Segmentation, classification and denoising of time series of images by a variational method

Piero Barone, Maria Francesca Carfora, Riccardo March

*Istituto per le Applicazioni del Calcolo “M. Picone”, Consiglio Nazionale delle Ricerche, Viale del Policlinico 137, 00161 Roma, Italy*

Abstract

Many real life problems can be represented by an ordered sequence of digital images. At a given pixel a specific time course is observed which is morphologically related to the time courses at neighbor pixels. Useful information can be usually extracted from a set of such observations if we are able to classify pixels in groups, according to some features of interest for the final user. In a continuous setting we can formalize the problem by assuming to observe a noisy version of a positive real function defined on a bounded set $T \times \Omega \subset \mathbb{R} \times \mathbb{R}^2$, parameterized by a vector of unknown functions defined on $\mathbb{R}^2$ with discontinuities along regular curves in $\Omega$ which separate regions with different features. Suitable regularity conditions on the parameters are also assumed in order to take into account the physical constraints. The problem consists in estimating the parameter functions, segmenting $\Omega$ in subsets with
regular boundaries and assigning to each subset a label according to the values that the parameters assume on the subset. A global model is proposed which allows to address all of the above subproblems in the same framework. A variational approach is then adopted to compute the solution and an algorithm has been developed. Some numerical results obtained by using the proposed method to solve a dynamic Magnetic Resonance imaging problem are reported.

1 Introduction

We introduce a variational model for simultaneous classification, segmentation and denoising of a sequence of images, ordered in time with common domain. The model is based on the partition of the image domain in different subsets which are associated to classes. Each class is characterized by a vector of parameters. We consider supervised classification which means that the number of classes and the parameter vectors are a priori known.

The observed sequence of images is a positive real function defined on a bounded set $\mathcal{T} \times \Omega \subset \mathbb{R} \times \mathbb{R}^2$, and we adopt for the true, but unobserved function, a temporal model parameterized by a vector field defined on $\mathbb{R}^2$.

The variational model consists in the minimization of a functional that penalizes the $L^2$ distance between the image data and the parametric temporal model. The minimization of the functional forces the vector field to be close to the parameter vector of a class at each point of the image domain, thus yielding a classification. The functional penalizes the length of the boundaries of the partition of the image domain in order to achieve a partition with regular boundaries. A Mumford-Shah energy term is included in the functional in order to obtain the simultaneous segmentation of the parameter vector field.
The functional is defined on suitable classes of functions of bounded variation whose discontinuity sets represent the boundaries of the partition. Since the functional involves geometric terms corresponding to the length of the unknown discontinuity sets, the discretization of the functional is not a straightforward matter of using either finite elements or differences. Then we resort to the theory of Γ-convergence to approximate in a variational sense (i.e., convergence of minimizers) the functional of the model by means of functionals defined on Sobolev spaces. The approximating functionals are numerically more tractable, in the sense that the corresponding Euler equations now admit a straightforward finite element discretization. A numerical scheme based on such a discretization is then designed.

A related variational approach for image classification and restoration has been proposed in [16] and analyzed mathematically in [4].

In order to illustrate the behavior of the proposed method, a problem arising in dynamic Magnetic Resonance studies is considered. A dynamic MR clinical exam produces a set of images which monitor in time the concentration of a contrast agent in the vessels of an organ. The portion of tissue represented by each pixel can be classified either as normal or as a benign or malignant tumoral one, according to the qualitative behavior of the time series associated to it. The time series can be considered as the noisy output of a model with pixel dependent parameters. For technical reasons, a compromise must be done between the noise level and the observation time. This makes the parameter identification and the classification problems difficult especially if short computation time is required by physicians. In [10] this problem was addressed in a Bayesian framework by using two separate models, the first one for estimating the parameters and the second one to compute the classification. In [6] a pharmacokinetic model was assumed for the time courses and a global deterministic model was used to estimate at the same time the parameters
and the classification labels. Based on this model a fast algorithm has been devised, overcoming the main drawback of the Bayesian approach. However in none of the above approaches the regularity of the classified regions was controlled explicitly. This can be an important drawback because the shape of a tumor is an important diagnostic feature. The proposed method overcomes this problem still providing a good solution in a time comparable with that of the Bayesian approach.

2 Mathematical framework

We first recall the definition of functions of bounded variation. Let $\Omega$ be a bounded open subset of $\mathbb{R}^2$. A function $u \in L^1(\Omega)$ is called a function of bounded variation if the gradient of $u$ in the sense of distributions is a vector valued measure with finite total variation in $\Omega$. We denote by $BV(\Omega)$ the class of functions of bounded variation, and by $Du$ the distributional gradient of $u$. The total variation of the measure $Du$ is defined by

$$\int_{\Omega} |Du| = \sup \left\{ \int_{\Omega} u \text{div} \phi \, dx : \phi = (\phi_1, \phi_2) \in C^1_0(\Omega; \mathbb{R}^2), \ |\phi(x)| \leq 1 \text{ for all } x \in \Omega \right\}.$$ 

The distributional gradient of a function $u \in BV(\Omega)$ admits the decomposition

$$Du(B) = \int_B \nabla u \, dx + D^s u(B), \quad D^s u(B) = Ju(B) + Cu(B),$$

for any Borel set $B$, where $\nabla u$ here denotes the density of the absolutely continuous part of $Du$ with respect to the two-dimensional Lebesgue measure, and $D^s u$ denotes the singular part of the measure. The singular part of the distributional gradient is further decomposed into the jump part $Ju$ and the Cantor part $Cu$. The jump part is a measure which is concentrated on the discontinuity set of $u$, which is denoted by $S_u$. The discontinuity set is defined in a weak sense: $S_u$ is the complement of the
Lebesgue set of $u$, i.e.,

\[ S_u = \left\{ x \in \Omega : \exists z \in \mathbb{R} \text{ such that } \lim_{\rho \to 0^+} \rho^{-2} \int_{B_\rho(x)} |u - z| dx = 0 \right\}, \]

where $B_\rho(x)$ denotes the open ball with center $x$ and radius $\rho$. It can be shown that $S_u$ has negligible Lebesgue measure. The Cantor part takes into account possible pathological behaviour of $BV$ functions. For instance there exist nonconstant $BV$ functions which are continuous (i.e., $J_u = 0$) and have $\nabla u = 0$ almost everywhere, so that $Du = Cu$. An example in the one-dimensional case of a function of this kind is the Cantor-Vitali function, whose distributional derivative is a measure concentrated on Cantor’s middle third set.

