

A Multi-task Learning Approach for Compartmental Model Parameter Estimation in DCE-CT Sequences

Blandine Romain^{1,2,5}, Véronique Letort¹, Olivier Lucidarme³,
Laurence Rouet², and Florence d'Alché-Buc^{4,5}

¹ MAS, Ecole Centrale Paris, Chatenay-Malabry

² Philips Research, Suresnes

³ Hospital La Pitie-Salpetriere, AP-HP, Paris

⁴ INRIA-Saclay, LRI CNRS 8623, Orsay, France

⁵ IBISC, University of Evry, Evry, France

Abstract. Today's follow-up of patients presenting abdominal tumors is generally performed through acquisition of dynamic sequences of contrast-enhanced CT. Estimating parameters of appropriate models of contrast intake diffusion through tissues should help characterizing the tumor physiology, but is impeded by the high level of noise inherent to the acquisition conditions. To improve the quality of estimation, we consider parameter estimation in voxels as a multi-task learning problem (one task per voxel) that takes advantage from the similarity between two tasks. We introduce a temporal similarity between tasks based on a robust distance between observed contrast-intake profiles of intensity. Using synthetic images, we compare multi-task learning using this temporal similarity, a spatial similarity and a single-task learning. The similarities based on temporal profiles are shown to bring significant improvements compared to the spatial one. Results on real CT sequences also confirm the relevance of the approach.

Keywords: Multi-task learning, CT perfusion, model parameter estimation.

1 Introduction

Providing earlier assessment of drugs efficiency is a major challenge for the improvement of patient care in oncology. Today's follow-up of patients presenting tumors is performed using dynamic sequences of contrast-enhanced CT (DCE-CT) [1]. The goal is to associate the temporal behavior of contrast intake to physiological parameters that characterize vascularisation of the tissues.

A possible approach relies on modeling the diffusion of contrast agent in a spatially discretized region [2]. However, estimation of physiological parameters from the intensity profile of enhancement curves remains a challenging task. First, sequences acquired in free-breathing condition are contaminated by errors resulting from their registration and noise due to low-dose settings required to

limit X-Ray exposure of the patient. Second, physiological models are highly nonlinear and may induce non-convexity in the least square criterion. Several studies have examined the use of prior knowledge in order to reduce variance in estimation and focus on relevant local minima, especially in DCE-MR [3,4]. They introduce spatial knowledge under the form of a smoothing prior in a Bayesian approach to impose that parametric models attached to voxels in the same neighborhood have close parameters. However this raises the issue of the tissue boundaries where the constraint must be locally relaxed. [3] developed an adaptive Gaussian Random Field for which they also update locally the variance of each prior to be robust on tissue boundary. In the case of [4], a Generalized Gaussian Random Field (GGRF) was used with a smoothing spatial prior. In our work, we adopt a compatible but different angle using a multi-task learning approach, considering one parametric model estimation task per voxel. Expressed in the regularization framework, the smoothness constraint involved in multi-task learning is based on the definition of a similarity between tasks. It allows us to consider other similarity candidates than the spatial one: a temporal similarity between tasks from a robust distance between intensity time-course of two voxels. Numerical results on synthetic and real image sequences show the relevance of this approach compared to the use of spatial similarity.

2 Framework of Two-Compartment Models of Perfusion

In oncology, the objective of DCE-CT imaging is to assess tissue microvasculature and microcirculation in abdominal regions using the temporal change of the image signal following the administration of a diffusive contrast agent. Several types of compartmental models have been developed [5,2], that describe the temporal evolution of the tissue concentration of contrast agent ($C(t)$) in a region of interest as a function of its physiological properties. Most of them can be written under the following generic formalism of a set of first-order non-linear ordinary differential equations (ODEs) including hidden state variables:

