

Convergence Properties of Piecewise Power Approximations

Arcady Ponosov¹, Anna Machina², Valeria Tafintseva¹

¹Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, Ås, Norway

²Department of Mathematics and Statistics, University of Victoria, Victoria, Canada

Email: arkadi@nmbu.no, annama@uvic.ca, valeria.tafintseva@nmbu.no

Received 18 June 2016; accepted 8 August 2016; published 11 August 2016

Copyright © 2016 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We address the problem of convergence of approximations obtained from two versions of the piecewise power-law representations arisen in Systems Biology. The most important cases of mean-square and uniform convergence are studied in detail. Advantages and drawbacks of the representations as well as properties of both kinds of convergence are discussed. Numerical approximation algorithms related to piecewise power-law representations are described in Appendix.

Keywords

Power-Law Representations, Piecewise Nonlinear Approximations, Least-Squares Minimization, Mean-Square and Uniform Convergence

1. Introduction

For a given function $v(x_1, \dots, x_n) > 0$, defined in a domain $\Omega \in \mathbb{R}^n$, let us calculate its partial derivatives in the logarithmic space:

$$f_j(P) = \left(\frac{\partial v}{\partial x_j} \frac{x_j}{v} \right)_P = \frac{\partial \ln v}{\partial \ln x_j}, \quad (1)$$

where $P = (x_1, \dots, x_n)$ is an arbitrary point. If $f_j(P) = \text{const}$ in the domain Ω for all j , then v clearly is a power function in Ω of the form $\alpha \prod_{j=1}^n x_j^{\alpha_j}$ where $\alpha > 0$. In this paper we study piecewise constant approximations of the quantities (1) or, in other words, nonlinear approximations of the function v by piecewise power functions.

This study is first of all motivated by applications in Systems Biology, where many networks can be described via compartment models

$$\dot{x}_i = V_i^+(x_1, \dots, x_n) - V_i^-(x_1, \dots, x_n), \quad i = 1, \dots, n, \quad (2)$$

with the influx and efflux functions $V_i^+ \geq 0$ and $V_i^- \geq 0$, respectively.

For instance, in a typical metabolic network used in Biochemical Systems Theory the index i ($i = 1, \dots, n$) refers to the n internal metabolites $x_i \geq 0$. The influx $V_i^+(x_1, \dots, x_n) \geq 0$, resp. efflux $V_i^-(x_1, \dots, x_n) \geq 0$ function accounts for the rate (velocity) of a production (synthesis), resp. degradation of the metabolite x_i .

Another important example is gene regulatory networks which in many cases can be described as a system of nonlinear ordinary differential equations of the form

$$\frac{dx_i}{dt} = F_i(z_1, \dots, z_n) - G_i(z_1, \dots, z_n)x_i, \quad i = 1, \dots, n, \quad (3)$$

where $x_i(t)$ is the gene concentration ($i = 1, \dots, n$) at time t , while the regulatory functions F_i and G_i depend on the response functions $z_k = z_k(x_k)$, which control the activity of gene k and which are assumed to be sigmoid-type functions [1].

The derivatives $f_j(P)$ in the logarithmic space are very important local characteristics of biological networks. In Biochemical Systems Theory these derivatives are known as *the kinetic orders* of the function v , while in Metabolic Control Analysis (see e.g. [2]) they are called *elasticities*. From the mathematical point of view, these quantities measure the local response of the function v to changes in the dependent variable (for instance, the local response of enzyme or other chemical reaction to changes in its environment). Thus, they describe *local sensitivity* of the function v , the terminology which is widespread in e.g. engineering sciences.

If all influx and efflux functions in (2) have constant kinetic orders, one obtains the so-called “synergetic system”, or briefly “S-system”:

$$\dot{x}_i = \alpha_i \prod_{j=1}^n x_j^{g_{ij}} - \beta_i \prod_{j=1}^n x_j^{h_{ij}}, \quad i = 1, \dots, n, \quad (4)$$

where the exponents g_{ij} , h_{ij} represent all the (constant) kinetic orders associated with (4). The right-hand side of an S-system, thus, contains power functions, and analysis based on S-systems is, therefore, called “Power-Law (PL) Formalism”, see e.g. [3]-[7]).

The Power-Law Formalism has been successfully applied to a wide number of problems, for example, to metabolic systems [8], gene circuits [9], signalling networks [10]. Such systems are very advantageous in biological applications, as the systems’ format considerably simplifies mathematical and numerical analysis such as steady state analysis, sensitivity, stability analysis, etc. For instance, calculation of steady states for the S-systems is a linear problem (see [7]). By these and other biological and mathematical reasons, it was suggested in [11] to classify such systems as “a canonical nonlinear form” in systems biology.

In many models, however, the kinetic orders may vary considerably. A typical example is a model coming from Generalized Mass Action

$$\dot{x}_i = \sum_{r=1}^p \mu_{ir} V_r(x_1, \dots, x_n), \quad i = 1, \dots, n, \quad (5)$$

where the power functions $V_r(x_1, \dots, x_n)$ describe the rates of the process no. r , while μ_{ir} is a stoichiometric factor that stands for the number of molecules of x_i produced, *i.e.* $\mu_{ir} = 1, 2, \dots$ or $\mu_{ir} = -1, -2, \dots$. Collecting the processes in (5) in a net process of synthesis V_i^+ (positive terms) and a net process of degradation V_i^- (negative terms) results in an aggregated system (2), which is not an S-system.

Another example of generic systems with non-constant kinetic orders stems from Saturable and Cooperativity Formalism [12] reflecting two essential features of biological systems, which gave the name to this formalism (see [13] for more details). In this case, the system (5) becomes

$$\dot{x}_i = \frac{v \prod_{j=1}^n x_j^{n_j}}{\prod_{j=1}^n (K_j^{n_j} + x_j^{n_j})} - \frac{w \prod_{j=1}^n x_j^{m_j}}{\prod_{j=1}^n (L_j^{m_j} + x_j^{m_j})}, \quad (6)$$

where $i = 1, \dots, n$ and n_j , m_j , K_j and L_j are real numbers.

Another version of Saturable and Cooperativity Formalism, which is mentioned in [12], is defined as follows:

$$\dot{x}_i = \frac{\alpha^{-1} \prod_{j=1}^n x_j^{n_j}}{\prod_{j=1}^n x_j^{n_j} + \alpha^{-1} \sum_{i=1}^n b_i \prod_{j=1, j \neq i}^n x_j^{n_j}} - \frac{\beta^{-1} \prod_{j=1}^n x_j^{m_j}}{\prod_{j=1}^n x_j^{m_j} + \beta^{-1} \sum_{i=1}^n c_i \prod_{j=1, j \neq i}^n x_j^{n_j}}, \tag{7}$$

where $i = 1, \dots, n$ and $n_j, m_j, b_i, c_i, \alpha$ and β are real numbers.

In the case of gene regulatory networks (3) the sensitivities (1) are non-constant as well, even if one considers the functions F_i and G_i to be multilinear in z_k . In addition, the usage of non-multilinear functions are also known in this theory [14].

Taking into account the importance of kinetic orders/elasticities/sensitivities (1) in Systems Biology, one one hand, and convenience of the well-developed analysis of S -systems (stability theory [7], parameter estimation routines [15], software packages) on the other, a new kind of generic representations of compartment systems (2) was suggested in [16] (see also [17] for further applications of this representation). According to this idea, the entire operating domain is divided to partition subsets where all kinetic orders can be viewed as constants. In other words, the system (2) is approximated by a set of S -systems, each being only active in its own partition subset. This way of representing (2) is called ‘‘Piecewise Power-Law Formalism’’ [18].

