



HAL
open science

Derivative formulas and gradient of functions with non-independent variables

Matieyendou Lamboni

► **To cite this version:**

Matieyendou Lamboni. Derivative formulas and gradient of functions with non-independent variables. Axioms, 2023, 12 (9), pp.845. 10.3390/axioms12090845 . hal-04621210

HAL Id: hal-04621210

<https://hal.science/hal-04621210>

Submitted on 23 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Derivative formulas and gradient of functions with non-independent variables

Matieyendou Lamboni^{1a,b}

^aUniversity of Guyane, Department DFRST, 97346 Cayenne, French Guiana, France

^b228-UMR Espace-Dev: University of Guyane, University of Réunion, IRD, University of Montpellier, France.

Abstract

Stochastic characterizations of functions subject to constraints result in treating them as functions with non-independent variables. Using the distribution function or copula of the input variables that comply with such constraints, we derive two types of partial derivatives of functions with non-independent variables (i.e., actual and dependent derivatives) and argue in favor of the latter. Dependent partial derivatives of functions with non-independent variables rely on the dependent Jacobian matrix of non-independent variables, which is also used to define a tensor metric. The differential geometric framework allows for deriving the gradient, Hessian and Taylor-type expansion of functions with non-independent variables.

Keywords: Copulas, Dependent and correlated variables, Gradient, Hessian, Riemannian Tensors

1. Introduction

We use to work with models defined through functions that include non-independent variables such as correlated input variables. It is also the case of models defined via functions with independent input variables and equations or inequations connecting such inputs; functions subject to constraints involving input variables and/or the model output. Knowing the key role of partial derivatives at a given point in i) mathematical analysis of functions and convergence, ii) Poincaré inequalities ([1, 2]) and equalities ([3, 4]), iii) optimization and active subspaces ([5, 6]), iv) implicit functions ([7, 8]) and v) differential geometry (see e.g., [9, 10, 11]), it is interesting and relevant to have formulas that enable the calculations of partial derivatives of functions in presence of non-independent variables, including the gradients. Of course, such formulas must account for the dependency structures among the model inputs, including the constraints imposed on such inputs.

Actual partial derivatives aim at calculating the partial derivatives of functions, which take account of the relationship between the input variables ([12]). For instance, let us consider the function $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ given by $f(x, y, z) = 2xy^2z^3$ under the constrained equation $h(x, y, z) = 0$ with h any smooth function. Using f'_x for the formal partial derivative of f w.r.t. x , that is, the partial derivative of f by considering other inputs as constant or independent, the chain rule gives the following partial derivative:

$$\begin{aligned} \frac{\partial f}{\partial x}(x, y, z) &= f'_x(x, y, z) + f'_y(x, y, z) \frac{\partial y}{\partial x} + f'_z(x, y, z) \frac{\partial z}{\partial x} \\ &= 2y^2z^3 + 4xyz^3 \frac{\partial y}{\partial x} + 6xy^2z^2 \frac{\partial z}{\partial x}. \end{aligned}$$

¹Corresponding author: matieyendou.lamboni[at]gmail.com, matieyendou.lamboni[at]univ-guyane.fr, June 18, 2023

In probability and statistics, the point (x, y, z) stands for a realization or a sample value of the implicit random vector (X, Y, Z) . The quantities $\frac{\partial y}{\partial x}$, $\frac{\partial z}{\partial x}$ and $\frac{\partial f}{\partial x}$ do not have a definite meaning when the variables (X, Y, Z) are npr-independent. So far, $\frac{\partial y}{\partial x}$, $\frac{\partial z}{\partial x}$ and $\frac{\partial f}{\partial x}$ are often hard to determine without supplementary assumptions such as the choice of directions or paths. When (X, Z) or (X, Y) are independent and using the equation $h(x, y, z) = x + 2y - z = 0$, we can write ([12])

$$\frac{\partial f}{\partial x}(x, y, z) = \begin{cases} 2y^2z^3 - 2xyz^3 & \text{if } X \text{ and } Z \text{ are independent or } z \text{ is being held fixed} \\ 2y^2z^3 + 6xy^2z^2 & \text{if } X \text{ and } Y \text{ are independent or } y \text{ is being held fixed} \end{cases} .$$

It is clear that the actual partial derivatives are not unique since such derivatives rely on two different paths or assumptions. While each supplementary assumption can make sense in some cases, it cannot be always guaranteed by the constrained equation $h(x, y, z) = 0$ in general. Indeed, when all the initial input variables are dependent or correlated, the above partial derivatives are no longer valid even for a linear function h , and it is worth finding the right relationship between the input variables and the partial derivatives, including the gradient.

In differential geometry (see e.g., [9, 10, 11]), using the differential of the function f , that is,

$$df = 2y^2z^3 dx + 4xyz^3 dy + 6xy^2z^2 dz ,$$

the gradient of f is defined as the dual of df w.r.t. a given tensor metric. Obviously, different tensor metrics will give different gradients of the same function. While the Euclidean metric given by $\gamma := dx^2 + dy^2 + dz^2$ is more appropriate for independent inputs, finding the appropriate metrics is challenging in general. Indeed, the first fundamental form (in differential geometry) requires the Jacobian matrix of the inputs to define the associated tensor metric.

For non-independent variables having F as the joint cumulative distribution function (CDF), the bi-variate dependency models ([13]) and the multivariate dependency models ([3, 14, 15]), including the conditional and inverse Rosenblatt transformation ([16, 17]), establish formal and analytical relationships among such variables using either CDFs or the corresponding copulas or new distributions that look like and behave as a copula ([18]). A dependency function characterizes the probabilistic dependency structures among such variables. For a d -dimensional random vector of non-independent variables, the dependency models express a subset of $d - 1$ variables as a function of independent variables, consisted of the remaining input and new independent variables.

In this paper, we propose a new approach for calculating the partial derivatives of functions, which account for the dependency structures among the input variables. Our approach relies on dependency models. By providing known relationships between the dependent inputs (including constraints imposed on inputs or outputs), dependency models can be seen as the global and probabilistic implicit functions (see Section 2.2). Such dependency models are used for determining the dependent Jacobian and the tensor metric for non-independent variables. The contributions of this paper are threefold:

- provide a generalization of the actual partial derivatives of functions with non-independent variables and give its limits;
- introduce the general derivative formulas of functions with non-independent variables (called dependent partial derivatives) without any additional assumption;

- provide the gradient, Hessian and the Taylor-type expansions of functions with non-independent variables that comply with the dependency structures among the input variables.

In Section 2, we firstly recall dependency models of dependent input variables, including correlated variables. Secondly, we derive interesting properties of such models regarding the calculus of partial derivatives and the probabilistic implicit functions. By coupling dependency functions with the function of interest, we extend the actual partial derivatives of functions with non-independent variables in Section 3. To avoid its drawbacks, the dependent partial derivatives of functions with non-independent variables are provided in Section 4. The gradient and the Hessian matrix of such functions are derived in Section 5 by using the framework of differential geometric. We provide an application in Section 6 and conclude this work in Section 7.

General notation

For an integer $d > 0$, let $\mathbf{X} := (X_1, \dots, X_d)$ be a random vector of continuous variables having F as the joint cumulative distribution function (CDF) (i.e., $\mathbf{X} \sim F$). For any $j \in \{1, \dots, d\}$, we use F_{x_j} or F_j for the marginal CDF of X_j and F_j^{-1} for its inverse. Also, we use $(\sim j) := (w_1, \dots, w_{d-1})$ for an arbitrary permutation of $\{1, \dots, d\} \setminus \{j\}$ and $\mathbf{X}_{\sim j} := (X_{w_1}, \dots, X_{w_{d-1}})$. For a function f that includes \mathbf{X} as inputs, we use f'_{x_j} for the formal partial derivative of f w.r.t. X_j , that is, by considering other inputs as constant or independent of X_j , and $\nabla f := f'_{\mathbf{x}} := [f'_{x_1}, \dots, f'_{x_d}]^T$. We use $\frac{\partial f}{\partial x_j}$ for the partial derivative of f w.r.t. X_j that accounts for the dependencies among inputs. We also use $\frac{\partial f}{\partial \mathbf{x}} := \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d} \right]^T \in \mathbb{R}^d$. Of course, we have $\frac{\partial f}{\partial \mathbf{x}} = \nabla f$ for independent inputs.

2. Probabilistic characterization of functions with non-independent variables

In probability theory, it is common to treat input variables as random vectors having some CDFs. For instance, for the inputs that take their values within known domain or space, the Bayesian framework allows for assigning a joint distribution known as *a priori* distribution to such inputs. Without additional information about the inputs, it is common to use non-informative *a priori* distributions such as uniform distributions or Gaussian distributions associated with higher values of the variances (see e.g., [19]).

Functions with non-independent variables include many types of models encountered in practice. An example is the models defined via a given function and equations or inequations connecting its inputs. The resulting inputs that comply with such constraints are often dependent or correlated. In what follows, we are going to use the probability theory for characterizing non-independent variables (see Definition 1).

Definition 1. Consider $n \in \mathbb{N} \setminus \{0\}$ and a function $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$ that includes $\mathbf{X} \sim F$ as inputs.

Then, f is said to be a function with non-independent variables whenever there exists at least a pair $j_1, j_2 \in \{1, \dots, d\}$ with $j_1 \neq j_2$ such that

$$F_{j_1, j_2}(x_{j_1}, x_{j_2}) \neq F_{j_1}(x_{j_1})F_{j_2}(x_{j_2}).$$

Using $\mathcal{N}(\mathbf{0}, \Sigma)$ for the multivariate normal distribution, we can check that a function that includes $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ as inputs with Σ a non-diagonal covariance matrix is a member of the class

of functions defined by

$$\mathcal{D}_{d,n} = \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R}^n : \mathbf{X} \sim F, F(\mathbf{x}) \neq \prod_{j=1}^d F_j(x_j); \mathbf{x} \in \mathbb{R}^d \right\}.$$

2.1. New insight into dependency functions

In this section, we recall useful results about generic dependency models of non-independent variables (see [13, 15, 3, 14, 18]). For a d -dimensional random vector of non-independent variables (i.e., $\mathbf{X} \sim F$), a dependency model of \mathbf{X} consists in expressing a subset of $d - 1$ variables (i.e., $\mathbf{X}_{\sim j}$) as a function of independent variables, including X_j .