If $u \in C^1(\Omega)$ then $\nabla u$ is the gradient, and one can show using integration by parts that

\[ \int_{\Omega} |Du| = \int_{\Omega} |\nabla u| dx. \]

A function $u \in BV(\Omega)$ belongs to the class of special functions of bounded variation, and we write $u \in SBV(\Omega)$, if $Cu = 0$. In the $SBV$ class pathological functions, having the distributional derivative with a Cantor part, are then not included and this is a natural requirement in applications to image processing. For further information concerning functions of bounded variation we refer to [1, 12].

A set $A \subset \Omega$ is called a set of finite perimeter if $\chi_A \in BV(\Omega)$, where $\chi_A$ denotes the characteristic function of $A$, i.e., $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$. The perimeter of $A$ in $\Omega$ is given by the total variation of $D\chi_A$, and the following equality holds:

\[ \int_{\Omega} |D\chi_A| = \mathcal{H}^1(\partial^* A \cap \Omega), \]

where $\partial^* A$ denotes the reduced boundary of $A$ [12], and $\mathcal{H}^1$ denotes the one-dimensional Hausdorff measure on $\mathbb{R}^2$. The use of the term perimeter for the total
variation of $D \chi_A$ is justified by the fact that if $A$ is a set with smooth boundary, then $\partial^* A$ is the usual topological boundary $\partial A$, and $\mathcal{H}^1(\partial A)$ is the length of $\partial A$.

For any integer $m$, we say that a function $u$ with values in $\mathbb{R}^m$ is a vector valued $BV$ function, and we write $u \in BV(\Omega; \mathbb{R}^m)$, if $u_i \in BV(\Omega)$ for any component $u_i$ of $u$. We set

$$\int_\Omega |Du| = \sum_{i=1}^m \int_\Omega |Du_i|.$$  

Analogously, we write $u \in SBV(\Omega; \mathbb{R}^m)$, if $u_i \in SBV(\Omega)$ for any component $u_i$ of $u$. The discontinuity set of $u$ is defined by $S_u = \bigcup_{i=1}^m S_{u_i}$.

We now recall the definition of $\Gamma$-convergence [9]. Let $X$ be a metric space, and let $Y \subseteq X$. A family of functionals $F_\varepsilon : Y \to [0, +\infty]$ is said to $\Gamma$-converge to $F : X \to [0, +\infty]$ with respect to the metric of $X$ if, for every $v \in X$,

$$\forall v_\varepsilon \to v, \quad \liminf_{\varepsilon \to 0^+} F_\varepsilon(v_\varepsilon) \geq F(v) \quad (1)$$

and

$$\exists v_\varepsilon \to v : \quad \limsup_{\varepsilon \to 0^+} F_\varepsilon(v_\varepsilon) \leq F(v). \quad (2)$$

The $\Gamma$-limit $F$, if it exists, is unique and lower semicontinuous. Using (1), the inequality (2) can be replaced with

$$\exists v_\varepsilon \to v : \quad \lim_{\varepsilon \to 0^+} F_\varepsilon(v_\varepsilon) = F(v). \quad (3)$$

$\Gamma$-convergence is stable under continuous perturbations, that is, $F_\varepsilon + G$ $\Gamma$-converges to $F + G$ if $F_\varepsilon$ $\Gamma$-converges to $F$ and if the functional $G$ is continuous with respect to the metric in $X$.

The most important property of $\Gamma$-convergence is variational convergence: if a minimizer $v_\varepsilon$ for $F_\varepsilon$ exists for every $\varepsilon > 0$ and if there is a sequence $h \mapsto \varepsilon_h$, with $\varepsilon_h \to 0$, such that the corresponding $v_{\varepsilon_h}$ converge to $v$, then $v$ is a minimizer for $F$. 
We recall two $\Gamma$-convergence theorems that will be used in the sequel. We denote by $W^{1,2}(\Omega; \mathbb{R}^m)$ the Sobolev space of the vector valued functions $u \in L^2(\Omega; \mathbb{R}^m)$ whose distributional partial derivatives of first order all belong to $L^2(\Omega; \mathbb{R}^m)$.

We say that $k$ Borel sets $A_1, \ldots, A_k$ define a partition of $\Omega$ if
\[ A_i \cap A_j = \emptyset \quad \forall i, j \in \{1, \ldots, k\}, \quad i \neq j, \quad \text{meas}(\Omega \setminus \bigcup_{i=1}^k A_i) = 0, \]
where $\text{meas}(\cdot)$ denotes the Lebesgue measure in $\mathbb{R}^2$. Let $\{e_1, \ldots, e_k\}$ denote the vectors of the canonical basis of $\mathbb{R}^k$ and let $u \in BV(\Omega; \{e_1, \ldots, e_k\})$. We set
\[ A_i = \{ x \in \Omega : u(x) = e_i \} \]
for any $i = 1, \ldots, k$. In order to avoid a cumbersome notation, we drop the dependence of the sets $A_i$ on the function $u$. Then the sets $A_1, \ldots, A_k$ define a partition of $\Omega$ into sets with finite perimeter. We set
\[ X(\Omega) = L^2(\Omega; \mathbb{R}^k), \quad Y(\Omega) = W^{1,2}(\Omega; \mathbb{R}^k). \]

We define the functional $E : X(\Omega) \to [0, +\infty]$ by
\[ E(u) = \begin{cases} \sum_{i,j=1}^{k} H^1(\partial^* A_i \cap \partial^* A_j \cap \Omega) & \text{if } u \in BV(\Omega; \{e_1, \ldots, e_k\}), \\ +\infty & \text{elsewhere in } X(\Omega), \end{cases} \]
and, for any $\varepsilon > 0$, the functional $E_\varepsilon : Y(\Omega) \to [0, +\infty]$ by
\[ E_\varepsilon(u) = \int_\Omega \left[ \varepsilon|\nabla u|^2 + \frac{W(u)}{\varepsilon} \right] dx, \]
where $|\nabla u|^2 = \sum_{i=1}^{k} |\nabla u_i|^2$, and $W : \mathbb{R}^k \to \mathbb{R}$ is a non-negative function that vanishes only in the points $e_1, \ldots, e_k$:
\[ W(u) = \prod_{i=1}^{k} |u - e_i|^2. \]

The following theorem was proved by Baldo [5].
Theorem 2.1 The family of functionals $E_\varepsilon \Gamma$-converges to $\eta E$ with respect to the $L^2(\Omega; \mathbb{R}^k)$ metric as $\varepsilon \to 0$, where $\eta$ is a positive constant.