From the biological point of view, piecewise power-law representations are useful in many respects, when compared to other ways of approximations, as they take into account biologically relevant characteristics (kinetic orders) rather than the standard partial derivatives. Therefore, piecewise S -systems preserve important biological structures and, at the same time, do not destroy a relatively simple mathematical structure of plain S -systems. By this reason, approximations of a general target function by piecewise power approximations may be of a great importance for biological and other modelling. A rigorous mathematical justification of the idea of piecewise power-law approximations is the main purpose of the present paper. More precisely, we consider mean-square and uniform convergence of approximations by piecewise power functions to the target function provided that the associated partitions of the operating domain Ω satisfy some additional assumptions. One of the challenges is that partitions of the operating domain Ω may not be chosen freely in applications. For instance, the partitions may directly stem from biological properties of the model [17]. Other ways of constructing partitions can be dictated by optimality-oriented algorithms. In Appendix (see also [18]) we describe such a method which goes back to the paper [19] and which is based on an automatical procedure, allowing to obtain simultaneously the best possible polyhedral partition and the respective best possible piecewise linear approximation in the logarithmic space.

The main results of the paper are presented in Section 3 (mean-square convergence of piecewise power approximations) and in Section 4 (uniform convergence of piecewise power approximations). Several auxiliary results are proved in Appendices A.1-A.3, while Appendix A.4 presents an approximation algorithm which provides an automated partition and the respective best possible approximation in the logarithmic space for a given number of subdomains. Finally, in Appendix A.5 we explain by example why a direct piecewise power-law fitting is ill-posed.

2. Preliminaries

Throughout the paper we use the following notations (see **Table 1**). Let $v : \mathbb{R}_+^n \setminus \{0\} \rightarrow \mathbb{R}_+$, and let $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $\psi(y) = \log v(x)$, $y = \log x$.

Let Ω be a domain in the space $\mathbb{R}_+^n \setminus \{0\}$, which we call Cartesian. We assume Ω to be closed and bounded (*i.e.* compact) subset of \mathbb{R}^n . Let Δ be its image in the logarithmic space \mathbb{R}^n and $\{\Delta_i^N\}_{i=1}^N$ be a measurable partition of Δ . This means that Δ_i^N are all Borel measurable subsets of \mathbb{R}^n , $\Delta_i^N \cap \Delta_j^N = \emptyset$ for every $i, j = 1, \dots, N$ and $\bigcup_{i=1}^N \Delta_i^N = \Delta$ for any natural N . In some results and algorithms Δ will be a polyhedron domain in the logarithmic space, and Δ_i^N will be a polyhedral partition.

Below $\{\Omega_i^N\}_{i=1}^N$ is the measurable partition of Ω which is the image of the partition $\{\Delta_i^N\}_{i=1}^N$ under the inverse logarithmic transformation.

Table 1. Overview of the basic terminology and notation used in the paper (LS-least-squares).

	Cartesian space	Logarithmic space
Space	$\mathbb{R}_+^n \setminus \{0\}$	\mathbb{R}^n
Target function	$v(x)$	$\psi(y) = \log v(x)$
Domain	Ω	Δ
Measurable partition	$\{\Omega_i^N\}_{i=1}^N$	$\{\Delta_i^N\}_{i=1}^N$
LS power approximation	$v^N(x)$	$\Rightarrow \psi^N(y) = \log(v^N(x))$
LS linear approximation	$V^N(x) = \exp(\Psi^N(y)) \Leftarrow$	$\Psi^N(y)$

We also put $\delta^N = \max_{1 \leq i \leq N} (\text{diam}(\Delta_i^N))$.

Let $v_i^N(x) = \gamma_i^N \prod_{j=1}^n x_j^{f_{ij}^N}$ be a least-squares (LS) power-law fitting to the function v on Ω_i^N , $i = 1, \dots, N$. For $x \in \Omega$ we consider the piecewise power function $v^N(x) = v_i^N(x)$, whenever $x \in \Omega_i^N$. We put $\psi^N(y) = \log(v^N(x))$.

Let $\Psi_i^N(y) = c^N + \sum_{j=1}^n g_{ij}^N y_j$ be a LS linear approximation to the function ψ on Δ_i^N , $i = 1, \dots, N$. For $y \in \Delta$ we consider the piecewise linear function $\Psi^N(y) = \Psi_i^N(y)$, whenever $y \in \Delta_i^N$. We put also $V^N(x) = \exp(\Psi^N(y))$.

We remind that the parameters c^N and g_{ij}^N of the linear functions Ψ_i^N are uniquely obtained from the following minimization criterion in the logarithmic space:

$$\sum_{i=1}^N \int_{\Delta_i^N} [\psi(y_1, \dots, y_n) - \Psi_i^N(y_1, \dots, y_n)]^2 dy_1 \cdots dy_n \rightarrow \min. \quad (8)$$

Alternatively, one can define approximations of the target function v by power functions minimizing the distance in the space Ω :

$$\sum_{i=1}^N \int_{\Omega_i^N} [v(x_1, \dots, x_n) - v_i^N(x_1, \dots, x_n)]^2 dx_1 \cdots dx_n \rightarrow \min. \quad (9)$$

Our last minimization criterion looks similar to (6), but is, in fact, very different

$$\sum_{i=1}^N \int_{\Delta_i^N} [\psi(y_1, \dots, y_n) - \Psi_i^N(y_1, \dots, y_n)]^2 dy_1 \cdots dy_n \rightarrow \min, \quad (10)$$

as the minimum here is taken over *all polyhedral partitions* $\{\Delta_i^N\}_{i=1}^N$ of the polyhedral domain Δ , and *all* corresponding linear functions Ψ_i^N ($i = 1, \dots, N$).

The main advantage of the criteria (8) and (10) is their linearity that provides the uniqueness of the solution and also makes the process of finding the solution computationally cheap, as it is based on explicit matrix formulas. On the other hand the use of the logarithmic transformation requires caution. The influences of the data values will change, as will the error structure of the model. Yet, the criterion (8) only requires a standard linear regression, while the criterion (10) requires a special regression algorithm, still linear, but much more involved (see Appendix A.4 for details).

The criterion (9) gives best possible approximation in terms of the LS error in the Cartesian space. However, a nonlinear regression algorithm should be used in this case, which is less advantageous, especially when the number of the estimated parameters is big. In addition, the nonlinear regression may have other drawbacks, one of which is ill-posedness (see Appendix A.5).

3. Mean-Square Convergence of Piecewise Power Approximations

The results of this section provide the mean-square convergence (L^2 -convergence) of piecewise approximations by power functions. The involved parameters may be e.g. obtained according to one of the minimization criteria (8) or (9).

The main technical challenges stemming from the nature of these minimization algorithms can be summarized as follows: 1) the L^2 -convergence of the approximations in the logarithmic space may not imply the L^2 -convergence of their images in the Cartesian space (and *vice versa*); 2) it is not evident that automatic dissections of the operating domain, as e.g. in the algorithms based on the minimization criterion (10), make the diameters of the partition subsets go to zero even if the number of partition subsets tends to ∞ .

Three propositions below deal with L^2 -convergence in the logarithmic domain.

Proposition 1. *Let the target function $v > 0$ be measurable and bounded on Ω and $\psi = \log v$. Suppose that the measurable partitions $\{\Delta_i^N\}_{i=1}^N$ satisfy the property $\delta^N \rightarrow 0$ ($N \rightarrow \infty$). Then for the corresponding LS approximations Ψ^N in Δ and v^N in Ω we have $\Psi^N \rightarrow \psi$, $v^N \rightarrow v$ in the respective L^2 -norms, if $N \rightarrow \infty$.*

To prove this proposition we need the following lemma, the proof of which can be found in Appendix A.1:

Lemma 1. *Let v be measurable and $0 < m \leq v(y) \leq M$ on Ω for some constants m, M . Let $\psi = \log v$ and the measurable partitions $\{\Delta_i^N\}_{i=1}^N$ satisfy the property $\delta^N \rightarrow 0$ ($N \rightarrow \infty$). Then there exists a sequence $\{\theta^N\}$ of continuous on Δ functions satisfying the properties $\theta^N(y) = \text{const}$ on any Δ_i^N , $\theta^N(y) \leq \psi(y)$ for all $y \in \Delta$ and $\{\theta^N\}$ (resp. $\{\Theta^N\} = \{\exp(\theta^N)\}$) L^2 -converges to ψ (resp. v) on Δ (resp. Ω).*

Proof of Proposition 1. We use the sequences

$$\{\theta^N\} \quad \text{and} \quad \{\Theta^N\} = \{\exp(\theta^N)\}$$

from the lemma 1, which both converge in the L^2 -sense in the respective domains.

