



How to choose biomarkers in view of parameter estimation

Jean-Frédéric Gerbeau, Damiano Lombardi, Eliott Tixier

► To cite this version:

Jean-Frédéric Gerbeau, Damiano Lombardi, Eliott Tixier. How to choose biomarkers in view of parameter estimation. 2018. hal-01811158

HAL Id: hal-01811158

<https://inria.hal.science/hal-01811158>

Preprint submitted on 8 Jun 2018

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.

How to choose biomarkers in view of parameter estimation

Jean-Frédéric Gerbeau, Damiano Lombardi, Elliott Tixier

Inria Paris & Laboratoire Jacques-Louis Lions (Sorbonne Université),
2 rue Simone Iff, F-75012 Paris, France

Abstract

In numerous applications in biophysics, physiology and medicine, the system of interest is studied by monitoring quantities, called biomarkers, extracted from measurements. These biomarkers convey some information about relevant hidden quantities, which can be seen as parameters of an underlying model. In this paper we propose a strategy to automatically design biomarkers to estimate a given parameter. Such biomarkers are chosen as the solution of a sparse optimization problem given a user-supplied dictionary of candidate features. The method is in particular illustrated with two realistic applications, one in electrophysiology and the other in hemodynamics. In both cases, our algorithm provides composite biomarkers which improve the parameter estimation problem.

Keywords: feature selection, sparse optimization, inverse problems, electrophysiology, hemodynamics

1. Introduction

In numerous applications in biophysics, physiology and medicine, the system of interest is studied by monitoring a number of quantities, called biomarkers. For example in electrophysiology, an action potential can be viewed through simple quantities like the action potential duration, the amplitude, the rate of depolarization, *etc.* In hemodynamics, the systolic and diastolic pressures, or the pulse wave velocity, are typical biomarkers extracted from pressure measurements.

The biomarkers are obtained by applying a nonlinear map to the signal measured during experiments or clinical observations. They convey some information on hidden quantities, that are not directly measured. For example in hemodynamics, the pulse wave velocity can be linked to the arterial stiffness. When performing parameter estimation for a biophysical model, it is often much more convenient to work with biomarkers than with the whole signals. A natural question is therefore: which biomarker should be chosen to estimate a given parameter?

Biomarkers are usually proposed by the community, based on physical intuition and experimental observations. They are often relevant in qualitatively describing the hidden quantities. However, in most practical applications, although the biomarkers exhibit a good correlation with respect to the hidden quantity they are designed to monitor, they have a non-negligible correlation with respect to others, making them less robust or of difficult interpretation.

In the present work, we propose a strategy to automatically design biomarkers. The basic ideas of our approach are: (1) to design composite biomarkers that are maximally correlated with the hidden quantities they have to reveal, and minimally correlated with respect to all the others; (2) to provide a set of quantities making the parameter estimation better conditioned.

The biomarker design problem may be interpreted as a feature selection problem. Most of the literature considers the problem of selecting features in the input space in order to predict a given output (that may be the output of a computational model). Even though the aim of this work is reverse we will momentarily consider, for the sake of comparison, the biomarkers to be inputs and the parameters of interests to be the outputs. A common strategy to select a subset of the available features is by ranking or eliminating them according to a given criterion or score. This score may be based on a sensitivity analysis (e.g. first-order sensitivity indices [1]), based on information theory (e.g. Fisher information matrix [2] or mutual information [3]) or on the input covariance matrix [4]. For other feature selection techniques and an overview of the field, the interested reader is referred to [5]. Other methods consist in selecting directions in the parameter space. In other words, instead of selecting a subset of features, linear combinations of the features are sought. In Principle Component Analysis directions of maximum variance in the parameter space are sought [6]. The same principle holds for functional-PCA [7], its counterpart applied to the case where the input space is a function space. Neither of these approaches take into account the relationship between inputs and outputs. In Active Subspaces [8], directions in the input space are sought so that the gradient of the output with respect to these directions is maximum. In Partial Least Squares [9], directions are sought in the parameter space so that their covariance with the output is maximum. In this regard, this approach bears some similarities with the present work. Indeed, we look for biomarkers that are maximally correlated with their respective parameters. The main difference is that we also add the constraint that they are minimally correlated with all remaining parameters. Another way to reduce the input space is to perform a sparse linear regression using the Lasso algorithm [10]. The use of the ℓ^1 norm penalization makes this approach similar to the present work even though the cost function to be minimized is different.

Another aspect which makes the present approach different from previous works is that the feature, or biomarker, selection is performed in order to simplify future inverse problems. This issue has been addressed in [11] but is rarely the focus of

feature selection studies. For a comprehensive review of inverse problem techniques, the reader is referred to [12].

Our method is based on a semi-empirical approach. A mathematical model of the system of interest is considered and a database of simulations is built, by taking meaningful scenarios into account. Then, a dictionary of linear and nonlinear forms of the observable is considered. Such a dictionary is user-defined and therefore the efficiency of the proposed method relies on a relevant choice of its entries. Typically, such a dictionary is built by considering first common or “classical” biomarkers defined in the literature and then by incorporating agnostic features to enrich the space spanned by the dictionary entries. The composite biomarker is defined as a linear combination of the elements of the dictionary. The linear combination is sought such that the resulting biomarker is maximally correlated to the hidden quantity it refers to, and minimally correlated with respect to all the others.

From a practical point of view, at the computational cost of one single *offline* database computation (done once for all), the expansion coefficients of the biomarkers on a dictionary of observable forms are computed, for a given experimental setup or physical system. Then this result can be exploited for an unlimited number of experiments. As a by-product, when doing parameter identification, the ℓ^2 -distance in the space of the biomarkers defines a metric which is the ℓ^2 -distance in the space of the hidden quantities, up to a controlled perturbation (see Proposition 1). This makes the inverse problem less ill-conditioned and, in general, easier to solve.

The structure of the work is as follows. In section 2, the notation is introduced and the method is described. In section 3, a property of the proposed composite biomarkers is analyzed. The numerical methods used to solve the problem in practice are described in Section 4. In section 5, three numerical experiments are shown. In the first one, a synthetic case is considered to illustrate the approach and highlight its features and performances. Then, two realistic examples are considered, one in the context of electrophysiology and the other in hemodynamics. In both cases, our algorithm provides composite biomarkers which lead to more efficient inverse problems.

2. Problem setting

2.1. Notation

Let $(\Theta, \mathcal{A}, \mathcal{P})$ be a complete probability space, Θ being the set of outcomes, \mathcal{A} a σ -algebra and \mathcal{P} a probability measure. Let $\mathbb{E}[\cdot]$ denote the expectation operator. In the following, random variables are denoted by boldface letters.

We assume we have an ODE or PDE model, depending on p parameters $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$, which, for the sake of simplicity, are assumed to be zero-mean unit-variance and mutually uncorrelated (see remark 1).

The observable of the model is defined as a function v from \mathbb{R}^p to \mathbb{R}^M which corresponds to the output of the parameterized equations. The observable can be viewed as a random vector \mathbf{v} of the following form:

$$\mathbf{v} : \begin{cases} (\Theta, \mathcal{A}, \mathcal{P}) & \longrightarrow \mathbb{R}^M \\ (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p) & \longmapsto v(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p). \end{cases} \quad (1)$$

Let us now define a biomarker g as a function from \mathbb{R}^M to \mathbb{R} . A biomarker can also be viewed as a random variable \mathbf{g} defined as follows:

$$\mathbf{g} : \begin{cases} (\Theta, \mathcal{A}, \mathcal{P}) & \longrightarrow \mathbb{R} \\ (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p) & \longmapsto g(v(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)). \end{cases} \quad (2)$$

We introduce the *biomarkers dictionary*, a set of N_g biomarkers $\mathcal{G} = \{g_1, \dots, g_{N_g}\}$. The biomarker g_k is referred to as the k -th entry of the dictionary.

The purpose of this work is to derive p *composite biomarkers* corresponding to the p parameters. The composite biomarker $y^{(j)}$ associated with the parameter θ_j is sought as a linear combination of the dictionary entries:

$$y^{(j)} = \sum_{k=1}^{N_g} w_k^{(j)} g_k, \quad \forall j \in \{1, \dots, p\}, \quad (3)$$

where $w_k^{(j)}$, $1 \leq k \leq N_g$, $1 \leq j \leq p$, are the weights of the dictionary decomposition and the unknowns of the problem. Analogously, a random composite biomarker is defined as follows:

$$\mathbf{y}^{(j)} = \sum_{k=1}^{N_g} w_k^{(j)} \tilde{\mathbf{g}}_k, \quad \forall j \in \{1, \dots, p\}, \quad (4)$$

where $\tilde{\mathbf{g}}_k = \mathbf{g}_k - \mathbb{E}[\mathbf{g}_k]$, \mathbf{g}_k is the k -th random dictionary entry as defined in (2). The vector of the unknowns $(w_1^{(j)}, \dots, w_{N_g}^{(j)})$ is denoted by $w^{(j)}$ in what follows.

Example: To illustrate the notation, consider the following parametric ODE model:

$$\begin{cases} \dot{u}(t; \theta_1, \theta_2) &= \theta_1 u(t; \theta_1, \theta_2), \quad t \in [0, T], \\ u(0; \theta_1, \theta_2) &= \theta_2. \end{cases} \quad (5)$$

The output of the model is assumed to be made of the values of the solution u at M time instants t_1, \dots, t_M . Thus the observable $v = (v_1, \dots, v_M)$ is defined as:

$$v : \begin{cases} \mathbb{R}^2 & \longrightarrow \mathbb{R}^M \\ (\theta_1, \theta_2) & \longmapsto (u(t_1; \theta_1, \theta_2), \dots, u(t_M; \theta_1, \theta_2)). \end{cases} \quad (6)$$

The dictionary is assumed to be made of $N_g = 3$ biomarkers $\mathcal{G} = \{g_1, g_2, g_3\}$, $g_i :$

$\mathbb{R}^M \longrightarrow \mathbb{R}$, defined for example by:

$$\begin{aligned} g_1 : (v_1, \dots, v_M) &\longmapsto \max(v_1, \dots, v_M), \\ g_2 : (v_1, \dots, v_M) &\longmapsto \min(v_1, \dots, v_M), \\ g_3 : (v_1, \dots, v_M) &\longmapsto \frac{1}{M} \sum_{i=1}^M v_i. \end{aligned}$$

We will not discuss the choice of the entries of the dictionary: we just assume that they are given and can be applied to the measurements of the phenomenon of interest. Our wish is to build a quantity $y^{(1)}$ (resp. $y^{(2)}$) that is well suited to the identification of the parameter θ_1 (resp. θ_2). More precisely, the algorithm proposed in this work will find two vectors $w^{(1)}$ and $w^{(2)}$ of \mathbb{R}^3 which define the two composite biomarkers as:

$$\begin{aligned} y^{(1)} &= w_1^{(1)} g_1 + w_2^{(1)} g_2 + w_3^{(1)} g_3, \\ y^{(2)} &= w_1^{(2)} g_1 + w_2^{(2)} g_2 + w_3^{(2)} g_3. \end{aligned}$$

Once the problem is solved, Proposition 1 below can give an a posteriori hint about the quality of the dictionary with respect to the parameter estimation problem.