Let $\{u_\varepsilon\} \subset Y(\Omega)$ be a family of functions such that, for any $\varepsilon > 0$, $|u_\varepsilon| \leq M_1$ for almost every $x \in \Omega$ and $E_\varepsilon(u_\varepsilon) \leq M_2$, with $M_1$ and $M_2$ positive constants independent of $\varepsilon$. Then there exist a function $u \in BV(\Omega; \{e_1, \ldots, e_k\})$ and a sequence $\varepsilon_h \to 0$ such that $u_{\varepsilon_h}$ converges to $u$ in $L^2(\Omega; \mathbb{R}^k)$.

Now we recall the approximation by $\Gamma$-convergence of the Mumford-Shah functional [14] in the vectorial case:

$$\int_\Omega |\nabla \theta|^2 dx + \alpha H^1(S_{\theta}),$$

where $\theta \in SBV(\Omega; \mathbb{R}^m)$, $\nabla \theta$ is the density of the absolutely continuous part of $D\theta$ with respect to Lebesgue measure, and $\alpha$ is a positive weight. We now set

$$X(\Omega) = L^2(\Omega; \mathbb{R}^m) \times L^\infty(\Omega; [0, 1]), \quad Y(\Omega) = W^{1,2}(\Omega; \mathbb{R}^m) \times W^{1,2}(\Omega; [0, 1]).$$

We define the functional $G : X(\Omega) \to [0, +\infty]$ by

$$G(\theta, s) = \begin{cases} 
\int_\Omega |\nabla \theta|^2 dx + \alpha H^1(S_{\theta}) & \text{if } \theta \in SBV(\Omega; \mathbb{R}^m), \ s \equiv 1, \\
+\infty & \text{elsewhere in } X(\Omega),
\end{cases}$$

and, for any $\gamma > 0$, the functional $G_\gamma : Y(\Omega) \to [0, +\infty]$ by

$$G_\gamma(\theta, s) = \int_\Omega (s^2 + \kappa_\gamma)|\nabla \theta|^2 dx + \alpha \int_\Omega \left[ \gamma|\nabla s|^2 + \frac{(1-s)^2}{4\gamma} \right] dx,$$

where $\kappa_\gamma$ is a positive infinitesimal of order $o(\gamma)$. The function $s$ is related to the discontinuity set $S_{\theta}$, in the sense that $s$ constitutes an approximation of $1 - \chi_{S_{\theta}}$. The link between $s$ and $S_{\theta}$ is discussed below.

The following theorem was proved by Ambrosio and Tortorelli [2, 3], and then extended to the vectorial case in a more general context by Focardi [11].
Theorem 2.2 The family of functionals $G_{\gamma}$ $\Gamma$-converges to $G$ with respect to the $L^2(\Omega; \mathbb{R}^m) \times L^2(\Omega)$ metric as $\gamma \to 0$.

Let $\{(\theta_{\gamma}, s_{\gamma})\} \subset Y(\Omega)$ be a family of pairs such that, for any $\gamma > 0$, $|\theta_{\gamma}| \leq M_1$ for almost every $x \in \Omega$ and $G_{\gamma}(\theta_{\gamma}, s_{\gamma}) \leq M_2$, with $M_1$ and $M_2$ positive constants independent of $\gamma$. Then there exist a function $\theta \in SBV(\Omega; \mathbb{R}^m)$ and a sequence $\gamma_h \to 0$ such that $(\theta_{\gamma_h}, s_{\gamma_h})$ converges to the pair $(\theta, 1)$ in $L^2(\Omega; \mathbb{R}^m) \times L^2(\Omega)$.

The functions $s_{\gamma}$ of the above theorem provide a diffuse representation of the discontinuity set of the limit function $\theta$. Indeed, for sufficiently small $\gamma$, $s_{\gamma}$ becomes close to zero in a tubular neighborhood of the discontinuity set $S_{\theta}$, and close to one outside. In the limit, the neighborhood shrinks to $S_{\theta}$, so that $s_{\gamma}$ converges pointwise to 1 on $\Omega \setminus S_{\theta}$ and to 0 on $S_{\theta}$.

In the next section the two $\Gamma$-convergence theorems will be used to build a numerical approximation scheme for the variational model of classification and segmentation.

3 The variational model

Let $\Omega$ be a bounded open subset of $\mathbb{R}^2$, $\mathcal{T}$ a compact subset of $\mathbb{R}$, and let $g \in L^\infty(\mathcal{T} \times \Omega)$ be the observed function. Hence $g(t, x)$ represents the observed intensity profile at time $t \in \mathcal{T}$ and point $x \in \Omega$. Let $f : \mathcal{T} \times \mathbb{R}^m \to \mathbb{R}^+$ be a continuous function. We adopt a parametric temporal model $f(t, \theta(x))$ for the true, but unobserved function to be computed, where $\theta : \Omega \to \mathbb{R}^m$ is a vector field which represents the parameter vector at point $x \in \Omega$. We require that $\theta \in SBV(\Omega; \mathbb{R}^m)$.

The classification is defined by means of a partition of $\Omega$ with sets of finite perimeter $A_1, \ldots, A_k$, such that

$$ A_i = \{x \in \Omega : u(x) = e_i\} $$
for any \( i = 1, \ldots, k \), where \( u \in BV(\Omega; \{e_1, \ldots, e_k\}) \). A class is characterized by a parameter vector belonging to \( \mathbb{R}^m \). The integer \( k \) denotes the number of classes and the vectors \( c_i \in \mathbb{R}^m, i = 1, \ldots, k \), denote the parameter vectors characterizing each class. We denote by \( C \in \mathbb{R}^{m \times k} \) the matrix \( C = (c_1, \ldots, c_k) \). A classification is defined by a bijective association between the sets \( A_1, \ldots, A_k \) and the parameter vectors \( c_1, \ldots, c_k \). We assume that the number \( k \) and the vectors \( c_1, \ldots, c_k \) are a priori known, i.e., we consider a supervised classification.