Since $\Psi^N(y)$ is the LS piecewise linear approximation in Δ , we have

$$\|\Psi^N - \psi\|_{L^2(\Delta)} \leq \|\theta^N - \psi\|_{L^2(\Delta)} \rightarrow 0$$

as $N \rightarrow \infty$. □

In the next proposition we do not assume that $\delta^N \rightarrow 0$.

Proposition 2. *Let Δ be a polyhedral domain in \mathbb{R}^n , the function ψ be square integrable in Δ and $\{\Delta_i^N\}_{i=1}^N$ be the optimal polyhedral partition of Δ obtained by the algorithm described in Appendix A.4. Then for the corresponding LS approximations Ψ^N in Δ we have $\Psi^N \rightarrow \psi$ in the L^2 -norm, if $N \rightarrow \infty$.*

Proof. Evidently, for the L^2 -function ψ there exists a sequence of polyhedral partitions $\{\bar{\Delta}_i^N\}_{i=1}^N$ of Δ such that $\bar{\delta} = \max_i(\text{diam}(\bar{\Delta}_i^N)) \rightarrow 0$ as $N \rightarrow \infty$ and a sequence of piecewise constant functions $\{\bar{\Psi}^N(y)\}$ given by $\bar{\Psi}^N(y) = \bar{C}_i^N$, whenever $y \in \bar{\Delta}_i^N$ for which $\bar{\Psi}^N \rightarrow \psi$ in the L^2 -norm if $N \rightarrow \infty$.

For the optimal polyhedral approximation $\Psi^N(y)$ we obtain

$$\|\Psi^N - \psi\|_{L^2(\Delta)} \leq \|\bar{\Psi}^N - \psi\|_{L^2(\Delta)} \rightarrow 0$$

as $N \rightarrow \infty$. □

In particular, the assumption on ψ is fulfilled if the target function v is measurable and bounded on Ω .

The case of the L^2 -convergence of the approximations V^N , given as $V^N = \exp(\Psi^N)$, is more involved. The reason for that is that the L^2 -convergence of the sequence $\{\Psi^N\}$ does not necessarily imply the L^2 -convergence of the sequence $\{\exp(\Psi^N)\}$.

We introduce the following notation. Given a partition subset Δ_i^N of Δ we put

$$e_i^N(0, y) = 1, \quad e_i^N(k, y) = y_k - \bar{y}_i^N(k) \quad (k = 1, \dots, n), \tag{11}$$

where the point $(\bar{y}_i^N(k))_{k=1}^n \in \Delta_i^N$ is the center of mass of the convex set Δ_i^N given by

$$\bar{y}_i^N(k) = \frac{1}{\text{mes } \Delta_i^N} \int_{\Delta_i^N} y_k dy, \quad k = 1, \dots, n. \tag{12}$$

Let $A_i^N = (a_i^N(k, l))$ be the symmetric $n \times n$ -matrix with the entries defined as

$$a_i^N(k, l) = \frac{1}{\text{mes } \Delta_i^N} \int_{\Delta_i^N} e_i^N(k, y) e_i^N(l, y) dy, \quad k, l = 1, \dots, n, \tag{13}$$

Below we fix a matrix norm $\|\cdot\|$. All matrix norms are equivalent. One of the norms is Euclidean, which is defined via the maximal eigenvalues: $\|A\| = \sqrt{\lambda_{\max}(A^T A)}$. In the case of symmetric, positive definite matrices (like A_i^N above) we can write that $\|A\| = \lambda_{\max}(A)$.

We say that the sequence of partitions $\{\Delta_i^N\}_{i=1}^N$ ($N = 1, 2, \dots$) of Δ satisfies the condition (Δ) if there exists a constant $c_0 > 0$ such that

$$\text{diam } \Delta_i^N \sqrt{\|(A_i^N)^{-1}\|} \leq c_0.$$

If the chosen norm is Euclidean, then the latter estimate can be rewritten as

$$\frac{\text{diam } \Delta_i^N}{\sqrt{\lambda_i^N}} \leq c_0,$$

where λ_i^N is the least (positive) eigenvalue of the matrix A_i^N ($i = 1, \dots, N$; $N = 1, 2, \dots$).

Informally speaking, this property means that the partition subsets cannot be too different from each other in the shape. Assume, for instance, that the partition sets are enclosed in rectangular boxes. The result below says that if the ratio of the longest and the shortest edges of the boxes is bounded above, *i.e.* boxes are not “too thin”, then the sequence of such boxes satisfies the property (Δ) .

Proposition 3. *A sequence of rectangular boxes $\{P^N\}$ satisfies the property (Δ) if and only if $\sup_{N=1,2,\dots} \frac{a^N}{b^N} < \infty$, where a^N (resp. b^N) is the length of the smallest (resp. biggest) edge of the box P^N .*

Proof. We calculate the matrix (13).

We fix N and the N th rectangular box $P = P^N$ given by

$$y_1^l \leq y_1 \leq y_1^r, y_2^l \leq y_2 \leq y_2^r, \dots, y_n^l \leq y_n \leq y_n^r.$$

Let $\bar{y}_i = \frac{y_i^l + y_i^r}{2}$, $i = 1, \dots, n$, be the center of mass and $e(i, y) = y_i - \bar{y}_i$.

The substitution

$$y_i = y_i^l + \delta_i z_i, \text{ where } \delta_i = y_i^r - y_i^l, z_i \in [0, 1], y_i - \bar{y}_i = \delta_i (z_i - 1/2),$$

yields

$$\int_P e(k, y) e(l, y) dy = \int_{\Sigma} \delta_1 \dots \delta_n \delta_k \delta_l (z_k - 1/2)(z_l - 1/2) dz \quad (1 \leq k, l \leq n)$$

where $\Sigma = [0, 1] \times [0, 1] \times \dots \times [0, 1] \in \mathbb{R}^n$. Since

$$\int_{\Sigma} (z_k - 1/2)(z_l - 1/2) dz = 0 \quad (k \neq l), \int_{\Sigma} (z_k - 1/2)^2 dz = \frac{1}{12},$$

and $\text{mes } \Delta_i^N = \delta_1 \dots \delta_n$, the matrix (13) becomes $\text{diag} \left[\frac{\delta_1^2}{12}, \dots, \frac{\delta_n^2}{12} \right]$. The least eigenvalue of the matrix is equal

to $1/12(\min\{\delta_1^2, \dots, \delta_n^2\})$, i.e. to $1/12(a^N)^2$. The diameter of the box can be estimated above by the constant $\sqrt{nb^N}$, which also dominates the asymptotics of the diameter. Therefore the condition (Δ) is fulfilled for the given sequence of rectangular boxes if and only if the sequence $\{b^N/a^N\}$ is bounded above. \square

The next lemma is proved in Appendix A.2.

Lemma 2. Assume that the target function $v > 0$ is measurable and bounded on Ω and $\psi = \log v$. Assume further that the sequence of partitions $\{\Delta_i^N\}_{i=1}^N$ ($N = 1, 2, \dots$) of Δ satisfies the condition (Δ) . Then the corresponding LS approximations $V^N(x)$ and $\Psi^N(y)$ are uniformly bounded on Ω and Δ , respectively, i.e. there exist constant $C_1 > 0$ and $C_2 > 0$ such that

$$|V^N(x)| \leq C_1 \sup_{x \in \Delta} |v(x)|$$

for all $N = 1, 2, \dots$ and all $x \in \Delta$,

$$|\Psi^N(y)| \leq C_2 \sup_{y \in \Delta} |\psi(y)|$$

for all $N = 1, 2, \dots$ and all $y \in \Delta$.