Formally, if $\mathbf{X} \sim F$ with $F(\mathbf{x}) \neq \prod_{j=1}^d F_j(x_j)$, then there exists ([13, 15, 3, 14, 18])

- (i) new independent variables $\mathbf{Z} := (Z_{w_1}, \dots, Z_{w_{d-1}})$, which are independent of X_j ;
- (ii) a dependency function $r_j : \mathbb{R}^d \rightarrow \mathbb{R}^{d-1}$;

such that

$$\mathbf{X}_{\sim j} \stackrel{d}{=} r_j(X_j, \mathbf{Z}); \quad \text{and} \quad (X_j, \mathbf{X}_{\sim j}) \stackrel{d}{=} (X_j, r_j(X_j, \mathbf{Z})), \quad (1)$$

where $\mathbf{X}_{\sim j} := (X_{w_1}, \dots, X_{w_{d-1}})$, and $A \stackrel{d}{=} B$ means that the random variables A and B have the same CDF.

It is worth noting that the dependency model is not unique in general. The uniqueness can be obtained under additional conditions given in Proposition 1, which enable the inversion of the dependency function r_j .

Proposition 1. *Consider a dependency model of the continuous random vector $\mathbf{X} \sim F$ given by $\mathbf{X}_{\sim j} = r_j(X_j, \mathbf{Z})$ with a prescribed order (w_1, \dots, w_{d-1}) .*

If X_j is the explanatory variable and the distribution of \mathbf{Z} is prescribed, then

- (i) the dependency model is uniquely defined;*
- (ii) the dependency model is invertible, and the unique inverse is given by*

$$\mathbf{Z} = r_j^{-1}(\mathbf{X}_{\sim j} | X_j). \quad (2)$$

Proof. See Appendix A.

□

It is to be noted that the dependency models (DMs) are vector-valued functions of independent input variables. Thus, DMs ease the calculus of partial derivatives such as the partial derivatives of \mathbf{X} w.r.t. X_j . Moreover, the inverse of a DM avoids working with \mathbf{Z} . A natural choice of the order (w_1, \dots, w_{d-1}) is $(1, \dots, j-1, j+1, \dots, d)$.

2.2. Enhanced implicit functions: dependency functions

In this section, we provide a probabilistic version of the implicit function using DMs.

Consider $\mathbf{X} \sim F$, a sample value of \mathbf{X} given by $\mathbf{x} \in \mathbb{R}^d$ and a function $h : \mathbb{R}^d \rightarrow \mathbb{R}^p$ with $p \leq d$ an integer. When connecting the input variables by p compatible equations, that is,

$$h(\mathbf{X}) = \mathbf{0},$$

the well-known theorem of the implicit function (see Theorem 1 below) claims that for each sample value \mathbf{x}^* satisfying $h(\mathbf{x}^*) = \mathbf{0}$, a subset of \mathbf{X} can be expressed as a function of the others in the neighborhood of \mathbf{x}^* . To recall this theorem, we use $u \subseteq \{1, \dots, d\}$ with $|u| := \text{card}(u) = d - p$, $\mathbf{X}_u := (X_j, \forall j \in u)$, $\mathbf{X}_{\sim u} := (X_j, \forall j \in \{1, \dots, d\} \setminus u)$ and $B(\mathbf{x}_u^*, r_1) \subseteq \mathbb{R}^{d-p}$ (resp. $B(\mathbf{x}_{\sim u}^*, r_2)$) for an open ball centered on \mathbf{x}_u^* (resp. $\mathbf{x}_{\sim u}^*$) with radius r_1 (resp. r_2). Again, $h'_{\mathbf{x}_{\sim u}}(\mathbf{x}^*)$ (resp. $h'_{\mathbf{x}_u}$) is the formal Jacobian of h w.r.t. to $\mathbf{x}_{\sim u}$ (resp. \mathbf{x}_u).

Theorem 1. (*implicit function*)

Assume that $h(\mathbf{x}^*) = \mathbf{0}$ and $h'_{\mathbf{x}_{\sim u}}(\mathbf{x}^*)$ is invertible. Then, there exists a function $g : B(\mathbf{x}_u^*, r_1) \rightarrow B(\mathbf{x}_{\sim u}^*, r_2)$ such that

$$\mathbf{x}_{\sim u} = g(\mathbf{x}_u); \quad \mathbf{x}_{\sim u}^* = g(\mathbf{x}_u^*); \quad \frac{\partial \mathbf{x}_{\sim u}}{\partial \mathbf{x}_u}(\mathbf{x}_u^*) = - [h'_{\mathbf{x}_{\sim u}}(\mathbf{x}^*)]^{-1} h'_{\mathbf{x}_u}(\mathbf{x}^*).$$

While Theorem 1 is useful, it turns out that the implicit function theorem gives a local relationship among the variables. Remark that the DMs derived in Section 2.1 provide the global relationships once the CDFs of the input variables are known. The distribution function of the variables that complies with the constraints given by $h(\mathbf{X}) = \mathbf{0}$ is needed for building the global implicit function. To derive such distribution function, we assume that

(A1) all the constraints $h(\mathbf{X}) = \mathbf{0}$ are compatible.

Under (A1), the constraints $h(\mathbf{X}) = \mathbf{0}$ introduce new dependency structures on the initial CDF F , which matter for our analysis. The probability theory ensures the existence of a distribution function that captures such dependencies.

Proposition 2. Let $\mathbf{X} \sim F$ and $\mathbf{X}^c := \{\mathbf{X} \sim F : h(\mathbf{X}) = \mathbf{0}\} \sim F^c$ be the constrained variables. If (A1) holds, then we have

$$\begin{cases} \mathbf{X} \sim F \\ \text{s.t. } h(\mathbf{X}) := \mathbf{0} \end{cases} \stackrel{d}{=} \mathbf{X}^c,$$

where $\stackrel{d}{=}$ denotes the equality in distribution.

It comes out that introducing constraints on initial variables leads us to work with constrained variables following a new CDF, that is, $\mathbf{X}^c \sim F^c$. Some generic and constrained variables \mathbf{X}^c and their corresponding distribution functions can be found in [14, 20, 15]. When analytical derivations of the CDF of \mathbf{X}^c are hard or impossible, a common practice consists in fitting a distribution function to the observations of \mathbf{X}^c by means of numerical simulations (See [21, 22, 23, 14, 24, 25] for examples of the distributions and densities estimations). Using the new distributions of the input variables, Corollary 1 provides the probabilistic version of the implicit function.

Definition 2. A distribution G is said to be a degenerate CDF whenever G is the CDF of the Dirac measure having δ_a as the probability mass function with $a \in \mathbb{R}$.

Corollary 1. ([14, 15])

Consider a random vector $\mathbf{X}^c := \{\mathbf{X} \sim F : h(\mathbf{X}) = \mathbf{0}\}$ that follows F^c as CDF. Assume (A1) holds and F^c is a non degenerate CDF.

Then, there exists a function $r_j : \mathbb{R}^d \rightarrow \mathbb{R}^{d-1}$ and $d - 1$ independent variables \mathbf{Z} such that \mathbf{Z} is independent of X_j^c and

$$\mathbf{X}_{\sim j}^c = r_j(X_j^c, \mathbf{Z}).$$

Proof. Such result is the DM for the distribution F^c (see Section 2.1).

□

While Corollary 1 gives the explicit function that links $\mathbf{X}_{\sim j}$ to X_j , we can sometime extend that result as follows:

$$\mathbf{X}_{\sim u}^c = r_u(\mathbf{X}_u^c, \mathbf{Z}_u), \quad (3)$$

where \mathbf{Z}_u is a vector of $d - |u|$ independent variables, \mathbf{Z}_u is independent of \mathbf{X}_u^c and $r_u : \mathbb{R}^d \rightarrow \mathbb{R}^{d-|u|}$ (see Section 2.1 and [14]).

Remark 1. We can easily generalize the above process to cope with i) constrained inequations such as $h(\mathbf{X}) < \mathbf{0}$ or $h(\mathbf{X}) > \mathbf{0}$ (see Section 6 and [15]), ii) a mixture of constrained equations and inequations involving different variables.

Remark 2. For a continuous random vector \mathbf{X} having C as the copula and F_j , $j = 1, \dots, d$ as the marginal distributions, an expression of its DM is given by ([3, 14])

$$\begin{aligned} X_1 &= F_1^{-1} \left(C_{1|j}^{-1} (Z_1 | F_j(X_j)) \right) =: r_{j,1} (X_j, Z_1) \\ X_2 &= F_2^{-1} \left(C_{2|j,1}^{-1} (Z_2 | F_j(X_j), F_1(r_{j,1}(X_j, Z_1))) \right) =: r_{j,2} (X_j, Z_1, Z_2) \quad , \\ &\vdots \end{aligned} \quad (4)$$

where $C_{1|j}$ is the conditional copula, $C_{1|j}^{-1}$ is the inverse of $C_{1|j}$, and $r_{j,1}, r_{j,2}$ are real-valued functions.

2.3. Representation of functions with non-independent variables

In general a function may include a group of independent variables as well as groups of non-independent variables such as correlated variables and/or dependent variables. We can then organize such input variables as follows:

(O): the random vector $\mathbf{X} := (X_1, \dots, X_d)$ is consisted of K independent random vector(s) given by $\mathbf{X} = (\mathbf{X}_{\pi_1}, \dots, \mathbf{X}_{\pi_K})$ where the sets π_1, \dots, π_K form a partition of $\{1, \dots, d\}$. The random vector $\mathbf{X}_{\pi_{k_1}} := (X_j, \forall j \in \pi_{k_1})$ is independent of $\mathbf{X}_{\pi_{k_2}} := (X_j, \forall j \in \pi_{k_2})$ for every pair $k_1, k_2 \in \{1, \dots, K\}$ with $k_1 \neq k_2$. Without loss of generality, we use \mathbf{X}_{π_1} for a random vector of $d_1 \geq 0$ independent variable(s); \mathbf{X}_{π_k} with $k \geq 2$ for a random vector of $d_k \geq 2$ dependent variables.

