A TWO-DIMENSIONAL POISSON EQUATION FORMULATION OF NON-PARAMETRIC STATISTICAL NON-LINEAR MODELING*

SIMONE FIORI†

Abstract. The present paper deals with a Poisson equation arising in statistical modeling of semi-deterministic non-linear systems with two independent (input) variables and one dependent (output) variable. Statistical modeling is formulated in terms of a differential equation that relates the second-order joint probability density functions of the model's input/output random variables with the sought nonlinear model transference. The discussed modeling procedure makes no prior assumptions on the functional structure of the model, except for monotonicity and continuity with respect to both input variables. In particular, the method is non-parametric. Results of numerical tests are presented and discussed in order to get an insight into the behavior of the devised statistical modeling procedure. The results of numerical tests confirm that the proposed statistical modeling approach is able to cope with both synthetic and real-world data sets and, in particular, with underlying systems and data that exhibit strong hidden nuisance variables and measurement disturbances.
Poisson equation; Statistical modeling; Nonlinear system; Nonparametric modeling; Isotonic modeling.

1. Introduction. Real-world phenomena may rarely be described accurately by a mathematical model to be evaluated analytically. Statistical modeling provides a useful tool to build up a model of a phenomenon under observation, on the basis of the statistical features of the variables describing such phenomenon. Statistical modeling finds applications in as diverse fields as social and behavioral sciences [2], biomedical research [8], computer vision and content-based image retrieval [14, 15] and ecomometrics [20]. As intended here, statistical modeling is based on four main assumptions: 1) The physical phenomenon under observation relates two independent variables with a single dependent variable. Namely, it is assumed that three variables of interest in a modeling problem are related by $y = \Phi(x_1, x_2)$, where $y \in \mathcal{Y}$ represents the *dependent variate* and $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \subset \mathbb{R}^2$ represent the *independent variates*. The function $\Phi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ is semi-deterministic as it includes nuisance variables that cannot be directly measured. The joint statistical features of the input/output variates are described by the joint probability density function $p_{y,x_1,x_2}(y, x_1, x_2)$, where, by a slight abuse of notation, the variates have been confused with their realizations. 2) The sought model is monotonically increasing or decreasing with respect to both input variables (or, equivalently, it is of *dose-response* type). The hypothesis of monotonicity in data modeling occurs frequently in applied fields such as data regression and data mining [18]. 3) Statistical modeling is based on the estimation of second-order joint probability density functions: It is assumed

*S. FIORI, "A TWO-DIMENSIONAL POISSON EQUATION FORMULATION OF NON-PARAMETRIC STATISTICAL NON-LINEAR MODELING", *COMPUTERS AND MATHEMATICS WITH APPLICATIONS* (ELSEVIER), VOL. 67, NO. 5, PP. 1171 – 1185, MARCH 2014.

†Dipartimento di Ingegneria dell'Informazione (DII), Facoltà di Ingegneria, Università Politecnica delle Marche, Via Brecce Bianche, Ancona I-60131, Italy, E-mail: s.fiori@univpm.it.

that the number of available observations of the triples (x_1, x_2, y) is large enough to get meaningful statistical estimates. 4) The model is non-linear, non-parametric, namely, there is no assumption on its shape (except for the assumptions of monotonicity and continuity), hence, the devised modeling technique may cope with arbitrary dependencies, albeit restricted to be monotonic.

Non-parametric modeling is a form of analysis in which the predictor does not take a predetermined form but is constructed according to information derived from the data. Nonparametric modeling requires larger data sets than modeling based on parametric models because the data must supply the model structure as well as the model estimates.

The isotonic modeling problem is a special case of the general modeling problem that arises in various fields, such as production planning, inventory control and psychometry [6]. Isotonic modeling occurs whenever it is known that the dependency between the output variable and the input variables is either monotonically increasing or monotonically decreasing. Typically, isotonic modeling in a single-input/single-output setting is formulated as a constrained quadratic programming problem. The notion of isotonic modeling may be traced back to the seminal contribution [4]. The formulation in terms of least-squares may be generalized to L_1 norm [1] and to L_∞ norm (which gives rise to the notion of strict isotonic modeling) [19]. Note that isotonic modeling is often applied to nonmonotonic data, which justifies the assumption on the underlying system to be semi-deterministic. An example is given in [12]: Consider the problem of measuring the viscosity of a fluid at different temperatures. Viscosity is a nonincreasing function of temperature; however, *due to measurement error*, the observed viscosity may not be nonincreasing when ordered by temperature. In this case, the purpose of isotonic modeling is to replace the observed viscosities with a set of values that are nonincreasing when ordered by the temperature.

The present research work aims at extending the previous contribution on one-to-one statistical modeling summarized in the conference paper [10] and explained in details in the publication [11]. A feature of the present work that is inherited from the above contribution is that the inferred model is non-parametric. An example of the importance of non-parametric isotonic modeling is offered in [19]: Researchers are less willing to impose strong assumptions in their modeling. For instance, applied researchers may be willing to make the weak assumption that the expected height of a woman is an increasing function of the height of her father and of her mother, but be unwilling to make parametric assumptions such as linearity.

Statistical isotonic modeling is about determining a relationship between the two independent variates $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ and the dependent variate $y \in \mathcal{Y}$, described by the model:

$$y = f(x_1, x_2), \tag{1.1}$$

with $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ denoting a nonlinear deterministic model, by making use

of pooled information only, summarized by the joint probability distributions of the variates y, x_1, x_2. The statistical modeling method proposed in the present paper is based on known formulas to quantify the distortion of the statistical distribution of the input values operated by a deterministic monotonic nonlinear system.

PAPER ORGANIZATION: The present paper is organized as follows. Section 2 explains the fundamental principles behind the proposed bivariate isotonic statistical modeling approach. In particular, it shows that an appropriate choice of a two-input/two-output non-linear system to probe the sought two-input/one-output model allows formulating the statistical isotonic modeling problem in terms of conservation of probability measures in the model's second-order joint probability space. The result is a system of partial differential equations having the sought model as unknown. Section 3 of the present paper moves forward the formulation in terms of a two-dimensional Poisson equation derived on the basis of a variational principle applied to a functional least-squares-error formulation. In such section, the problem of setting up appropriate boundary conditions is also discussed on and the numerical implementation used to solve the Poisson equation is described briefly. Section 4 summarizes the quantities used for measuring objectively the features of the devised bivariate isotonic statistical modeling technique. Section 4 also illustrates and discusses the results of modeling synthetic data sets as well as real-world data sets, namely, a data set arising from robotic arm dynamics, a data set arising from a food toxicology research and a data set from quantitative palynology research. The results of modeling the synthetic as well as the real-world data sets are encouraging and show that the devised statistical modeling technique can cope with underlying systems and data that include strong hidden nuisance variables and measurement disturbances. Section 5 concludes the paper.

2. Statistical bivariate isotonic modelling: Fundamental principles. Monotonic dependencies are common in physical systems. For instance, the rates of biogeochemical processes can be monotonic functions of factors like temperature and humidity. For such systems, it is of prime importance to infer monotonic relationships from a given data set by constructing a model that is consistent with monotonicity, namely, that is isotonic. The majority of the challenging applied isotonic modeling problems are characterized by very large data sets. An example is offered in [1]: In the analysis of large-scale microarray data, which is one of the most important tools in biology, the same procedure is used for studying the fit of tens of thousands of genes to a given partial order. A further example is offered in [17], which concerns the classification of large portions of texts extracted from the World Wide Web.

2.1. System-theoretic grounds of the proposed statistical modeling method. Consider a non-linear system with two input variables and two output variables:

$$(y, z) = \varphi(x_1, x_2), \tag{2.1}$$

where $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2 \subset \mathbb{R}^2$, $(y, z) \in \mathcal{Y} \times \mathcal{Z} \subset \mathbb{R}^2$ and $\varphi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y} \times \mathcal{Z}$. The non-linear system (2.1) is supposed to be invertible in the domain of interest and its inverse is denoted by $\varphi^{-1} : \mathcal{Y} \times \mathcal{Z} \to \mathcal{X}_1 \times \mathcal{X}_2$.