The following quantities will be useful to the definition of the algorithm:

$$\begin{aligned} C_{k,l} &:= \mathbb{E} [\tilde{\mathbf{g}}_k \boldsymbol{\theta}_l], & \text{for } 1 \leq k \leq N_g, 1 \leq l \leq p, \\ V_{k,l} &:= \mathbb{E} [\tilde{\mathbf{g}}_k \tilde{\mathbf{g}}_l], & \text{for } 1 \leq k \leq N_g, 1 \leq l \leq N_g, \\ \nu(w^{(j)}) &= w^{(j)T} V w^{(j)}, & \text{for } 1 \leq j \leq p, \\ e^{(j)} &= (e_l^{(j)})_{l=1,\dots,p}, \text{ with } e_l^{(j)} := \delta_{lj}, & \text{for } 1 \leq j \leq p, 1 \leq l \leq p, \end{aligned} \tag{7}$$

where δ_{lj} is 1 if $l = j$ and 0 if $l \neq j$.

2.2. Problem formulation

For $j \in \{1, \dots, p\}$, we define the composite biomarker $y^{(j)}$ by $\sum_{k=1}^{N_g} w_k^{(j)} g_k$, where $(w_k^{(j)})_{k=1,\dots,N_g}$ is solution to the minimization problem:

$$\boxed{w_*^{(j)} = \arg \min_{w^{(j)} \in \mathbb{R}^{N_g}} \mathcal{J}_\lambda(w^{(j)})} \tag{8}$$

with

$$\mathcal{J}_\lambda(w^{(j)}) = \mathcal{J}(w^{(j)}) + \frac{\lambda^{(j)}}{N_g} \|w^{(j)}\|_{\ell^1(\mathbb{R}^{N_g})}, \tag{9}$$

and

$$\mathcal{J}(w^{(j)}) = \frac{1}{2} \|C^T w^{(j)} - e^{(j)}\|_{\ell^2(\mathbb{R}^p)}^2 + \frac{\xi}{2} (\nu(w^{(j)}) - 1)^2, \tag{10}$$

where the parameters $\lambda^{(j)}$ and ξ are assumed to be given for now. Note that the first part of $\mathcal{J}(w^{(j)})$ is convex while the second one is not.

It follows from the definition of $y^{(j)}$ and the definitions in (7) that

$$\begin{aligned} C^T w^{(j)} &= \mathbb{E} [\boldsymbol{\theta} \mathbf{y}^{(j)}], \\ \nu(w^{(j)}) &= \mathbb{E} [(\mathbf{y}^{(j)})^2]. \end{aligned} \tag{11}$$

Let us explain the rationale of problem (8). We remind the reader that the parameters $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$ are supposed to be zero-mean unit-variance random variables. Thus, the minimization of $\|C^T w^{(j)} - e^{(j)}\|_{\ell^2(\mathbb{R}^p)}^2$ implies that $\mathbf{y}^{(j)}$, defined by $\sum_{k=1}^{N_g} w_k^{(j)} \tilde{\mathbf{g}}_k$, has a covariance with parameter $\boldsymbol{\theta}_j$ (i.e. $\mathbb{E} [\mathbf{y}^{(j)} \boldsymbol{\theta}_j]$) close to 1, and a covariance with the other parameters (i.e. $|\mathbb{E} [\mathbf{y}^{(j)} \boldsymbol{\theta}_l]|, \forall l \neq j$) close to 0. The penalty term $\frac{\epsilon}{2} (\nu(w^{(j)}) - 1)^2$ will force the variance of $\mathbf{y}^{(j)}$ to be close to one. The penalty term $\frac{\lambda^{(j)}}{N_g} \|w^{(j)}\|_{\ell^1(\mathbb{R}^{N_g})}$ tends to promote sparsity. The motivation behind the sparsity promotion is twofold. First, the obtained biomarkers, if most values of $w^{(j)}$ are zero, can be more easily interpreted. Second, ℓ^1 -penalized models tend to be more predictive and less prone to overfitting than ℓ^2 -penalized ones, as proved for example in the case of logistic regression in the comparative study [13].

Remark 1. *The parameters $\boldsymbol{\theta}_i$ have been assumed to be zero-mean unit-variance and mutually uncorrelated. In the general case, the mean and the covariance of the parameters read:*

$$\begin{aligned} \overline{\boldsymbol{\theta}_i} &:= \mathbb{E}[\boldsymbol{\theta}_i], \\ S_{ij}^\vartheta &:= \mathbb{E}[(\boldsymbol{\theta}_j - \overline{\boldsymbol{\theta}_j})(\boldsymbol{\theta}_i - \overline{\boldsymbol{\theta}_i})]. \end{aligned}$$

When the matrix S^ϑ is positive definite, there exists an affine change of coordinates such that the transformed parameters have the properties assumed in the derivation of the method.

Remark 2. *Problem (8) corresponds to a penalization of a minimization problem which has a unique solution (minimization of the strictly convex function \mathcal{J}_λ over the compact set defined by $\nu(w^{(j)}) = 1$).*

3. Analysis of a property of the composite biomarkers

In this Section, we propose an estimation which illustrates the benefits of the composite biomarkers as defined above. For the sake of simplicity, the analysis is carried out in absence of ℓ^1 -penalization, i.e. it is assumed that $\lambda^{(j)} = 0$.

From a parameter identification perspective, an ideal configuration would be to have the j -th composite biomarker $\mathbf{y}^{(j)}$ proportional to the parameter $\boldsymbol{\theta}_j$. So we

propose here to assess how far we are from this configuration and to relate this distance to the value of the functional $\mathcal{J}^{(j)}$. The result can also be viewed as an indication about the quality of the biomarkers dictionary \mathcal{G} .

The composite biomarker $\mathbf{y}^{(j)}$, after the coefficients $w^{(j)}$ have been found, can be seen as a function of the random variables $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p$. We remind that the $\boldsymbol{\theta}_j$ are zero-mean unit-variance and mutually uncorrelated random variables, *i.e* $\mathbb{E}[\boldsymbol{\theta}_j] = 0$, $\mathbb{E}[\boldsymbol{\theta}_j \boldsymbol{\theta}_l] = \delta_{jl}$, $\forall j, l$. The expression of the biomarker $\mathbf{y}^{(j)}$ can be decomposed into a linear part with respect to the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)$ and a non-linear part as follows:

$$\mathbf{y}^{(j)} = \boldsymbol{\theta}^T \boldsymbol{\alpha}^{(j)} + \mathbf{q}^{(j)}, \quad (12)$$

where $\boldsymbol{\alpha}^{(j)} \in \mathbb{R}^p$ and $\mathbf{q}^{(j)} : (\Theta, \mathcal{A}, \mathcal{P}) \rightarrow \mathbb{R}$ are such that $\mathbb{E}[\mathbf{q}^{(j)} \boldsymbol{\theta}_l] = 0, \forall l$.

Proposition 1. *Decomposing the biomarker $\mathbf{y}^{(j)}$ as in (12), we have:*

$$\|\mathbf{q}^{(j)}\|_2 \leq K (\mathcal{J}^{(j)})^{1/4},$$

where K is a constant, $\mathcal{J}^{(j)} = \mathcal{J}(w^{(j)})$ is defined as the minimum of (10), and $\|\mathbf{v}\|_2^2 = \mathbb{E}[\mathbf{v}^2]$.

Proof. For convenience, the vector $\boldsymbol{\alpha}^{(j)}$ is decomposed in $\mathbf{e}^{(j)} + \boldsymbol{\gamma}^{(j)}$, where $e_t^{(j)} = \delta_{lj}$. Let $\boldsymbol{\Delta}^{(j)} = \boldsymbol{\theta}^T \boldsymbol{\gamma}^{(j)}$. Inserting the formulas in (11) into Eq.(10) leads to:

$$\mathcal{J}^{(j)} = \frac{1}{2} \|\mathbb{E}[\boldsymbol{\theta} \mathbf{y}^{(j)}] - \mathbf{e}^{(j)}\|_{\ell^2(\mathbb{R}^p)}^2 + \frac{\xi}{2} \left(\mathbb{E}[(\mathbf{y}^{(j)})^2] - 1 \right)^2, \quad (13)$$

$$= \frac{1}{2} \|\mathbb{E}[\boldsymbol{\theta} \boldsymbol{\theta}_j] + \mathbb{E}[\boldsymbol{\theta} \boldsymbol{\theta}^T] \boldsymbol{\gamma}^{(j)} - \mathbf{e}^{(j)}\|_{\ell^2(\mathbb{R}^p)}^2 + \frac{\xi}{2} \left(\mathbb{E}[\boldsymbol{\theta}_j^2 + \boldsymbol{\Delta}^{(j)2} + \mathbf{q}^{(j)2} + 2\boldsymbol{\Delta}^{(j)} \boldsymbol{\theta}_j] - 1 \right)^2, \quad (14)$$

$$= \frac{1}{2} \|\boldsymbol{\gamma}^{(j)}\|_{\ell^2(\mathbb{R}^p)}^2 + \frac{\xi}{2} \left(\mathbb{E}[\boldsymbol{\Delta}^{(j)2}] + \mathbb{E}[\mathbf{q}^{(j)2}] + 2\mathbb{E}[\boldsymbol{\Delta}^{(j)} \boldsymbol{\theta}_j] \right)^2, \quad (15)$$

$$= \frac{1}{2} \|\boldsymbol{\Delta}^{(j)}\|_2^2 + \frac{\xi}{2} \left(\|\mathbf{q}^{(j)}\|_2^2 + \|\boldsymbol{\Delta}^{(j)}\|_2^2 + 2\mathbb{E}[\boldsymbol{\Delta}^{(j)} \boldsymbol{\theta}_j] \right)^2. \quad (16)$$

First, we notice that $\|\boldsymbol{\Delta}^{(j)}\|_2 \leq \sqrt{2} (\mathcal{J}^{(j)})^{1/2}$. Then, applying the Cauchy-Schwarz inequality and observing that $\|\boldsymbol{\theta}_j\|_2 = 1, \forall j$, we have:

$$\begin{aligned} \|\mathbf{q}^{(j)}\|_2^2 &\leq \frac{1}{\sqrt{\xi}} \left(2\mathcal{J}^{(j)} - \|\boldsymbol{\Delta}^{(j)}\|_2^2 \right)^{1/2} - \|\boldsymbol{\Delta}^{(j)}\|_2^2 - 2\mathbb{E}[\boldsymbol{\Delta}^{(j)} \boldsymbol{\theta}_j] \\ &\leq \sqrt{\frac{2}{\xi}} (\mathcal{J}^{(j)})^{1/2} - \|\boldsymbol{\Delta}^{(j)}\|_2^2 + 2\|\boldsymbol{\Delta}^{(j)}\|_2 \leq \sqrt{\frac{2}{\xi}} (\mathcal{J}^{(j)})^{1/2} + 2\|\boldsymbol{\Delta}^{(j)}\|_2, \\ &\leq K^2 (\mathcal{J}^{(j)})^{1/2} \end{aligned} \quad (17)$$

with $K^2 = \sqrt{2/\xi} + 2\sqrt{2}$. \diamond

The result of the proposition shows that $\mathbf{q}^{(j)}$, which is the nonlinear part (with respect to the parameters) of the biomarker $\mathbf{y}^{(j)}$, has a norm $\|\cdot\|_2$ controlled by the value of the functional $\mathcal{J}^{(j)}$, which is minimized. If the value of the functional is close to zero, then, the norm of $\mathbf{q}^{(j)}$ is close to zero. Thus, the biomarker, which is a nonlinear map of the observable, is close to a linear map of the parameters. In addition, this linear map is a perturbation of the identity. The amplitude of this perturbation is measured by the quantity $\|\Delta^{(j)}\|_2$ introduced in the proof of Proposition 1, and it was noticed that $\|\Delta^{(j)}\|_2 \leq \sqrt{2} (\mathcal{J}^{(j)})^{1/2}$. This shows that if the value of $\mathcal{J}^{(j)}$ is small, then the ℓ^2 -distance of the biomarkers is close to the ℓ^2 -distance of the parameters. This property is particularly appealing when considering parameter identification problems, as will be shown below.