We use $(\pi_{1,k}, \dots, \pi_{d_k,k})$ for the ordered permutation of π_k (i.e., $\pi_{1,k} < \pi_{2,k} \dots < \pi_{d_k,k}$). For any $\pi_{j,k} \in \pi_k$, we use $X_{\pi_{j,k}}$ for an element of \mathbf{X}_{π_k} ; $(\sim \pi_{j,k}) := (\pi_{1,k}, \dots, \pi_{j-1,k}, \pi_{j+1,k}, \dots, \pi_{d_k,k})$ and $\mathbf{Z}_k := (Z_{\pi_{\ell,k}}, \forall \pi_{\ell,k} \in (\sim \pi_{j,k}))$. Bearing in mind the DMs (see Section 2.1), we can represent \mathbf{X}_{π_k} by

$$\mathbf{X}_{\sim \pi_{j,k}} \stackrel{d}{=} r_{\pi_{j,k}} (X_{\pi_{j,k}}, \mathbf{Z}_k), \quad \forall k \in \{2, \dots, K\}, \quad (5)$$

where $r_{\pi_{j,k}}(\cdot) = (r_{\pi_{1,k}}(\cdot), \dots, r_{\pi_{j-1,k}}(\cdot), r_{\pi_{j+1,k}}(\cdot), \dots, r_{\pi_{d_k,k}}(\cdot))$; \mathbf{Z}_k is a random vector of $d_k - 1$ independent variables, and $X_{\pi_{1,k}}$ is independent of \mathbf{Z}_k . Based on the above DM of \mathbf{X}_{π_k} with $k = 2, \dots, K$, let us introduce new functions, that is, $c_{\pi_{j,k}} : \mathbb{R}^{d_k} \rightarrow \mathbb{R}^{d_k}$ given by $c_{\pi_{j,k}}(\cdot) = (r_{\pi_{1,k}}(\cdot), \dots, r_{\pi_{j-1,k}}(\cdot), r_{\pi_{j,k}}(\cdot) = X_{\pi_{j,k}}, r_{\pi_{j+1,k}}(\cdot), \dots, r_{\pi_{d_k,k}}(\cdot))$ and

$$\mathbf{X}_{\pi_k} = c_{\pi_{j,k}}(X_{\pi_{j,k}}, \mathbf{Z}_k) =: (X_{\pi_{j,k}}, r_{\pi_{j,k}}(X_{\pi_{j,k}}, \mathbf{Z}_k)) .$$

The function $c_{\pi_{j,k}}$ maps independent variables $(X_{\pi_{j,k}}, \mathbf{Z}_k)$ onto \mathbf{X}_{π_k} , and the chart

$$\begin{array}{ccc} \mathbb{R}^d & \xrightarrow{c} & \mathbb{R}^d & \xrightarrow{f} & \mathbb{R}^n \\ \begin{pmatrix} \mathbf{X}_{\pi_1} \\ X_{\pi_{j,2}} \\ \mathbf{Z}_2 \\ \vdots \\ X_{\pi_{j,K}} \\ \mathbf{Z}_K \end{pmatrix} & \mapsto & \begin{pmatrix} \mathbf{X}_{\pi_1} \\ \mathbf{X}_{\pi_2} \\ \vdots \\ \mathbf{X}_{\pi_K} \end{pmatrix} = \mathbf{X} & \mapsto & \begin{pmatrix} f_1(\mathbf{X}) \\ \vdots \\ f_n(\mathbf{X}) \end{pmatrix}, \end{array}$$

leads to a new representation of functions with non-independent variables. Indeed, composing f by c yields

$$f(\mathbf{X}_{\pi_1}, \mathbf{X}_2, \dots, \mathbf{X}_{\pi_k}) \stackrel{d}{=} f \circ c(\mathbf{X}_{\pi_1}, X_{\pi_{j,2}}, \mathbf{Z}_2, \dots, X_{\pi_{j,K}}, \mathbf{Z}_K), \quad (6)$$

with $c(\mathbf{X}_{\pi_1}, X_{\pi_{j,2}}, \mathbf{Z}_2, \dots, X_{\pi_{j,K}}, \mathbf{Z}_K) := (\mathbf{X}_{\pi_1}, c_{\pi_{j,2}}(X_{\pi_{j,2}}, \mathbf{Z}_2), \dots, c_{\pi_{j,K}}(X_{\pi_{j,K}}, \mathbf{Z}_K))$.

The equivalent representation of f given by (6) relies on the innovation variables $\mathbf{Z} := (\mathbf{Z}_2, \dots, \mathbf{Z}_K)$. Recall that for the continuous random vector \mathbf{X}_{π_k} , the DM $r_{\pi_{j,k}}$ given by (5) is always invertible (see Proposition 1), and therefore $c_{\pi_{j,k}}$ is also invertible. Such inversions are helpful for working with \mathbf{X} only.

3. Actual partial derivatives

This section deals with the calculus of partial derivatives of functions with non-independent variables using only one relationship among inputs such as the DM given by Equation (5). The usual assumptions made are

(A2) the joint (resp. marginal) CDF is continuous and has a density function $\rho > 0$ on its open support;

(A3) each component of the dependency function $r_{\pi_{j,k}}$ is differentiable w.r.t. $X_{\pi_{j,k}}$;

(A4) each component of the function f , that is, f_ℓ with $\ell \in \{1, \dots, n\}$ is differentiable w.r.t. each input.

Without loss of generality and for the sequel of simplicity, we suppose that $n = 1$ in what follows. Namely, we use $\mathcal{I}_{d \times d} \in \mathbb{R}^{d \times d}$ for the identity matrix; $\mathbf{O}_{d \times d_1} \in \mathbb{R}^{d \times d_1}$ for a null matrix. It is common to use $\nabla f_{\mathbf{x}_{\pi_k}}$ for the formal partial derivatives of f w.r.t. each input of \mathbf{X}_{π_k} (i.e., the derivatives obtained by considering inputs as independent) with $k = 1, \dots, K$. Thus, the formal gradient of f (i.e., the gradient w.r.t. the Euclidean metric) is given by

$$\nabla f := \left[\nabla f_{\mathbf{x}_{\pi_1}}^T \quad \nabla f_{\mathbf{x}_{\pi_2}}^T \quad \dots \quad \nabla f_{\mathbf{x}_{\pi_K}}^T \right]^T. \quad (7)$$

Keeping in mind the function $c_{\pi_{j,k}}(\cdot)$, the partial derivatives of each component of \mathbf{X}_{π_k} w.r.t.

$X_{\pi_{j,k}}$ are given by

$$J^{(\pi_{j,k})} := \frac{\partial c_{\pi_{j,k}}}{\partial \mathbf{x}_{\pi_{j,k}}} = \left[\frac{\partial X_{\pi_{1,k}}}{\partial x_{\pi_{j,k}}} \cdots \frac{\partial X_{\pi_{d_k,k}}}{\partial x_{\pi_{j,k}}} \right]^T = \left[\frac{\partial r_{\pi_{1,k}}}{\partial x_{\pi_{j,k}}} \cdots \underbrace{1}_{j^{\text{th}} \text{ position}} \cdots \frac{\partial r_{\pi_{d_k,k}}}{\partial x_{\pi_{j,k}}} \right]^T. \quad (8)$$

We use $J_i^{(\pi_{j,k})}$ for the i^{th} element of $J^{(\pi_{j,k})}$. For instance, $J_j^{(\pi_{j,k})} = 1$ and $J_{d_k}^{(\pi_{j,k})}$ represents the partial derivative of $X_{\pi_{d_k,k}}$ w.r.t. $X_{\pi_{j,k}}$. It is worth recalling that $J^{(\pi_{j,k})}$ is a vector-valued function of $(X_{\pi_{j,k}}, \mathbf{Z}_k)$, and Lemma 1 expresses $J^{(\pi_{j,k})}$ as a function of \mathbf{x}_{π_k} only.

Lemma 1. *Let \mathbf{x}_{π_k} be a sample value of \mathbf{X}_{π_k} . If assumptions (A2)-(A4) hold, then the partial derivatives of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{j,k}}$ evaluated at \mathbf{x}_{π_k} is given by*

$$J^{(\pi_{j,k})}(\mathbf{x}_{\pi_k}) := \left[\frac{\partial r_{\pi_{1,k}}}{\partial x_{\pi_{j,k}}} \cdots \underbrace{1}_{j^{\text{th}} \text{ position}} \cdots \frac{\partial r_{\pi_{d_k,k}}}{\partial x_{\pi_{j,k}}} \right]^T \left(x_{\pi_{j,k}}, r_{\pi_{j,k}}^{-1}(\mathbf{x}_{\sim \pi_{j,k}} | x_{\pi_{j,k}}) \right). \quad (9)$$

Proof. See Appendix B. □

Again, $J_\ell^{(\pi_{j,k})}(\mathbf{x}_{\pi_k})$ with $\ell \in \{1, \dots, d_k\}$ stands for the ℓ^{th} component of $J^{(\pi_{j,k})}(\mathbf{x}_{\pi_k})$ provided in Lemma 1. Using such components and the chain rule, Theorem 2 provides the actual partial derivatives of functions with non-independent variables (i.e., $\frac{\partial_a f}{\partial \mathbf{x}} =: J_{\pi_k}^a(\mathbf{x}_{\pi_k})$), that is, the derivatives obtained by making use of only one dependency function given by Equation (5).