$$\forall i \in \Omega, \forall t > 0, \quad \begin{cases} \frac{d\mathbf{x}^i}{dt}(t) = f(\mathbf{x}^i(t), t, \boldsymbol{\theta}^i) \\ \mathbf{x}^i(0) = \mathbf{c}_0^i \\ C_{tiss}^i(t) = g(\mathbf{x}^i(t), \boldsymbol{\theta}^i) \\ y^i(t_l) = C^i(t_l) + \epsilon^i(t_l), \forall t_l \in \Omega_T \end{cases} \quad (1)$$

where $\Omega = \{1 \dots, N\}$ refers to the set of voxels; $\mathbf{x}^i(t)$ is a set of state variables, for instance the concentrations of contrast agent in the tissue; g expresses tissue concentration from the internal state variables, and $\boldsymbol{\theta}^i \in \mathbb{R}^m$ is a vector of m parameters related to the physiological properties of the tissue. These parameters have to be estimated for the multiple local models (C^i) _{$i \in \Omega$} from noisy contrast-intake curves $y^i(t_l)$, where $t_l \in \Omega_T = \{t_0 \dots, t_{T-1}\}$ is the sequence of acquisition times and where $\epsilon^i(t_l)$ denotes the noise in the i^{th} zone at time t_l . A particular set of initial conditions, \mathbf{c}_0^i , is associated to each voxel.

In this paper, we assume that there exists an analytical solution for C^i (Eq. 1), noted $C^i(t, \boldsymbol{\theta}^i, \mathbf{c}_0^i)$ as the concentration at time t .

3 Parametric Estimation Using a Multi-task Learning Approach

A direct approach to parameter estimation for each voxel $i \in \Omega$ consists in minimizing the square loss function under constraints $0 \leq \theta^i \leq \theta_{max}$:

$$\hat{\theta}^i = \arg \min_{\theta_p^i \in \mathbb{R}^m} \frac{1}{T} \sum_{t \in \Omega_T} (C^i(t; \theta^i, \mathbf{c}_0^i) - y^i(t))^2, \forall p \in \{1, \dots, m\} \quad (2)$$

Such voxelwise estimation of each model raises different issues depending on the nature of the function f (Eq. 1) and the corresponding analytical solution of the ODE. A complex function C can lead to a solution with some practically non-identifiable parameters due to the lack of data [6]. Another issue concerns the loss function itself, that can exhibit multiple local minima. To circumvent these difficulties, we can add additional information. For that purpose, we introduce a multi-task learning approach to jointly estimate all the models while benefiting from the similarity between tasks. Multi-task learning has been developed in other fields using nonparametric models. In this case [7], a smoothness constraint forcing two models to be close when the underlying tasks are close is added to the empirical loss function (Eq. 2). Since our models are parametric, the constraint will directly use their parameters: two voxels in similar tissue will correspond to similar parameters. Sharing parameters can be expressed as an additional smoothing constraint under the form of a weighted norm of the two parameters. Thus, the complete loss functional $\mathcal{L}(\theta)$ writes as a penalized least square loss:

$$\mathcal{L}(\theta) = \frac{1}{T \cdot N} \sum_{\substack{i \in \Omega \\ t \in \Omega_T}} (C^i(t; \theta^i, \mathbf{c}_0^i) - y^i(t))^2 + \lambda \frac{1}{N^2} \sum_{(j, \ell) \in \Omega^2} w_{j, \ell} \|\theta^j - \theta^\ell\|_2^2 \quad (3)$$

where λ weights the trade-off between the square loss and the regularization term; $W = (w_{ij})$ is a $N \times N$ matrix encoding the similarity between the i^{th} and the j^{th} tasks. $\|\cdot\|_2$ corresponds to norm ℓ_2 .