Assume that x_1, x_2 are random variables with joint probability density function denoted by $p_{x_1,x_2}(x_1, x_2)$. Such random variables get transformed into two random variables y, z that are distributed according to the probability density function $p_{y,z}(y, z)$. Define the Jacobian of the system φ as:

$$J(x_1, x_2) \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} \\ \frac{\partial z}{\partial x_1} & \frac{\partial z}{\partial x_2} \end{bmatrix}. \tag{2.2}$$

The joint probability density function of the output variates is related to the joint probability density function of the input variates by the relationship:

$$p_{y,z}(y, z) = \left. \frac{p_{x_1,x_2}(x_1, x_2)}{|\det J(x_1, x_2)|} \right|_{(x_1,x_2)=\varphi^{-1}(y,z)}, \tag{2.3}$$

as illustrated in the Figure 2.1. Recall that the function $\det J$ has the following plain form:

$$\det J(x_1, x_2) = \frac{\partial y}{\partial x_1}\frac{\partial z}{\partial x_2} - \frac{\partial y}{\partial x_2}\frac{\partial z}{\partial x_1}. \tag{2.4}$$

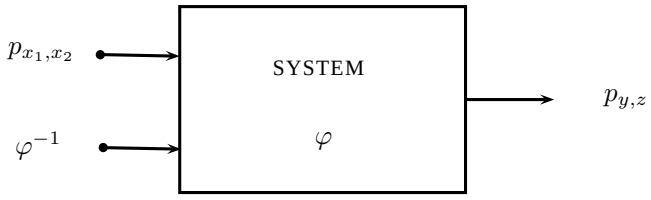


FIG. 2.1. *System-theoretic relationship between the joint probability density function of two variates x_1, x_2 and the joint probability density function of two variates $(y, z) = \varphi(x_1, x_2)$ induced by a deterministic invertible system φ.*

Since the non-linear function φ is invertible, by hypothesis, in the domain $\mathcal{X}_1 \times \mathcal{X}_2$, the matrix-function $J(x_1, x_2)$ is non-singular in $\mathcal{X}_1 \times \mathcal{X}_2$ and hence the scalar-function $\det J(x_1, x_2)$ is non-zero in $\mathcal{X}_1 \times \mathcal{X}_2$. The fundamental equation (2.3) may be rewritten in plain form according to the following two cases:

- **Positive Jacobian case:** If it holds that $\det J(x_1, x_2) > 0$, then the fundamental equation (2.3) may be rewritten in plain form as:

$$\left(\frac{\partial y}{\partial x_1}\frac{\partial z}{\partial x_2} - \frac{\partial y}{\partial x_2}\frac{\partial z}{\partial x_1}\right)(p_{y,z} \circ \varphi)(x_1, x_2) = p_{x_1,x_2}(x_1, x_2). \tag{2.5}$$

- **Negative Jacobian case:** If it holds that $\det J(x_1, x_2) < 0$, then the fundamental equation (2.3) may be rewritten in plain form as:

$$\left(\frac{\partial y}{\partial x_2}\frac{\partial z}{\partial x_1} - \frac{\partial y}{\partial x_1}\frac{\partial z}{\partial x_2}\right)(p_{y,z} \circ \varphi)(x_1, x_2) = p_{x_1,x_2}(x_1, x_2). \tag{2.6}$$

The main idea behind statistical modelling based on the above considerations is that, whenever the probability density functions p_{x_1,x_2} and $p_{y,z}$ are *known* and a non-linear model that links the four variates x_1, x_2, y, z as in equation (2.1) is sought, the equation (2.5) and the equation (2.6) provide (differential) constraints to be satisfied by the model.

In the case of interest in the present paper, the non-linear model (1.1) includes three variates x_1, x_2, y. In order to exploit the relationships between four variates, construct the non-linear system:

$$\varphi : \begin{cases} y = f(x_1, x_2), \\ z = ax_1 + bx_2, \end{cases} \tag{2.7}$$

with $a, b \in \mathbb{R}$ known, which apparently may be regarded as an extension of the non-linear regression model (1.1) where the variable z may be regarded as an auxiliary variable.

The det-Jacobian of the extended model (2.7) takes on values:

$$\det J(x_1, x_2) = b\frac{\partial f}{\partial x_1} - a\frac{\partial f}{\partial x_2}. \tag{2.8}$$

Assume that the model $y = f(x_1, x_2)$ is such that $\frac{\partial f}{\partial x_1} > 0$ and $\frac{\partial f}{\partial x_2} > 0$. (The following equations modify in a rather straightforward way if one or both partial derivatives possess a reverse sign.) Consider the following two cases:

- **Case $b = 1$ and $a = 0$:** With such choice of the constant parameters, the condition on the Jacobian reads $\frac{\partial f}{\partial x_1} > 0$, the auxiliary variable is such that $z = x_2$, and thus the equation (2.5) becomes:

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = \frac{p_{x_1,x_2}(x_1, x_2)}{p_{y,x_2}(f(x_1, x_2), x_2)}. \tag{2.9}$$

- **Case $b = 0$ and $a = 1$:** With such choice of the constant parameters, the condition on the Jacobian reads $-\frac{\partial f}{\partial x_2} < 0$, the auxiliary variable is such that $z = x_1$, and thus the equation (2.6) becomes:

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = \frac{p_{x_1,x_2}(x_1, x_2)}{p_{y,x_1}(f(x_1, x_2), x_1)}. \tag{2.10}$$

The above two conditions must be satisfied at the same time, therefore the non-linear 2-to-1 model f may be inferred by solving the system of two differential equations (2.9)-(2.10), equipped with suitable boundary conditions. The complete system of partial differential equations in the case of mixed monotonically increasing/decreasing behavior with respect to the model's independent variables reads:

$$\begin{cases} \frac{\partial f(x_1,x_2)}{\partial x_1} = \pm\frac{p_{x_1,x_2}(x_1,x_2)}{p_{y,x_2}(f(x_1,x_2),x_2)}, \\ \frac{\partial f(x_1,x_2)}{\partial x_2} = \pm\frac{p_{x_1,x_2}(x_1,x_2)}{p_{y,x_1}(f(x_1,x_2),x_1)}. \end{cases} \tag{2.11}$$

2.2. Discussion on the proposed technique and comparison with existing methods. The present statistical modeling method by a 2-to-1 non-linear model extends the 1-to-1 statistical modeling technique discussed in details in the previous contribution [11]. The previous method does not require the input/output variables to be necessarily paired and it allows estimating a model in presence of missing values in both variables' records. Such benefits are lost in the present formulation because pairing is necessary to estimate the joint probability density functions p_{x_1,x_2}, p_{y,x_1}, p_{y,x_2}.

The proposed method relies on the fundamental relationship (2.3), which holds only is the system transfer function φ is monotonically increasing (or decreasing) with respect to both input variables. If such assumption does not hold, the relationship between the joint probability density function of the input variables and the joint probability density function of the output variables does not allow recovering the non-linear model φ univocally.

Although extensions of the proposed statistical modeling method to more than four variables is possible, in principle, there exists a practical obstruction given by the intrinsic difficulty related with the estimation of the joint probability density functions of several variables. As long as histogram-based estimation is invoked (as is the case in the present contribution), an accurate estimate requires volumes of data that are hardly available, in practice. It appears clear that an extension to more than four variables would require a companion method for the estimation of joint probability density function that is non-parametric, computationally simple and that provides a sufficiently accurate estimate with the volume of data available in common applications.

The proposed statistical modeling technique exhibits peculiar features that distinguish it from existing modeling methods and algorithms.

An important feature of the proposed statistical modeling method is the absence of any assumption on the shape and on the functional structure of the model f (except for monotonicity and continuity). It is interesting to compare such feature with the ones of a well-known modeling technique based on artificial neural networks. Several neural network-based models are similar to well-known statistical techniques such as generalized linear models, polynomial models and projection pursuit models, while some neural network models, such as self-organizing maps, have no precise statistical equivalents. All the mentioned methods, however, require some assumptions on the model, such as number of hidden layers, structure of the artificial neurons and structure of the learning rule, which affect severely the result.

Likewise, the flexibility exhibited by the proposed modeling method compares favorably with the parametric method that consists in choosing a model from a *catalog* of predefined models, or as a combination of a set of models extracted from a *dictionary* of predefined models, on the basis of a fitting criterion. It should be noted, on the other hand, that fitting a predefined model does not rely on the assumption of monotonicity.

The low computational complexity of the proposed modeling technique is likewise a prominent feature. Once the joint second-order probability density functions of the data are estimated, the computational burden of the modeling technique depends essentially on the desired accuracy of the model: In fact, the computational burden essentially depends on the number of subdivisions of the domain of the x_1, x_2, y variables. Although the model is nonlinear, the mathematical operations that are necessary to construct the model are essentially sums and multiplications of real-valued numbers and the histogram-based approximation of the joint probability density functions of two variables are represented by two-dimensional tables, whose entries are essentially counters that get increased once a sample hits the corresponding 'bin' [11].

A further prominent feature of the devised method is robustness against measurement errors on the values of the involved independent/dependent variables. In fact, the devised modeling method does not treat explicitly the triples (x_1, x_2, y), as the devised modeling method method relies on the joint statistical features inferred by pooling the data in pairwise records (x_1, x_2), (y, x_1) and (y, x_2). Each joint second-order probability density function is estimated by subdividing the rectangles $\mathcal{X}_1 \times \mathcal{X}_2$, $\mathcal{Y} \times \mathcal{X}_1$ and $\mathcal{Y} \times \mathcal{X}_2$ in small rectangles and by counting how many pairs (x_1, x_2), (y, x_1) and (y, x_2) fall in each subdivision. If a pair is affected by measurement noise, for example, the actual pair (x_1, x_2) is measured as $(x_1 + \nu_1, x_2 + \nu_2)$, where ν_1 and ν_2 represent measurement errors, a modeling algorithm that insists directly on the data will be directly affected by the measurements errors, while the proposed method is unaffected, provided that the errors ν_1 and ν_2 are sufficiently small (or the extent of the subdivisions in the histogram-based probability density function estimation algorithm is sufficiently large) so that the pair $(x_1 + \nu_1, x_2 + \nu_2)$ will fall in the same bin as the actual pair (x_1, x_2).