Theorem 2. *Let $\mathbf{x} \in \mathbb{R}^d$ be a sample value of \mathbf{X} and $\pi_{j,k} \in \pi_k$ with $k = 2, \dots, K$. If assumptions (A2)-(A4) hold, then*

(i) *the actual Jacobian matrix of $c_{\pi_{j,k}}$ is given by*

$$J_{\pi_k}^a(\mathbf{x}_{\pi_k}) := \left[\frac{J^{(\pi_{j,k})}(\mathbf{x}_{\pi_k})}{J_1^{(\pi_{j,k})}(\mathbf{x}_{\pi_k})} \cdots \frac{J^{(\pi_{j,k})}(\mathbf{x}_{\pi_k})}{J_{d_k}^{(\pi_{j,k})}(\mathbf{x}_{\pi_k})} \right], \quad \forall k \in \{2, \dots, K\}. \quad (10)$$

(ii) *The actual Jacobian matrix of c or $\frac{\partial_a \mathbf{X}}{\partial \mathbf{x}}$ is given by*

$$J^a(\mathbf{x}) := \begin{bmatrix} \mathcal{I}_{d_1 \times d_1} & \mathbf{O}_{d_1 \times d_2} & \cdots & \mathbf{O}_{d_1 \times d_K} \\ \mathbf{O}_{d_2 \times d_1} & J_{\pi_2}^a(\mathbf{x}_{\pi_2}) & \cdots & \mathbf{O}_{d_2 \times d_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{d_K \times d_1} & \mathbf{O}_{d_K \times d_2} & \cdots & J_{\pi_K}^a(\mathbf{x}_{\pi_K}) \end{bmatrix}, \quad (11)$$

(iii) *The actual partial derivatives of f are given by*

$$\frac{\partial_a f}{\partial \mathbf{x}}(\mathbf{x}) := J^a(\mathbf{x})^T \nabla f(\mathbf{x}). \quad (12)$$

Proof. see Appendix C. □

Results from Theorem 2 are based only on one dependency function, which uses $X_{\pi_{j,k}}$ as the explanatory input. Thus, the actual Jacobian $J^a(\mathbf{x})$ and the actual partial derivatives of f provided in (10)-(12) are going to change with the choice of the explanatory input $X_{\pi_{j,k}}$ for every $j \in \{1, \dots, d_k\}$. All these possibilities are not surprising. Indeed, while no additional explicit assumption is necessary for calculating the partial derivatives of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{j,k}}$ (i.e., $J^{(\pi_{j,k})}$),

we implicitly keep the other variables fixed when calculating the partial derivative of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{i,k}}$, that is, $\frac{J_i^{(\pi_{j,k})}}{J_i^{(\pi_{j,k})}}$ for each $i \neq j$. Such an implicit assumption is due to the reciprocal rule used to derive the results (see Appendix C). In general, the components of $\mathbf{X}_{\sim\pi_{j,k}}$ such as $X_{\pi_{i_1,k}}$ and $X_{i_2,k}$ are both functions of $X_{\pi_{j,k}}$ and $Z_{\pi_{1,k}}$ at least. Thus, different possibilities of the actual Jacobians are based on different implicit assumptions, and it becomes difficult to use the actual partial derivatives. Further drawbacks of the actual partial derivatives of f are illustrated in Example 1.

Example 1

We consider the function $f(X_1, X_2) = X_1 + X_2 + X_1X_2$, which includes two correlated inputs $\mathbf{X} \sim \mathcal{N}_2\left(\mathbf{0}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$. We see that $\nabla f(\mathbf{X}) = \begin{bmatrix} 1 + X_2 \\ 1 + X_1 \end{bmatrix}$. Using the DM of \mathbf{X} given by (see [3, 14, 15])

$$X_2 = \rho X_1 + \sqrt{1 - \rho^2} Z_2 \iff Z_2 = (X_2 - \rho X_1) / \sqrt{1 - \rho^2},$$

the actual Jacobian matrix of c and the actual partial derivatives of f are given by

$$J^a(\mathbf{X}) = \begin{bmatrix} 1 & \frac{1}{\rho} \\ \rho & 1 \end{bmatrix}; \quad \frac{\partial_a f}{\partial \mathbf{X}}(\mathbf{X}) = \begin{bmatrix} 1 + X_2 + \rho(1 + X_1) \\ \frac{1 + X_2}{\rho} + 1 + X_1 \end{bmatrix}.$$

When $\rho = 1$, both inputs are perfectly correlated, and we have $X_1 = X_2$, which also implies $f(X_1, X_2) = f(X_1) = 2X_1 + X_1^2 = f(X_2) = 2X_2 + X_2^2$. We can check that $\frac{\partial_a f}{\partial \mathbf{X}}(\mathbf{X}) = \begin{bmatrix} 2 + 2X_1 \\ 2 + 2X_2 \end{bmatrix}$. However, when $\rho = 0$, both inputs are independent, and we should expect the actual partial derivatives to be equal to the formal gradient ∇f , but it is not the case. Moreover, using the second DM, that is,

$$X_1 = \rho X_2 + \sqrt{1 - \rho^2} Z_1 \iff Z_1 = (X_1 - \rho X_2) / \sqrt{1 - \rho^2},$$

it comes out that $J^a(\mathbf{X}) = \begin{bmatrix} 1 & \rho \\ \frac{1}{\rho} & 1 \end{bmatrix}$; $\frac{\partial_a f}{\partial \mathbf{X}}(\mathbf{X}) = \begin{bmatrix} 1 + X_2 + \frac{1 + X_1}{\rho} \\ \rho(1 + X_2) + 1 + X_1 \end{bmatrix}$, which differ from the previous results. All these drawbacks are due to the implicit assumptions made (e.g., keeping some variables fixed), which can be avoided (see Section 4).

4. Dependent Jacobian and partial derivatives

This section aims at deriving the first and second-order partial derivatives of functions with non-independent variables without using any additional assumption neither explicit nor implicit. Basically, we are going to calculate or compute the partial derivatives of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{i,k}}$ using only the dependency function that includes $X_{\pi_{i,k}}$ as explanatory input, that is,

$$\mathbf{X}_{\sim\pi_{i,k}} = r_{\pi_{i,k}}(X_{\pi_{i,k}}, \mathbf{Z}_{\pi_{i,k}}), \quad \forall i = 1, \dots, d_k; \quad \forall k = 2, \dots, K.$$

Using the above dependency function, the partial derivatives of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{i,k}}$ is given as follows (see (9)):

$$J^{(\pi_{i,k})}(\mathbf{x}_{\pi_k}) := \frac{\partial \mathbf{X}_{\pi_k}}{\partial x_{\pi_{i,k}}} = \begin{bmatrix} \frac{\partial r_{\pi_{1,k}}}{\partial x_{\pi_{i,k}}} & \dots & \underbrace{1}_{i^{\text{th}} \text{ position}} & \dots & \frac{\partial r_{\pi_{d_k,k}}}{\partial x_{\pi_{i,k}}} \end{bmatrix}^T \left(x_{\pi_{i,k}}, r_{\pi_{i,k}}^{-1}(\mathbf{x}_{\sim\pi_{i,k}} | x_{\pi_{i,k}}) \right).$$

It is to be noted that $J^{(\pi_{i,k})}$ does not require any supplementary assumption, as $X_{\pi_{i,k}}$ and $\mathbf{Z}_{\pi_{i,k}}$ are independent. Thus, d_k different DMs are necessary to derive the dependent Jacobian and partial derivatives of f (see Theorem 3).

Theorem 3. Let $\mathbf{x} \in \mathbb{R}^d$ be a sample value of \mathbf{X} , and assume that (A2)-(A4) hold.

(i) For all $k \geq 2$, the dependent Jacobian matrix $\frac{\partial \mathbf{X}_{\pi_k}}{\partial \mathbf{x}_{\pi_k}}$ is given by

$$J_{\pi_k}^d(\mathbf{x}_{\pi_k}) := \left[J^{(\pi_{1,k})}(\mathbf{x}_{\pi_k}) \quad \dots \quad J^{(\pi_{d_k,k})}(\mathbf{x}_{\pi_k}) \right]. \quad (13)$$

(ii) The dependent Jacobian matrix $\frac{\partial \mathbf{X}}{\partial \mathbf{x}}$ is given by

$$J^d(\mathbf{x}) := \begin{bmatrix} \mathcal{I}_{d_1 \times d_1} & \mathbf{O}_{d_1 \times d_2} & \dots & \mathbf{O}_{d_1 \times d_K} \\ \mathbf{O}_{d_2 \times d_1} & J_{\pi_2}^d(\mathbf{x}_{\pi_2}) & \dots & \mathbf{O}_{d_2 \times d_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{d_K \times d_1} & \mathbf{O}_{d_K \times d_2} & \dots & J_{\pi_K}^d(\mathbf{x}_{\pi_K}) \end{bmatrix}, \quad (14)$$

(iii) The partial derivatives of f are given by

$$\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) := J^d(\mathbf{x})^T \nabla f(\mathbf{x}), \quad (15)$$

Proof. See Appendix D. □

Although the results from Theorem 3 require different DMs, such results are more comprehensive than the actual partial derivatives because no supplementary assumption is available for each non-independent variables.

To derive the second-order partial derivatives of f , we use $f''_{x_i x_j}$ for the formal cross-partial derivative of f w.r.t. x_i, x_j and $H_{\pi_k} := \left(f''_{x_{j_1} x_{j_2}} \right)_{j_1, j_2 \in \pi_k}$ for the formal or ordinary Hessian matrices of f restricted to \mathbf{X}_{π_k} with $k = 1, \dots, K$. In the same sense, we use $H_{\pi_{k_1}, \pi_{k_2}} := \left(f''_{x_{j_1} x_{j_2}} \right)_{j_1 \in \pi_{k_1}, j_2 \in \pi_{k_2}}$ for the formal cross-Hessian matrix of f restricted to $(\mathbf{X}_{\pi_{k_1}}, \mathbf{X}_{\pi_{k_2}})$ for every pair $k_1, k_2 \in \{1, \dots, K\}$ with $k_1 \neq k_2$. To ensure the existence of the second-order partial derivatives, we assume that

(A5): the function f is twice (formal) differentiable w.r.t. each input;

(A6): every dependency function $r_{\pi_{j,k}}$ is twice differentiable w.r.t. $X_{\pi_{j,k}}$.