In multi-task learning problems, W is defined prior to the data if some knowledge is available or defined from the data. We discuss two choices of data-based similarity. A first choice is a similarity between tasks, based on spatial proximity between voxels. Let d_s denote the Euclidean distance between two voxels. The similarity matrix $W_s = (w_{i,j}^s)$ is defined as the Gram matrix of a Gaussian kernel

$$\forall (i, j) \in \{1, \dots, N\} \times \{1, \dots, N\}, w_{i,j}^s = \exp(-\gamma_s d_s(i, j)^2) \quad (4)$$

where γ_s is the kernel parameter.

The constraint based on the spatial similarity encourages that two voxels in the same neighborhood share the same parameters. However, when two voxels are spatially close but reflect different temporal behaviors or belong to different organs, this constraint must be relaxed. So, we introduce a second distance between the estimation tasks that compares two observed temporal intensity profiles C^i and C^j , noted $d_t(i, j)$. To ensure robustness to noise, this distance is

defined between two nonparametric kernel-based models of the observed data. In the Reproducing Kernel Hilbert Space \mathcal{H} , built from a Gaussian kernel k , we define the distance $d_t(i, j) = \|h^i - h^j\|_{\mathcal{H}}^2$ where h^i (resp. h^j) is a nonparametric kernel-based model built from the data C^i (resp. C^j). For a given $i \in \Omega$, the kernel-based function h^i is obtained by minimizing the ridge loss $\sum_{\ell \in \Omega_T} (h^i(t_\ell) - y^i(t_\ell))^2 + \beta_i \|h^i\|_{\mathcal{H}}^2$. This equation admits a closed-form solution

$$h^i(t) = \sum_{\ell \in \Omega_T} \alpha_\ell^i k(t, t_\ell)$$

where $\alpha^i = (K + \beta_i I)^{-1} \mathbf{y}^i$; $K = (k(t_\ell, t_m))_{\ell, m}$ the Gram matrix of the kernel on the data, e.g. the time points and I , the identity matrix of same size $\{T \times T\}$. Given that the squared norm of $h^i(\cdot) = \sum_{\ell} \alpha_\ell^i k(\cdot, t_\ell)$ is $\|h^i\|_{\mathcal{H}}^2 = (\boldsymbol{\alpha}^i)^T K \boldsymbol{\alpha}^i$, this leads to the following distance d_t and the temporal similarity matrix $W_t = (w_{i,j}^t)$, with γ_t is the kernel parameter:

$$\forall (i, j) \in \{1, \dots, N\} \times \{1, \dots, N\} \quad d_t(i, j)^2 = (\boldsymbol{\alpha}^i - \boldsymbol{\alpha}^j)^T K (\boldsymbol{\alpha}^i - \boldsymbol{\alpha}^j) \quad (5)$$

$$w_{i,j}^t = \exp(-\gamma_t d_t(i, j)^2)$$

Minimizing \mathcal{L} (Eq. 3) is a nonlinear optimization problem that we solve using the interior-point method [8] with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method to calculate the Hessian.

4 Results: Validation on Synthetic Sequences and Application to CT Abdominal Images

4.1 Synthetic Sequences

In order to test our joint estimation method over voxelwise estimations, we use a simple parametric model for the concentration intake curve [9]:

$$\forall i \in \Omega, \forall t \in \Omega_T, \quad \frac{dC^i(t)}{dt} = \theta_1^i C_a(t) - \theta_2^i C^i(t) \quad (6)$$

where C_a represents the arterial input function and C^i the tracer concentration which varies with time in tissue at the voxel i . Vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is a vector of parameters to estimate, with $\boldsymbol{\theta}_1 = (\theta_1^i)_{i \in \Omega}$ and $\boldsymbol{\theta}_2 = (\theta_2^i)_{i \in \Omega}$. These parameters have biophysical meaning [9]: θ_1 is related to the rate of molecule transfer between the blood plasma and the extravascular extracellular space (EES) while θ_2 include the influence of the relative EES volume within the considered voxel.

We compare our proposed framework of multi-task parameter estimation with the voxelwise method (Eq. 2), noted E . The multi-task estimation can be derived with a spatial similarity (Eq. 4) or a temporal one (Eq. 5). These estimation methods are respectively noted SE and TE .