3. Formulation in terms of a two-dimensional Poisson equation and numerical details. The system of partial differential equations (2.11) is an instance of the structured differential system:

$$\begin{cases} \frac{\partial y}{\partial x_1} = F_1(y, x_1, x_2) \stackrel{\text{def}}{=} \pm \frac{p_{x_1,x_2}(x_1,x_2)}{p_{y,x_2}(y,x_2)}, \\ \frac{\partial y}{\partial x_2} = F_2(y, x_1, x_2) \stackrel{\text{def}}{=} \pm \frac{p_{x_1,x_2}(x_1,x_2)}{p_{y,x_1}(y,x_1)}. \end{cases} \tag{3.1}$$

In what follows, it is assumed that the sets $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$ are bounded intervals.

3.1. Reformulation as a Poisson equation. A system of two first-order differential equations in one unknown may be reduced to a single second-order differential equation as follows. The problem represented by the differential system (3.1) may be reformulated as a functional least-squares problem by the help of the error functional:

$$\mathbb{E}(y) \stackrel{\text{def}}{=} \frac{1}{2} \int_{\mathcal{X}_1 \times \mathcal{X}_2} \left[\left(\frac{\partial y}{\partial x_1} - F_1(y, x_1, x_2) \right)^2 + \left(\frac{\partial y}{\partial x_2} - F_2(y, x_1, x_2) \right)^2 \right] \mathrm{d}x_1 \, \mathrm{d}x_2. \tag{3.2}$$

Any solution $y : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ of the differential system (3.1) is also a minimizer of the error functional $\mathbb{E}$. By the calculus of variation, it is known that a minimizer $y^\star$ of the error functional $\mathbb{E}$ satisfies the equation:

$$\left.\frac{\mathrm{d}\mathbb{E}(y^\star + \varepsilon\eta)}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0, \tag{3.3}$$

for every perturbation $\eta : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ such that $\eta(x_1, x_2) = 0$ for all $(x_1, x_2) \in \partial(\mathcal{X}_1 \times \mathcal{X}_2)$ (namely, for every perturbation that vanishes to zero at the boundary of the domain $\mathcal{X}_1 \times \mathcal{X}_2$). From the definition of the error functional $\mathbb{E}$, it follows that:

$$\mathbb{E}(y) = \frac{1}{2}\int_{\mathcal{X}_1\times\mathcal{X}_2} \left[\left(\frac{\partial y}{\partial x_1}\right)^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 - 2\frac{\partial y}{\partial x_1}F_1 - 2\frac{\partial y}{\partial x_2}F_2 + F_1^2 + F_2^2\right] \mathrm{d}x_1\,\mathrm{d}x_2. \tag{3.4}$$

Note that:

$$\begin{aligned} F_1(y + \varepsilon\eta, x_1, x_2) &= F_1(y, x_1, x_2) + \varepsilon\eta\frac{\partial F_1}{\partial y}(y, x_1, x_2) + o(\varepsilon), \\ F_2(y + \varepsilon\eta, x_1, x_2) &= F_2(y, x_1, x_2) + \varepsilon\eta\frac{\partial F_2}{\partial y}(y, x_1, x_2) + o(\varepsilon), \end{aligned}$$

where $\lim_{\varepsilon\to 0} o(\varepsilon)/\varepsilon = 0$. Then, the variational principle (3.3) leads to the equation:

$$\begin{aligned} \lim_{\varepsilon\to 0}\frac{\mathbb{E}(y+\varepsilon\eta) - \mathbb{E}(y)}{\varepsilon} = \int_{\mathcal{X}_1\times\mathcal{X}_2} \Bigg(&\frac{\partial y}{\partial x_1}\frac{\partial \eta}{\partial x_1} + \frac{\partial y}{\partial x_2}\frac{\partial \eta}{\partial x_2} - \frac{\partial \eta}{\partial x_1}F_1 - \frac{\partial y}{\partial x_1}\frac{\partial F_1}{\partial y}\eta \\ &- \frac{\partial \eta}{\partial x_2}F_2 - \frac{\partial y}{\partial x_2}\frac{\partial F_2}{\partial y}\eta + \frac{\partial F_1}{\partial y}F_1\eta + \frac{\partial F_2}{\partial y}F_2\eta\Bigg)\,\mathrm{d}x_1\,\mathrm{d}x_2 = 0, \end{aligned} \tag{3.5}$$

in the unknown $y : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$. By invoking the integration-by-parts method, one gets:

$$\int_{\mathcal{X}_1\times\mathcal{X}_2} \frac{\partial y}{\partial x_1}\frac{\partial \eta}{\partial x_1}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{\mathcal{X}_2} \left.\frac{\partial y}{\partial x_1}\eta\right|_{\partial\mathcal{X}_1} \mathrm{d}x_2 - \int_{\mathcal{X}_1\times\mathcal{X}_2} \frac{\partial^2 y}{\partial x_1^2}\eta\,\mathrm{d}x_1\,\mathrm{d}x_2, \tag{3.6}$$

$$\int_{\mathcal{X}_1\times\mathcal{X}_2} \frac{\partial y}{\partial x_2}\frac{\partial \eta}{\partial x_2}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{\mathcal{X}_1} \left.\frac{\partial y}{\partial x_2}\eta\right|_{\partial\mathcal{X}_2} \mathrm{d}x_1 - \int_{\mathcal{X}_1\times\mathcal{X}_2} \frac{\partial^2 y}{\partial x_2^2}\eta\,\mathrm{d}x_1\,\mathrm{d}x_2, \tag{3.7}$$

$$\begin{aligned} \int_{\mathcal{X}_1\times\mathcal{X}_2} \frac{\partial \eta}{\partial x_1}F_1\,\mathrm{d}x_1\,\mathrm{d}x_2 = &\int_{\mathcal{X}_2} F_1\eta|_{\partial\mathcal{X}_1}\,\mathrm{d}x_2 \\ &- \int_{\mathcal{X}_1\times\mathcal{X}_2} \left(\frac{\partial F_1}{\partial y}\frac{\partial y}{\partial x_1} + \frac{\partial F_1}{\partial x_1}\right)\eta\,\mathrm{d}x_1\,\mathrm{d}x_2, \end{aligned} \tag{3.8}$$

$$\begin{aligned} \int_{\mathcal{X}_1\times\mathcal{X}_2} \frac{\partial \eta}{\partial x_2}F_2\,\mathrm{d}x_1\,\mathrm{d}x_2 = &\int_{\mathcal{X}_1} F_2\eta|_{\partial\mathcal{X}_2}\,\mathrm{d}x_1 \\ &- \int_{\mathcal{X}_1\times\mathcal{X}_2} \left(\frac{\partial F_2}{\partial y}\frac{\partial y}{\partial x_2} + \frac{\partial F_2}{\partial x_2}\right)\eta\,\mathrm{d}x_1\,\mathrm{d}x_2. \end{aligned} \tag{3.9}$$

As the perturbation function η vanishes on the border $\partial(\mathcal{X}_1 \times \mathcal{X}_2)$, the first term on the right-hand side of the equations (3.6), (3.7), (3.8) and (3.9) is zero. Therefore, the equation (3.5) simplifies to:

$$\int_{\mathcal{X}_1\times\mathcal{X}_2} \left[-\frac{\partial^2 y}{\partial x_1^2} - \frac{\partial^2 y}{\partial x_2^2} + \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{1}{2}\frac{\partial}{\partial y}(F_1^2 + F_2^2)\right]\eta\,\mathrm{d}x_1\,\mathrm{d}x_2 = 0. \tag{3.10}$$

The perturbation η is arbitrary, hence the above equation leads to the second-order partial differential equation

$$\frac{\partial^2 y}{\partial x_1^2} + \frac{\partial^2 y}{\partial x_2^2} = \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{1}{2}\frac{\partial}{\partial y}(F_1^2 + F_2^2). \tag{3.11}$$

Defining the function $G = G(y, x_1, x_2)$ as

$$G(y, x_1, x_2) \stackrel{\text{def}}{=} \frac{\partial F_1}{\partial x_1} + \frac{\partial F_2}{\partial x_2} + \frac{1}{2}\frac{\partial}{\partial y}(F_1^2 + F_2^2), \tag{3.12}$$

and introducing the Laplace operator ∇^2 (in rectangular coordinates), the equation (3.11) takes on the form

$$\nabla^2 y = G, \tag{3.13}$$

which is a *Poisson equation*. Poisson's equation is a partial differential equation of elliptic type with broad applications in electrostatics, in mechanical engineering and biology (e.g., to model the motion of biological organisms in a solution), in theoretical physics (e.g., in the study of gravitation), in applied physics (e.g., in the study of *n-p* junctions in semiconductor devices) [9] as well as, e.g., in computer graphics [13]. A Poisson equation is typically defined on a two-dimensional or a three-dimensional domain, although it may be formulated on high-dimensional manifolds. A Poisson equation may be solved using a Green's function [5, 16], but there are various known methods to approach a Poisson equation numerically.