By considering the d_k DMs of \mathbf{X}_{π_k} (i.e., $\mathbf{X}_{\sim \pi_{i,k}} = r_{\pi_{i,k}}(X_{\pi_{i,k}}, \mathbf{Z}_{\pi_{i,k}})$ with $i = 1, \dots, d_k$) used to derive the dependent Jacobian, we can write

$$\frac{\partial J^{(\pi_{i,k})}}{\partial x_{\pi_{i,k}}}(\mathbf{x}_{\pi_k}) := \frac{\partial^2 \mathbf{X}_{\pi_k}}{\partial^2 x_{\pi_{i,k}}} = \left[\frac{\partial^2 r_{\pi_{1,k}}}{\partial^2 x_{\pi_{1,k}}} \quad \dots \quad \underbrace{0}_{i^{\text{th}} \text{ position}} \quad \dots \quad \frac{\partial^2 r_{\pi_{d_k,k}}}{\partial^2 x_{\pi_{i,k}}} \right]^T \left(x_{\pi_{i,k}}, r_{\pi_{i,k}}^{-1}(\mathbf{x}_{\sim \pi_{i,k}} | x_{\pi_{i,k}}) \right),$$

for the second partial derivatives of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{i,k}}$. Using $\text{diag}(\mathbf{x}) \in \mathbb{R}^{d \times d}$ for a diagonal matrix

with \mathbf{x} the diagonal elements and

$$\mathcal{D}J_{\boldsymbol{\pi}_k}^d(\mathbf{x}) := \text{diag} \left(\left[\frac{\partial J^{(\pi_{1,k})}}{\partial x_{\pi_{1,k}}}(\mathbf{x}_{\boldsymbol{\pi}_k}), \dots, \frac{\partial J^{(\pi_{d_k,k})}}{\partial x_{\pi_{d_k,k}}}(\mathbf{x}_{\boldsymbol{\pi}_k}) \right]^T \nabla f_{\mathbf{x}_{\boldsymbol{\pi}_k}}(\mathbf{x}) \right) J_{\boldsymbol{\pi}_k}^d(\mathbf{x}_{\boldsymbol{\pi}_k}), \quad (16)$$

for all $k \in \{2, \dots, K\}$, Theorem 4 provides the dependent second-order partial derivatives (i.e., $\frac{\partial^2 f}{\partial^2 \mathbf{x}}$).

Theorem 4. *Let \mathbf{x} be a sample value of \mathbf{X} . If (A2), (A5) and (A6) hold, then*

$$\frac{\partial^2 f}{\partial^2 \mathbf{x}}(\mathbf{x}) := \begin{bmatrix} H_{\boldsymbol{\pi}_1}(\mathbf{x}) & H_{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2}(\mathbf{x}) J_{\boldsymbol{\pi}_2}^d(\mathbf{x}_{\boldsymbol{\pi}_2}) & \dots & H_{\boldsymbol{\pi}_1, \boldsymbol{\pi}_K}(\mathbf{x}) J_{\boldsymbol{\pi}_K}^d(\mathbf{x}_{\boldsymbol{\pi}_K}) \\ J_{\boldsymbol{\pi}_2}^d(\mathbf{x}_{\boldsymbol{\pi}_2})^T H_{\boldsymbol{\pi}_2, \boldsymbol{\pi}_1}(\mathbf{x}) & J_{\boldsymbol{\pi}_2}^d(\mathbf{x}_{\boldsymbol{\pi}_2})^T H_{\boldsymbol{\pi}_2}(\mathbf{x}) J_{\boldsymbol{\pi}_2}^d(\mathbf{x}_{\boldsymbol{\pi}_2}) + \mathcal{D}J_{\boldsymbol{\pi}_2}^d(\mathbf{x}) & \dots & J_{\boldsymbol{\pi}_2}^d(\mathbf{x}_{\boldsymbol{\pi}_2})^T H_{\boldsymbol{\pi}_2, \boldsymbol{\pi}_K}(\mathbf{x}) \times J_{\boldsymbol{\pi}_K}^d(\mathbf{x}_{\boldsymbol{\pi}_K}) \\ \vdots & \vdots & \ddots & \vdots \\ J_{\boldsymbol{\pi}_K}^d(\mathbf{x}_{\boldsymbol{\pi}_K})^T H_{\boldsymbol{\pi}_K, \boldsymbol{\pi}_1}(\mathbf{x}) & J_{\boldsymbol{\pi}_K}^d(\mathbf{x}_{\boldsymbol{\pi}_K})^T H_{\boldsymbol{\pi}_K, \boldsymbol{\pi}_2}(\mathbf{x}) J_{\boldsymbol{\pi}_2}^d(\mathbf{x}_{\boldsymbol{\pi}_2}) & \dots & J_{\boldsymbol{\pi}_K}^d(\mathbf{x}_{\boldsymbol{\pi}_K})^T H_{\boldsymbol{\pi}_K}(\mathbf{x}) \times J_{\boldsymbol{\pi}_K}^d(\mathbf{x}_{\boldsymbol{\pi}_K}) + \mathcal{D}J_{\boldsymbol{\pi}_K}^d(\mathbf{x}) \end{bmatrix}.$$

Proof. See Appendix E. □

Example 1 (revisited)

Since $f(X_1, X_2) = X_1 + X_2 + X_1 X_2$ and the DMs of \mathbf{X} are given by

$$X_2 = \rho X_1 + \sqrt{1 - \rho^2} Z_2 \implies Z_2 = (X_2 - \rho X_1) / \sqrt{1 - \rho^2},$$

$$X_1 = \rho X_2 + \sqrt{1 - \rho^2} Z_1 \implies Z_1 = (X_1 - \rho X_2) / \sqrt{1 - \rho^2},$$

we can check that

$$J^d = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}; \quad \frac{\partial f}{\partial \mathbf{x}}(\mathbf{X}) = \begin{bmatrix} 1 + X_2 + \rho(1 + X_1) \\ \rho(1 + X_2) + 1 + X_1 \end{bmatrix}; \quad \frac{\partial^2 f}{\partial^2 \mathbf{x}}(\mathbf{X}) = \begin{bmatrix} 2\rho & 1 + \rho^2 \\ 1 + \rho^2 & 2\rho \end{bmatrix}.$$

For instance, when $\rho = 1$, we have $\frac{\partial f}{\partial \mathbf{x}}(\mathbf{X}) = \frac{\partial_a f}{\partial \mathbf{x}}(\mathbf{X})$, and when $\rho = 0$ we have $\frac{\partial f}{\partial \mathbf{x}}(\mathbf{X}) = \nabla f(\mathbf{X})$ and $\frac{\partial^2 f}{\partial^2 \mathbf{x}}(\mathbf{X}) = H(\mathbf{X})$. Thus, the dependent partial derivatives of f are coherent with the formal gradient and Hessian matrix when the inputs are independent.

5. Expansion of functions with non-independent variables

Although Section 4 provides the partial derivatives and cross-partial derivatives of f , it is misleading to think that infinitesimal increment of f given by $f(\mathbf{X}_{\boldsymbol{\pi}_k} + \epsilon \mathbf{e}_{\pi_{j,k}}) - f(\mathbf{X}_{\boldsymbol{\pi}_k})$ should result

in the individual effect quantified by $\frac{\partial f(\mathbf{X}_{\boldsymbol{\pi}_k})}{\partial x_{\pi_{j,k}}} \epsilon$ with $\mathbf{e}_{\pi_{j,k}} := \left[0, \dots, 0, \underbrace{1}_{\pi_{j,k}^{\text{th}} \text{ position}}, 0, \dots, 0 \right]^T$ and

$\epsilon > 0$. Indeed, moving $X_{\pi_{j,k}}$ turns out to move the other variables partially, and the effects we obtain (i.e., $\nabla_{J_{\mathbf{x}_{\boldsymbol{\pi}_k}}^T} J^{(\pi_{j,k})}$) can be imputed to other variables as well. The dependence structures of such effects are given by the dependent Jacobian matrix $J_{\boldsymbol{\pi}_k}^d(\mathbf{x}_{\boldsymbol{\pi}_k})$ (see Equation (13)). Therefore,

the definition of gradient and the Hessian of f with non-independent variables need an introduction of the tensor metric or the Riemannian tensor.

In differential geometry, the function of the form

$$c_{\pi_{j,k}} : \begin{array}{ccc} \mathbb{R}^{d_k} & \rightarrow & \mathbb{R}^{d_k} \\ (X_{\pi_{j,k}}, \mathbf{Z}_{\pi_{j,k}}) & \mapsto & \mathbf{X} := (X_{\pi_{1,k}}, \dots, X_{\pi_{j,k}}, \dots, X_{\pi_{d_k,k}}); \quad \mathbf{X} \sim_{\pi_{j,k}} = r_{\pi_{j,k}} (X_{\pi_{j,k}}, \mathbf{Z}_{\pi_{j,k}}) \end{array},$$

for every $\pi_{i,k} \in \pi_k$ can be seen as a parametrization of a manifold \mathcal{M}_k in \mathbb{R}^{d_k} . The d_k column entries of the dependent Jacobian matrix $J_{\pi_k}^d(\mathbf{x}_{\pi_k}) \in \mathbb{R}^{d_k \times d_k}$ span a local m_k -dimensional vector space a.k.a. the tangent space at \mathbf{x}_{π_k} with m_k the rank of $J_{\pi_k}^d(\mathbf{x}_{\pi_k})$, that is, the number of linearly independent columns of $J_{\pi_k}^d(\mathbf{x}_{\pi_k})$.

By considering all the K groups of inputs and the corresponding dependent Jacobian matrix $J^d(\mathbf{x})$, we see that the support of the random vector \mathbf{X} is a m -dimensional manifold \mathcal{M} in \mathbb{R}^d with m the rank of $J^d(\mathbf{x})$. When $m \leq d$, we are going to work in the tangent space $T\mathbb{R}^m$ (or local coordinate system) spanned out by the m column entries of $J^d(\mathbf{x})$ that are linearly independent. Working in $T\mathbb{R}^m$ rather than $T\mathbb{R}^d$ ensures that the Riemannian tensor induced by $\mathbf{x} \mapsto \mathbf{x}$ using the dot product is invertible. Since the Riemannian tensor metric is often symmetric, the Moore-Penrose generalized inverse of symmetric matrices ([26, 27, 28]) allows us to keep working in $T\mathbb{R}^d$ in what follows. Using the first fundamental form (see e.g., [9, 10, 11]), the induced tensor metric is defined as the inner-product between the column entries of the dependent Jacobian matrix of the dependency functions, that is,

$$G(\mathbf{x}) := J^d(\mathbf{x})^T J^d(\mathbf{x}). \quad (17)$$

Based on these elements, the gradient and the Hessian matrix are provided in Corollary 2. To that end, we use G^{-1} for the inverse of the metric G given by Equation (17) when $m = d$ or the generalized inverse of G for every $m < d$ ([26, 27, 28]). For any $k \in \{1, \dots, d\}$, the Christoffel symbols are defined by ([29, 9, 11, 30])

$$\Gamma_{ij}^k := \frac{1}{2} \sum_{\ell=1}^{m=d} G_{k\ell}^{-1}(\mathbf{x}) \left(G'_{i\ell, x_j}(\mathbf{x}) + G'_{j\ell, x_i}(\mathbf{x}) - G'_{ij, x_\ell}(\mathbf{x}) \right); \quad \forall i, j = 1, \dots, d,$$

with $G'_{i\ell, x_j}$ the formal partial derivative of $G_{i\ell}$ w.r.t. x_j .