The main result of this section is the following theorem:

Theorem 4. Let the target function v be measurable and bounded on Ω .

1) If the measurable partitions $\{\Delta_i^N\}_{i=1}^N$ have the property $\delta^N \rightarrow 0$ ($N \rightarrow \infty$), then $v^N \rightarrow v$ and $V^N \rightarrow v$ in the $L^2(\Omega)$ -norm as $N \rightarrow \infty$.

2) Assume that Δ is a polyhedral domain in \mathbb{R}^n , while a sequence of polyhedral partitions $\{\Delta_i^N\}_{i=1}^N$ of Δ and associated LS piecewise linear approximations Ψ^N satisfy the criterion (10) for each $N = 1, 2, \dots$. Assume further that the partitions $\{\Delta_i^N\}_{i=1}^N$ ($N = 1, 2, \dots$) satisfy the condition (Δ) . Then $V^N \rightarrow v$ in the $L^2(\Omega)$ -norm as $N \rightarrow \infty$.

Proof. To prove the first part of the theorem, we apply Lemma 1 and obtain

$$\|v^N - v\|_{L^2(\Omega)} \leq \|\Theta^N - \psi\|_{L^2(\Omega)} \rightarrow 0$$

as $N \rightarrow \infty$, since $v^N(x)$ is the LS piecewise power approximation in Ω .

In the second part of the theorem, we use either Proposition 1 or Proposition 2, which yields the L^2 -convergence of the LS approximations Ψ^N to the function $\psi = \log v$. Applying Lemma 2 we obtain the uniform boundedness of the approximations: $|\psi(y)|, |\Psi^N(y)| \leq M$ for some M and any $N = 1, 2, \dots$. Then we have

$$\begin{aligned} \|V^N - v\|_{L^2(\Omega)} &= \int_{\Omega} |V^N(x) - v(x)|^2 dx \\ &= \int_{\Delta} |\exp(\Psi^N(y)) - \exp(\psi(y))|^2 \exp(y_1 + \dots + y_n) dy \\ &\leq C \|\Psi^N - \psi\|_{L^2(\Delta)}, \end{aligned}$$

where $C = \exp(M) \max \exp(y_1 + \dots + y_n)$. The latter estimate is due to the uniform Lipschitz continuity of the function $\exp(u)$ on the interval $[0, M]$: $|\exp(u) - \exp(v)| \leq \exp(M)|u - v|$.

This estimate proves the L^2 -convergence of the LS approximations V^N to the target function v . \square

4. Uniform Convergence of Approximations

In the previous section we studied convergence of LS approximations in the L^2 -norm. In many applications, however, it is desirable to consider their uniform convergence. This may be, for instance, of interest if we include the obtained approximations into the models based on differential equations, as it is well-known that convergence of (approximations of) solutions is only guaranteed by the uniform convergence of (approximations

of) the right-hand sides.

The main result of this section is formulated in terms of kinetic orders $f_j(x)$ and $f_j^N(x)$ of the target function $v(x_1, \dots, x_n)$ and its piecewise power approximations w^N .

Theorem 5. *Let the target function $v > 0$ be a C^1 -function (i.e. differentiable with the continuous partial derivatives). Let the sequence of partitions $\{\Omega_i^N\}_{i=1}^N$ ($N = 1, 2, \dots$) of Ω have the following two properties:*

- 1) *The closure of each Ω_i^N coincides with the closure of its interior $\text{int } \Omega_i^N$.*
- 2) *$\delta^N \rightarrow 0$ ($N \rightarrow \infty$).*

Assume, in addition, that for any $i = 1, \dots, N$, $N = 1, 2, \dots$ there exist points $P_i^N \in \Omega_i^N$, $Q_i^N \in \text{int } \Omega_i^N$ such that piecewise power approximations w^N ($= w_i^N$ on Ω_i^N) satisfy

$$v(P_i^N) = w^N(P_i^N), \quad f_j(Q_i^N) = f_j^N(Q_i^N). \tag{14}$$

Then $w^N \rightarrow v$ uniformly on Ω as $N \rightarrow \infty$.

Proof. We fix N and consider the corresponding partition Δ_i^N of the domain Δ . Let $\psi = \log v$, $\Phi^N = \log w^N$, $y_i^N = \log P_i^N$, $c_i^N = \log Q_i^N$. Clearly, $\nabla \psi(c_i^N) = (f_1(P_i^N), \dots, f_n(P_i^N))$. By assumption, for $y \in \Delta_i^N$ we have $\Phi_i^N(y) = \nabla \psi(c_i^N) \cdot (y - y_0) + \psi(y_i^N)$, where $y_i^N \in \Delta_i^N$, $c_i^N \in \text{int } \Delta_i^N$. On the other hand, the mean value theorem yields $\psi(y) - \psi(y_i^N) = \nabla \psi(c) \cdot (y - y_i^N)$ where $c \in \Delta_i^N$ depends on y . Therefore

$$|\Phi_i^N(y) - \psi(y)| \leq |\nabla \psi(c) - \nabla \psi(c_i^N)| |y - y_i^N|.$$

The uniform continuity of the continuous vector function $\nabla \psi(y)$ on Δ and the property that

$$\delta^N = \max_{1 \leq i \leq N} \text{diam } \Delta_i^N \rightarrow 0 \quad (N \rightarrow \infty)$$

imply that, given an $\varepsilon > 0$, the estimate

$$|\Phi_i^N(y) - \psi(y)| < \varepsilon. \tag{15}$$

is fulfilled for sufficiently large N .

Since (15) holds for any $i \in \{1, 2, \dots, N\}$, we also obtain that for sufficiently large N $|\Phi^N(y) - \psi(y)| < \varepsilon$, i.e. $\Phi^N \rightarrow \psi$ uniformly on Δ as $N \rightarrow \infty$.

As the uniform convergence of the sequence $\{\Phi^N\}$ implies its uniform boundedness, there is M such that $|\Phi^N(y)|, |\psi(y)| \leq M$ for all $y \in \Delta$. Therefore,

$$|w^N(x) - v(x)| = |\exp \Phi^N(y) - \exp \psi(y)| \leq C |\Phi^N(y) - \psi(y)|,$$

where $C = \exp M$. This gives the uniform convergence of w^N to v as $N \rightarrow \infty$. □

Our last result shows that the LS approximations converge uniformly in the scalar case. This is due to the fact that in the scalar case the equalities (14) are always fulfilled.

Corollary 1. *Let the target function v be continuous on $\Omega = [A, B]$ ($A > B > 0$) and $\psi = \log v$. Assume that the sequence of partitions $\{\Omega_i^N\}_{i=1}^N$ ($N = 1, 2, \dots$) of Ω has the property $\delta^N \rightarrow 0$ ($N \rightarrow \infty$).*

Then for the corresponding LS power approximations V^N and v^N we have $V^N \rightarrow v$, $v^N \rightarrow v$ ($N \rightarrow \infty$) uniformly on Ω .

The proof of the theorem follows directly from the previous theorem and the following lemma, the proof of which is given in Appendix A.3.

Lemma 3. *Let a linear function $l: [a, b] \rightarrow \mathbb{R}$ be the LS approximation of a C^1 function $\psi: [a, b] \rightarrow \mathbb{R}$ on the entire interval $[a, b]$. Then there exist $y_0 \in [a, b]$ and $c \in (a, b)$ such that $l(y_0) = \psi(y_0)$, $l'(c) = \psi'(c)$.*

5. Discussion and Conclusions

Piecewise power-law representations may be very useful as practical approximations to target functions which are defined analytically or numerically. However, a strict mathematical justification of these approximations is not always paid attention to. Unfortunately, such an analysis is not always straightforward, especially if one puts additional a priori assumptions on the approximations, which is quite common in many applications.

