	$\sigma_a = 0.00002$ $\sigma_b = 0.0001$
Astigmatism	$a = -0.0008$ $b = 0.986$	$\sigma_a = 0.0003$ $\sigma_b = 0.002$
Defocus	$a = 0.0030$ $b = 1.002$	$\sigma_a = 0.0003$ $\sigma_b = 0.002$
Distorsion	$a = -0.00029$ $b = 0.9980$	$\sigma_a = 0.00002$ $\sigma_b = 0.0001$

whereas spherical aberration, astigmatism and defocus are affected by larger systematic errors and larger dispersion.

5. – Conclusions.

In this paper we use a neural network to recognize aberrations in astronomical images. We test the performances of the network encoding the problem through a compact set of image descriptors, selected among the statistical moments up to the fourth order.

The relation between network outputs and targets is quite close to the ideal linear case; standard improvement methods as increasing the statistical sample for training and the number of internal nodes, may further improve the performance.

We plan in our future work the optimization of input moment selection and detailed investigations of noise propagation properties.

REFERENCES

[1] M. BORN - E. WOLF, *Principles of optics*, Pergamon, New York, 1985.

[2] R. CANCELLIERE - M. GAI, *A Comparative Analysis of Neural Network Performances in Astronomy Imaging*, Applied Numerical Mathematics, **45**, n. 1 (2003), 87-98.

[3] G. CYBENKO, ∞ *approximation by superpositions of a sigmoidal function*, Math. Control-Signals Systems, **2** (1989), 303-314.

[4] S. W. ELLACOTT, *Aspects of the numerical analysis of neural networks*, Acta Numerica, 1994, 145-202.

[5] K. I. FUNAHASHI, *On the approximate realization of continuous mappings by neural networks*, Neural Networks, **2** (1989), 183-192.
[6] M. GAI - S. CASERTANO - D. CAROLLO - M. G. LATTANZI, *Location estimators for interferometric fringe*, Publ. Astron. Soc. Pac., **110** (1998), 848-862.
[7] M. GAI - D. CAROLLO - M. DELBÓ - M. G. LATTANZI - G. MASSONE - F. BERTINETTO - G. MANA - S. CESARE, *Location accuracy limitations for CCD cameras*, A&A, **367** (2001), 362-370.
[8] S. HAYKIN, *Neural Networks*, a Comprehensive Foundation IEEE Computer Society Press, 1994.
[9] K. HORNIK - M. STINCHCOMBE - H. WHITE, *Multilayer feedforward networks are universal approximators*, Neural Networks, **2** (1989), 359-366.
[10] M. LOYD-HART - P. WIZINOWICH - B. MCLEOD - D. WITTMAN - D. COLUCCI - R. DEKANY - D. MCCARTHY - J. R. P. ANGEL - D. SANDLER, *First Results of an On-line Adaptive Optics System with Atmospheric Wavefront Sensing by an Artificial Neural Network*, ApJ, **390** (1992), L41-44.
[11] M. MINSKY - S. PAPERT, *Perceptrons*, Cambridge, MA:MIT Press, 1969.
[12] D. RUMELHART - G. E. HINTON - R. J. WILLIAMS, *Learning internal representation by error propagation. Parallel Distribuited Processing (PDP): Exploration in the Microstructure of Cognition*, MIT Press, Cambridge, Massachussetts, **1** (1986), 318-362.
[13] P. WIZINOWICH - M. LOYD-HART - R. ANGEL, *Adaptive Optics for Array Telescopes Using Neural Networks Techniques on Transputers*, Transputing '91, IOS Press, Washington D.C., **1** (1991), 170-183.

Rossella Cancelliere: Dipartimento di Matematica, Università di Torino
v. C. Alberto 10, 10123 Torino, Italy. E-mail: cancelliere@dm.unito.it

Mario Gai: Osservatorio Astronomico di Torino, str. Osservatorio 20
10025 Pino Torinese, Torino, Italy. E-mail: gai@to.astro.it

Pervenuta in Redazione
il 4 marzo 2003 e in forma rivista il 9 maggio 2003