Corollary 2. *Let \mathbf{x} be a sample value of \mathbf{X} , and assume that (A2), (A5)-(A6) hold.*

(i) *The gradient of f is given by*

$$\text{grad}(f)(\mathbf{x}) := G^{-1}(\mathbf{x}) \nabla f(\mathbf{x}), \quad (18)$$

(ii) *The Hessian matrix of f is given by*

$$\text{Hess}_{ij}(f)(\mathbf{x}) := f''_{x_i x_j}(\mathbf{x}) - \sum_{k=1}^{m=d} \Gamma_{ij}^k(\mathbf{x}) f'_{x_k}(\mathbf{x}). \quad (19)$$

Proof. *Points (i)-(ii) result from the definition of the gradient and the Hessian matrix in a Riemannian geometric endowed with the metric G (see [9, 11, 10, 31]).*

□

Taylor's expansion is widely used for approximating functions with independent variables. In what follows, we are concerned with the approximation of a function with non-independent variables. The Taylor-type expansion of a function with non-independent variables is provided in Corollary 3 using the gradient and the Hessian matrix.

Corollary 3. *Let \mathbf{x} , \mathbf{x}_0 be two sample values of \mathbf{X} , and assume (A2), (A5)-(A6) hold. Then, we have*

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^T \text{grad}(f)(\mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \text{Hess}(f)(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0), \quad (20)$$

provided that \mathbf{x} is close to \mathbf{x}_0 .

Proof. *The proof is straightforward using the dot product induced by the tensor metric G in the tangent space and bearing in mind the Taylor expansion provided in [11].*

□

Example 1 (revisited)

For the function in Example 1, we can check that the tensor metric is

$$G(\mathbf{X}) = \begin{bmatrix} 1 + \rho^2 & 2\rho \\ 2\rho & 1 + \rho^2 \end{bmatrix}; \quad G^{-1}(\mathbf{X}) = \frac{1}{(\rho^2 - 1)^2} \begin{bmatrix} 1 + \rho^2 & -2\rho \\ -2\rho & 1 + \rho^2 \end{bmatrix}; \text{ and the gradient is}$$

$$\text{grad}(f)(\mathbf{X}) = \frac{1}{(\rho^2 - 1)^2} \begin{bmatrix} (1 - \rho)^2 + X_2(1 + \rho^2) - 2\rho X_1 \\ (1 - \rho)^2 + X_1(1 + \rho^2) - 2\rho X_2 \end{bmatrix},$$

which comes down to $\nabla f(\mathbf{X})$ when the variables are independent (i.e., $\rho = 0$).

6. Application

In this section, consider three independent input factors $X_j \sim \mathcal{N}(0, 1)$ with $j = 1, 2, 3$, $\mathbf{X} := (X_1, X_2, X_3)$, a constant $c \in \mathbb{R}_+$ and the function

$$f(\mathbf{X}) = X_1^2 + X_2^2 + X_3^2.$$

Also, consider the constraint $f(\mathbf{X}) \leq c$. It is known in [15] (Corollary 4) that the DM of

$$\mathbf{X}^w := (X_1^w, X_2^w, X_3^w) := \{X_j \sim \mathcal{N}(0, 1), j = 1, 2, 3 : X_1^2 + X_2^2 + X_3^2 \leq c\}$$

is given by

$$X_2^w = R_2 Z_2 \sqrt{c - (X_1^w)^2}; \quad X_3^w = R_3 Z_3 \sqrt{c - (X_1^w)^2} \sqrt{1 - Z_2^2}.$$

We can then write

$$\begin{aligned} \frac{\partial X_2^w}{\partial X_1^w} &= -X_1^w R_2 Z_2 \frac{1}{\sqrt{c - (X_1^w)^2}} = \frac{-X_1^w X_2^w}{c - (X_1^w)^2}, \\ \frac{\partial X_3^w}{\partial X_1^w} &= -X_1^w R_3 Z_3 \frac{\sqrt{1 - Z_2^2}}{\sqrt{c - (X_1^w)^2}} = \frac{-X_1^w X_3^w}{c - (X_1^w)^2}. \end{aligned}$$

Using the above derivatives and symmetry among the inputs, the dependent Jacobian and the tensor metric are given by

$$J^d(\mathbf{X}^w) = \begin{bmatrix} 1 & \frac{-X_1^w X_2^w}{c-(X_2^w)^2} & \frac{-X_1^w X_3^w}{c-(X_3^w)^2} \\ \frac{-X_1^w X_2^w}{c-(X_2^w)^2} & 1 & \frac{-X_2^w X_3^w}{c-(X_3^w)^2} \\ \frac{-X_1^w X_3^w}{c-(X_3^w)^2} & \frac{-X_2^w X_3^w}{c-(X_2^w)^2} & 1 \end{bmatrix}, \quad G(\mathbf{X}^w) = [J^d(\mathbf{X}^w)]^T J^d(\mathbf{X}^w).$$

The following partial derivatives of f can be deduced:

$$\frac{\partial f}{d\mathbf{x}^w} = \begin{bmatrix} 1 & \frac{-X_1^w X_2^w}{c-(X_1^w)^2} & \frac{-X_1^w X_3^w}{c-(X_1^w)^2} \\ \frac{-X_1^w X_2^w}{c-(X_2^w)^2} & 1 & \frac{-X_2^w X_3^w}{c-(X_2^w)^2} \\ \frac{-X_1^w X_3^w}{c-(X_3^w)^2} & \frac{-X_2^w X_3^w}{c-(X_3^w)^2} & 1 \end{bmatrix} \begin{bmatrix} 2X_1^w \\ 2X_2^w \\ 2X_3^w \end{bmatrix} = \begin{bmatrix} 2X_1^w \left(1 - \frac{(X_2^w)^2}{c-(X_1^w)^2} - \frac{(X_3^w)^2}{c-(X_1^w)^2}\right) \\ 2X_2^w \left(1 - \frac{(X_1^w)^2}{c-(X_2^w)^2} - \frac{(X_3^w)^2}{c-(X_2^w)^2}\right) \\ 2X_3^w \left(1 - \frac{(X_1^w)^2}{c-(X_3^w)^2} - \frac{(X_2^w)^2}{c-(X_3^w)^2}\right) \end{bmatrix}.$$

For given values of X_1^w , X_2^w and X_3^w and when $c \rightarrow \infty$, we can see that $\frac{\partial f}{dx_1^w} = 2X_1^w$, which is exactly the partial derivative of f when the inputs are independent. Note that $c \rightarrow \infty$ implies that the inputs are independent, as the constraint imposed on \mathbf{X} is always satisfied.

Keeping in mind Equation (6), it is worth noting that the partial derivatives of f can be directly derived by making use of an equivalent DM of \mathbf{X}^w , that is, $(X_2^w)^2 = Z_2(c - (X_1^w)^2)$, $(X_3^w)^2 = Z_3(c - (X_1^w)^2)(1 - Z_2)$, where $(X_1^w)^2 \sim B1(c, 1/2, 2)$, $Z_2 \sim Beta(1/2, 3/2)$ and $Z_3 \sim Beta(1/2, 1)$ are independent with $B1$ the beta distribution of first-kind (see [15], Corollaries 2). Indeed, we have

$$f(\mathbf{X}^w) = (X_1^w)^2(1 - Z_2 - Z_3(1 - Z_2)) + cZ_2 + cZ_3(1 - Z_2) = (X_1^w)^2(1 - Z_2)(1 - Z_3) + cZ_2 + cZ_3(1 - Z_2);$$

and

$$\frac{\partial f}{dx_1^w} = 2X_1^w(1 - Z_2)(1 - Z_3) = 2X_1^w \frac{c - (X_1^w)^2 - (X_2^w)^2 - (X_3^w)^2}{c - (X_1^w)^2},$$

because

$$Z_2 = \frac{(X_2^w)^2}{c - (X_1^w)^2}; \quad 1 - Z_2 = \frac{c - (X_1^w)^2 - (X_2^w)^2}{c - (X_1^w)^2}; \quad Z_3 = \frac{(X_3^w)^2}{c - (X_1^w)^2 - (X_2^w)^2}.$$

As a matter of fact, we obtain the same partial derivatives of f bearing in mind the symmetry among the inputs.

7. Conclusion

A new approach for calculating and computing the partial derivatives, gradient and Hessian of functions with non-independent variables is proposed and studied in this paper. It relies on i) dependency functions that model the dependency structures among dependent variables, including correlated variables, ii) the dependent Jacobian of the dependency functions, and iii) the tensor metric using the dependent Jacobian. Based on the unique tensor metric thanks to the first fundamental form, the unique gradient of a function of non-independent variables is provided. Since the so-called dependent partial derivatives and the dependent Jacobian do not require any additional assumption (which is always the case), such derivatives (including the gradient) should be used.

The results obtained depend on the parameters of the distribution function or the density function of non-independent variables. For the values of such parameters that lead to independent variables, the proposed gradient and partial derivatives come down to the formal gradient or

the gradient w.r.t. the Euclidean metric. In the same sense, the proposed tensor metric comes down to the Euclidean metric using the above values of the parameters of the distribution function.

Using the proposed gradient and Hessian matrix, the Taylor-type expansion of a function with non-independent variables is provided. Although, the generalized inverse of a symmetric matrix is used in some cases, more investigation is needed for the gradient calculus when the tensor metric is not invertible. The proposed gradient is going to be used for i) the development of the active subspaces of functions with non-independent variables, and ii) enhancing the optimization of functions subject to constraints.