In conclusion, it is desirable that the value of $\mathcal{J}^{(j)}$ be as small as possible. Thus, the value reached by $\mathcal{J}^{(j)}$ once the minimization has been performed can be seen as an *a posteriori* indicator of the quality of the dictionary.

Remark 3. *If the smallness of $\mathcal{J}^{(j)}$ is not deemed satisfactory, the biomarkers dictionary may be augmented with additional entries. Knowing a priori the typical number of entries in the dictionary required to obtain a small enough cost function is a challenging task. This important question goes beyond the scope of the present paper.*

4. Numerical methods

4.1. Discretization

The expectation operators appearing in the definition (7) are approximated by a Monte-Carlo method. In the *offline phase*, N samples of the random parameters $(\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_p)$ are drawn and stored in the matrix $P \in \mathbb{R}^{N \times p}$. For each sample $(\theta_{1,i}, \dots, \theta_{p,i})$, $i \in \{1, \dots, N\}$, the computational model is evaluated and $v(\theta_{1,i}, \dots, \theta_{p,i})$ is computed. The entries of the matrix $G \in \mathbb{R}^{N \times N_g}$ are then defined by:

$$G_{i,k} = g_k(v(\theta_{1,i}, \dots, \theta_{p,i})) - \frac{1}{N} \sum_{i=1}^N g_k(v(\theta_{1,i}, \dots, \theta_{p,i})). \quad (18)$$

4.2. Solution of the optimization problem

The ℓ^1 penalty term makes the objective function \mathcal{J}_λ not differentiable. A standard way to deal with this difficulty is to introduce the change of variables $w^{(j)} = w_+^{(j)} - w_-^{(j)}$, where $w_+^{(j)}$ (resp. $w_-^{(j)}$) is the positive (resp. negative) part of

$w^{(j)}$ [14]. The minimization problem can then be recast as follows:

$$\begin{cases} \min_{w_+^{(j)}, w_-^{(j)}} & \mathcal{J}(w_+^{(j)} - w_-^{(j)}) + \frac{\lambda^{(j)}}{N_g} (\mathbb{1}^T w_+^{(j)} + \mathbb{1}^T w_-^{(j)}) \\ \text{s.t.} & w_+^{(j)} \geq 0 \\ & w_-^{(j)} \geq 0 \end{cases}, \quad (19)$$

where $\mathbb{1}$ denotes the vector of ones in \mathbb{R}^{N_g} . Note that while the non-differentiability issue of the ℓ^1 -norm is avoided, the number of unknowns is doubled and $2N_g$ inequality constraints are added.

Many algorithms can be used to solve the constrained optimization problem in (19). In the present work, we choose the Nesterov accelerated gradient descent method [15, 16] in which, at each step, the unknowns are projected onto the constraint set if they do not satisfy the inequality constraints. Given the type of constraints, this projection is straightforward. The gradient to be used in the Nesterov gradient descent iterations reads:

$$\begin{aligned} \partial_{w_+^{(j)}} \mathcal{J}_\lambda &= \frac{1}{N} G^T P \left(\frac{1}{N} P^T G w^{(j)} - e^{(j)} \right) + \frac{2\xi}{N} \left(\frac{1}{N} w^{(j)T} G^T G w^{(j)} - 1 \right) G^T G w^{(j)} + \frac{\lambda^{(j)}}{N_g} \mathbb{1}, \\ \partial_{w_-^{(j)}} \mathcal{J}_\lambda &= -\frac{1}{N} G^T P \left(\frac{1}{N} P^T G w^{(j)} - e^{(j)} \right) - \frac{2\xi}{N} \left(\frac{1}{N} w^{(j)T} G^T G w^{(j)} - 1 \right) G^T G w^{(j)} + \frac{\lambda^{(j)}}{N_g} \mathbb{1}. \end{aligned} \quad (20)$$

4.3. Choice of the penalty parameter ξ

A very simple automatic tuning of the penalty parameter ξ is performed along the resolution of the minimization problem (19): starting from an initial guess, it is increased until the constraint is verified up to a certain tolerance parameter ϵ_c . The strategy is summarized in Algorithm 1 in the Annex A1.

4.4. Choice of the penalty parameter $\lambda^{(j)}$

Two strategies are proposed to set the regularization parameters $\lambda^{(1)}, \dots, \lambda^{(p)}$: “L-curve” criterion and training error threshold.

“L-curve” criterion. A common way to set the regularization parameters involves computing the so-called “L-curve” which represents the balance between the sparsity of the solution and the corresponding training error. Here, the sparsity of the solution is measured by $\|w^{(j)}\|_{\ell^1(\mathbb{R}^{N_g})}$. The training error corresponding to the biomarker $y^{(j)}$ is defined as $\|\frac{1}{N} P^T G w^{(j)} - e^{(j)}\|_{\ell^2(\mathbb{R}^p)}$ and quantifies how well the discrete biomarker $y^{(j)} = G w^{(j)}$ fulfills the objectives for a given training set $\{P, G\}$. The problem in (19) is solved for different values of $\lambda^{(j)}$ and therefore one L-curve is obtained for each biomarker $y^{(j)}$ (i.e. for each parameter θ_j) for $j = 1, \dots, p$. In

Figure 1 an example of L-curves is provided when considering the first two parameters of the model presented in Section 5.3. A natural way to determine $\lambda^{(j)}$ from the

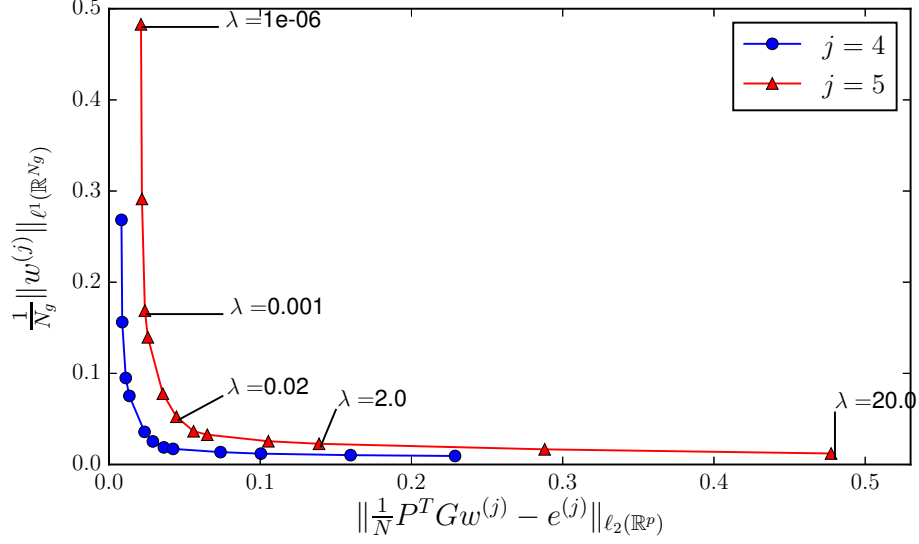


Figure 1: The “L-curve” is obtained for the first two parameters of the test case presented in Section 5.3. The ℓ^1 -norm of $w^{(j)}$ is plotted against the training error for different values of $\lambda^{(j)}$. Here, $\lambda^{(j)}$ varies between 10^{-6} and 20.

L-curve is to choose the $\lambda^{(j)}$ that maximizes its curvature. It is, in practice, a good compromise between an over-fitted ($\lambda^{(j)}$ is too small) and an over-regularized ($\lambda^{(j)}$ is too large) solution, as detailed in [17] where several approaches are compared. It also has the advantage of not requiring other hyperparameters.

Training error threshold. Another way to set the regularization parameters is to choose the largest $\lambda^{(j)}$ so that a user-defined training error is respected. While easy to implement, this requires an additional tuning parameter.

5. Numerical Experiments

5.1. Parameter identification algorithm

One of the outcomes of the proposed methodology is to make the parameter estimation easier. Before illustrating this feature in numerical experiments, we briefly present in this section the strategy used to identify parameters.

The parameter estimation is formulated as an optimization problem, in which the misfit between measurements and the simulated observables is minimized. The cost function has the following form:

$$J_1(\theta) = \frac{1}{2} \|v(\theta) - v^*\|_{\ell^2(\mathbb{R}^M)}^2, \quad (21)$$

where $v^* \in \mathbb{R}^M$ denotes the vector of measurements from which the parameters true values θ^* are to be estimated. In the following numerical experiments, v^* is simply a noisy evaluation of the model:

$$v^* = v(\theta^*) + (\eta_1, \dots, \eta_M) \quad (22)$$

where the η_i are independent realizations of $\boldsymbol{\eta} \sim \mathcal{N}(0, \tau^2)$.

When considering the composite biomarkers, the cost function reads:

$$J_2(\theta) = \frac{1}{2} \|y(v(\theta)) - y(v^*)\|_{\ell^2(\mathbb{R}^p)}^2. \quad (23)$$

The minimization is performed with a simplification of Powell's BOBYQA algorithm [18] which relies on successive quadratic approximations of the cost function. Starting from an initial guess, the cost function is locally evaluated and approximated by a quadratic function from which the minimum is easily computed. This iterative process does not require any gradient evaluation and converges, in practice, faster than the classical gradient descent. The algorithm, later referred to as gradient-free minimization, is presented in Algorithm 2 in the Annex A2.