In the present context, the non-linear function $G : \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ takes on the form:

$$G = \frac{1}{p_{y,x_2}}\frac{\partial p_{x_1,x_2}}{\partial x_1} + \frac{1}{p_{y,x_1}}\frac{\partial p_{x_1,x_2}}{\partial x_2} - \frac{p_{x_1,x_2}^2}{p_{y,x_2}^3}\frac{\partial p_{y,x_2}}{\partial y} - \frac{p_{x_1,x_2}^2}{p_{y,x_1}^3}\frac{\partial p_{y,x_1}}{\partial y}. \tag{3.14}$$

The function G is defined in the set $\mathcal{D} \subset \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2$ given by:

$$\mathcal{D} \stackrel{\text{def}}{=} \{(y, x_1, x_2) \in \mathcal{Y} \times \mathcal{X}_1 \times \mathcal{X}_2 \mid p_{y,x_1}(y, x_1) \neq 0,\ p_{y,x_2}(y, x_2) \neq 0\}. \tag{3.15}$$

To be evaluated, it needs the computation of the first-order partial derivatives $\frac{\partial p_{x_1,x_2}}{\partial x_1}$, $\frac{\partial p_{x_1,x_2}}{\partial x_2}$, $\frac{\partial p_{y,x_2}}{\partial y}$ and $\frac{\partial p_{y,x_1}}{\partial y}$. It might be also noted that, by defining:

$$P_{x_1,x_2} \stackrel{\text{def}}{=} \log p_{x_1,x_2},\ P_{y,x_1} \stackrel{\text{def}}{=} \log p_{y,x_1},\ P_{y,x_2} \stackrel{\text{def}}{=} \log p_{y,x_2}, \tag{3.16}$$

on the set $\mathcal{D}$, the function G may be rewritten as:

$$\begin{aligned} G = {} & e^{(P_{x_1,x_2} - P_{y,x_2})}\frac{\partial P_{x_1,x_2}}{\partial x_1} + e^{(P_{x_1,x_2} - P_{y,x_1})}\frac{\partial P_{x_1,x_2}}{\partial x_2} \\ & - e^{2(P_{x_1,x_2} - P_{y,x_2})}\frac{\partial P_{y,x_2}}{\partial y} - e^{2(P_{x_1,x_2} - P_{y,x_1})}\frac{\partial P_{y,x_1}}{\partial y}. \end{aligned} \tag{3.17}$$

A special set of boundary conditions is devised as follows. Assume again monotonically-increasing dependency and define $\underline{y} \stackrel{\text{def}}{=} \min\{\mathcal{Y}\}$, $\underline{x}_1 \stackrel{\text{def}}{=} \min\{\mathcal{X}_1\}$ and $\underline{x}_2 \stackrel{\text{def}}{=} \min\{\mathcal{X}_2\}$.

As the function $y = y(x_1, x_2)$ is monotonically increasing with respect to both variables, the identity $f(\underline{x}_1, \underline{x}_2) = \underline{y}$ holds. Usable boundary conditions are:

$$\begin{cases} y(\underline{x}_1, \underline{x}_2) = \underline{y}, \\ \frac{\partial y}{\partial x_1}(x_1, \overline{x}_2) = F_1(y(x_1, \overline{x}_2), x_1, \overline{x}_2), \ x_1 \in \mathcal{X}_1, \\ \frac{\partial y}{\partial x_1}(x_1, \underline{x}_2) = F_1(y(x_1, \underline{x}_2), x_1, \underline{x}_2), \ x_1 \in \mathcal{X}_1, \\ \frac{\partial y}{\partial x_2}(\overline{x}_1, x_2) = F_2(y(\overline{x}_1, x_2), \overline{x}_1, x_2), \ x_2 \in \mathcal{X}_2, \\ \frac{\partial y}{\partial x_2}(\underline{x}_1, x_2) = F_2(y(\underline{x}_1, x_2), \underline{x}_1, x_2), \ x_2 \in \mathcal{X}_2, \end{cases} \tag{3.18}$$

where $\overline{x}_1 \stackrel{\text{def}}{=} \max\{\mathcal{X}_1\}$ and $\overline{x}_2 \stackrel{\text{def}}{=} \max\{\mathcal{X}_2\}$. Such boundary conditions prescribe the values at the bottom-left corner of the domain and on the four borders of the domain.

Another special boundary condition, that retraces the center-of-mass-to-center-of-mass-mapping condition utilized in [10] is as follows. Define:

$$\begin{cases} x_1^{\text{m}} \stackrel{\text{def}}{=} \int_{\mathcal{X}_1 \times \mathcal{X}_2} p_{x_1, x_2}(x_1, x_2) x_1 \, \mathrm{d}x_1 \, \mathrm{d}x_2, \\ x_2^{\text{m}} \stackrel{\text{def}}{=} \int_{\mathcal{X}_1 \times \mathcal{X}_2} p_{x_1, x_2}(x_1, x_2) x_2 \, \mathrm{d}x_1 \, \mathrm{d}x_2, \\ y^{\text{m}} \stackrel{\text{def}}{=} \int_{\mathcal{Y}} p_y(y) y \, \mathrm{d}y. \end{cases} \tag{3.19}$$

A center-of-mass-to-center-of-mass-mapping condition casts as follows:

$$\begin{cases} y(x_1^{\text{m}}, x_2^{\text{m}}) = y^{\text{m}}, \\ \frac{\partial y}{\partial x_1}(x_1, x_2^{\text{m}}) = F_1(y(x_1, x_2^{\text{m}}), x_1, x_2^{\text{m}}), \ x_1 \in \mathcal{X}_1, \\ \frac{\partial y}{\partial x_2}(x_1^{\text{m}}, x_2) = F_2(y(x_1^{\text{m}}, x_2), x_1^{\text{m}}, x_2), \ x_2 \in \mathcal{X}_2, \end{cases} \tag{3.20}$$

Such boundary conditions prescribe the values on a 'cross' in the middle of the domain.

3.2. Numerical implementation. The statistical modeling problem is solved by implementing a numerical scheme based on a discretization of the Poisson equation (3.11). Denote the set $\mathcal{X}_1 = [\underline{x}_1, \ \overline{x}_1]$ and the set $\mathcal{X}_2 = [\underline{x}_2, \ \overline{x}_2]$. The interval $\mathcal{X}_1$ is subdivided into B_1 sub-intervals of equal width $h_1 \stackrel{\text{def}}{=} (\overline{x}_1 - \underline{x}_1)/B_1$ and the interval $\mathcal{X}_2$ is subdivided into B_2 sub-intervals of equal width $h_2 \stackrel{\text{def}}{=} (\overline{x}_2 - \underline{x}_2)/B_2$. The solution $y = f(x_1, x_2)$ is represented by the $(B_2 + 1) \times (B_1 + 1)$ matrix $\mathbf{y}$ of entries $y_{i,j}$, with $i = 1, \ldots, B_2 + 1$ and $j = 1, \ldots, B_1 + 1$. (Note that the index i represents the rows, hence is associated with the variable x_2, while the index j represents the columns, hence it is associated with the variable x_1.) The quantity $y_{i,j}$ represents an approximation of the true value $y(\underline{x}_1 + (j-1)h_1, \underline{x}_2 + (i-1)h_2)$. A fixed-point iterative scheme generates a sequence $\mathbf{y}^{(k)}$, with $k = 0, 1, 2, \ldots$ of increasingly-refined approximations of the true solution to the Poisson equation, where it is understood that the matrix $\mathbf{y}^{(0)}$ denotes a suitably-chosen initial guess.

It is assumed that the boundary conditions (3.18) hold. In order to approximate the second-order partial derivatives involved in the non-linear Poisson equation, the following *central difference approximation scheme* is made use of:

$$\frac{\partial^2 y}{\partial x_1^2}(\underline{x}_1 + (j-1)h_1, \underline{x}_2 + (i-1)h_2) \approx \frac{y_{i,j+1} - 2y_{i,j} + y_{i,j-1}}{h_1^2}, \tag{3.21}$$

$$\frac{\partial^2 y}{\partial x_2^2}(\underline{x}_1 + (j-1)h_1, \underline{x}_2 + (i-1)h_2) \approx \frac{y_{i+1,j} - 2y_{i,j} + y_{i-1,j}}{h_2^2}, \tag{3.22}$$

that holds in the central part of the subdivision grid, namely, for $i = 2, \ldots, B_2$ and $j = 2, \ldots, B_1$. It is also convenient to define the matrix of functional components $G_{i,j} : \mathcal{Y} \to \mathbb{R}$ defined by:

$$G_{i,j}(y) \stackrel{\text{def}}{=} G(y, \underline{x}_1 + (j-1)h_1, \underline{x}_2 + (i-1)h_2). \tag{3.23}$$

Placing the above approximations into the Poisson equation (3.11) yields the *five-point scheme*:

$$2\left(\frac{1}{h_1^2} + \frac{1}{h_1^2}\right) y_{i,j} + G_{i,j}(y_{i,j}) = \frac{1}{h_1^2}(y_{i,j+1} + y_{i,j-1}) + \frac{1}{h_1^2}(y_{i+1,j} + y_{i-1,j}), \tag{3.24}$$

which again holds in a non-peripheral point of the grid, as illustrated in the Figure 3.1.