To evaluate the accuracy of estimation, we build synthetic perfusion sequences. The simulations used to build the test sets are based on realistic values of the

perfusion parameters $\theta = (\theta_1, \theta_2)$ and on C_a values extracted from a real sequence. A reference sequence (RS) of (6×6) frames is simulated with $T = 66$ and real time intervals ranging from 2.5 sec to 10 sec. Each frame is composed of 3 regions, with two types of tissue. To reflect the noisy conditions of real acquisitions, we generate noisy sequences NS from our RS with a zero-mean gaussian white noise of local variance ($\sigma = 0.01$). For statistical validation of parameter estimation, we create 50 instances of NS , noted $NS_s, s \in \{1, \dots, 50\}$.

Selection of Hyperparameters. Hyperparameter γ_s (resp. γ_t) is fixed to the inverse of the empirical variance of the distance d_s (resp. d_t). These values correspond to the maximum of the corresponding Gram matrix entropies. Hyperparameters β_i involved in the one-dimensional nonparametric modeling for definition of d_t are fixed using Generalized Cross Validation [10]. Hyperparameter λ , the most important one, is fixed by studying the sensitivity of estimators from NS inputs on a grid of candidate values for λ .

To this end, we define the images of errors $IE_{p,s}$ for each parameter θ_p on each image NS_s : $IE_{p,s}^i = |\hat{\theta}_{p,s}^i - \theta_p^i|/\theta_p^i$ with $p = \{1; 2\}$ and $\hat{\theta}_{p,s}$ the estimated parameter of θ_p . We then define a global estimation error for each sequence NS_s by: $EF_s = \sum_{i \in \Omega} \sum_{p=1}^2 \left(\hat{\theta}_{p,s}^i - \theta_p^i \right)^2$.

Each method of multi-task parameter estimation (SE and TE) is run using different values of λ on the 50 samples NS_s . Figure 1 shows the behavior of EF_s as a function of λ . The value $\lambda = 10^5$ corresponds to a minimum of EF_s for all the tested estimation approaches, and is thus selected in the experiments.

Results on Synthetic Sequences. To evaluate our method, we compute the empirical means of $IE_{1,s}$ and $IE_{2,s}$, respectively noted $\overline{IE_1}$ and $\overline{IE_2}$, and the empirical variances $var(IE_1)$ and $var(IE_2)$ for each image NS_s . The improvement for parameter 1 is about 0.2% in variance and 1% in average which seems modest but was expected given that even voxelwise estimates were relatively accurate. Results (Fig. 2) show that multi-task estimation allows in all cases (SE and TE) to reduce the variance on the error of parameter estimations. The temporal similarity provides an additional improvement on the estimation accuracy. More gain is obtained for the second parameter. The improvement is about 1% in variance and 5% in average. In addition, the overall variances of the estimates are much reduced by multi-task estimation as shown in the boxplots.

4.2 Application to Real CT Abdominal Images

In real sequences, patients are on free-breathing during the whole acquisition (~ 5 minutes). A registration algorithm [11] is therefore first applied. An additional DCT-based temporal denoising is then used to limit the influence of residual patient motions and of high noise level due to CT low-dose.

Figure 3 shows the perfusion parameters that are estimated from 5 patients in specific regions of interest (ROI) with the three methods (E , SE and TE). For computational performance reasons, estimations are performed at lower image resolutions. For Patient 1 and 2 (ROI at the boundary of the left kidney), the parameters in each zone are more homogeneous using our joint estimation meth-

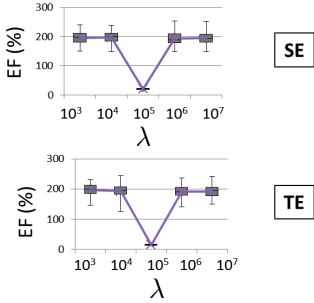


Fig. 1. Sensitivity of the different estimators to λ . Results are averaged on 50 samples.