5.2. A synthetic model

To start with, we illustrate our algorithm with a simple algebraic model defined as follows:

$$\begin{aligned} u(x; \theta_1, \theta_2) = & x^2 \exp\left(-\frac{(\theta_1^2 + \theta_2^2 - R^2)^2}{2\sigma_0^2}\right) + \theta_1 \exp\left(-\frac{(x - (x_1 + \alpha\theta_1))^2}{2\sigma_1^2}\right) \\ & + \theta_2 \exp\left(-\frac{(x - (x_2 + \alpha\theta_2))^2}{2\sigma_2^2}\right), \quad x \in [-1, 1] \end{aligned} \quad (24)$$

with the following fixed parameters: $R = 0.75$, $x_1 = -0.5$, $x_2 = 0.5$, $\alpha = 0.1$, $\sigma_0 = 0.2$ and $\sigma_1 = \sigma_2 = 0.1$. Here, x is the state variable and the uncertain parameters are θ_1 and θ_2 . The segment $[-1, 1]$ is discretized into a uniform grid of $M = 201$ nodes x_1, \dots, x_M . The observable v is defined as follows:

$$v : \begin{cases} \mathbb{R}^2 & \longrightarrow \mathbb{R}^M \\ (\theta_1, \theta_2) & \longmapsto (u(x_1; \theta_1, \theta_2), \dots, u(x_M; \theta_1, \theta_2)) \end{cases} \quad (25)$$

The biomarkers dictionary is composed of $N_g = M$ entries defined as follows:

$$g_k : \begin{cases} \mathbb{R}^M & \longrightarrow \mathbb{R} \\ (v(\theta_1, \theta_2)_1, \dots, v(\theta_1, \theta_2)_M) & \longmapsto v(\theta_1, \theta_2)_k \end{cases}, \quad 1 \leq k \leq N_g. \quad (26)$$

Note that, in this particular example, $(g_1(v), \dots, g_{N_g}(v))$ and v coincide but this is not always the case. More complex transformations of the observable will be presented

in the next numerical examples.

The biomarkers dictionary entries are the values of the model at each node of the discretized domain $u(x_1), \dots, u(x_{N_g})$. The training set is generated with $N = 10^3$ samples of (θ_1, θ_2) drawn from the normal distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu = (0, 0)$ and $\Sigma = 2.5 \cdot 10^{-1} \times I_2$.

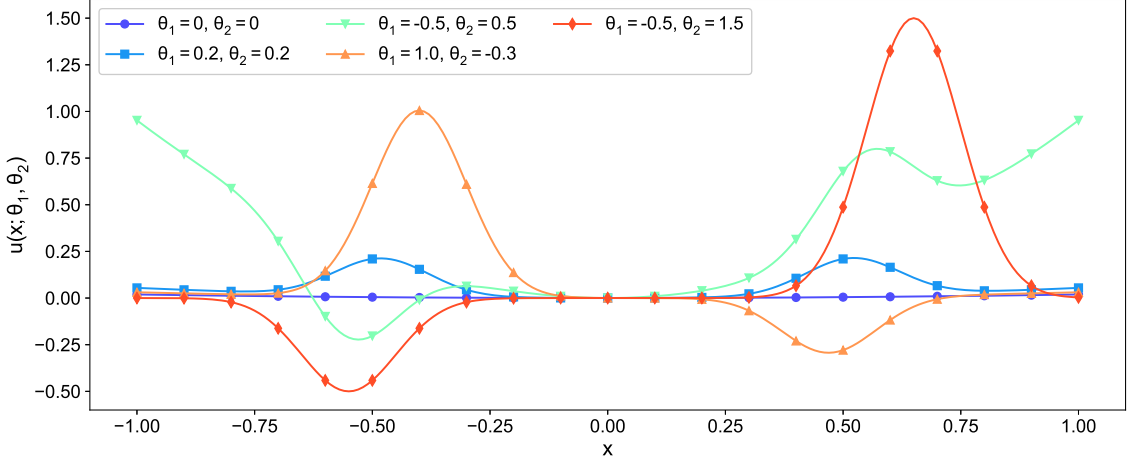


Figure 2: Model outputs for different values of θ_1, θ_2 .

5.2.1. Biomarkers computation

The procedure described in Section 4 is applied to the present test case. The thresholding method is used to choose the regularization parameters $\lambda^{(1)}$ and $\lambda^{(2)}$. The method in Algorithm 1 is used with a constraint tolerance $\epsilon_c = 10^{-3}$. The threshold on the training error is set to 10^{-2} which yields $\lambda^{(1)} = 0.11$ and $\lambda^{(2)} = 0.09$. The obtained biomarkers weights $w^{(1)}$ and $w^{(2)}$ are represented in Fig. 3 as well as the composite biomarkers $y^{(1)}$ and $y^{(2)}$ with respect to the parameters. Note that most coefficients are zero, as expected with the sparsity promoting ℓ^1 penalty.

5.2.2. Parameter estimation

Here, we highlight how using the previously obtained composite biomarkers leads to easier inverse problems. In the context of a parameter estimation problem, one seeks the true values of the uncertain parameters θ^* from a measurement v^* as defined in (22) with $\theta^* = (0.3, 0.3)$ and $\tau = 2 \cdot 10^{-2}$. We propose to perform a parameter estimation using J_1 and J_2 with a gradient-free minimization. The cost functionals J_1 and J_2 as function of θ_1, θ_2 are plotted in Fig. 4. While J_1 features local minima and is non convex, the functional J_2 has only one global minimum, corresponding to the true solution. The results of the parameter estimations are summarized in Table 1. As expected, the parameter estimation fails when using J_1 in the sense that the distance to the true solution at the end of the minimization

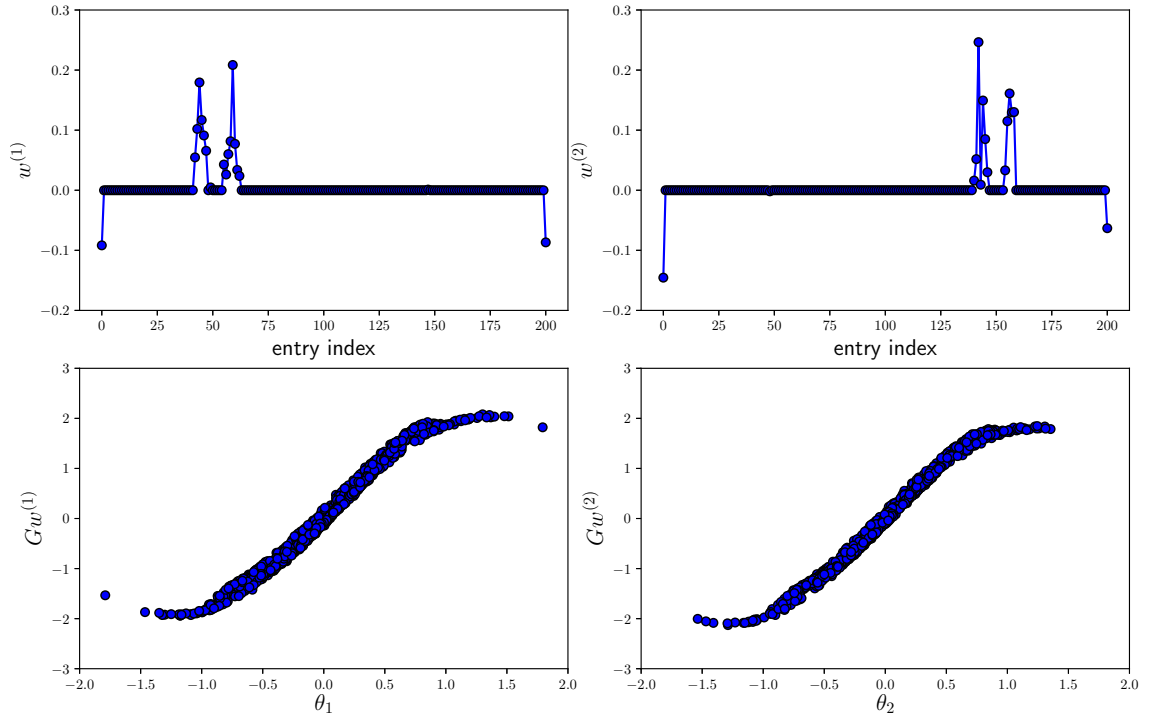


Figure 3: Analytical model: biomarkers weights (top) and corresponding biomarkers (bottom).

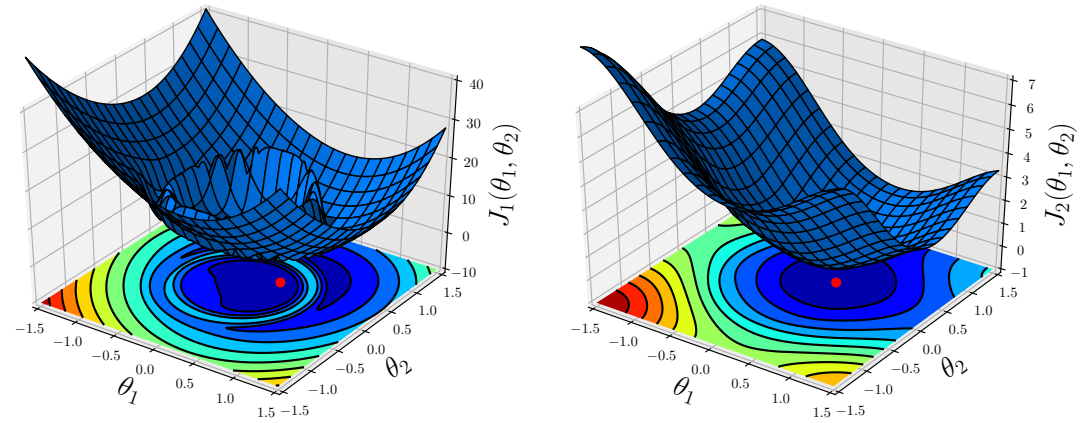


Figure 4: Contours of the inverse problem cost function with true solution $\theta^* = (0.3, 0.3)$ represented by a red dot. J_1 using the raw model outputs (left) and J_2 using the composite biomarkers (right).

procedure is large. The minimization method falls, for J_1 , into a local minimum as seen in Fig. 4. On the contrary, the minimization succeeds in finding the true solution when using J_2 (using the composite biomarkers).

5.2.3. Influence of hyperparameters

Using the analytical model, we now investigate the influence of the regularization parameters $\lambda^{(1)}, \lambda^{(2)}$ and the training set parameters samples (denoted in matrix form by P). For the sake of simplicity and because θ_1 and θ_2 have essentially symmetrical effects, the study will be restricted to the cases where $\lambda^{(1)} = \lambda^{(2)}$. In Figure 5, the cost function $J_2(\theta_1, \theta_2)$ associated with the inverse problem in (23) is represented for a regularization parameter increasing from 10^{-5} to 10^1 . For high values of $\lambda^{(j)}$, the cost function loses its convexity since the penalty term in (8) becomes predominant.

In Figure 6 is represented $J_2(\theta_1, \theta_2)$ for different training set parameters samples P . Using the same number of samples ($N = 10^3$) and same standard deviation, the samples are drawn from a normal distribution with different means. The method is robust with respect to the choice of training set in the sense that the cost function remains convex even when the training set center is far from the true solution, except in the last case ($\mu = (-2, -2)$).