Appendix A Proof of Proposition 1

For continuous random variables and prescribed (w_1, \dots, w_{d-1}) , the Rosenblatt transform of $\mathbf{X}_{\sim j}|X_j$ is unique and strictly increasing ([16]). Therefore, the inverse of the Rosenblatt transform of $\mathbf{X}_{\sim j}|X_j$ is also unique ([17]), and we can write

$$\mathbf{X}_{\sim j} \stackrel{d}{=} r'_j(X_j, \mathbf{U}),$$

where $\mathbf{U} \sim \mathcal{U}(0, 1)^{d-1}$. For the prescribed distribution of the $d - 1$ innovation variables $\mathbf{Z} = (Z_{w_i} \sim F_{Z_{w_i}}, i = 1, \dots, d - 1)$, the above model becomes

$$\mathbf{X}_{\sim j} \stackrel{d}{=} r'_j(X_j, U_1, \dots, U_{d-1}) \stackrel{d}{=} r'_j\left(X_j, F_{Z_{w_1}}(Z_{w_1}), \dots, F_{Z_{w_{d-1}}}(Z_{w_{d-1}})\right) = r_j(X_j, \mathbf{Z}),$$

because $Z_{w_i} \stackrel{d}{=} F_{Z_{w_i}}^{-1}(U_i) \iff U_i = F_{Z_{w_i}}(Z_{w_i})$ for continuous variable. Thus, Point (i) holds.

For Point (ii), since $\mathbf{X}_{\sim j} \stackrel{d}{=} r'_j(X_j, \mathbf{U})$ is the inverse of the Rosenblatt transform of $\mathbf{X}_{\sim j}|X_j$, we then have the unique inverse

$$\mathbf{U} = r_j'^{-1}(\mathbf{X}_{\sim j}|X_j),$$

which yields to the unique inverse of the DM. Indeed,

$$(F_{Z_1}(Z_1), \dots, F_{Z_{d-1}}(Z_{d-1})) = r_j'^{-1}(\mathbf{X}_{\sim j}|X_j) \implies \mathbf{Z} = r_j^{-1}(\mathbf{X}_{\sim j}|X_j).$$

Appendix B Proof of Lemma 1

Using the partial derivatives $\frac{\partial r_{\pi_i, k}}{\partial x_{\pi_j, k}}(X_{\pi_j, k}, \mathbf{Z})$ with $i = 1, \dots, d_k$ given by Equation (8) and the unique inverse of $\mathbf{X}_{\sim \pi_j, k} = r_{\pi_j, k}(X_{\pi_j, k}, \mathbf{Z}_k)$ for any distribution of \mathbf{Z}_k given by $\mathbf{Z}_k = r_{\pi_j, k}^{-1}(\mathbf{X}_{\sim \pi_j, k}|X_{\pi_j, k})$ (see Proposition 1), the result is immediate.

Appendix C Proof of Theorem 2

Firstly, using the partial derivatives of \mathbf{X}_{π_k} w.r.t $X_{\pi_j, k}$ in (9), that is, $J^{(\pi_j, k)} = \left[\frac{\partial X_{\pi_1, k}}{\partial x_{\pi_j, k}} \dots \frac{\partial X_{\pi_{d_k}, k}}{\partial x_{\pi_j, k}} \right]^T$ with $\frac{\partial X_{\pi_i, k}}{\partial x_{\pi_j, k}} = \frac{\partial r_{\pi_i, k}}{\partial x_{\pi_j, k}} = J_i^{(\pi_j, k)}$ for any $\pi_i, k \in \pi_k$, the reciprocal rule allows for writing

$$\frac{\partial X_{\pi_j, k}}{\partial X_{\pi_i, k}} = \frac{1}{\frac{\partial X_{\pi_i, k}}{\partial X_{\pi_j, k}}} = \frac{1}{J_i^{(\pi_j, k)}}.$$

Applying the chain rule yields

$$\frac{\partial X_{\pi_{i_1,k}}}{\partial X_{\pi_{i_2,k}}} = \frac{\partial X_{\pi_{i_1,k}}}{\partial X_{\pi_{j,k}}} \frac{\partial X_{\pi_{j,k}}}{\partial X_{\pi_{i_2,k}}} = \frac{J_{i_1}^{(\pi_{j,k})}}{J_{i_2}^{(\pi_{j,k})}}.$$

Thus, the partial derivatives of \mathbf{X}_{π_k} w.r.t. $X_{\pi_{i,k}}$ are given by

$$\frac{\partial \mathbf{X}_{\pi_k}}{\partial X_{\pi_{i,k}}} := \begin{bmatrix} J_1^{(\pi_{j,k})} & \cdots & J_{d_k}^{(\pi_{j,k})} \\ J_i^{(\pi_{j,k})} & \cdots & J_i^{(\pi_{j,k})} \end{bmatrix}^T = \frac{J^{(\pi_{j,k})}}{J_i^{(\pi_{j,k})}},$$

and the actual Jacobian matrix of \mathbf{X}_{π_k} (i.e., $\frac{\partial_a \mathbf{X}_{\pi_k}}{\partial \mathbf{X}_{\pi_k}}$) is given by

$$J_{\pi_k}^a := \begin{bmatrix} J_1^{(\pi_{j,k})} & \cdots & J_{d_k}^{(\pi_{j,k})} \\ J_1^{(\pi_{j,k})} & \cdots & J_{d_k}^{(\pi_{j,k})} \end{bmatrix}.$$

Secondly, under the organization of input variables (O) and for every pair $k_1, k_2 \in \{1, \dots, K\}$ with $k_1 \neq k_2$, we have the following cross-Jacobian matrices:

$$\frac{\partial \mathbf{X}_{\pi_{k_1}}}{\partial \mathbf{X}_{\pi_{k_2}}} = \mathbf{O}_{d_{k_1} \times d_{k_2}}; \quad J_{\pi_1} := \frac{\partial \mathbf{X}_{\pi_1}}{\partial \mathbf{x}_{\pi_1}} = \mathcal{I}_{d_1 \times d_1}.$$

Therefore, the actual Jacobian matrix of \mathbf{X} is given by

$$J^a := \begin{bmatrix} \frac{\partial \mathbf{X}_{\pi_1}}{\partial \mathbf{x}_{\pi_1}} & \frac{\partial \mathbf{X}_{\pi_1}}{\partial \mathbf{x}_{\pi_2}} & \cdots & \frac{\partial \mathbf{X}_{\pi_1}}{\partial \mathbf{x}_{\pi_K}} \\ \frac{\partial \mathbf{X}_2}{\partial \mathbf{x}_{\pi_1}} & \frac{\partial \mathbf{X}_2}{\partial \mathbf{x}_{\pi_2}} & \cdots & \frac{\partial \mathbf{X}_2}{\partial \mathbf{x}_{\pi_K}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{X}_{\pi_K}}{\partial \mathbf{x}_{\pi_1}} & \frac{\partial \mathbf{X}_{\pi_K}}{\partial \mathbf{x}_{\pi_2}} & \cdots & \frac{\partial \mathbf{X}_{\pi_K}}{\partial \mathbf{x}_{\pi_K}} \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{d_1 \times d_1} & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & J_{\pi_2}^a & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & J_{\pi_K}^a \end{bmatrix}.$$

Finally, using the formal gradient of f in (7), that is, $\nabla f := [\nabla f_{\mathbf{x}_{\pi_1}}^T \ \nabla f_{\mathbf{x}_{\pi_2}}^T \ \cdots \ \nabla f_{\mathbf{x}_{\pi_K}}^T]^T$ and bearing in mind the cyclic rule, we can write

$$\frac{\partial f}{\partial \mathbf{x}_{\pi_1}} = \left(\nabla f^T \frac{\partial \mathbf{X}}{\partial \mathbf{x}_{\pi_1}} \right)^T = \nabla f_{\mathbf{x}_{\pi_1}}, \quad \frac{\partial f}{\partial \mathbf{x}_{\pi_k}} = \left(\nabla f^T \frac{\partial \mathbf{X}}{\partial \mathbf{x}_{\pi_k}} \right)^T = \left(\nabla f_{\mathbf{x}_{\pi_k}}^T J_{\pi_k}^a \right)^T,$$

and the actual partial derivatives of f are then given by $\frac{\partial_a f}{\partial \mathbf{x}} = J^a(\mathbf{x})^T \nabla f(\mathbf{x})$.

Appendix D Proof of Theorem 3

For Point (i), building the d_k dependency functions for every explanatory input $X_{\pi_{i,k}}$ with $i = 1, \dots, d_k$, the partial derivatives of \mathbf{X}_{π_k} w.r.t $X_{\pi_{i,k}}$ evaluated at \mathbf{x}_{π_k} is given by (see Equation (9))

$$J^{(\pi_{i,k})}(\mathbf{x}_{\pi_k}) = J^{(\pi_{i,k})} \left((x_{\pi_{i,k}}, r_{\pi_{i,k}}^{-1}(\mathbf{x}_{\pi_{i,k}} | x_{\pi_{i,k}})) \right); \quad i = 1, \dots, d_k.$$

Points (ii)-(iii) are similar to those of Theorem 2 using the dependent Jacobian matrix given by Point (i).

Appendix E Proof of Theorem 4

First, using Equation (15) given by $\frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) := J^d(\mathbf{x})^T \nabla f(\mathbf{x})$, we can extract

$$\frac{\partial f}{\partial \mathbf{x}_{\pi_1}}(\mathbf{x}) = \nabla f_{\mathbf{x}_{\pi_1}}(\mathbf{x}); \quad \frac{\partial f}{\partial \mathbf{x}_{\pi_k}}(\mathbf{x}) = J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x}),$$

By applying the vector-by-vector derivatives of $\frac{\partial f}{\partial \mathbf{x}_{\pi_1}}(\mathbf{x})$ w.r.t. \mathbf{x}_{π_1} and \mathbf{x}_{π_k} , we have

$$\frac{\partial^2 f}{\partial^2 \mathbf{x}_{\pi_1}}(\mathbf{x}) = \frac{\partial [\nabla f_{\mathbf{x}_{\pi_1}}]}{\partial \mathbf{x}_{\pi_1}}(\mathbf{x}) = H_{\pi_1}(\mathbf{x}); \quad \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}_{\pi_1} \partial \mathbf{x}_{\pi_k}} = \frac{\partial [\nabla f_{\mathbf{x}_{\pi_1}}(\mathbf{x})]}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}_{\pi_k}} = H_{\pi_1, \pi_k}(\mathbf{x}) J_{\pi_k}^d,$$

as \mathbf{X}_{π_1} is a vector of independent variables and $\frac{\partial \mathbf{x}}{\partial \mathbf{x}_{\pi_k}} = \begin{bmatrix} \mathbf{O}_{d_1 \times d_k} \\ \vdots \\ J_{\pi_k}^d \\ \vdots \\ \mathbf{O}_{d_K \times d_k} \end{bmatrix}$, bearing in mind the depen-

dent Jacobian matrix provided in Equation (14).