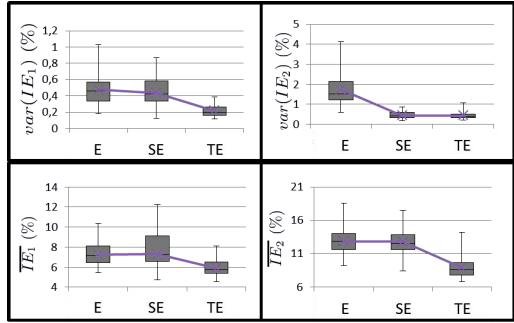


Fig. 2. Boxplots on the 50 images $NS_s, s \in \{1, \dots, 50\}$ for the error variances (top) and means (bottom). Left: Param. θ_1 , Right: θ_2 .

ods, with organ borders slightly less blurred with the temporal similarity. For Patient 3, the ROI consists a part of the tumor, a part of the right kidney and its boundaries. These regions can be better identified when including a regularization term, with similar performance of the *SE* and *TE* methods. For Patient 4, the ROI is in a tumor near the left kidney. The tumor is well drawn when we display the parameters that are estimated with *TE* method, in particular for θ_2 . For Patient 5, the ROI is a tumor inside the liver and its low vascularization level makes the estimation task particularly difficult. The parametric images are therefore very noisy, but the form of the tumor can be distinguished on those obtained with *TE* method, in particular with θ_2 .

To have more quantitative validation, we also computed an index, defined as the number of connected regions obtained after a threshold segmentation into 3 parameter classes. We found in average 22.6 regions for *E*, 8.6 for *SE* and 7 for *TE*, thus quantifying the gain provided by the regularization. To further differentiate *SE* and *TE*, we counted the “artifact” regions, consisting of 1 or 2 pixels: there were in average 5.2 for *SE* and 4 for *TE*.

5 Discussion

Other works [3,4] have proposed to use spatial prior knowledge in parametric estimation by encoding this knowledge as a smoothing prior in the Bayesian estimation of Gaussian Random Fields.

The multitask angle we adopt in this paper differs from those two approaches. First, instead of developing a Bayesian approach, we use a regularization framework to encode a smoothness constraint that only requires to define a similarity matrix between tasks. Second, we have chosen a similarity that expresses how close are the input data. Such a similarity measures how two estimation tasks are close whatever the position of a voxel and consequently, the corresponding constraint tends to reinforce the continuity of the parametric estimator. The