5.3. A model in electrophysiology

In this section, the proposed methodology is applied to a cardiomyocyte (heart muscle cell) electrophysiology model. Cardiomyocytes are contractible cells that react to an electrical activation. Detailed electrophysiology models have been developed to reproduce the cardiomyocytes complex electrical activity. Such models consist in general in a set of nonlinear coupled ODEs of the following form:

$$\begin{cases} \frac{dV}{dt} &= \sum_{i=1}^m \Phi_i(V, t), \\ \Phi_i(V, t) &= \kappa_i f_i(V, z_i) & \text{for } 1 \leq i \leq m, \\ \frac{dz_i}{dt} &= \zeta_i(V, z_i) & \text{for } 1 \leq i \leq m. \end{cases} \quad (27)$$

	$\ \theta^{(n_{\text{iter}})} - \theta^*\ _{\ell^2(\mathbb{R}^p)}$	n_{iter}
J_1	1.31×10^0	7
J_2	6.14×10^{-3}	18

Table 1: Parameter estimation using Powell’s method. Both approaches (J_1 and J_2) are compared in terms of number of iterations to reach convergence and distance from the converged solution $\theta^{(n_{\text{iter}})}$ to the true solution θ^*

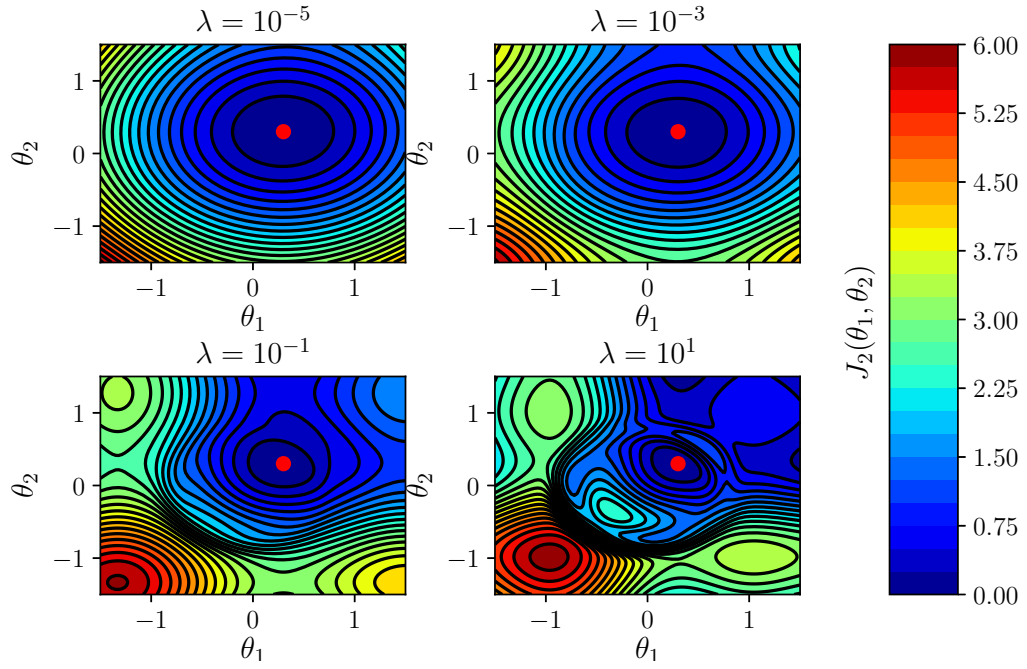


Figure 5: Cost function $J_2(\theta_1, \theta_2)$ with different regularization parameters $\lambda = \lambda^{(1)} = \lambda^{(2)}$ and parameters samples centered on $\mu = (0, 0)$. The red dot materializes the position of the true solution $\theta^* = (0.3, 0.3)$.

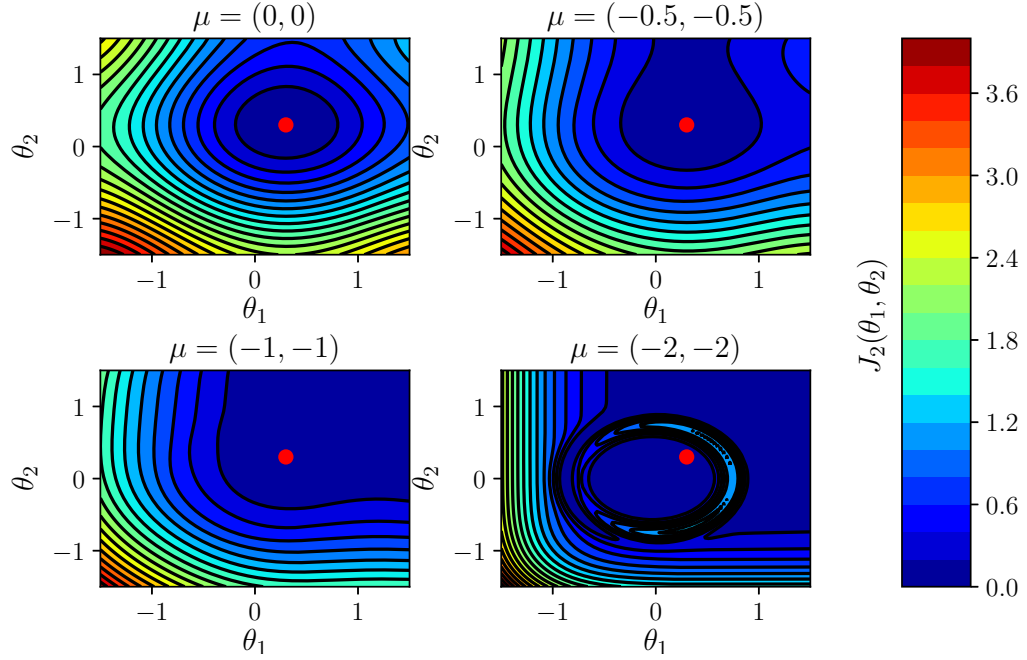


Figure 6: Cost function $J_2(\theta_1, \theta_2)$ with different training sets and $\lambda^{(1)} = \lambda^{(2)} = 10^{-3}$. The red dot materializes the position of the true solution $\theta^* = (0.3, 0.3)$.

Despite a wide variety of models, most of them share this common structure. The integer m corresponds to the number of different kinds of ionic currents and ranges from 1 to more than 30 depending on the model's precision level. In this paper, the model by Davies *et al.* [19] was chosen. It counts $m = 17$ ionic currents, 29 ODEs and is designed to model cardiomyocytes belonging to the ventricular region of the canine heart. The observable quantity is the action potential (AP), denoted by V in (27). It corresponds to the electrical potential difference between the inside and the outside of the cell. It drives the heart contraction and its alteration is at the origin of many cardiac diseases. In Figure 8 is plotted a typical AP along with some biomarkers that are commonly extracted from it. In our case, the parameters of interest are the conductances, κ_i in (27). They are the parameters commonly calibrated to fit different kind of cells or to model diseases and drug-induced AP alterations. We chose to retain 5 of these conductances as the parameters of interest, namely D_{Kr} , D_{K1} , C_{to1} , $D_{Ca,L}$ and D_{NaCa} . Each parameter θ_j , $1 \leq j \leq 5$ is a multiplicative factor of the corresponding parameter reference value, which is equivalent to making the following modifications in the original model:

$$\begin{aligned} D_{Kr} &= \theta_1 \bar{D}_{Kr}, & D_{K1} &= \theta_2 \bar{D}_{K1}, & C_{to1} &= \theta_3 \bar{C}_{to1}, & D_{Ca,L} &= \theta_4 \bar{D}_{Ca,L}, \\ D_{NaCa} &= \theta_5 \bar{D}_{NaCa}, \end{aligned}$$

where the bar symbol refers to the published values. Estimating this kind of parameters has many applications in electrophysiology, ranging from fitting models to experiments [20, 21, 22] to obtaining insights into disease-induced cell modifications [23]. Most often, this inverse problem is solved by using biomarkers.

Figure 9 shows how correlated these 5 parameters of interest are with their classically associated biomarkers. The latter are, respectively, APD_{90} , RMP , V_{notch} , APD_{50}/APD_{90} , APD_{50} (see Figure 8). This choice of biomarkers reflects the common associations made in the literature [19, 24, 25]. Without surprise, this result shows that the classically adopted biomarkers are indeed correlated with the parameters of interest. Nevertheless, it also shows that each feature is not maximally correlated with its corresponding parameter and that it is also correlated with the other parameters. There is therefore room for improvement, and we propose to test if our strategy to define composite biomarkers can lead to a better parameter estimation.

We apply our method to a dictionary of 104 entries computed from $V(t) + \eta$, where $\eta \sim \mathcal{N}(0, \tau)$ with $\tau = 0.5$. More details about the dictionary entries may be found in the Annex A3. This corresponds to a SNR of 43 dB. The dictionary entries are the AP values at different times, the products of AP values at different times, the time derivatives, the integral over time, the PCA coefficients and the “classical” biomarkers (see Figure 8). The training set is generated with $N = 5 \times 10^4$ samples drawn from a uniform distribution over the hypercube $[0.5, 3]^5$. This means the

conductances κ_i are allowed to vary from half to three times their reference values. The penalization parameter ξ is fixed and set to 1. *A posteriori* verifications show that the error on the variance constraint does not exceed 10^{-2} in practice.

The regularization parameters are set using the L-curve criterion.

As an example, Figure 7 shows the resulting biomarker for parameter θ_3 . Interestingly, the algorithm finds that the composite biomarker is a correction of V_{notch} , which is indeed the biomarkers classically associated with parameter θ_3 .

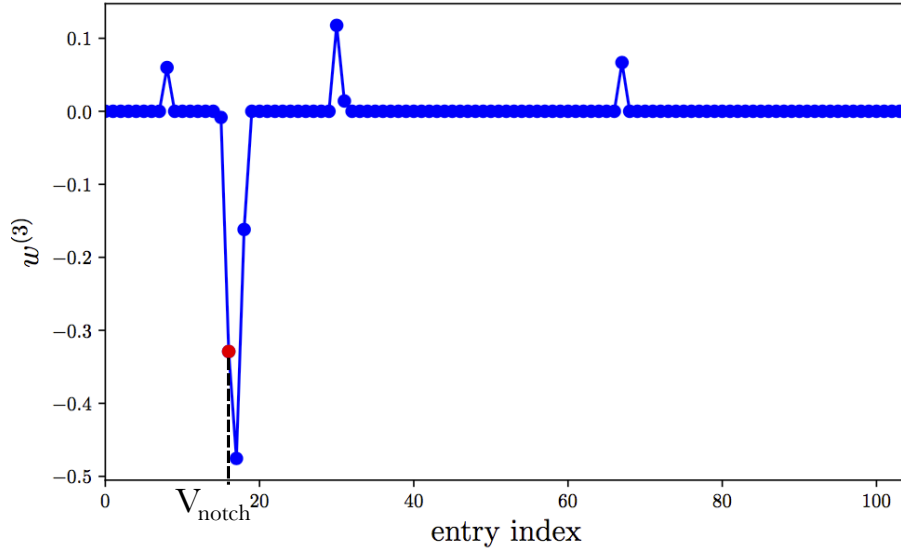


Figure 7: Biomarker weights $w^{(3)}$ associated with θ_3 . The composite biomarker is actually a correction of the classically associated one V_{notch} .