We showed in the present paper that under additional assumptions power approximations do converge to the target function. We studied least-squares and uniform convergence, both of which are widely used (explicitly or implicitly) in applications.

Our analysis dealt with two types of regression: linear regression in the logarithmic space and power-law regression in the Cartesian space. The first procedure has all the advantages of the linear regression, but the transformation back to the Cartesian space distorts the error structure of the problem; the least squares error for the resulting piecewise power-law fitting is in general less accurate than the corresponding error for a power-law regression of the original data. As a partial remedy, it may be advantageous to apply power-law regression to the original data over each of the partition subsets back in the Cartesian space. Yet, being nonlinear regression this procedure is essentially ill-posed. Thus, both kinds of regression have their strong and weak sides, so that the choice between them must be undertaken by modeling consideration.

In many cases, it may also be advantageous to use the classical linear regression in combination with optimal partitions of the operating domain. In the logarithmic space this procedure is again linear and can be automated, but this may also cause several technical problems when proving the convergence of the corresponding approximations.

In the present paper, we offered a partial mathematical justification of the analysis based on piecewise power approximations, stemming from both kinds of regression, by verifying their convergence in the mean-square (L^2) and uniform sense. Uniform convergence is e.g. important if target functions are included in differential equations, as it is the uniform, and not L^2 -convergence, which is inherited by the solutions of the equations. However, a comprehensive analysis of convergence of solutions of differential equations, approximated by piecewise S-systems, is beyond the scope of this paper and will be discussed in a separate publication.

Acknowledgements

The work of the first author was partially supported by a EEA grant coordinated by Universidad Complutense de Madrid, Spain, and by the grant #239070 of the Norwegian Research Council.

References

- [1] de Jong, H. (2002) Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*, **9**, 67-104. <http://dx.doi.org/10.1089/10665270252833208>
- [2] Kascr, H. and Burns, J. (1973) The Control of Flux. *Symposia of the Society for Experimental Biology*, **27**, 65-104.
- [3] Savageau, M.A. (1969) Biochemical Systems Analysis. I. Some Mathematical Properties of the Rate Law for the Component Enzymatic Reactions. *Journal of Theoretical Biology*, **25**, 365-369. [http://dx.doi.org/10.1016/S0022-5193\(69\)80026-3](http://dx.doi.org/10.1016/S0022-5193(69)80026-3)
- [4] Savageau, M.A. (1969) Biochemical Systems Analysis. II. The Steady-State Solutions for an n-Pool System Using a Power-Law Approximation. *Journal of Theoretical Biology*, **25**, 370-379. [http://dx.doi.org/10.1016/S0022-5193\(69\)80027-5](http://dx.doi.org/10.1016/S0022-5193(69)80027-5)
- [5] Savageau, M.A. (1970) Biochemical Systems Analysis. III. Dynamic Solutions Using a Power-Law Approximation. *Journal of Theoretical Biology*, **26**, 215-226. [http://dx.doi.org/10.1016/S0022-5193\(70\)80013-3](http://dx.doi.org/10.1016/S0022-5193(70)80013-3)
- [6] Voit, E.O., *et al.* (1991) Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity. Van Nostrand Reinhold, NY.
- [7] Voit, E.O. (2000) Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists. Cambridge University Press, New York.
- [8] Alvarez-Vasquez, F., Sims, K.J., Cowart, L.A., Okamoto, Y., Voit, E.O. and Hannun, Y.A. (2005) Simulation and Validation of Modeled Sphingolipid Metabolism in *Saccharomyces cerevisiae*. *Nature*, **433**, 425-430. <http://dx.doi.org/10.1038/nature03232>
- [9] Atkinson, M.R., Savageau, M.A., Myers, J.T. and Ninfa, A.J. (2003) Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in *Escherichia coli*. *Cell*, **113**, 597-607. [http://dx.doi.org/10.1016/S0092-8674\(03\)00346-5](http://dx.doi.org/10.1016/S0092-8674(03)00346-5)
- [10] Vera, J., Rath, O., Balsa-Canto, E., Banga, J.R., Kolch, W. and Wolkenhauer, O. (2010) Investigating Dynamics of Inhibitory and Feedback Loops in ERK Signalling Using Power-Law Models. *Molecular BioSystems*, **6**, 2174-2191. <http://dx.doi.org/10.1039/c0mb00018c>
- [11] Savageau, M.A. and Voit, E. (1987) Recasting Nonlinear Differential Equations as S-Systems: A Canonical Nonlinear

- Form. *Mathematical Biosciences*, **87**, 83-115. [http://dx.doi.org/10.1016/0025-5564\(87\)90035-6](http://dx.doi.org/10.1016/0025-5564(87)90035-6)
- [12] Hernandez-Bermejo, B., Fairen, V. and Sorribas, A. (2000) Power-Law Modeling Based on Least-Squares Criteria: Consequences for System Analysis and Simulation. *Mathematical Biosciences*, **167**, 87-107. [http://dx.doi.org/10.1016/S0025-5564\(00\)00039-0](http://dx.doi.org/10.1016/S0025-5564(00)00039-0)
- [13] Sorribas, A., Hernández-Bermejo, B., Vilaprinyo, E. and Alves, R. (2007) Cooperativity and Saturation in Biochemical Networks: A Saturable Formalism Using Taylor Series Approximations. *Biotechnology and Bioengineering*, **97**, 1259-1277. <http://dx.doi.org/10.1002/bit.21316>
- [14] Tafinsteva, V., Machina, A. and Ponosov, A. (2013) Polynomial Representations of Piecewise-Linear Differential Equations Arising from Gene Regulatory Networks. *Nonlinear Analysis: Real World Applications*, **14**, 1732-1754. <http://dx.doi.org/10.1016/j.nonrwa.2012.11.008>
- [15] Chou, I.C. and Voit, E.O. (2009) Recent Developments in Parameter Estimation and Structure Identification of Biochemical and Genomic Systems. *Mathematical Biosciences*, **219**, 57-83. <http://dx.doi.org/10.1016/j.mbs.2009.03.002>
- [16] Savageau, M.A. (2002) Alternative Designs for a Genetic Switch: Analysis of Switching Times Using the Piecewise Power-Law Representation. *Mathematical Biosciences*, **180**, 237-253. [http://dx.doi.org/10.1016/S0025-5564\(02\)00113-X](http://dx.doi.org/10.1016/S0025-5564(02)00113-X)
- [17] Coelho, P.M.B.M., Salvador, A. and Savageau, M.A. (2010) Relating Mutant Genotype to Phenotype via Quantitative Behavior of the NADPH Redox Cycle in Human Erythrocytes. *PLoS ONE*, **5**, e13031. <http://dx.doi.org/10.1371/journal.pone.0013031>
- [18] Machina, A., Ponosov, A. and Voit, E.O. (2010) Automated Piecewise Power-Law Modeling of Biological Systems. *Journal of Biotechnology*, **149**, 154-165. <http://dx.doi.org/10.1016/j.jbiotec.2009.12.016>
- [19] Ferrari-Trecate, G. and Muselli, M. (2002) A New Learning Method for Piecewise Linear Regression. *Lecture Notes in Computer Science*, **2415**, 444-449. http://dx.doi.org/10.1007/3-540-46084-5_72

Appendix

A.1. Proof of Lemma 1, Section 1

Let us prove the lemma for the domain Δ . Notice that the function $\psi = \log v$ is measurable on the compact set Δ and satisfies the estimate $\log m \leq \psi(y) \leq \log M$ on Δ . By Lusin's theorem, for any N , there is a uniformly continuous function φ_N on Δ such that $\log m \leq \varphi_N \leq \log M$ for all N and the Lebesgue measure of the set $y \in \Delta : \varphi_N(y) \neq \psi(y)$ is less than $1/N$. Now, there exists a $\delta > 0$ for which $|y - y'| < \delta$, $y, y' \in \Delta$ implies $|\varphi_N(y) - \varphi_N(y')| < 1/N$. We define $\theta^N(y) = \varphi_N(y_i^N)$ for some $y_i^N \in \Delta_i^N$.