The classification and segmentation of the image data \( g \) is accomplished by minimizing the following functional:

\[
\mathcal{F}(u, \theta) = \lambda_0 \int_{\Omega} \int_{T} |g - f(t, \theta)|^2 dt \, dx + \int_{\Omega} \left| \sqrt{\Lambda_1} (\theta - Cu) \right|^2 dx \\
+ \lambda_2 \sum_{i,j=1}^{k} \mathcal{H}^1 (\partial^* A_i \cap \partial^* A_j \cap \Omega) \\
+ \int_{\Omega} \left| \sqrt{\Lambda_3} \nabla \theta \right|^2 dx + \lambda_4 \mathcal{H}^1 (S_\theta),
\]

where \( \Lambda_1, \Lambda_3 \in \mathbb{R}^{m \times m} \) are positive diagonal matrices of weights, \( \sqrt{\Lambda_i} \) denotes the diagonal matrix having as coefficients the square roots of the coefficients of \( \Lambda_i \) for \( i = 1, 3 \), and \( \lambda_0, \lambda_2, \lambda_4 \) are positive weights.

The meaning of the terms of the functional is the following:

- the first term penalizes the discrepancy between the image data \( g \) and the parametric temporal model \( f(t, \theta) \);
- since \( u \in BV(\Omega; \{e_1, \ldots, e_k\}) \), the term with the weight \( \Lambda_1 \) forces the vector \( \theta(x) \) to be close to the parameter vector \( c_i \) for any point \( x \in A_i \) and for any \( i = 1, \ldots, k \), thus yielding a classification.
• the term with the weight $\lambda_2$ penalizes the length of the boundaries of the partition of $\Omega$, thus encouraging a partition with regular boundaries;

• the terms with the weights $\Lambda_3$ and $\lambda_4$ are a Mumford-Shah functional defined on vector valued functions, so that they yield a segmentation with regular boundaries of the parameter vector field $\theta$.

The numerical minimization of functional (5) presents a number of technical difficulties. In particular, due to the presence of the terms

$$\sum_{i,j=1}^{k} \mathcal{H}^1(\partial^* A_i \cap \partial^* A_j \cap \Omega) \quad \text{and} \quad \mathcal{H}^1(S_\theta),$$

the discretization of the functional $\mathcal{F}$ is not a straightforward matter of using finite elements/differences. For this reason, we consider the approximation of the functional in the sense of $\Gamma$-convergence, by a family of numerically more tractable functionals defined over Sobolev spaces. We set

$$X(\Omega) = L^2(\Omega; K_1) \times L^2(\Omega; K_2) \times L^\infty(\Omega; [0, 1]),$$

where $K_1 \subset \mathbb{R}^k$, $K_2 \subset \mathbb{R}^m$ are non-empty compact sets such that $\{e_1, \ldots, e_k\} \subset K_1$ and $\{c_1, \ldots, c_k\} \subset K_2$. Then we define the functional $F : X(\Omega) \to [0, +\infty]$ by

$$F(u, \theta, s) = \begin{cases} 
\mathcal{F}(u, \theta) & \text{if } u \in BV(\Omega; \{e_1, \ldots, e_k\}), \theta \in SBV(\Omega; \mathbb{R}^m), s \equiv 1, \\
+\infty & \text{elsewhere in } X(\Omega).
\end{cases}$$

Then we set

$$Y(\Omega) = W^{1,2}(\Omega; K_1) \times W^{1,2}(\Omega; K_2) \times W^{1,2}(\Omega; [0, 1]),$$

11
and, for any $\varepsilon, \gamma > 0$, we define the functional $F_{\varepsilon, \gamma} : Y(\Omega) \to [0, +\infty]$ by

$$F_{\varepsilon, \gamma}(u, \theta, s) = \lambda_0 \int_{\Omega} \int_{T} |g - f(t, \theta)|^2 dt \, dx + \int_{\Omega} \left| \sqrt{\Lambda_1 (\theta - Cu)} \right|^2 \, dx$$

$$+ \left( \frac{\lambda_2}{\eta} \right) \int_{\Omega} \left[ \varepsilon |\nabla u|^2 + \frac{W(u)}{\varepsilon} \right] \, dx$$

$$+ \int_{\Omega} (s^2 + \kappa_\gamma) \left| \sqrt{\Lambda_3 \nabla \theta} \right|^2 \, dx + \lambda_4 \int_{\Omega} \left[ |\nabla s|^2 + \frac{(1 - s)^2}{4\gamma} \right] \, dx,$$

where $\eta$ is the constant appearing in Theorem 2.1 and $\kappa_\gamma = o(\gamma)$. The following $\Gamma$-convergence result holds.

**Theorem 3.1** Let $h \mapsto \varepsilon_h, h \mapsto \gamma_h$, be sequences with $\varepsilon_h \to 0, \gamma_h \to 0$. Then the family of functionals $F_{\varepsilon_h, \gamma_h}$ $\Gamma$-converges to $F$ with respect to the $L^2(\Omega; \mathbb{R}^k) \times L^2(\Omega; \mathbb{R}^m) \times L^2(\Omega)$ metric as $\varepsilon_h, \gamma_h \to 0$.

The functional $F_{\varepsilon, \gamma}$ has a minimizer in $Y(\Omega)$ for any $\varepsilon, \gamma > 0$. Let $\{(u_{\varepsilon}, \theta_{\gamma}, s_{\gamma})\} \subset Y(\Omega)$ be a family of triplets such that, for any $\varepsilon, \gamma > 0$, $F_{\varepsilon, \gamma}(u_{\varepsilon}, \theta_{\gamma}, s_{\gamma}) \leq M$, with $M$ positive constant independent of $\varepsilon, \gamma$. Then there exist a pair

$$(u, \theta) \in BV(\Omega; \{e_1, \ldots, e_k\}) \times SBV(\Omega; K_2)$$

and sequences $\varepsilon_h, \gamma_h \to 0$ such that $(u_{\varepsilon_h}, \theta_{\gamma_h}, s_{\gamma_h})$ converges to the triplet $(u, \theta, 1)$ in $L^2(\Omega; \mathbb{R}^k) \times L^2(\Omega; \mathbb{R}^m) \times L^2(\Omega)$. Moreover, if $\{(u_{\varepsilon}, \theta_{\gamma}, s_{\gamma})\} \subset Y(\Omega)$ is a family of minimizers of $F_{\varepsilon, \gamma}$, then the limit triplet $(u, \theta, 1)$ minimizes the functional $F$ in $X(\Omega)$.