The obtained composite biomarkers are now used to solve parameter estimation problems. Figure 10 shows an example of such an inverse problem. Given a certain true value θ^* and initial guess $\theta^{(0)}$, both in $[0.5, 3]^5$, the parameter estimation is carried out by minimizing $J_1(\theta)$ and $J_2(\theta)$ with the gradient-free method described in Algorithm 2. In this specific example, minimizing J_1 fails to reach the true solution because it falls into a local minimum, whereas minimizing J_2 leads to the true solution.

We now propose to illustrate the advantage of using composite biomarkers instead of the time series to identify parameters. Since the efficiency and the accuracy of the parameter estimation strongly depends on the initial guess and the sought values, we carry out a statistical survey: 25 samples are randomly drawn from $[0.5, 3]^5$ and an inverse problem is executed for each two-point combination. This amounts to 625 inverse problems. Each inverse problem is carried out twice by minimizing J_1 and J_2 and the estimation error is stored. Here, the estimation error is simply defined as the 2-norm between the true solution and the solution of the minimization

problem. Figure 11 shows the histogram of estimation errors in both cases. When using the composite biomarkers instead of the whole time series, the estimation error is in average smaller and is less spread (lower standard deviation). In the first case, the estimation error has mean 0.12 and standard deviation 0.16. In the second case, the estimation error has mean 0.44 and standard deviation 0.43.

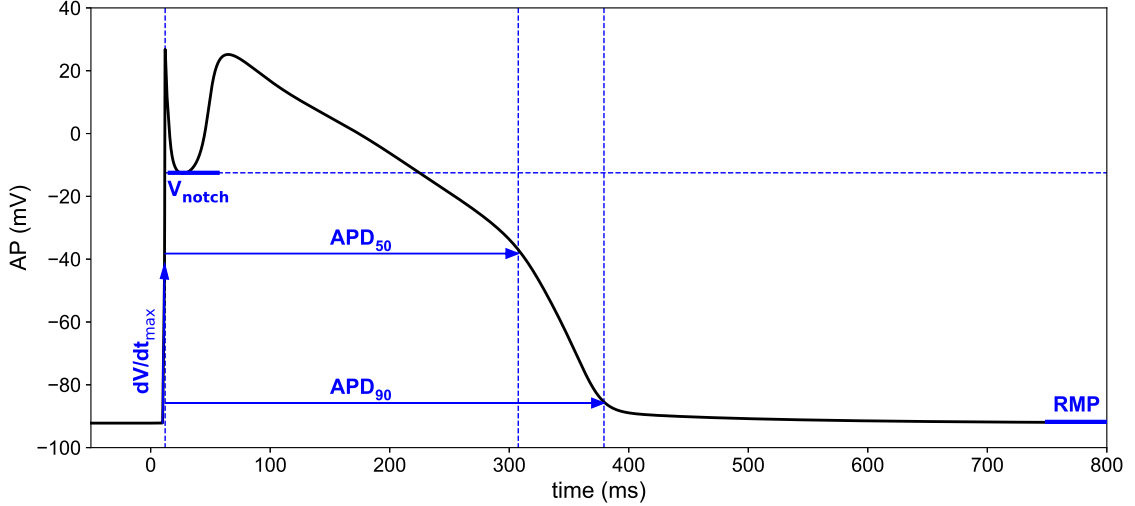


Figure 8: Action potential ($V(t)$ in (27)) computed using the Davies model. Common features associated with the parameters of interest used in the literature are also represented.

5.4. A model of systemic blood flow circulation

A last an application is provided concerning the hemodynamics in a networks of arteries. The motivation is for example the monitoring of hypertensive patients, for which it is important to assess arteries stiffening from non-invasive measurements. The biomarker which is typically used in practice is the Pulse Wave Velocity (PWV), which is the average speed of the pressure waves travelling from carotid to femoral arteries. But arterial stiffness is not the only property affecting the PWV. Daily fluctuations of PWV can be explained as the interplay of different phenomena occurring in the cardiovascular system, such as the different metabolic need of the peripheral organs or the fluctuation in the heart rate due to physical activity. In what follows, the possibility of correcting PWV is studied, in order to have a biomarker which is more intrinsically related to arteries stiffness.

The main 55 arteries of the human body are considered, and the fluid-solid interaction occurring in blood vessels is described by means of a simplified 1-D model (see e.g. [26, 27, 28]). The resulting system is a set of coupled hyperbolic partial differential equations in the unknowns (A_i, q_i) , where $A_i(x, t)$ is the cross-sectional area of the i -th vessel and $q_i(x, t)$ is the sectional averaged blood velocity,

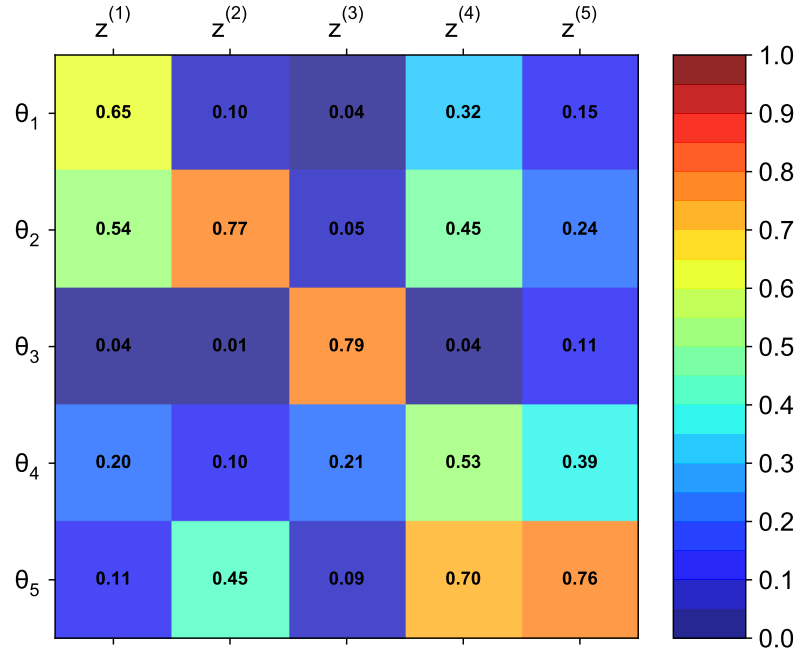


Figure 9: Covariance matrix between the five parameters of interest of the Davies model and their classically associated biomarkers.

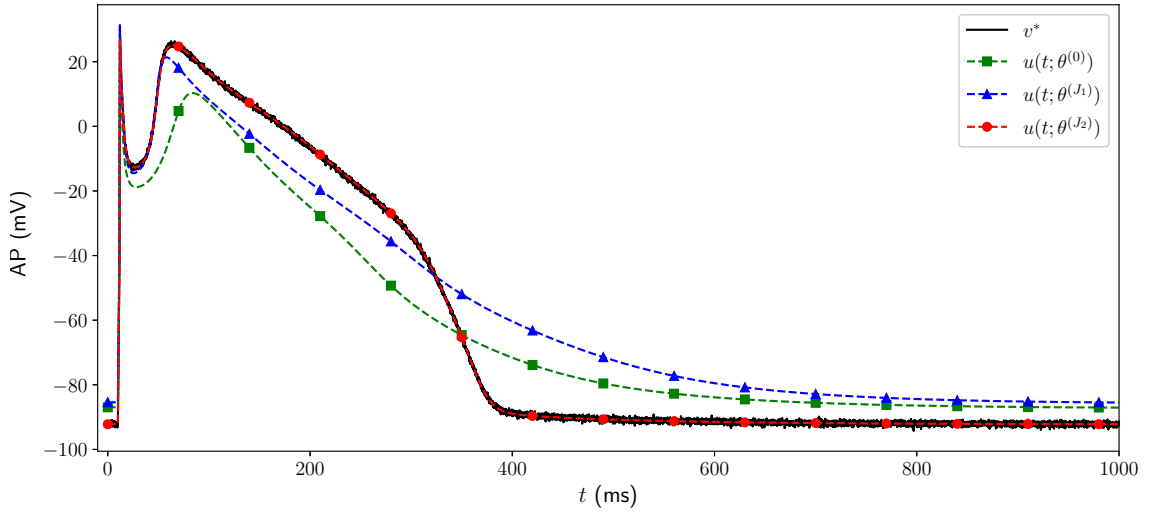


Figure 10: Example of a parameter estimation from a synthetic AP using the gradient-free minimization algorithm. v^* is the synthetic measurement, $\theta^{(0)}$ the initial guess and $\theta^{(J_1)}$ (resp. $\theta^{(J_2)}$) the solution of the minimization of $J_1(\theta)$ (resp. $J_2(\theta)$).

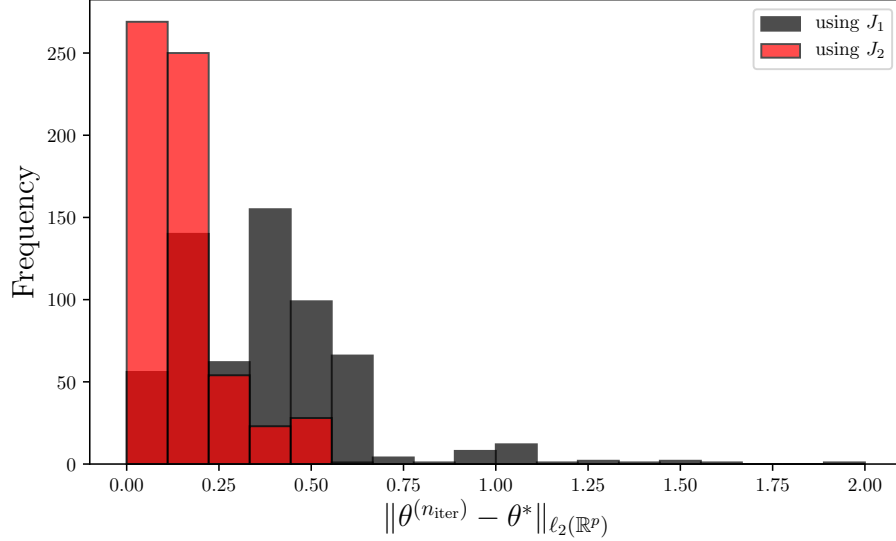


Figure 11: Histograms of the parameter estimation error for the 625 inverse problems carried out with both J_1 and J_2 as the inverse problem cost function.