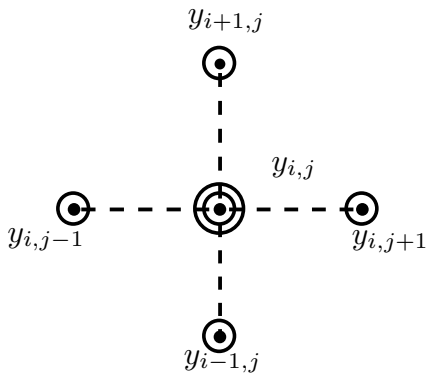


FIG. 3.1. *Five-points numerical scheme for integrating the Poisson equation.*

(Similar equations might be written to implement the boundary conditions and are omitted from here.) The equations (3.24) constitute a set of non-linear equations in the unknown $y_{i,j}$. It is convenient to write such a non-linear system by the conventional representation:

$$\mathbf{y} + L(\mathbf{y}) = N(\mathbf{y}), \tag{3.25}$$

where the matrix $\mathbf{y}$ represents the whole set of unknowns, the symbol L represents a linear operator and the symbol N represents a non-linear operator. A fixed-point iterative scheme generates a sequence $\mathbf{y}^{(k)}$, with $k = 0, 1, 2, \ldots$ of increasingly-refined approximations of the true solution to the Poisson equation, computed by the recurrence rule:

$$\mathbf{y}^{(k+1)} = -L(\mathbf{y}^{(k)}) + N(\mathbf{y}^{(k)}), \tag{3.26}$$

where it is understood that the quantity $\mathbf{y}^{(0)}$ denotes a suitably-chosen initial guess.

A pseudocode to implement a possible instance of the above numerical scheme is reported in the Algorithm 1. Note that the top-right corner's value is computed

as the average of the value obtained by integrating the upper-side boundary ordinary differential equation and of the value obtained by integrating the right-hand-side boundary ordinary differential equation. (The purpose of the pseudo-code explained

Algorithm 1 Pseudocode to implement the numerical solution of the differential system (3.1) of partial differential equations reformulated as the Poisson equation (3.11) with boundary conditions (3.18).

▷ Input domain boundaries $\underline{x}_1, \overline{x}_1, \underline{x}_2, \overline{x}_2$, numbers of subdivisions B_1, B_2, functions F_1, F_2, G and boundary value $\underline{y}$
Compute widths $h_1 = \frac{\overline{x}_1 - \underline{x}_1}{B_1}$ and $h_2 = \frac{\overline{x}_2 - \underline{x}_2}{B_2}$
Define constant $\gamma = 2\left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right)$
Set $y_{1,1}^{(0)} = \underline{y}$
for $i = 2$ to $B_2 + 1$ **do**
  Compute $y_{i,1}^{(0)} = y_{i-1,1}^{(0)} + h_2 F_2(y_{i-1,1}^{(0)}, \underline{x}_1, \underline{x}_2 + (i-1)h_2)$ [*Left-hand border*]
end for
for $j = 2$ to $B_1 + 1$ **do**
  Compute $y_{1,j}^{(0)} = y_{1,j-1}^{(0)} + h_1 F_1(y_{1,j-1}^{(0)}, \underline{x}_1 + (j-1)h_1, \underline{x}_2)$ [*Lower-side border*]
end for
for $i = 2$ to B_2 **do**
  Compute $y_{i,B_1+1}^{(0)} = y_{i-1,B_1+1}^{(0)} + h_2 F_2(y_{i-1,B_1+1}^{(0)}, \overline{x}_1, \underline{x}_2 + (i-1)h_2)$ [*Right-hand border excluded the top-right corner*]
end for
for $j = 2$ to B_1 **do**
  Compute $y_{B_2+1,j}^{(0)} = y_{B_2+1,j-1}^{(0)} + h_1 F_1(y_{B_2+1,j-1}^{(0)}, \underline{x}_1 + (j-1)h_1, \overline{x}_2)$ [*Upper-side border excluded the top-right corner*]
end for
Set $y_{B_2+1,B_1+1}^{\mathrm{a}} = y_{B_2+1,B_1}^{(0)} + h_1 F_1(y_{B_2+1,B_1}^{(0)}, \underline{x}_1 + B_1 h_1, \overline{x}_2)$
Set $y_{B_2+1,B_1+1}^{\mathrm{b}} = y_{B_2,B_1+1}^{(0)} + h_2 F_2(y_{B_2,B_1+1}^{(0)}, \overline{x}_1, \underline{x}_2 + B_2 h_2)$
Set $y_{B_2+1,B_1+1}^{(0)} = \frac{1}{2}\left(y_{B_2+1,B_1+1}^{\mathrm{a}} + y_{B_2+1,B_1+1}^{\mathrm{b}}\right)$ [*Top-right corner*]
Set $y_{i,j}^{(0)} = \underline{y}$ for every $i = 1, \ldots, B_2$ and $j = 1, \ldots, B_1$ [*Initial guess*]
for $k = 0, 1, 2, \ldots$ **do**
  for $i = 2$ to B_2 **do**
    for $j = 2$ to B_1 **do**
      Compute $y_{i,j}^{(k+1)} = (y_{i+1,j}^{(k)} + y_{i-1,j}^{(k)})/(\gamma h_2^2) + (y_{i,j+1}^{(k)} + y_{i,j-1}^{(k)})/(\gamma h_1^2) - G_{i,j}(y_{i,j}^{(k)})/\gamma$ [*Central part of the grid*]
    end for
  end for
end for
▷ Output results $y_{i,j}$

in the present section is to illustrate and clarify the basic notions behind the implementation of the discussed numerical procedures. More refined and better optimized versions could be implemented, indeed.)

The second-order joint probability density functions p_{x_1,x_2}, p_{y,x_1} and p_{y,x_2} are

estimated by a histogram method and log-warped to obtain the functions P_{x_1,x_2}, P_{y,x_1} and P_{y,x_2} defined in the equations (3.16). The functions F_1 and F_2 are thus evaluated numerically and the function G is evaluated by the equation (3.17) upon numerically approximating the required first-order partial derivatives.

4. Numerical tests. In the numerical tests, the *average relative error* at iteration k refers to the quantity

$$\frac{\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|_{\mathrm{F}}}{(B_1+1)(B_2+1)}, \tag{4.1}$$

while the *average absolute error* refers to the quantity

$$\frac{\|(\mathbf{y}^{(k+1)} - \mathbf{y}^{\star}) \oslash \mathbf{y}^{\star}\|_{\mathrm{F}}}{(B_1+1)(B_2+1)}, \tag{4.2}$$

where $\mathbf{y}^{\star}$ denotes the matrix of values corresponding to the actual solution f (which is known only in those tests about synthetic data sets), symbol $\oslash$ denotes component-wise division and symbol $\|\cdot\|_{\mathrm{F}}$ denotes Frobenius norm. Likewise, the *point-wise absolute error* refers to the quantities

$$\left| \frac{y_{i,j}^{(k+1)} - y_{i,j}^{\star}}{y_{i,j}^{\star}} \right|, \tag{4.3}$$

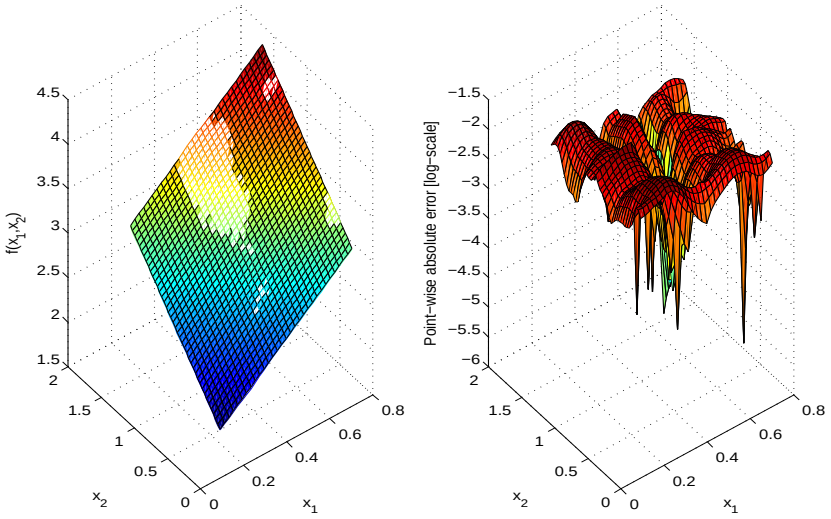
evaluated for each pair of index (i, j) separately.