**Proof.** The proof is a straightforward application of Theorems 2.1 and 2.2 and we only sketch the main steps. Let $h \mapsto \varepsilon_h, h \mapsto \gamma_h$, be sequences with $\varepsilon_h \to 0, \gamma_h \to 0$. Let $\{(u_{\varepsilon_h}, \theta_{\gamma_h}, s_{\gamma_h})\} \subset Y(\Omega)$ be a sequence converging to $(u, \theta, s) \in X(\Omega)$.
in $L^2(\Omega; \mathbb{R}^k) \times L^2(\Omega; \mathbb{R}^m) \times L^2(\Omega)$. Using Theorem 2.1, Theorem 2.2 and (1), we have
\[
\liminf_{h \to +\infty} \left[ \frac{\lambda_2}{\eta} \varepsilon_h(u_{\varepsilon h}) + G_{\gamma h}(\theta_{\gamma h}, s_{\gamma h}) \right] \geq \frac{\lambda_2}{\eta} \liminf_{h \to +\infty} E_{\varepsilon h}(u_{\varepsilon h}) + \liminf_{h \to +\infty} G_{\gamma h}(\theta_{\gamma h}, s_{\gamma h}) \\
\geq \lambda_2 E(u) + G(\theta, s),
\] (7)
where $\alpha$ has been replaced with $\lambda_4$ in the definition of the functionals $G$ and $G_\gamma$.

Using (3), for any $(u, \theta, s) \in X(\Omega)$ there exists a sequence $\{(u_{\varepsilon h}, \theta_{\gamma h}, s_{\gamma h})\} \subset Y(\Omega)$ converging to $(u, \theta, s)$ such that
\[
\lim_{h \to +\infty} \left[ \frac{\lambda_2}{\eta} \varepsilon_h(u_{\varepsilon h}) + G_{\gamma h}(\theta_{\gamma h}, s_{\gamma h}) \right] = \frac{\lambda_2}{\eta} \lim_{h \to +\infty} E_{\varepsilon h}(u_{\varepsilon h}) + \lim_{h \to +\infty} G_{\gamma h}(\theta_{\gamma h}, s_{\gamma h}) \\
= \lambda_2 E(u) + G(\theta, s).
\] (8)
The inequalities (7) and (8) still hold true with the introduction of the matrix $\Lambda_3$ of weights with a slight modification of the proof.

It follows the $\Gamma$-convergence of the sum of the term with the weight $(\lambda_2/\eta)$ and the term with the weights $\Lambda_3$ and $\lambda_4$ in the functional $F_{\varepsilon, \gamma}$ to the sum of the corresponding terms in the functional $F$

Since the term with the weight $\lambda_0$ and the term with the weight $\Lambda_1$ are continuous with respect to the $L^2$ metric, the full $\Gamma$-convergence of $F_{\varepsilon, \gamma}$ to $F$ follows from the stability property of $\Gamma$-convergence with respect to continuous perturbations.

By standard results in Sobolev spaces, for any $\varepsilon, \gamma > 0$ there exists a triplet $(u_{\varepsilon}, \theta_{\gamma}, s_{\gamma})$ which minimizes $F_{\varepsilon, \gamma}$ in $Y(\Omega)$. Let now $\{(u_{\varepsilon h}, \theta_{\gamma h}, s_{\gamma h})\} \subset Y(\Omega)$ be a family of triplets such that, for any $\varepsilon, \gamma > 0$, $F_{\varepsilon, \gamma}(u_{\varepsilon}, \theta_{\gamma}, s_{\gamma}) \leq M$, with $M$ positive constant independent of $\varepsilon, \gamma$. Since $u_{\varepsilon}(x) \in K_1$, which is compact, for any $x \in \Omega$, using Theorem 2.1, there exist $u \in BV(\Omega; \{e_1, \ldots, e_k\})$ and a sequence $\varepsilon_h \to 0$ such that $u_{\varepsilon h}$ converges to $u$ in $L^2(\Omega; \mathbb{R}^k)$. Since $\theta_{\gamma}(x) \in K_2$, which is compact, for any $x \in \Omega$, using Theorem 2.2 there exist $\theta \in SBV(\Omega; \mathbb{R}^m)$ and a sequence $\gamma_h \to 0$ such that the pair $(\theta_{\gamma h}, s_{\gamma h})$ converges to $(\theta, 1)$ in $L^2(\Omega; \mathbb{R}^m) \times L^2(\Omega)$. 13
Eventually, if \( \{(u_\varepsilon, \theta_\gamma, s_\gamma)\} \) is a family of minimizers of \( F_{\varepsilon, \gamma} \), from the variational property of \( \Gamma \)-convergence it follows that the limit triplet \( (u, \theta, 1) \) minimizes \( F \).

4 Euler equations

The approximating functionals \( F_{\varepsilon, \gamma} \) are more convenient for the purpose of numerical computations, since the corresponding Euler equations admit a straightforward discretization by means of finite elements. Since the functional \( F_{\varepsilon, \gamma} \) is not convex, the solution of the system of Euler equations yields a stationary point for the problem which need not necessarily be a minimizer of the functional.

If we assume that the function \( f \) is continuously differentiable, the Euler equations for functional (6) are the following. The first variation with respect to the vector variable \( \theta \) yields the equations

\[
\lambda_0 \int_T (g - f(t, \theta)) \frac{\partial f}{\partial \theta_p} \, dt - \Lambda_1(p, p) \left( \theta_p - \sum_{j=1}^{k} C_{pj} u_j \right) + \text{div} \left( (s^2 + \kappa) \Lambda_3 \nabla \theta_p \right) = 0, \tag{9}
\]
for \( p = 1, \ldots, m \).

The first variation with respect to the vector variable \( u \) yields the equations

\[
\sum_{i=1}^{m} \Lambda_1(i, i) C_{ih} \left( \theta_i - \sum_{j=1}^{k} C_{ij} u_j \right) + \lambda_2 \Delta u_h - \lambda_2 \frac{\partial W}{\partial u_h} = 0, \tag{10}
\]
for \( h = 1, \ldots, k \), where the constant \( \eta \) has been absorbed in the weight \( \lambda_2 \).