$x \in [0, \ell_i]$ being the axial coordinate of the i -th vessel, $t \in \mathbb{R}^+$ being the time. The system takes the following form:

$$\partial_t A_i + \partial_x (A_i q_i) = 0, \quad (28)$$

$$\partial_t q_i + \partial_x \left(\frac{q_i^2}{2} + \frac{p_i}{\varrho} \right) = -\kappa \frac{q_i}{A_i}, \quad (29)$$

where ϱ is the blood density, κ a friction coefficient, $p_i(x, t)$ is the pressure in the vessels which is related to the cross-sectional area through the following algebraic relation: $p_i = p_0 + \beta \left(A_i^{1/2} - A_i^{(0)1/2} \right)$, where β is the elastic coefficient [26] and $A_i^{(0)}$ is the cross sectional area of the i -th vessel, at rest. The boundary conditions for the model are of three natures. First, the flow $Q(t)$ is imposed at the inlet of the ascending aorta as follows:

$$A_1(0, t)q_1(0, t) = Q(t). \quad (30)$$

$Q(t)$ is user-defined and is supposed to mimic the heart activity. Second, the effects of the organs and micro-circulation are modeled by standard three-element Windkessel models at the outlets [28]. Third, boundary conditions at the bifurcations are set by enforcing mass conservation and continuity of total pressure. The model is discretized and solved following [29] where a kinetic scheme is considered.

The model, having more than 100 free parameters, suffers from a severe identifiability issue when using realistic data. Hence, the model has been reparametrised

by introducing 8 hyper-parameters, accounting for the main source of variability in the observed data. In particular, the main goal is to be able to represent (in a simple way) the daily fluctuations of the blood flow in the network as well as the variability in the material properties of the arteries. The stiffness of the different segments has been parametrized as follows:

$$\beta_i = \theta_1 + \theta_2 \beta_i^{(ref)}, \quad (31)$$

where $\beta_i^{(ref)}$ is a reference elastic coefficient for the i -th vessel, available in the literature [27]. The three-element Windkessel model parameters have been grouped into three sets: the first one gathers the terminal models of the upper part of the body, the second one gathers those of the central part and the last one the lumped models of the lower body. For each, the distal resistance is rescaled, and the capacitance is set in such a way that the characteristic time of the circuit remains unchanged:

$$R_{u,j} = \theta_3 R_{u,j}^{(ref)}, \quad C_{u,j} = \frac{C_{u,j}^{(ref)}}{\theta_3}, \quad (32)$$

$$R_{c,j} = \theta_4 R_{c,j}^{(ref)}, \quad C_{c,j} = \frac{C_{c,j}^{(ref)}}{\theta_4}, \quad (33)$$

$$R_{l,j} = \theta_5 R_{l,j}^{(ref)}, \quad C_{l,j} = \frac{C_{l,j}^{(ref)}}{\theta_5}. \quad (34)$$

Three additional parameters account for variations in the cardiac output and cardiac rhythm:

$$Q(t) = \begin{cases} \theta_6 \sin\left(\frac{\pi}{\theta_7 T_s} t\right) & 0 \leq t \leq \theta_7 T_s \\ 0 & \theta_7 < t \leq \theta_8 T \end{cases}, \quad (35)$$

where $T_s = 0.25$ s is the systolic period and $T = 0.8$ s is a reference cardiac period.

The training set is generated by drawing $N = 1024$ samples of $\theta = (\theta_1, \dots, \theta_8)$ from the normal distribution $\mathcal{N}(\mu, \Sigma)$, where $\mu = (0, 1, 1, 1, 1, 1, 1, 1)$ and $\Sigma = 0.05^2 \times I_8$. The flow and pressure signals are computed by integrating the model for these samples. With the variability considered, the fluctuation in terms of pressure and flow are about 20%, which is enough to cover meaningful scenarios of daily fluctuations.

The dictionary is built as follows. In three distinct locations – the carotid, the subclavian artery and the femoral artery – average in time, maximum and minimum value of pressure, flow, and mechanical power $P_m = pAq$ are taken. The last three elements of the dictionary are PWV taken from carotid to femoral, from carotid to

subclavian and from subclavian to femoral. The dictionary is ordered as follows:

$$\mathcal{G} = \{(Aq)_0, \dots, (Aq)_8, p_0, \dots, p_8, P_{m,0}, \dots, P_{m,8}, PWV_0, PWV_1, PWV_2\}.$$

The method is used with $\lambda^{(j)} = 10^{-3}$, $\forall j \in \{1, \dots, 8\}$. The penalization parameter ξ is fixed and set to 1. *A posteriori* verifications show that the error on the variance constraint does not exceed 10^{-2} in practice.

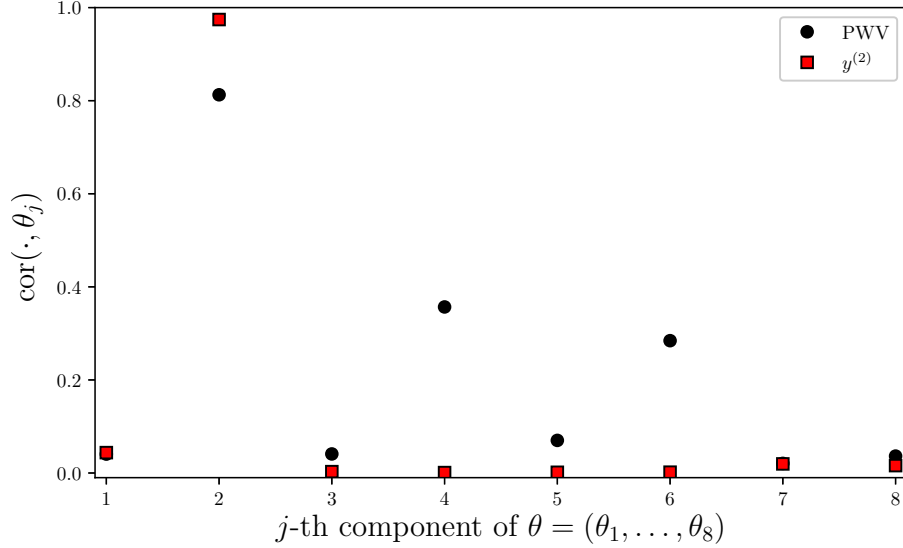


Figure 12: Correlation of PWV with respect to all parameters (black) and correlation of $y^{(2)}$, the composite biomarker associated with θ_2 , with respect to all parameters (red).

In Fig.12, the red squares (resp. the black dots) show the correlation coefficients between the biomarker y_1 provided by our algorithm (resp. the PWV) and each of the 9 parameters. The PWV is significantly correlated to θ_1 , which is not a surprise since the arterial stiffness is known to be linked to the PWV. But the PWV also has a significant correlation (of about 0.3 – 0.4) with the peripheral resistances of the central body and with the cardiac output. This provides a possible explanation for the observed daily fluctuations of the PWV. On the contrary, the biomarker y_1 provided by our method has a higher correlation with respect to θ_1 and a negligible correlation with respect to all the other parameters.

Interestingly, the decomposition of the biomarker onto the dictionary sheds some light onto the physical interpretation of the correction. In Fig.13 the expression of $w^{(2)}$ is shown. We can observe that it is sparse. As expected, the PWV from carotid to subclavian is selected. Then, the correction is provided by a combination of the mechanical power. It is remarkable that the relevant quantity proves to be the product of the flow and the pressure, but not the flow and the pressure separately.

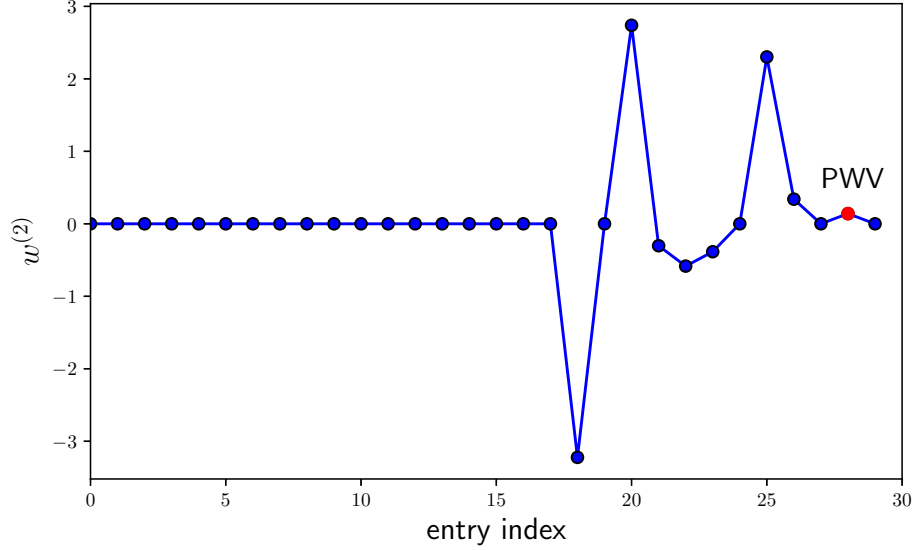


Figure 13: Biomarker expression on the dictionary, for $y^{(2)}$

In conclusion, if we are interested in estimating the arterial stiffness, the algorithm suggests that it could be more relevant to consider a combination of PWV and mechanical power rather than just the PWV.

This result is preliminary and would require a deeper investigation. Nevertheless, it shows an example of possible applications of the proposed method to correct the expression of standard biomarkers used in hemodynamics.

6. Conclusions and Perspectives

In this work a method is proposed to define composite biomarkers. A semi-empirical approach is used, that consists in building a database of simulations and exploiting the results in order to define biomarkers which are maximally correlated to the parameters they are meant to monitor and minimally correlated with respect to all the others. Such an objective is translated into an optimization problem and the biomarkers are sought as the solution of its discretized version. The biomarkers are, by construction, linear combinations of dictionary entries that comprise relevant features found in the literature and additional quantities computed from the observable signal. Furthermore, a sparsity-promoting penalization ensures that the obtained biomarker representation on the dictionary entries are sparse and, often, easily interpretable. As a by-product, a metric in the space of the observable is obtained which is equivalent, up to a small perturbation, to the ℓ^2 -distance in the space of parameters. The final cost function of the optimization problem can be used as an *a posteriori* estimate of the biomarkers dictionary quality. In case this

cost is deemed to high, additional entries may be added to the dictionary using agnostic or model-driven quantities. The method can be exploited in order to enhance the robustness of the biomarkers which are used in biology and medicine by using numerical simulations. Furthermore, it can be used to regularize inverse problems and make them less ill-conditioned.

The proposed approach is particularly suited to biomedical applications where biomarkers are often preferred to raw measurements. However, the authors believe the methodology presented in this work could find applications beyond this scope. For all the systems in which several signals are measured and predictions, identification, statistical learning tasks have to be performed, the proposed approach could be of some help. For instance, applications could be considered in environmental sciences, ecology, climate monitoring or geophysics.

7. Acknowledgments

Elliott Tixier was funded by a doctoral grant from the French Ministry of Research and Higher Education.