Let k_N be chosen in such a way that $\delta^{k_N} < \delta$. If $y \in \Delta_i^N$, then we have

$$|\varphi_N(y) - \theta^N(y)| = |\varphi_N(y) - \varphi_N(y_i^N)| < 1/N,$$

as $y, y_i^N \in \Delta_i^N$ and $\text{diam} \Delta_i^N < \delta^{k_N} < \delta$. Therefore, the Lebesgue measure of the set $y \in \Delta : |\theta^N(y) - \psi(y)| > 1/N$ is less than $1/N$, so that the sequence $\{\theta^N\}$ converges to ψ in measure, and due to its uniform boundedness, also in the L^2 -sense on Δ .

A similar argument applies to the sequence $\{\Theta^N\}$ on Ω , where we use the sequence $\{\exp(\theta^N)\}$ instead of $\{\theta^N\}$.

A.2. Proof of Lemma 2, Section 1

Clearly, $\psi(y)$ is measurable and bounded on Δ . Let $c = \sup |\psi(y)|$.

Let us fix a partition subset Δ_i^N . Our aim now is to find estimates for the norms of orthonormal basis functions $\{\tilde{e}_k\}$, $k = 0, \dots, n$ in the linear subspace of the space $L^2(\Delta_i^N)$ consisting of all linear functions and equipped with the scalar product

$$\xi \cdot \eta = \frac{1}{\text{mes} \Delta_i^N} \int_{\Delta_i^N} \xi(y) \eta(y) dy.$$

One basis is given by the set (11). However, this set is not necessarily orthogonal.

First of all, we choose $\tilde{e}_0 = 1$ and observe that its norm is equal to 1. Using the description (11) of the basis functions $e_i^N(k, y)$ defined via the center of mass we directly deduce from (12) that $\tilde{e}_0 = 1$ is orthogonal to any linear combination of the other basis functions. The challenge is therefore to estimate the norms of linear combinations $\tilde{e}_k = \sum_{l=1}^n \alpha_l e_i^N(l, y)$, where α_l are real numbers.

In the proof below we often omit one of the variables in $e_i^N(l, y)$, that is either l , or y , depending on a particular interpretation of this basis. Writing $e_i^N(y)$ means that we regard it as a vector for each particular y , i.e. $e_i^N(y) = (e_i^N(1, y), \dots, e_i^N(n, y))$ (the component $e_i^N(0, y)$ is excluded in further considerations). Omitting y ($e_i^N(l)$) means that we treat $e_i^N(l, y)$ as a function of y for a given l , i.e. as an element of the space $L^2(\Delta_i^N)$.

As $\|\tilde{e}_k\| = 1$, we require the following constraints on the coefficients:

$$1 = \frac{1}{\text{mes} \Delta_i^N} \int_{\Delta_i^N} |\tilde{e}_k(y)|^2 dy = \sum_{k, l=1}^n \frac{1}{\text{mes} \Delta_i^N} \int_{\Delta_i^N} e_i^N(k, y) e_i^N(l, y) \alpha_l \alpha_k dy = \alpha^T A_i^N \alpha,$$

where $\alpha = (\alpha_l)_{l=1}^n$. Therefore,

$$\max_{y \in \Delta_i^N} |\tilde{e}_k(y)| = \max_{y \in \Delta_i^N} |\alpha \cdot e_i^N(y)| \tag{16}$$

(where $|\cdot|$ is the Euclidean norm in \mathbb{R}^n and $a \cdot b$ is the scalar product of two vectors) with the constraint $\alpha^T A_i^N \alpha = 1$.

Diagonalization of the symmetric, positive definite matrix A_i^N with the help of an orthogonal matrix Q gives the matrix containing the eigenvalues $\lambda_k > 0$ of A_i^N on the diagonal. Putting $\beta = Q\alpha$ and using $|Qy| = |y|$, we obtain from (16) that

$$\max_{y \in \Delta_i^N} |\tilde{e}_k(y)| = \max_{y \in \Delta_i^N} |\beta \cdot Q e_i^N(y)| \leq |\beta| \max_{y \in \Delta_i^N} |e_i^N(y)| \leq \text{diam} \Delta_i^N \left(\sum_{k=1}^n \beta_k^2 \right)^{\frac{1}{2}}$$

with the constraint $\sum_{k=1}^n \lambda_k \beta_k^2 = 1$, where the constant $\text{diam} \Delta_i^N$ is evidently an upper estimate for the functions (11) on the partition subset Δ_i^N . The maximum value of the expression $\left(\sum_{k=1}^n \beta_k^2\right)^{\frac{1}{2}}$ under the above constraint is $\frac{1}{\sqrt{\lambda_i^N}}$, where λ_i^N is the minimal eigenvalue of the matrix A_i^N . Due to the condition (Δ) we get that $\max_{y \in \Delta_i^N} |e_i^N(k, y)| \leq c_0$, where the constant c_0 does not depend on i and N .

The final step in the proof of the lemma uses the explicit representation of the LS approximation Ψ^N :

$$\Psi^N(y) = \sum_{k=0}^n \mu_k \tilde{e}_k(y) \quad (y \in \Delta_i^N),$$

where

$$\mu_k = \frac{1}{\text{mes} \Delta_i^N} \int_{\Delta_i^N} \psi(y) \tilde{e}_k(y) dy.$$

Therefore $|\mu_0| \leq c$, $|\mu_k| \leq cc_0$ ($k = 1, \dots, n$) and

$$|\Psi^N(y)| \leq n(1+c)c_0^2 = C_2 \sup_{y \in \Delta} |\psi(y)|.$$

This implies also the uniform boundedness of the approximations $V_i^N(x)$ on Ω . The proof of the lemma is complete.

A.3. Proof of Lemma 3, Section 4

Let us first prove the existence of y_0 . Assume the converse, *i.e.* that $\theta(y) \neq l(y)$ for all $y \in [a, b]$. Let for instance $\theta(y) - l(y) > 0$ for all $y \in [a, b]$. Put $\varepsilon = \min_{y \in [a, b]} |\theta(y) - l(y)| > 0$. Then the linear function $l_1(y) = l(y) + \varepsilon/2$ satisfies the estimates $l(y) < l_1(y) < \theta(y)$. Therefore

$$\begin{aligned} \int_a^b (\theta(y) - l(y))^2 dy &= \int_a^b (\theta(y) - l_1(y))^2 dy + 2 \int_a^b (\theta(y) - l_1(y)) \frac{\varepsilon}{2} dy + \int_a^b \frac{\varepsilon^2}{4} dy \\ &> \int_a^b (\theta(y) - l_1(y))^2 dy. \end{aligned}$$

This, however, contradicts the definition of the least squares approximations. The case $\theta(y) - l(y) < 0$ is treated in a similar manner.

Assume now that $l'(y) \neq \theta'(y)$ for all $y \in [a, b]$. We shall prove that in this case the graph of the scalar linear function $l(y)$ intersects the graph of $\theta(y)$ in at least two points from the interval $[a, b]$.