4.1. Tests on isotonic statistical modeling: Synthetic data. The pairs (x_1, x_2) of independent variates are generated uniformly in $\mathcal{X}_1 \times \mathcal{X}_2$ and the variate y is generated by the rule $y = f(x_1, x_2) + \nu$, with ν being a zero-mean Gaussian noise. A total of $3,000$ samples were generated for the numerical simulations. Two tests are conducted, where the underlying non-linear model is described by the non-linear functions specified in the following test problems:

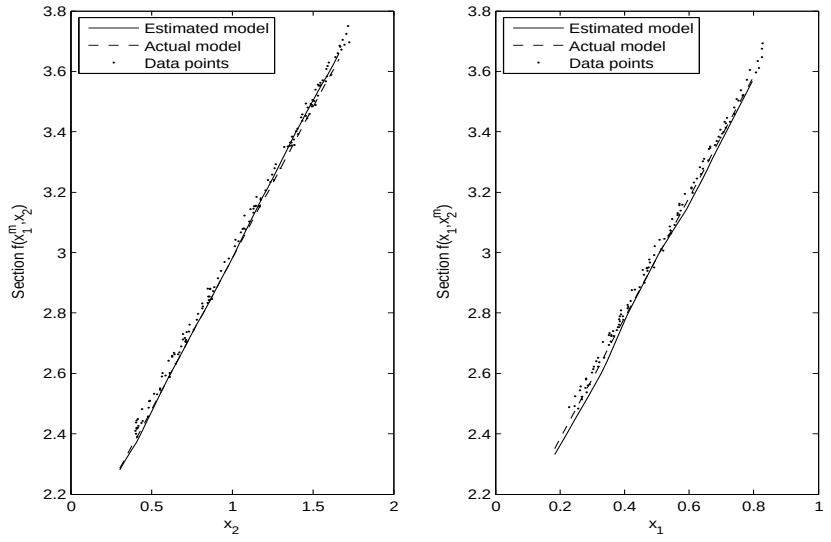
- **Test problem 1**: The problem is defined by the functions $F_1(y, x_1, x_2) = y - 2x_1 - x_2 + 1$ and $F_2(y, x_1, x_2) = y - 2x_1 - x_2$, by the intervals $\mathcal{X}_1 = [0,\ 1]$ and $\mathcal{X}_1 = [0,\ 2]$ and by the boundary condition $f(0, 0) = 1$. Such features give the function $G(y, x_1, x_2) = 2(y - 2x_1 - x_2 - 1)$ and the exact solution $f(x_1, x_2) = 2x_1 + x_2 + 1$.
- **Test problem 2**: The problem is defined by the functions $F_1(y, x_1, x_2) = 4x_1 e^{-y}$ and $F_2(y, x_1, x_2) = e^{-y}$, by the intervals $\mathcal{X}_1 = [0,\ 1]$ and $\mathcal{X}_1 = [0,\ 2]$ and by the boundary condition $f(0, 0) = 0$. Such features give the function $G(y, x_1, x_2) = 4e^{-y} - (16x_1^2 + 1)e^{-2y}$ and the exact solution $f(x_1, x_2) = \log(2x_1^2 + x_2 + 1)$.

The variance of the noise ν was set to 10^{-4}. The boundary conditions were chosen as in (3.20) and $B_1 = 50$ and $B_2 = 60$. Note that the errors (4.1), (4.2) and (4.3) refer to the actual value of the model, hence they take into account the measurement noise ν too.

The Figure 4.1 illustrates the numerical behavior on the Test problem 1, in terms of absolute error and of estimated model superimposed with the actual model and with the (noisy) data-points on two central sections.



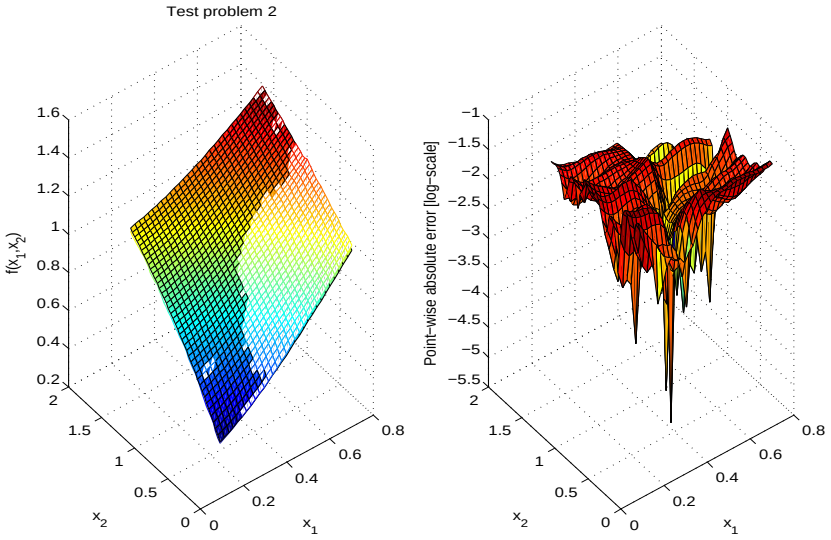
(a)



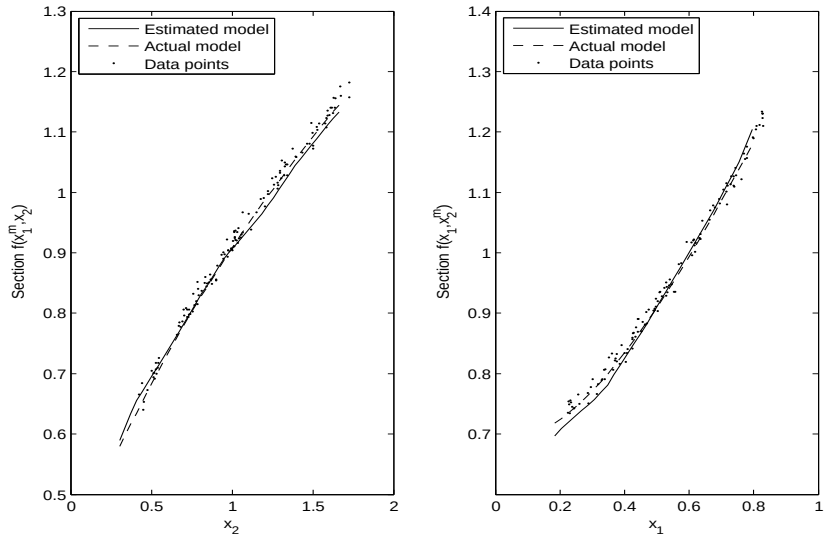
(b)

FIG. 4.1. *Test problem 1. (a) Left-hand panel: Estimated solution superimposed with the actual solution (lighter colors surface). Right-hand panel: Point-wise absolute error (in logarithmic scale). (b) Estimated model superimposed with the actual model and with the data-points on the section* $x_1 = x_1^{\mathrm{m}}$ *(left-hand panel) and on the section* $x_2 = x_2^{\mathrm{m}}$ *(right-hand panel).*

The Figure 4.2 illustrates the numerical behavior on the Test problem 2, in terms of absolute error and of estimated model superimposed with the actual model and with the (noisy) data-points on two central sections.



(a)



(b)

FIG. 4.2. *Test problem 2. (a) Left-hand panel: Estimated solution superimposed with the actual solution (lighter colors surface). Right-hand panel: Point-wise absolute error (in logarithmic scale). (b) Estimated model superimposed with the actual model and with the data-points on the section* $x_1 = x_1^{\mathrm{m}}$ *(left-hand panel) and on the section* $x_2 = x_2^{\mathrm{m}}$ *(right-hand panel).*

In the considered experiments, the statistical modeling technique returns the expected results. The accuracy of model estimation is better at the center of the domain,

due to the chosen boundary conditions, and are worst at the peripheral parts of the domain, due to numerical errors that occur in the estimation of probability density functions by finite-size data-sets, to numerical errors that occur in the estimation of the first-order derivatives of the non-linear functions and to numerical approximations.

4.2. Tests on isotonic statistical modeling: Robotic arm dynamics data. The data used in the present experiment come from a realistic simulation of the dynamics of a Puma 560 robot arm [3]. The data set contains 8 independent variables and 1 dependent variable. The relationship between the variables is known to be nonlinear and the system output measurement to be quite noisy (data set available on the LIACC Repository[1]). The original 8 independent variables were projected over the subspace spanned by the 2 minor components. Both the obtained variables x_1 and x_2 show negative correlation with the dependent variable y (namely, the coefficient of correlation between the variables x_1 and y is -0.5173, while the coefficient of correlation between the variables x_2 and y is -0.1548), which suggests a dependence of decreasing type from both inputs.

The Figure 4.3 illustrates the histogram-based estimates of the second-order joint probability density functions on the Puma 560 robot arm.