In the same sense, the derivatives of $\frac{\partial f}{\partial \mathbf{x}_{\pi_k}}(\mathbf{x})$ w.r.t. \mathbf{x}_{π_1} and \mathbf{x}_{π_ℓ} with $\ell \neq k$ are

$$\begin{aligned} \frac{\partial^2 f}{\partial \mathbf{x}_{\pi_k} \partial \mathbf{x}_{\pi_1}} &= \frac{\partial [J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}_{\pi_1}} = J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \frac{\partial [\nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}_{\pi_1}} \\ &= J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \frac{\partial [\nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}_{\pi_1}} = J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T H_{\pi_k, \pi_1}(\mathbf{x}), \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 f}{\partial \mathbf{x}_{\pi_k} \partial \mathbf{x}_{\pi_\ell}} &= \frac{\partial [J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}_{\pi_\ell}} = J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \frac{\partial [\nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}_{\pi_\ell}} \\ &= J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \frac{\partial [\nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{x}_{\pi_\ell}} = J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T H_{\pi_k, \pi_\ell}(\mathbf{x}) J_{\mathbf{x}_{\pi_\ell}}^d(\mathbf{x}_{\pi_\ell}). \end{aligned}$$

Finally, we have to derive the quantity $\frac{\partial^2 f}{\partial^2 \mathbf{x}_{\pi_k}} = \frac{\partial [J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}_{\pi_k}}$. For each $\pi_{\ell, k} \in \pi_k$, we can write

$$\begin{aligned} \frac{\partial^2 f}{\partial \mathbf{x}_{\pi_k} \partial x_{\pi_{\ell, k}}} &= \frac{\partial [J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial x_{\pi_{\ell, k}}} \\ &= \frac{\partial [J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T]}{\partial x_{\pi_{\ell, k}}} \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x}) + J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \frac{\partial [\nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial x_{\pi_{\ell, k}}} \\ &= \left[\frac{\partial J_{\pi_k}^{(\pi_1, k)}}{\partial x_{\pi_{\ell, k}}} \dots \frac{\partial J_{\pi_k}^{(\pi_{d_k, k})}}{\partial x_{\pi_{\ell, k}}} \right]^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x}) + J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T \frac{\partial [\nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x})]}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial x_{\pi_{\ell, k}}} \\ &= \left[J_1^{(\pi_{\ell, k})} \frac{\partial J_{\pi_k}^{(\pi_1, k)}}{\partial x_{\pi_1, k}} \dots J_{d_k}^{(\pi_{\ell, k})} \frac{\partial J_{\pi_k}^{(\pi_{d_k, k})}}{\partial x_{\pi_{d_k, k}}} \right]^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x}) + J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T H_{\pi_k}(\mathbf{x}) J^{(\pi_{\ell, k})}(\mathbf{x}_{\pi_k}), \end{aligned}$$

because for all $i \in \{1, \dots, d_k\}$, we can write (thanks to the chain rule)

$$\frac{\partial J^{(\pi_{i,k})}}{\partial x_{\pi_{\ell,k}}} := \left[\frac{\partial^2 X_{\pi_{1,k}}}{\partial^2 x_{\pi_{i,k}}} \frac{\partial x_{\pi_{i,k}}}{\partial x_{\pi_{\ell,k}}} \cdots \frac{\partial^2 X_{\pi_{d_k,k}}}{\partial^2 x_{\pi_{i,k}}} \frac{\partial x_{\pi_{i,k}}}{\partial x_{\pi_{\ell,k}}} \right]^T = \frac{\partial x_{\pi_{i,k}}}{\partial x_{\pi_{\ell,k}}} \frac{\partial J^{(\pi_{i,k})}}{\partial x_{\pi_{i,k}}} = J_i^{(\pi_{\ell,k})} \frac{\partial J^{(\pi_{i,k})}}{\partial x_{\pi_{i,k}}}.$$

Re-organizing the first element of the right-hand terms of the above equation yields

$$\frac{\partial^2 f}{\partial \mathbf{x}_{\pi_k} \partial x_{\pi_{\ell,k}}} = \text{diag} \left(\left[\frac{\partial J^{(\pi_{1,k})}}{\partial x_{\pi_{1,k}}} \cdots \frac{\partial J^{(\pi_{d_k,k})}}{\partial x_{\pi_{d_k,k}}} \right]^T \nabla f_{\mathbf{x}_{\pi_k}}(\mathbf{x}) \right) J^{(\pi_{\ell,k})} + J_{\pi_k}^d(\mathbf{x}_{\pi_k})^T H_{\pi_k}(\mathbf{x}) J^{(\pi_{\ell,k})}.$$

By running $\ell = 1, \dots, d_k$, we obtain the result.

Acknowledgements

We would like to thank the six reviewers and the associate editor for their comments and suggestions that have helped improving this paper.

Declarations

Competing interests

The author declares having no competing financial interests or personal relationships that can influence the work reported in this paper.

Funding

No funding received.

Availability of data and materials

Not applicable.

References

- [1] S. Bobkov, Isoperimetric and analytic inequalities for log-concave probability measures, *The Annals of Probability* 27 (1999) 1903–1921.
- [2] O. Roustant, F. Barthe, B. Iooss, Poincaré inequalities on intervals - application to sensitivity analysis, *Electron. J. Statist.* 11 (2) (2017) 3081–3119.
- [3] M. Lamboni, S. Kucherenko, Multivariate sensitivity analysis and derivative-based global sensitivity measures with dependent variables, *Reliability Engineering & System Safety* 212 (2021) 107519.
- [4] M. Lamboni, Weak derivative-based expansion of functions: ANOVA and some inequalities, *Mathematics and Computers in Simulation* 194 (2022) 691–718.
- [5] T. M. Russi, *Uncertainty quantification with experimental data and complex system models*, University of California, Berkeley, 2010.
- [6] P. Constantine, E. Dow, S. Wang, Active subspace methods in theory and practice: Applications to kriging surfaces, *SIAM Journal on Scientific Computing* 36 (2014) 1500–1524.

- [7] W. Zhang, S. S. Ge, A global implicit function theorem without initial point and its applications to control of non-affine systems of high dimensions, *Journal of Mathematical Analysis and Applications* 313 (1) (2006) 251–261.
- [8] M. Cristea, On global implicit function theorem, *Journal of Mathematical Analysis and Applications* 456 (2) (2017) 1290–1302.
- [9] J. J. Jost, *Riemannian geometry and geometric analysis*, Vol. 6, Springer, 2011.
- [10] P. Petersen, *Riemannian Geometry*, Springer International Publishing AG, 2016.
- [11] S. Sommer, T. Fletcher, X. Pennec, Introduction to differential and riemannian geometry, in: *Riemannian Geometric Statistics in Medical Image Analysis*, Elsevier, 2020, pp. 3–37.
- [12] MITOpenCourseWare, [Non-independent variables](https://ocw.mit.edu/courses/18-02-multivariable-calculus-fall-2007/8515a1da4c8e0b043740c8b7e32775f1_non_ind_variable.pdf), Open course, MIT institute, USA (2007).
URL https://ocw.mit.edu/courses/18-02-multivariable-calculus-fall-2007/8515a1da4c8e0b043740c8b7e32775f1_non_ind_variable.pdf
- [13] A. V. Skorohod, On a representation of random variables, *Theory Probab. Appl* 21 (3) (1976) 645–648.
- [14] M. Lamboni, On dependency models and dependent generalized sensitivity indices, arXiv preprint arXiv2104.12938 (2021).
- [15] M. Lamboni, Efficient dependency models: Simulating dependent random variables, *Mathematics and Computers in Simulation (MATCOM)* 200 (C) (2022) 199–217.
- [16] M. Rosenblatt, Remarks on a multivariate transformation, *Ann. Math. Statist.* 23 (3) (1952) 470–472.
- [17] G. L. O’Brien, The comparison method for stochastic processes, *The Annals of Probability* 3 (1) (1975) 80 – 88.
- [18] M. Lamboni, On Exact Distribution for Multivariate Weighted Distributions and Classification, *Methodology and Computing in Applied Probability* 25 (1) (2023) 1–26.
- [19] C. P. Robert, *From Prior Information to Prior Distributions*, Springer New York, New York, NY, 2007, pp. 105–163.
- [20] F. Durante, C. Ignazzi, P. Jaworski, On the class of truncation invariant bivariate copulas under constraints, *Journal of Mathematical Analysis and Applications* 509 (1) (2022) 125898.
- [21] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* 27 (1956) 832–837.
- [22] E. Parzen, On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33 (1962) 1065–1076.
- [23] V. Epanechnikov, Nonparametric estimation of a multidimensional probability density, *Theory Probab. Appl.* 14 (1969) 153–158.
- [24] A. J. McNeil, R. Frey, P. Embrechts, *Quantitative Risk Management*, Princeton University Press, Princeton and Oxford, 2015.
- [25] F. Durante, C. Sempi, *Principles of copula theory*, CRC/Chapman & Hall, London, 2015.

- [26] E. Moore, On the reciprocal of the general algebraic matrix, *Bulletin of American Mathematical Society* 26 (1920) 394–395.
- [27] E. Moore, General analysis, part 1, *Mem. Amer. Phil. Soc.* 1 (1935) 97–209.
- [28] R. Penrose, A generalized inverse for matrices, *Proc. Cambridge Phil. Soc.* 51 (1955) 406–413.
- [29] H. Rund, Differential-geometric and variational background of classical gauge field theories, *Aequationes mathematicae* 24 (1982) 121–174.
- [30] C. Vincze, On the extremal compatible linear connection of a generalized berwald manifold, *Aequationes mathematicae* 96 (2022) 53–70.
- [31] D. YiHua, Rigid properties for gradient generalized m-quasi-einstein manifolds and gradient shrinking ricci solitons, *Journal of Mathematical Analysis and Applications* 518 (2) (2023) 126702.