Annexes

A1. Algorithm to solve the minimization problem (19)

<p>Input:</p> <ul style="list-style-type: none"> • Sparsity promoting parameter: $\lambda^{(j)}$ • Initial guess: $w^{(j),0}$ • Stopping criteria: absTol and relTol • Constraint tolerance: ϵ_c • Constraint penalization parameter (initial guess): ξ • Constraint penalization parameter update factor: $\alpha > 1$ <p>$i \leftarrow 0$;</p> <p>$r^{(0)} := J_\lambda(w^{(j),0})$; // Residual</p> <p>while $r^{(i)} < \text{absTol}$ or $r^{(i)} - r^{(i-1)} < \text{relTol} \times r^{(i)}$ do</p> <table border="0" style="width: 100%;"> <tr> <td style="padding-left: 20px;"> if $\nu(w^{(j),i}) - 1 > \epsilon_c$ then </td> <td></td> </tr> <tr> <td style="padding-left: 40px;"> $\xi \leftarrow \alpha \xi$; </td> <td style="padding-left: 20px;">// increase constraint penalization parameter</td> </tr> <tr> <td style="padding-left: 20px;">end</td> <td></td> </tr> <tr> <td style="padding-left: 20px;"> Compute gradient: $\partial_{w_{\pm}^{(j)}} J_\lambda(w^{(j),i})$; </td> <td style="padding-left: 20px;">// using (20)</td> </tr> <tr> <td style="padding-left: 20px;"> Update weights: $w^{(j),i+1} := f(w^{(j),i}, \partial_{w^{(j)}} J_\lambda(w^{(j),i}))$; </td> <td style="padding-left: 20px;">// Nesterov formula</td> </tr> <tr> <td style="padding-left: 20px;"> Update residual: $r^{(i+1)} := J_\lambda(w^{(j),i+1})$; </td> <td></td> </tr> <tr> <td style="padding-left: 20px;"> $i \leftarrow i + 1$; </td> <td></td> </tr> </table> <p>end</p>	if $ \nu(w^{(j),i}) - 1 > \epsilon_c$ then		$\xi \leftarrow \alpha \xi$;	// increase constraint penalization parameter	end		Compute gradient: $\partial_{w_{\pm}^{(j)}} J_\lambda(w^{(j),i})$;	// using (20)	Update weights: $w^{(j),i+1} := f(w^{(j),i}, \partial_{w^{(j)}} J_\lambda(w^{(j),i}))$;	// Nesterov formula	Update residual: $r^{(i+1)} := J_\lambda(w^{(j),i+1})$;		$i \leftarrow i + 1$;	
if $ \nu(w^{(j),i}) - 1 > \epsilon_c$ then														
$\xi \leftarrow \alpha \xi$;	// increase constraint penalization parameter													
end														
Compute gradient: $\partial_{w_{\pm}^{(j)}} J_\lambda(w^{(j),i})$;	// using (20)													
Update weights: $w^{(j),i+1} := f(w^{(j),i}, \partial_{w^{(j)}} J_\lambda(w^{(j),i}))$;	// Nesterov formula													
Update residual: $r^{(i+1)} := J_\lambda(w^{(j),i+1})$;														
$i \leftarrow i + 1$;														

Algorithm 1: solve the minimization problem (19) with an automatic tuning of parameter ξ .

A2. Algorithm to solve the parameter identification problem

```

Input:
    • Cost function:  $J(\theta)$  (see Eqs. (21),(23))
    • Initial guess:  $\theta^{(0)}$ 
    • Bracket size:  $\gamma$ 
    • Stopping criteria: absTol and relTol

 $i \leftarrow 0$ ;
 $r^{(0)} := J(\theta^{(0)})$ ;
while  $r^{(i)} < \text{absTol}$  or  $|r^{(i)} - r^{(i-1)}| < \text{relTol} \times r^{(i)}$  do
    for  $j = 1$  to  $p$  do
        Fit quadratic model  $Q(s)$  to  $f(s) = J(\dots, \theta_{j-1}^{(i)}, s, \theta_{j+1}^{(i)}, \dots)$  for
         $s \in I_j^{(i)} = [\theta_j^{(i)} - \gamma, \theta_j^{(i)} + \gamma]$ ;
        Update:  $\theta_j^{(i+1)} := \arg \min_{s \in I_j^{(i)}} Q(s)$ ;
    end
    Update residual:  $r^{(i+1)} := J(\theta^{(i+1)})$ ;
     $i \leftarrow i + 1$  ;
end

```

Algorithm 2: Alternate directions descent quadratic approximations for the parameter identification problem (simplification of Powell’s BOBYQA algorithm [18]).

A3. Action potential biomarkers

The 104 dictionary entries for the action potential model used in Section 5.3 are hereby described.

Action potential durations. The first entries correspond to action potential durations (APD x) at $x\%$ repolarization. They are defined as the amount of time during which the action potential remains above its peak value minus $x\%$ of its amplitude. There are 8 APD values in the dictionary: APD90, APD80, ..., APD20.

Other classical biomarkers. Some classical biomarkers are also added, respectively APD50/APD90, APA (AP amplitude), RMP (resting membrane potential).

Action potential values. The following entries are action potential values at 61 different time steps between 0ms and 600 ms. V_{notch} is the 6th of these entries.

Derivatives and integrals. The following entries are dV/dt_{max} (maximum derivative during the upstroke), dV/dt_{min} (minimum derivative during the repolarization) and AUC (area under curve).

Auto-correlation. 21 entries correspond to products of the action potential values at different time steps of the form: $V(t_i) * V(t_j)$.

PCA coefficients. Finally, 8 entries correspond to the first 8 coefficients of a truncated PCA expansion of $V(t)$.

8. References

- [1] R. Gul, S. Bernhard, Parametric uncertainty and global sensitivity analysis in a model of the carotid bifurcation: Identification and ranking of most sensitive model parameters, *Mathematical Biosciences* 269 (2015) 104–116.
- [2] A. Cintrón-Arias, H. Banks, A. Capaldi, A. L. Lloyd, A sensitivity matrix based methodology for inverse problem formulation, *Journal of Inverse and Ill-Posed Problems* 17 (6) (2009) 545–564.
- [3] C. Krier, D. Francois, V. Wertz, M. Verleysen, Feature scoring by mutual information for classification of mass spectra, in: 7th International FLINS Conference on Applied Artificial Intelligence (FLINS 06), 2006, pp. –.
- [4] J. T. Ottesen, J. Mehlsen, M. S. Olufsen, Structural correlation method for model reduction and practical estimation of patient specific parameters illustrated on heart rate regulation, *Mathematical biosciences* 257 (2014) 50–59.
- [5] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, *Journal of machine learning research* 3 (Mar) (2003) 1157–1182.
- [6] D. G. Arquès, C. J. Michel, Study of a perturbation in the coding periodicity, *Mathematical biosciences* 86 (1) (1987) 1–14.
- [7] P. Gokulakrishnan, A. Lawrence, P. McLellan, E. Grandmaison, A functional-pca approach for analyzing and reducing complex chemical mechanisms, *Computers & chemical engineering* 30 (6) (2006) 1093–1101.
- [8] P. G. Constantine, *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*, SIAM, 2015.
- [9] S. Wold, A. Ruhe, H. Wold, W. Dunn, III, The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses, *SIAM Journal on Scientific and Statistical Computing* 5 (3) (1984) 735–743.
- [10] S. Young, M. E. Goddard, J. E. Pryce, G. Deng, Kernel methods and haplotypes used in selection of sparse dna markers for protein yield in dairy cattle, *Mathematical biosciences* 243 (1) (2013) 57–66.
- [11] S. M. Davidson, P. D. Docherty, R. Murray, The dimensional reduction method for identification of parameters that trade-off due to similar model roles, *Mathematical Biosciences* 285 (2017) 119–127.
- [12] J. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, Vol. 160, Springer Science & Business Media, 2006.

- [13] A. Y. Ng, Feature selection, l_1 vs. l_2 regularization, and rotational invariance, in: Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, p. 78.
- [14] F. Bach, R. Jenatton, J. Mairal, G. Obozinski, et al., Convex optimization with sparsity-inducing norms, *Optimization for Machine Learning* 5 (2011) 19–53.
- [15] Y. Nesterov, A method for unconstrained convex minimization problem with the rate of convergence $O(1/k^2)$, in: Doklady an SSSR, Vol. 269, 1983, pp. 543–547.
- [16] B. O’Donoghue, E. Candes, Adaptive restart for accelerated gradient schemes, *Foundations of computational mathematics* 15 (3) (2015) 715–732.
- [17] P. R. Johnston, R. M. Gulrajani, Selecting the corner in the l-curve approach to tikhonov regularization, *IEEE Transactions on biomedical engineering* 47 (9) (2000) 1293–1296.
- [18] M. J. Powell, The bobyqa algorithm for bound constrained optimization without derivatives, Cambridge NA Report NA2009/06, University of Cambridge, Cambridge.
- [19] M. R. Davies, H. B. Mistry, L. Hussein, C. E. Pollard, J.-P. Valentin, J. Swinton, N. Abi-Gerges, An in silico canine cardiac midmyocardial action potential duration model as a tool for early drug safety assessment, *American Journal of Physiology-Heart and Circulatory Physiology*.
- [20] Z. Syed, E. Vigmond, S. Nattel, L. Leon, Atrial cell action potential parameter fitting using genetic algorithms, *Medical and Biological Engineering and Computing* 43 (5) (2005) 561–571.
- [21] A. X. Sarkar, E. A. Sobie, Regression analysis for constraining free parameters in electrophysiological models of cardiac cells, *PLoS Comput Biol* 6 (9) (2010) e1000914.
- [22] J. Kaur, A. Nygren, E. J. Vigmond, Fitting membrane resistance along with action potential shape in cardiac myocytes improves convergence: application of a multi-objective parallel genetic algorithm, *PloS one* 9 (9) (2014) e107984.
- [23] C. Sánchez, A. Bueno-Orovio, E. Wettwer, S. Loose, J. Simon, U. Ravens, E. Pueyo, B. Rodriguez, Inter-subject variability in human atrial action potential in sinus rhythm versus chronic atrial fibrillation, *PloS one* 9 (8) (2014) e105897.

- [24] M. Courtemanche, R. J. Ramirez, S. Nattel, Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model, *American Journal of Physiology-Heart and Circulatory Physiology* 275 (1) (1998) H301–H321.
- [25] J. T. Koivumäki, G. Seemann, M. M. Maleckar, P. Tavi, In silico screening of the key cellular remodeling targets in chronic atrial fibrillation, *PLoS Comput Biol* 10 (5) (2014) e1003620.
- [26] L. Formaggia, A. Quarteroni, A. Veneziani, *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*, Vol. 1, Springer Science & Business Media, 2010.
- [27] P. Reymond, F. Merenda, F. Perren, D. Rüfenacht, N. Stergiopulos, Validation of a one-dimensional model of the systemic arterial tree, *American Journal of Physiology-Heart and Circulatory Physiology* 297 (1) (2009) H208–H222.
- [28] K. S. Matthys, J. Alastruey, J. Peiró, A. W. Khir, P. Segers, P. R. Verdonck, K. H. Parker, S. J. Sherwin, Pulse wave propagation in a model human arterial network: assessment of 1-d numerical simulations against in vitro measurements, *Journal of biomechanics* 40 (15) (2007) 3476–3486.
- [29] C. Audebert, P. Bucur, M. Bekheit, E. Vibert, I. E. Vignon-Clementel, J.-F. Gerbeau, Kinetic scheme for arterial and venous blood flow, and application to partial hepatectomy modeling, *Computer Methods in Applied Mechanics and Engineering*.