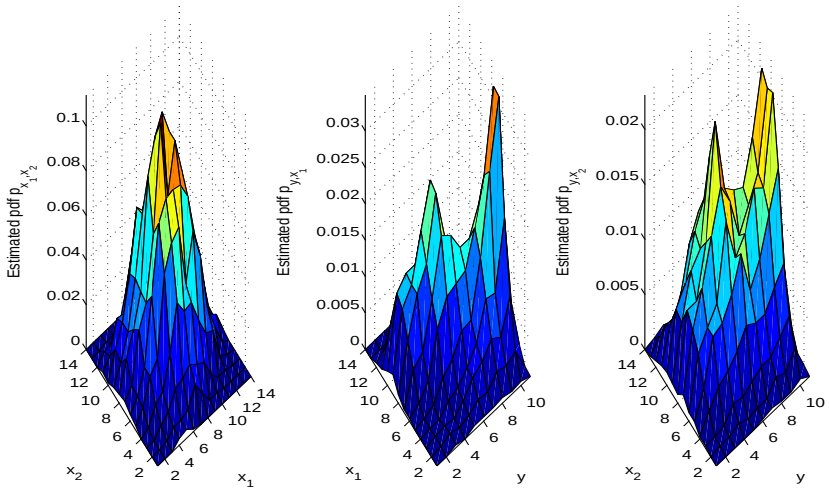
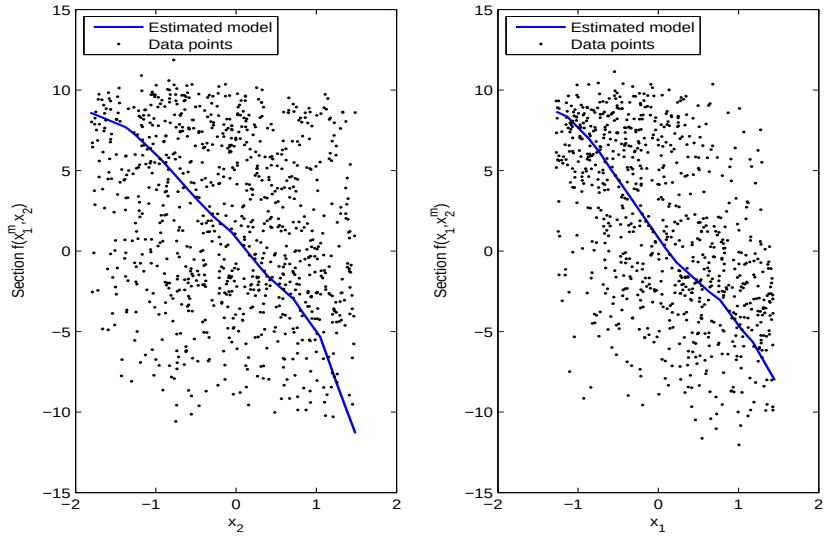


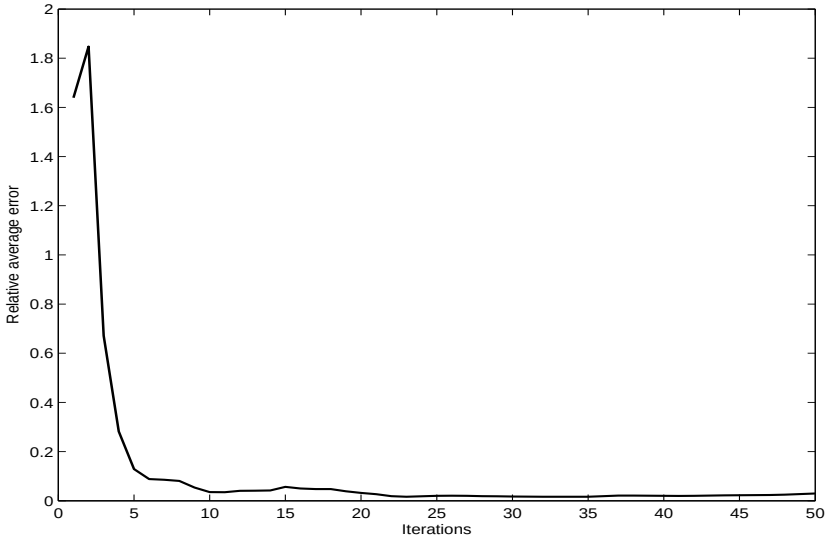
FIG. 4.3. *Puma 560 robot arm data set. Histogram-based estimates of the second-order joint probability density functions. Left-hand panel: Estimate of* p_{x_1,x_2}. *Central panel: Estimate of* p_{y,x_1}. *Right-hand panel: Estimate of* p_{y,x_2}.

[1] http://www.liaad.up.pt/ ltorgo/Regression/DataSets.html

The Figure 4.4 shows the accuracy of the achieved model in terms of estimated model superimposed with the data-points on two sections and in terms of relative average error during iteration. The partitions cardinality are $B_1 = 50$ and $B_2 = 60$.



(a)



(b)

FIG. 4.4. *Puma 560 robot arm data set. (a) Left-hand panel: Estimated model superimposed with the data-points on the section* $x_1 = x_1^{\mathrm{m}}$. *Right-hand panel: Estimated model superimposed with the data-points on the section* $x_2 = x_2^{\mathrm{m}}$. *(b) Relative average error during iteration.*

The purpose of the present experiment was to illustrate the behavior of the devised statistical nonlinear modeling procedure in presence of a monotonically decreasing relationship between the dependent variable and both independent variables. The data set looks noisy, as expected, and the obtained numerical results evidence that the estimated model is able to capture the mean dependence of the data, as the displayed model sections locate amidst the densest samples area of the parameter space.

4.3. Tests on isotonic statistical modeling: Acrylamide concentration data. The present experiment is about applying the devised statistical modeling procedure to the problem discussed in the contribution [7], which concerns the formation of *acrylamide* during the process of cooking French fries.

Acrylamide is a chemical compound with chemical formula C_3H_5NO. Human exposure to acrylamide through consumption of French fries and other foods has been recognized as a potential health concern. Documented studies have found that increased dietary intake of acrylamide is associated with increased risks of postmenopausal endometrial and ovarian cancer, particularly among nonsmokers. Consumption of French fries during preschool years is associated with a slightly increased risk of breast cancer later in life. Dietary acrylamide is also significantly associated with increased risk of oral cavity cancer in female nonsmokers. Acrylamide formation can be greatly influenced by food processing conditions. In particular, frying time and temperature are the most important key parameters influencing acrylamide formation.

In the *Acrylamide* data set, the predictor variable x_1 denotes the cooking temperature expressed in °C, the predictor variable x_2 denotes the cooking time in seconds, while the target variable y denotes the concentration of acrylamide, expressed in μg/kg. Both variables x_1 and x_2 show positive correlation with the dependent variable y (namely, the coefficient of correlation between the variable x_1 and the variable y is 0.6124, while the coefficient of correlation between the variable x_2 and the variable y is 0.5271). Such observation suggests to assume a dependence of increasing type with respect to both independent variables.

The Figure 4.5 shows the accuracy of the achieved model in terms of estimated model superimposed with the data-points on two sections. The partitions cardinality are $B_1 = 50$ and $B_2 = 60$. The model looks accurate, except at the border of the domain.

4.4. Tests on isotonic statistical modeling: Pollen grains data. The *Pollen grains* data set consists of four independent variables (three geometrical features of grains, namely 'ridge', 'nub' and 'crack', plus pollen grain weight) and the dependent variable to be predicted by the model is grain density. Such data set is available on the StatLib Repository[2].

[2] http://lib.stat.cmu.edu/datasets/

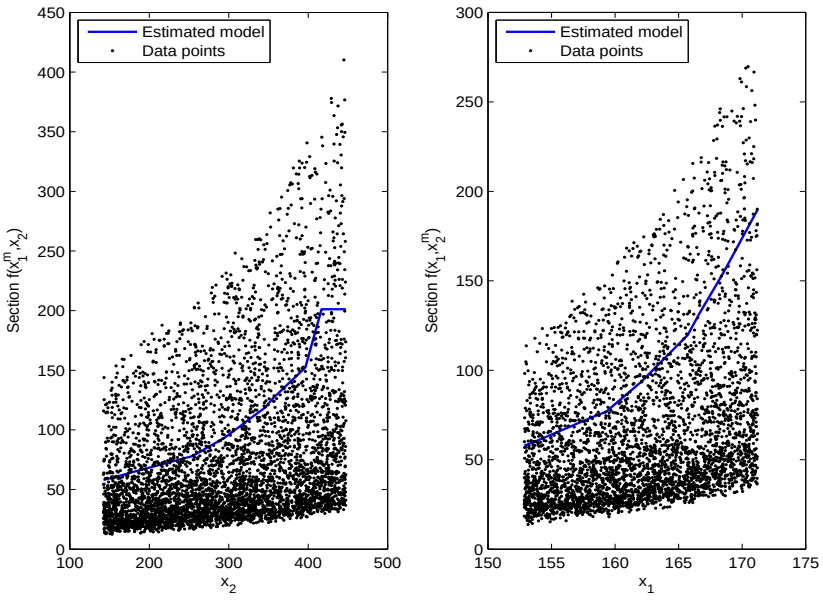


FIG. 4.5. *Numerical behavior of the statistical modeling technique on the Acrylamide data set. Left-hand panel: Estimated model superimposed with the data-points on the section* $x_1 = x_1^{\mathrm{m}}$. *Right-hand panel: Estimated model superimposed with the the data-points on the section* $x_2 = x_2^{\mathrm{m}}$.

The purpose of the present experiment is to show the behavior of the devised statistical nonlinear modeling procedure in presence of *mixed-monotonicity* type dependency between the dependent variable and the two independent variables. To this aim, the original 4 independent variables are projected over the subspace spanned by the 2 minor components. The obtained variables x_1 and x_2 show mixed-sign correlation with the independent variable y (namely, the coefficient of correlation between the variable x_1 and the variable y is 0.6407, while the coefficient of correlation between the variable x_2 and the variable y is -0.4874). Hence, mixed increasing/decreasing monotonic dependency is selected in the statistical isotonic modeling procedure.

The Figure 4.6 illustrates the histogram-based estimates of the second-order joint probability density functions. The Figure 4.7 shows the accuracy of the achieved model in terms of estimated model superimposed with the data-points on two sections. The partitions cardinality are $B_1 = 50$ and $B_2 = 60$.