The first variation with respect to the scalar variable \( s \) yields the equation

\[
s \sum_{i=1}^{m} \Lambda_3(i, i) |\nabla \theta_i|^2 - \lambda_4 \Delta s - \frac{\lambda_4}{4 \gamma} (1 - s) = 0. \tag{11}
\]
This is a set of \( k + m + 1 \) coupled nonlinear elliptic PDEs. Moreover, while Eq. (11) is linear in the variable \( s \), the other \( k + m \) equations, considered as equations
in the variables $\theta$ and $u$ with given $s$, contain two nonlinear terms involving the partial derivatives of the function $W$ and the integral in (9). We notice also that, given $\theta$, equations (10) and (11) are uncoupled. In order to cope with a possible ill conditioning of the stiffness matrix of the system of elliptic PDEs a ”divide et impera” approach can then be adopted. Given initial values for $\theta, u, s$, solve the subsystem (9) for $\theta$, plug the solution on subsystem (10) and on equation (11) and solve them separately and then iterate these steps. We have also that

$$\frac{\partial W}{\partial u_h} = 2 \left[ \hat{W}(u) u_h - w_h(u) \right], \quad h = 1, \ldots, k,$$

where

$$\hat{W}(u) = \sum_{i=1}^{k} \prod_{j=1, j\neq i}^{k} |u - e_j|^2$$

and

$$w_h(u) = \prod_{j=1, j\neq h}^{k} |u - e_j|^2.$$

Therefore one can resort to an iterative method where, at each iteration, $2\hat{W}(u)$ can be considered as a known coefficient of the linear part of the equation (10), and $-2w_h(u)$ is a part of its known term. Finally, if an analytic expression of the parametric temporal model $f$ is known, the integral $\lambda_0 \int_{T} (g - f(t, \theta)) \frac{\partial f}{\partial \theta} dt$ in equation (9) can be evaluated by first computing analytically $\frac{\partial f}{\partial \theta}$, and then considering the cubic spline which interpolates $(g - f(t, \theta)) \frac{\partial f}{\partial \theta}$ in a fixed finite set of points in $T$. The integral is then a function of the spline coefficients.

## 5 Estimation procedure

The functional (5) is fully determined after the weights $\lambda_0, \Lambda_1, \lambda_2, \Lambda_3, \lambda_4$ are fixed. This is a critical choice which can depend on the specific problem. However it is reasonable, as a first guess, to fix the weights in such a way that the five main terms
of the functional (5) are of the same order of magnitude in order that all of them are active during the minimization process. As we are looking for a solution of the Euler equations it is convenient to choose the weights that balance the terms of each equation. By using the initial values $\theta^{(0)}, u^{(0)}, s^{(0)}$, we then compute the weights as follows, dropping the superscript (0) for simplicity:

$$[\lambda_0^{(0)}]^{-1} = \frac{1}{m} \sum_{i=1}^{m} \left[ \int_{T} \int_{\Omega} (g - f(t, \theta)) \frac{\partial f}{\partial \theta_i} dt \right] dx,$$

$$[\Lambda_1^{(0)}(i, i)]^{-1} = \int_{\Omega} \left| \theta_i - \sum_{j=1}^{k} C_{ij} u_j \right| dx, \quad i = 1, \ldots, m,$$

$$[\lambda_2^{(0)}]^{-1} = \sum_{h=1}^{k} \frac{1}{\sum_{i,h} C_{ih}} \int_{\Omega} \left| \varepsilon \Delta u_h - \frac{1}{2\varepsilon} \frac{\partial W}{\partial u_h} \right| dx,$$

$$[\Lambda_3^{(0)}(i, i)]^{-1} = \int_{\Omega} \left| \text{div} \left[ (s^2 + \kappa \gamma) \nabla \theta_i \right] \right| dx, \quad i = 1, \ldots, m,$$

$$[\lambda_4^{(0)}]^{-1} = \int_{\Omega} \left| \gamma \Delta s + \frac{1}{4\gamma}(1 - s) \right| dx.$$

However this choice of weights might not allow to extract from the data all possible information. It turns out that in difficult problems, such as e.g. the dynamic MRI problem, more fine tuning of weights is necessary. To this aim we consider a family of functionals parameterized by the weights and exploit the information about the known statistical distribution of the noise affecting the data, to select a good member of the family. The idea is that if the functional is able to extract from the data all the structured information, the residual $g - f(t, \theta)$ should have the statistical distribution of the noise. In a simple discrete setting [7] where $f(t, \theta)$ is a linear function, only one scalar weight is present, and the functionals are convex, the choice of the weight and the minimization of the corresponding functional was formulated as a constrained optimization problem. It was proved that such an optimization problem admits one and only one solution, i.e., it exists one optimal
weight and the corresponding functional has one and only one minimum. Moreover
an heuristic fast algorithm was proposed to compute a solution. The basic idea is
to build an iteration where at each step the weights are chosen by unbalancing the
previous ones and the corresponding functional is minimized. If the initial values of
\( \theta \) give rise e.g. to a small residual, but non regularized \( f(t, \theta) \), the unbalancing of
the weights will be in the direction of increasing the regularity. In [6] the same ideas
were generalized to cope with non convex functionals, non linear function \( f(t, \theta) \)
and multiple weights and it was experimentally proved that the heuristic algorithm
works in this setting too. Therefore in the present setting we propose to use the
same algorithm where at each step instead of solving a local (in \( x \)) minimization
problem, a system of elliptic PDEs is solved.

For \( \mu_0, \mu_1 \in (0, 1) \) let be \( \mu_n = \mu_0 \mu_1^{n-1} \). Then the weights at iteration \( n \)-th are
unbalanced as follows

\[
\lambda_0^{(n)} = \mu_n \lambda_0^{(0)}, \\
\Lambda_1^{(n)}(i, i) = \mu_n \Lambda_1^{(0)}(i, i), \\
\lambda_2^{(n)} = (1 - \mu_n) \lambda_2^{(0)} \sum_{i,h} C_{ih}, \\
\Lambda_3^{(n)}(i, i) = (1 - \mu_n) \Lambda_3^{(0)}(i, i), \\
\lambda_4^{(n)} = (1 - \mu_n) \lambda_4^{(0)}.
\]

To quantify the statistical properties of residuals, let us assume to have a finite
number of observations

\[ g(t_j, x_l), \quad t_j = (j - 1)\Delta_t, \quad j = 1, \ldots, N_t, \quad x_l \in \mathcal{L}, \quad l = 1, \ldots, N_x^2, \]

where \( \mathcal{L} \) is a square lattice of dimension \( N_x \), such that

\[ g(t_j, x_l) = f(t_j, \theta^*(x_l)) + \nu(t_j, x_l), \]