From the first part of the proof we know that at least one intersection point does exist. Assume that there is exactly one point $d \in [a, b]$ such that $l(d) = \theta(d)$. Without loss of generality we may assume that $\Theta'(d) > 0$ where $\Theta(y) = \theta(y) - l(y)$. Since $\Theta(d) = 0$ and $\Theta(y) \neq 0$ for all $y \neq d$, we obtain that $\Theta(y) = \theta(y) - l(y) < 0$ for $y < d$ and $\Theta(y) = \theta(y) - l(y) > 0$ for $y > d$ (one of these sets may be empty). Consider a new linear approximation given by $l_1(y) = l(y) + \delta(y - d)$, where a sufficiently small $\delta > 0$ is chosen in such a way that the graphs of the functions $\theta(y)$ and $l_1(y)$ have still one intersection point in $[a, b]$ (namely, d by construction).

It is easy to see that such a δ does exist. Indeed, in a vicinity U of the point d we have that $\Theta'(y) \neq 0$, so that for small $\delta > 0$ we have $\theta'(y) > l_1'(y)$, $y \in U$ and hence d is the only intersection point of the graphs of the functions $\theta(y)$ and $l_1(y)$ in U . Outside U , *i.e.* inside the compact set $[a, b] \setminus U$ the continuous function Θ is non-zero, so that $m = \min_{y \in [a, b]} \Theta(y) > 0$. Choosing $\delta > 0$ in such a way that $\max_{y \in [a, b]} |l(y) - l_1(y)| < m$ guarantees that the graphs of the functions θ and l_1 meet only in d .

We complete now our analysis of the scalar case observing that for such δ

$$\int_a^b (\theta(y) - l(y))^2 dy > \int_a^b (\theta(y) - l_1(y))^2 dy,$$

simply because the graph of l_1 is closer to the graph of θ , than the graph of l . This contradicts the assumption that l is the LS approximation of θ . We have therefore proved that there exists $c \in [a, b]$ such that $\Theta'(c) = 0$.

A.4. Piecewise Power-Law Regression

In this subsection we describe a numerically stable way to find a best possible, in some sense, piecewise power approximation to an arbitrary target function. The method was suggested in [18] and was based on the piecewise linear regression from [19].

Let $v(x)$, $x \in \Omega$ be a target function (e.g. the in- and efflux functions in (2), possible only as a data set obtained from some measurements. The task is to find a set of power functions which approximate the target function in a “best possible” way given a number N of the partition sets. The problem is essentially non-convex being therefore not well-posed numerically. However, using the logarithmic space Δ one can convert this problem into a linear one. Performing an automated piecewise linear regression based on Artificial Neural Networks (ANN) introduced in [19], this algorithm returns an optimal polyhedral partition of the domain in logarithmic space and the corresponding optimal set of linear functions which are the best approximations to the image of the target function in the logarithmic space. Returning to the Cartesian space produces piecewise power functions that are not necessarily the best possible approximations in the sense of the metric in Ω , but this approximation procedure is numerically stable.

As before, we assume $\psi(y)$ to be the image of $v(x)$ under the logarithmic transformation, *i.e.* $\psi(y) = \log v(x)$, $y = \log x$. Let also N be a fixed number of partition subsets of a polyhedral set Δ . The task is to construct numerically the function $\Psi^N(y)$, $y \in \Delta$, which is piecewise linear and which is the best possible LS approximation to $\psi(y)$. The inverse logarithmic transformation of $\psi(y)$ will then give a piecewise power approximation V^N of the target function v , which is best possible with respect to the logarithmic distance. To simplify the notation we will below write $\Psi = \Psi^N$, thus removing the index N .

A polyhedral partition satisfies the following assumptions: $\Delta_i \cap \Delta_l = \emptyset$ for every $i, l = 1, \dots, N$, $i \neq l$, $\bigcup_{i=1}^N \Delta_i = \Delta$, and

$$\Delta_i = \left\{ y \in \Delta : d_{0j}^i + \sum_{k=1}^n d_{kj}^i y_k < 0 \right\} \quad (i = 1, \dots, N). \tag{17}$$

This partition gives rise to a partition of the original domain Ω . Applying the inverse logarithmic transformation, we obtain

$$\Omega^k = \left\{ x \in \Omega : d_{0j}^i + \sum_{k=1}^n d_{kj}^i \log x_k < 0 \right\} \quad (i = 1, \dots, N). \tag{18}$$

The piecewise linear function $\Psi(y)$ is assumed to have the following representation:

$$\Psi(y) = \Psi_i(y) = \alpha_0^i + \sum_{j=1}^n \alpha_j^i y_j, \quad y \in \Delta_i, \tag{19}$$

where Δ_i are all polyhedral sets defined by (17) and defining a partition of the logarithmic domain Δ .

Scalar weights $\alpha_0^i, \alpha_1^i, \dots, \alpha_n^i$, for $i = 1, \dots, N$ and the numbers d_{kj}^i for $j = 1, \dots, l^i$, $k = 0, 1, \dots, n$, $i = 1, \dots, N$ uniquely characterize the function ψ and the corresponding partition $\{\Delta_i\}$. Below the weights are collected in a vector α^i for every $i = 1, \dots, N$.

The piecewise linear algorithm has two targets.

Target 1: The weights vector α^i ($i = 1, \dots, N$), which should be reconstructed as soon as the partition is known.

Target 2: Scalars d_{kj}^i ($j = 1, \dots, l^i$, $k = 0, 1, \dots, n$) which should be estimated for every $i = 1, \dots, N$.

The aim of the piecewise linear regression: given a function $\psi(y_1, \dots, y_n)$ represented via a data set $\psi(y_1^k, \dots, y_n^k) = \psi_k$, a number of partition subsets N and a natural number c find a piecewise linear function Ψ and the polyhedral partition $\{\Delta_i\}_{i=1}^N$ such that

$$\int_{\Delta} [\Psi(y_1, \dots, y_n) - \psi(y_1, \dots, y_n)]^2 dy_1 \dots dy_n = \sum_{i=1}^N \int_{\Delta_i} [\Psi_i(y_1, \dots, y_n) - \psi(y_1, \dots, y_n)]^2 dy_1 \dots dy_n \rightarrow \min,$$

where the minimum is taken over all polyhedral partitions Δ_i and all linear functions Ψ_i ($i = 1, \dots, N$). The parameter c is the number of nearest neighboring points required by the implementation of the method. The neighboring points are used to aggregate approximations into clusters.

The algorithm consists of the following steps.

Step 1. Local regression: Define the set C_k containing the k th point (y_k, ψ_k) and the samples associated with $c-1$ nearest neighbors y to y_k and perform linear regression to obtain the weights u^k fitting the samples in C_k .

Step 2. Clustering: Perform a clustering process to subdivide the set of weight vectors u^k into N groups with similar features.

Step 3. Classification: Apply a classification algorithm based on a pattern-recognition method to produce the coefficients d_{kj}^k describing the partition subsets Δ_i .

Step 4. Regression: For every $i = 1, \dots, N$ perform linear regression on the samples (y, ψ) with $y \in \Delta_i$ to obtain the weights α^i for the i th approximation.

The inverse logarithmic transformation of Ω results in a piecewise power approximation of the function $v(x)$ and a partition $\{\Omega_i\}_{i=1}^N$ defined by (18).

A modification of this algorithm, which is suggested in [18], assumes a power-law regression to the original data over each of the N partition subsets of the optimal partition $\{\Omega_k\}$, so that the increase in difficulty is modest even though the regression is now nonlinear. The partitioning is found from the first three steps of the above algorithm and the new algorithm proceeds as follows:

Step 4A. Power-law regression: For every $i = 1, \dots, N$ perform power-law regression on the samples (x, v) with $x \in \Omega^k$ to obtain the power functions for each partition subset given by (18).

The algorithm is implemented in a free MatLab toolbox Hybrid Identification Toolbox (HIT).

A.5. Ill-Posedness of LS Power-Law Fitting

In this section we show that the nonlinear power-law regression, *i.e.* the one based on the minimization criterion (9), is ill-posed. This ill-posedness is caused by the fact that the set $\{\gamma x_1^{f_1} \dots x_n^{f_n}\}$ from (9) is a non-convex set.