Even in this experiment, the data set looks noisy and the inferred model is able to capture the mean dependence of the data, as it locates amid the densest portion of the parameter space. Noteworthy, the devised modeling procedure is able to effectively cope with mixed increasing/decreasing dependence of the independent variable with respect to the dependent variables.

5. Conclusions. The research endeavor illustrated in the present article took its moves from the following assumptions:

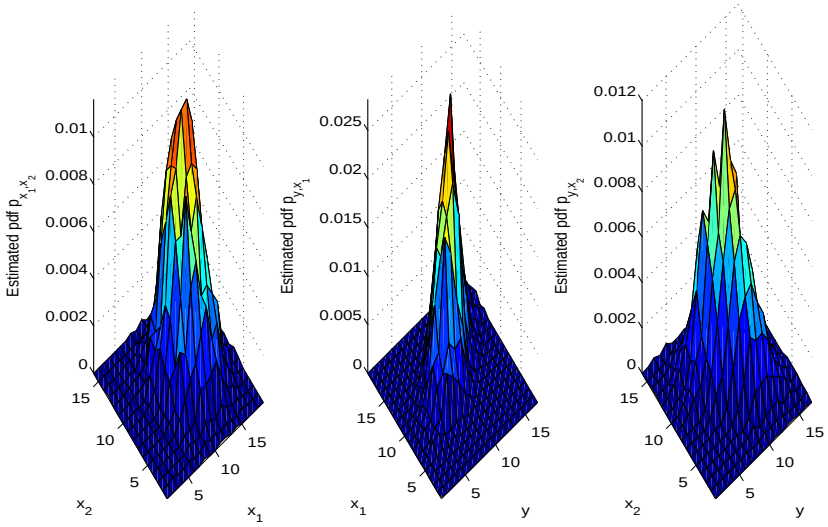


FIG. 4.6. *Pollen grains data set. Histogram-based estimates of the second-order joint probability density functions. Left-hand panel: Estimate of* p_{x_1,x_2}. *Central panel: Estimate of* p_{y,x_1}. *Right-hand panel: Estimate of* p_{y,x_2}.

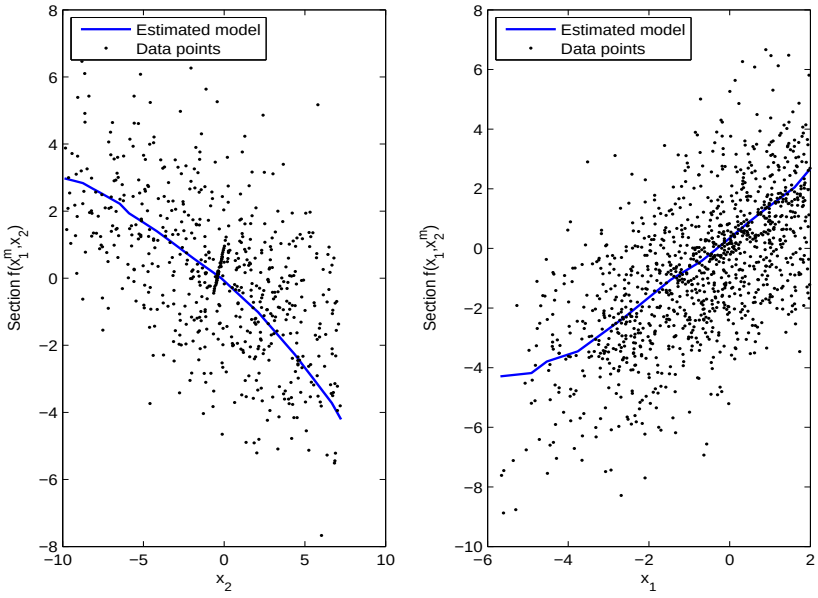


FIG. 4.7. *Numerical behavior of the statistical modeling technique on the Pollen grains data set. Left-hand panel: Estimated model superimposed with the data-points on the section* $x_1 = x_1^{\mathrm{m}}$. *Right-hand panel: Estimated model superimposed with the the data-points on the section* $x_2 = x_2^{\mathrm{m}}$.

- A statistical model describes how one random variable is related to one or more random variables. In particular, it was assumed that the physical phenomenon under observation relates two independent variables with a single de-

pendent variable. The joint statistical features of the independent/dependent variates are described by their second-order joint probability density functions.

- The statistical model is non-parametric, namely, there is no assumption on its shape (except that the model be monotonically increasing or decreasing and continuous).

The statistical modeling problem was formulated in terms of a system of first-order non-linear partial differential equations that relate the second-order joint probability density functions of the independent/dependent variates with the unknown model by a probability measure conservation principle. Such a system of partial differential equations was reduced to a single second-order non-linear differential equation of Poisson type.

In order to implement the devised statistical modeling method, a finite-difference numerical scheme of relaxation type was proposed and tested numerically. The numerical results illustrated the numerical features of the devised statistical isotonic modeling method. In particular, the devised method was tested on modeling synthetic data sets as well as real-world data sets, namely, a data set arising from robotic arm dynamics study, a data set arising from a food toxicology research and a data set from palynology research. The results of modeling suggest that the devised statistical modeling technique can cope with models that include strong hidden nuisance variables.

REFERENCES

[1] S. Angelov, B. Harb, S. Kannan and L.-S. Wang, "Weighted isotonic regression under the L_1 norm", in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), pp. 783 – 791, 2006

[2] G. Arminger, C.C. Clogg and M.E. Sobel (Editors), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Springer (1st Edition), 1994

[3] A. Benitez, I. Huitzil, A. Casiano, J. De La Calleja, and M.A. Medina, "PUMA 560: Robot Prototype with Graphic Simulation Environment", *Advances in Mechanical Engineering*, Vol. 2, No. 1, pp. 15 – 22, March 2012

[4] H.D. Brunk, "Maxmimum likelihood estimates of monotone parameters", *Annals of Mathematical Statistics*, Vol. 26, pp. 607 – 616, 1955

[5] A.G. Butkovskiy, *Green's Functions and Transfer Functions Handbook*, Halstead Press – John Wiley & Sons, New York, 1982

[6] N. Chakravarti, "Sensitivity analysis in isotonic regression", *Discrete Applied Mathematics*, Vol. 45, No. 3, pp. 183 – 196, September 1993

[7] M.-J. Chen, H.-T. Hsu, C.-L. Lin and W.-Y. Ju, "A statistical regression model for the estimation of acrylamide concentrations in French fries for excess lifetime cancer risk assessment", *Food and Chemical Toxicology*, Vol. 50, No. 10, pp. 3867 – 3876, October 2012

[8] W.D. Dupont, *Statistical modeling for biomedical researchers*, Cambridge University Press (2nd Edition), 2009

[9] L.C. Evans, *Partial Differential Equations*, American Mathematical Society, Providence, 1998

[10] S. Fiori, "Statistical nonparametric bivariate isotonic regression by look-up-table-based neural networks", Proceedings of the *2011 International Conference on Neural Information*

Processing (ICONIP 2011, Shanghai (China), November 14-17, 2011), B.-L. Lu, L. Zhang, and J. Kwok (Eds.), Part III, LNCS 7064, pp. 365 – 372, Springer, Heidelberg, 2011

[11] S. Fiori, "Fast statistical regression in presence of a dominant independent variable", *Neural Computing and Applications*, (Special issue of the 2011 International Conference on Neural Information Processing - ICONIP'2011), Vol. 22, No. 7, pp. 1367 – 1378, 2013

[12] A.J. Kearsley, R.A. Tapia and M.W. Trosset, "An approach to parallelizing isotonic regression", in *Applied Mathematics and Parallel Computing* (H. Fischer, B. Riedmüller and S. Schäffler Ed.s), pp. 141 – 147, Physica-Verlag Heidelberg, 1996

[13] L. Gorelick, M. Galun, E. Sharon, R. Basri and A. Brandt, "Shape representation and classification using the Poisson equation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, pp. 1991 – 2005, December 2006

[14] J. Li, "Automatic linguistic indexing of pictures by a statistical modeling approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, pp. 1075 – 1088, September 2003

[15] L. Li, "Statistical modeling of complex backgrounds for foreground object detection", *IEEE Transactions on Image Processing*, Vol. 13, No. 11, pp. 1459 – 1472, November 2004

[16] A.D. Polyanin, *Handbook of Linear Partial Differential Equations for Engineers and Scientists*, Chapman & Hall/CRC Press, 2002

[17] K. Punera and J. Gosh, "Enhanced hierarchical classification via isotonic smoothing", in *Proceedings of the* 17^{th} *International Conference on World Wide Web* (April 21-25, 2008, Beijing – China), pp. 151 – 160, 2008

[18] G. Salanti, "The isotonic regression framework: Estimating and testing under order restrictions", *Ph.D. Dissertation*, Fakultät für Matematik, Ludwig-Maximilians-Universität München, 2003

[19] Q.F. Stout, "Strict L_∞ isotonic regression", *Journal of Optimization Theory and Applications*, Vol. 152, No. 1, pp. 121 – 135, January 2012

[20] C.A. Sims, "Statistical modeling of monetary policy and its effects", Prize Lecture on 8 December 2011 at Aula Magna, Stockholm University [`http://www.nobelprize.org/nobel_prizes/economics/laureates/2011/sims-lecture.html`]