17
\( \theta^* \) are the true unknown values of the parameters, and \( \nu(t_j, x_l) \) are i.i.d. zero-mean Gaussian random variables with variance \( \sigma^2 \). Let

\[
\mathcal{E}(j, \theta) = \frac{1}{N_x} \sum_{l=1}^{N^2} I\{((j-1)B < \hat{\sigma}_l^2 \leq jB)\}, \quad B = \frac{N_t \Delta t}{N_b}, \tag{12}
\]

be the empirical distribution over the pixels of the statistic

\[
\hat{\sigma}_l^2 = \frac{\sum_{j=1}^{N_t} (g(t_j, x_l) - f(t_j, \theta(x_l)))^2}{\sigma^2}, \tag{13}
\]

where \( I(A) \) is the indicator function of the set \( A \) and \( N_b \) is the number of bins. Then define

\[
R(\theta) = \sum_{j=1}^{N_b} \left( \chi^2_{N_t}([j+1/2]B) - \mathcal{E}(j, \theta) \right)^2, \tag{14}
\]

where \( \chi^2_{N_t} \) is a chi-squared distribution with \( N_t \) degrees of freedom. When \( \theta = \theta^* \) the distribution of the residues is a \( \chi^2_{N_t} \) and \( R(\theta) \approx 0 \); therefore \( R(\theta) \) measures the adequacy of the model \( f \) to represent the data \( g \).

A sketch of the algorithm is given in Table 1 where we denoted by \( \text{pde}_{\theta}, \text{pde}_u, \text{pde}_s \) solvers of the elliptic systems (9), (10) and (11), respectively, and

\[
\Lambda^{(n)} = \left[ \lambda_0^{(n)}, \lambda_1^{(n)}, \lambda_2^{(n)}, \lambda_3^{(n)}, \lambda_4^{(n)} \right].
\]

The matrix \( C \) is computed from the initial values \( \theta^{(0)} \) as follows

\[
C_{ij} = \frac{1}{|Q_j|} \sum_{x \in Q_j} \theta^{(0)}_i(x), \quad i = 1, \ldots, m; j = 1, \ldots, k,
\]

where \( Q_j = \{x : u(x) = e_j\} \) and \( |Q_j| \) is the cardinality of \( Q_j \). More specifically at each inner iteration a damped Newton step with the Armijo-Goldstein line search strategy is performed. The Jacobian matrix is approximated by the stiffness matrix \( J \). The resulting linear system is solved by Generalized Minimum Residual method with restarts [15] using the LU factorization of \( J \) as preconditioners.


\[ i = 0; n = 0; R(\theta^{(-1)}) \gg 0; \]

fix \textit{max.iter.}; fix \textit{step};

Compute \( u^{(0)}, \theta^{(0)}, s^{(0)}, R(\theta^{(0)}), \Lambda^{(0)}, C; \)

while \( i < \text{max.iter.} \) do

\[ i = i + 1; \]

if \( \left\lfloor \frac{i+1}{\text{step}} \right\rfloor \text{step} = i + 1 \) then

\[ n = n + 1; \]

compute \( \Lambda^{(n)} \)

end

\[ s = \text{pde}_s(\theta^{(i-1)}, \Lambda^{(i-1)}); \]

\[ u = \text{pde}_u(\theta^{(i-1)}, \Lambda^{(i-1)}); \]

\[ \theta = \text{pde}_\theta(u^{(i)}, s^{(i)}, \Lambda^{(i-1)}); \]

if \( R(\theta) < R(\theta^{(i-1)}) \) then

\[ s^{(i)} = s; \quad u^{(i)} = u; \quad \theta^{(i)} = \theta; \]

end

stop

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{Table 1: The heuristic algorithm.} \\
\hline
\end{tabular}
\end{table}

Because a solution of Euler equations corresponds to a local minimum of the functional, initialization is a critical step. Moreover a wrong initialization has a bad feedback on the initial weights. However a reasonable initial value for \( \theta(x) \) can be obtained by solving for each \( x_l \) separately the nonlinear least squares problem

\[
\theta^{(0)}(x_l) = \arg\min_\theta \sum_{j=1}^{N_l} \|g(t_j, x_l) - f(t_j, \theta(x_l))\|^2.
\]

An initial value for \( u \) can then be obtained by any clustering method applied to \( \theta^{(0)} \) which produces a scalar function \( \tilde{u}(x) \in \{1, \ldots, k\} \). Therefore \( u^{(0)}(x) = e_j \) if
\( \tilde{u}(x) = j \). Despite the fact that equation (11) is linear in \( s \) and therefore does not need an initial value, it is important to start with a reasonable \( s \) because the initial weights will depend on it. We therefore define \( s^{(0)} = |\Delta \tilde{u}| \).

Some remarks on the tuning of the algorithm to choose the constants \( \varepsilon, \gamma, \kappa, \mu_0, \mu_1 \) are in order. The equality (3) of \( \Gamma \)-convergence is useful for the design of numerical algorithms. Particularly, the sequence \( s_\gamma \) used for the proof of (3) is constructed in the following way [3]. For any \( \gamma > 0 \) the function \( s_\gamma \) has a sharp transition from 1 to 0 in a tubular neighbourhood of the discontinuity set \( S_\theta \) of the parameters vector field \( \theta \). Let \( \tau(x) \) denote the distance function from the set \( S_\theta \), i.e., \( \tau(x) = \text{dist}(x, S_\theta) \), then in a tubular neighbourhood \( \{ x : \tau(x) < a_\gamma \} \) with \( a_\gamma = O(\varepsilon \log \varepsilon) \) positive infinitesimal, the function \( s_\gamma \) is given by

\[
\psi_\gamma \circ \tau, \quad \psi_\gamma(y) = 1 - \exp \left[ \frac{-y^2}{2\gamma} \right].
\]

The transition of the function \( s_\gamma \) from 1 to 0 takes place in a tubular neighbourhood of \( S_\theta \) having width approximately given by \( 20\gamma \). Because of the inequalities (1) and (3), which hold for any triplet \( (u, \theta, s) \in X(\Omega) \), the function \( s_\gamma \) given by (15) yields the asymptotic form of the optimal transition as \( \gamma \to 0 \). Analogously, it turns out that the function \( u_\varepsilon \) used for the proof of (3) has sharp transitions in tubular neighbourhoods of the sets \( \partial^* A_i \cap \partial^* A_j \cap \Omega \), for \( i, j \in \{1, \ldots, k\} \). The width of the neighbourhoods in this case are of order \( O(\varepsilon) \).