Let us assume that $v(x)$ is only known with a certain accuracy, as it is often the case. Mathematically, we will describe this situation by letting v depend on a (small) parameter λ , *i.e.* $v = v_\lambda(x)$ (and so becomes the function $\psi = \psi_\lambda(y)$ as well). But it turns out, as we will show below, that for certain values of λ small perturbations may cause a “jump” in the corresponding power-law representation, *i.e.* while functions $v_\lambda(x)$ remain close to each other, the least-squares minimization criterion (9) may produce the power-law representations that are very different.

Below we illustrate this fact analytically using a specific example. For the sake of simplicity, let us consider a target function $v(x)$ of one variable, so that

$$\int_L^U (v(x) - \gamma x^f)^2 dx \rightarrow \min. \tag{20}$$

If $v = v_\lambda(x)$, then γ and f that minimize (20) are also functions of λ .

We first consider a simpler problem assuming that $\gamma = \text{const}$, so that after rescaling we may assume that $\gamma = 1$ and rewrite (20) as

$$\int_L^U (v_\lambda(x) - x^{f_\lambda})^2 dx \rightarrow \min$$

After applying the substitution $y = \log x$, $a = \log U$, $b = \log L$, we obtain

$$\int_a^b (v_\lambda(e^y) - e^{yf_\lambda})^2 e^y dy \rightarrow \min$$

or

$$\int_a^b (H_\lambda(y) - e^{\beta_\lambda y})^2 dy \rightarrow \min, \tag{21}$$

where $H_\lambda(y) = v_\lambda(e^y)e^{y/2}$, $\beta_\lambda = f_\lambda + 1/2$.

Now, let $a = -1, b = 1$ and consider the function $H_\lambda(y) = 3\lambda y + 6y^2 + 1$ and its LS power approximation $e^{\beta(\lambda)y}$.

It is easily seen that the projection function $e^{\beta(\lambda)y}$ is discontinuous in λ (see Figure A1), while Figure A2 gives a graphical representation of this discontinuity in λ for one value of y .

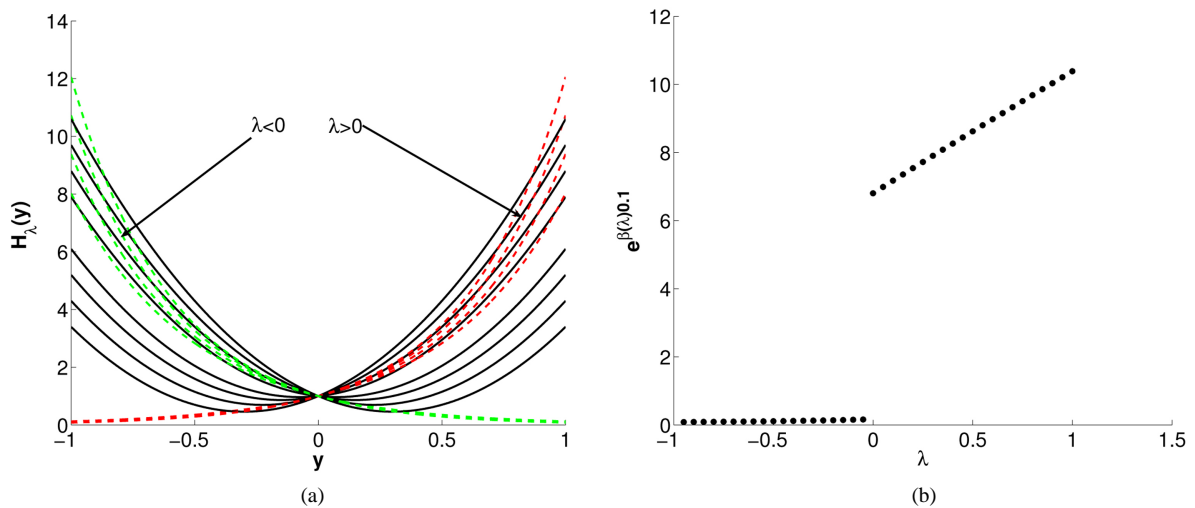


Figure A1. (a) The continuous lines represent the functions $H_\lambda(y)$ for different values of λ and the dotted lines give its LS approximations within the operating interval $[0.5, 1.5]$. The blue and green colors correspond to the values $\lambda > \lambda_0$, while the red and black colors describe the case of $\lambda < \lambda_0$. (b) This graph explains how the LS power approximations at $y = 1$, i.e. $e^{\beta(\lambda)+c(\lambda)}$, depend on λ . We see that $\beta(\lambda)+c(\lambda)$ is discontinuous at $\lambda_0 \approx -2.365$.

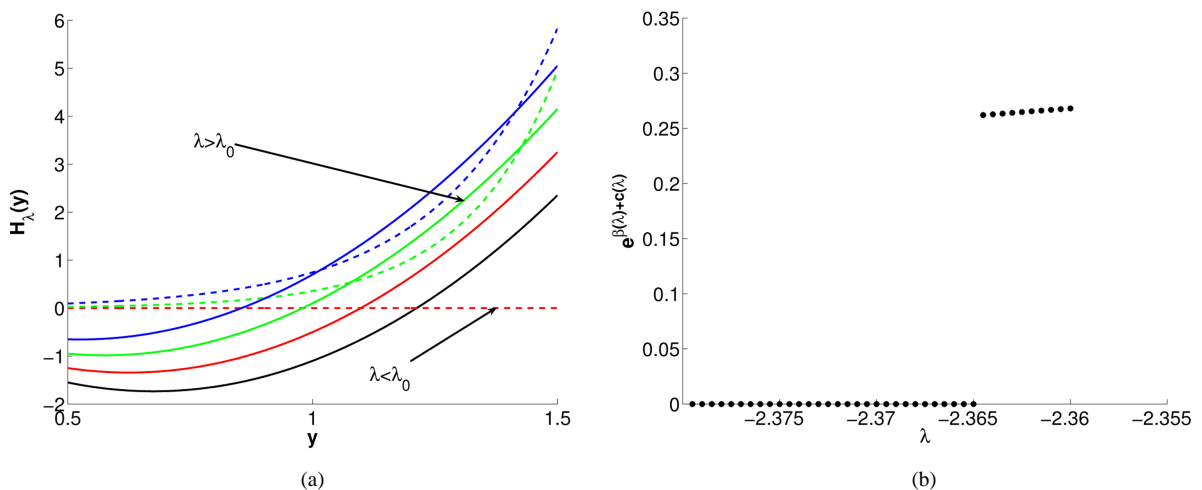


Figure A2. (a) The continuous lines represent the functions $H_\lambda(y)$ for different values of λ and the dotted lines give its LS approximations within the operating interval $[0.5, 1.5]$. The blue and green colors correspond to the values $\lambda > \lambda_0$, while the red and black colors describe the case of $\lambda < \lambda_0$. (b) This graph explains how the LS power approximations at $y = 1$, i.e. $e^{\beta(\lambda)+c(\lambda)}$, depend on λ . We see that $\beta(\lambda)+c(\lambda)$ is discontinuous at $\lambda_0 \approx -2.365$.

Going back to the variable x , the above function becomes $v_\lambda(x) = \frac{3\lambda \log x + 6\log^2 x + 1}{\sqrt{x}}$, $A = 1/e$, $B = e$, the discontinuity in λ being preserved.

This example shows that the criterion (9) may produce a LS power approximation that is not stable under small perturbations of the parameter λ and by this under small perturbations of the target function, which causes ill-posedness of the minimization problem. We stress also that this effect is generic, *i.e.* independent of the number of the involved parameters, as the comparison of **Figure A1(a)** (resp. **Figure A1(b)**) and **Figure A2(a)** (resp. **Figure A2(b)**) clearly demonstrates.



Scientific Research Publishing

Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>