Since the mesh of a finite element method must resolve the width of the transition regions, the mesh size \( h \) has to decrease when \( \varepsilon, \gamma \) decrease. Discretization by finite elements transforms the functionals \( F_{\varepsilon, \gamma} \) into functionals depending on both \( \varepsilon, \gamma \) and \( h \). For the case of the Mumford-Shah functional the \( \Gamma \)-convergence of the family of discretized functionals was proved by Bellettini and Coscia [8] under the condition that \( h \) approaches 0 faster than \( \gamma \). In practical applications the less restrictive condition \( h = O(\gamma) \) is generally used. In our case we set \( \gamma = h/10 \), \( \varepsilon = 3\gamma \), and
Finally, the balancing parameters $\mu_0, \mu_1$ are then chosen by trial and errors running the heuristic algorithm for several values of $\mu_0, \mu_1$. The best ones provide the less value of the adequacy test $R$ for given values of $\varepsilon$ and $\gamma$.

6 Numerical experiments

The proposed method is now used to solve the dynamic MRI problem described in the introduction with the pharmacokinetic model considered in [6]. The function $g(t, x)$ is defined as the output of a stochastic linear dynamic distributed system given by

$$
\dot{\zeta}(t, x) = A(\theta(x))\zeta(t, x) + \xi(t, x),
$$

$$
g(t, x) = b^T(x)\zeta(t, x) + \nu(t, x),
$$

$$
\zeta(0, x) = 0, \quad \theta_i(x) \geq 0, \quad i = 1, \ldots, m,
$$

(16)

$b(x) = [H(x), 0]^T, \quad \xi(t, x) = [0, \delta(t)]^T, \quad t \in T = [0, T]$ and $\nu(t, x)$ is a Gaussian white noise process with variance $\sigma^2$, independent of $\nu(t, y), \forall y \neq x$, representing the measurement noise. In [10] the conditions under which this hypothesis is valid are fully discussed. The matrix $A(\theta(x))$ is given by

$$
A = \begin{bmatrix}
-K_{21}(x) & K_{12}(x) \\
K_{21}(x) & -(K_{12}(x) + K_o(x))
\end{bmatrix}
$$

(17)

and $\theta(x) = [K_{21}(x), K_{12}(x), K_o(x), H(x)]^T$. The parameters $K_{21}(x), K_{12}(x), K_o(x)$ represent respectively the exchange rate of contrast agent entering the blood plasma from the extracellular space, that entering the extracellular space from the blood plasma, and that leaving the system. The parameter $H(x)$ depends on the intrinsic
longitudinal relaxation time of the tissue without contrast agent and its relaxivity [13]. The function $f(t, \theta(x))$ is given by the predictor

$$f(t, \theta(x)) = b^T(x) \left[ \int_0^t e^{(t-s)A(\theta(x))} \xi(s) ds \right].$$

(18)

We notice that in order to allow the function $f(t, \theta(x))$ to well represent the measured data the parameters $K_{21}(x)$ must be several order of magnitude lesser than $K_0(x)$. Coupled with the analytical form of $f(t, \theta(x))$, this makes the parameter identification problem very ill posed. Therefore numerical instability in the solution of the Euler equations can be expected. To cope with this problem, bounds for $\theta(x)$ have been computed in such a way that the measured data are roughly included in the envelope

$$\left\{ f(t, \theta(x)) : \theta_p(x) \in [\theta_{inf}^p(x), \theta_{sup}^p(x)], p = 1, \ldots, m \right\}.$$

Then the damped Newton step has been modified to take into account these constraints. The proposed procedure has been tested on the same data set used in [6, 10] in order to facilitate comparisons. A sequence of 20 dynamic magnetic resonance $256 \times 256$ images with a temporal resolution of 15 seconds has been considered. The first image has been acquired before the contrast agent injection. The data considered are the difference between the post contrast images and the pre contrast one. A region of interest (ROI) of $29 \times 29$ pixels was chosen enclosing a tumor. The measured time courses associated to the pixels in the ROI are shown in fig.1 where the segmentation and classification results obtained by the proposed procedure are also represented as illustrated below.

We set $\Omega = [-1, 1]^2$, $h = 2/29$, $\gamma = 2/290$, $\varepsilon = 6/290$, $\kappa_\gamma = 2/2900$, $\text{step} = 2$, $\text{max.iter} = 6$, $\mu_0 = 0.99$ and $\mu_1 = 0.995$. Moreover we chose $k = 3$ because we want to classify pixels either as normal or as benign or malignant tumoral ones.
In fig. 2 the segmentation, the classification and the empirical distribution \( \mathcal{E}(j, \theta) \), \( j = 1, \ldots, N_b \) and the corresponding \( \chi^2_{N_t}([j + 1/2]B), \ j = 1, \ldots, N_b \) obtained by the proposed method are reported as well as their corresponding initial values. Three classes are shown: normal (white), benign tumoral (gray) and malignant tumoral (black). To visually check the segmentation and the classification ability of the algorithm, in fig. 1 the measured time courses classified as malignant tumoral are plotted in red, the benign tumoral ones in black and the normal ones are not plotted. The tumoral region is consistent with the estimated segmentation. Moreover the main classification features such as the ring structure and the necrotic region at the centre of the tumor are well estimated too and are consistent with the observed time courses of the corresponding pixels given in fig. 1.

In fig. 3 the initial and final values of the pharmacokinetic parameters \( K_{21}, K_{12}, K_\circ, H \) are plotted. Some extra smoothing can be observed. However this is the price to pay to the intrinsic ill posedness of the problem discussed above.

**References**


Figure 1: Measured time courses inside a ROI in a patient affected by a breast tumor. Signals classified as malignant are in red and those classified as benign are in black. Non tumoral signals are not shown.
Figure 2: First row: segmentation, second row: classification, third row: the empirical distribution $\mathcal{E}(j, \theta)$, $j = 1, \ldots, N_b$, and the corresponding $\chi^2_{N_t}([j + 1/2]B)$, $j = 1, \ldots, N_b$. First column: initial estimates, second column: final estimates. $\mu_0 = 0.99$
Figure 3: Pharmacokinetic parameters $K_{21}, K_{12}, K_0, H$. Left: initial values. Right: final values. $\mu_0 = 0.99$