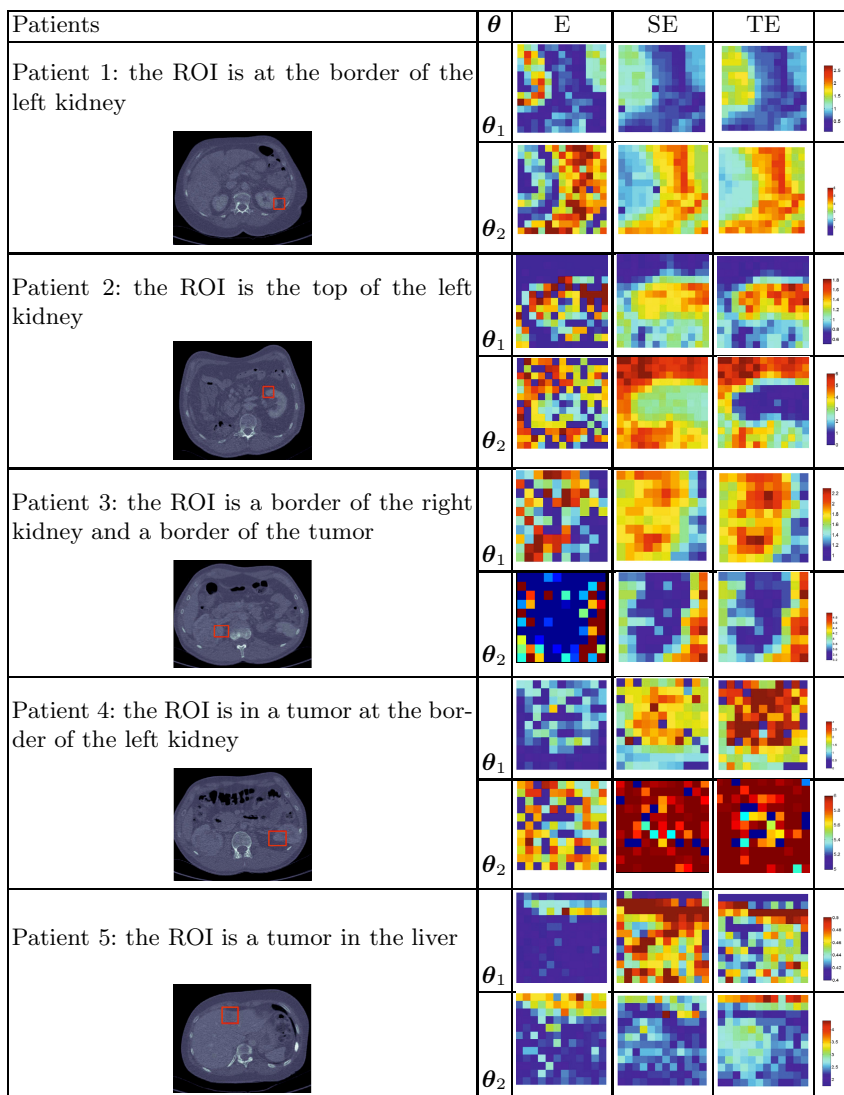


Fig. 3. Estimated parameters on 5 patients. For each patient, the first line represents θ_1 and the second line represents θ_2 , using 3 methods: single-voxel estimation E , joint estimation with spatial similarity SE , and with temporal similarity TE .

strength of this constraint is simply controlled by a single hyperparameter λ and does not require the definition of local priors. Clinically, two voxels located in a same organ, but at a large spatial distance, are expected to exhibit a high temporal similarity, and their perfusion parameters may be jointly estimated in a multi-task framework. Our results show that the estimation using temporal similarity instead of spatial one improves the quality of the parameters.

Interestingly, our multitask approach can be developed for other diffusion models that enjoy an analytical solution. In more complex models, such as those reviewed in [2], we expect that the approach will help with non-identifiability of some parameters (see for instance [5]). Moreover, even if the ODE system does not admit an analytical solution, methods of parametric estimation using numerical approximations of the ODE integration can also benefit from the same multitask approach but at the cost of increased computational complexity.

This regularized estimation based on a smoothing constraint can still be applied to larger images and the complexity order reduced: the similarity matrix W can be thresholded and quantized into binary values. Then the computational cost is related to the sparsity level of the matrix because only neighbors with non-zero weights are considered.

References

1. Miles, K., et al.: Current Status and Guidelines for the Assessment of Tumour Vascular Support with Dynamic Contrast-Enhanced Computed Tomography. *Eur. Radiol.* 22(7), 1430–1441 (2012)
2. Sourbron, S.P., Buckley, D.L.: Tracer Kinetic Modelling in MRI: Estimating Perfusion and Capillary permeability. *Phys. Med. Biol.* 57(2), R1–R33 (2012)
3. Schmid, V., Whitcher, B., et al.: Bayesian Methods for Pharmacokinetic Models in Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging* 25(12), 1627–1636 (2006)
4. Kelm, B.M., Menze, B.H., et al.: Estimating Kinetic Parameter Maps from Dynamic Contrast-Enhanced MRI using Spatial Prior Knowledge. *IEEE Trans. Med. Imaging* 28(10), 1534–1547 (2009)
5. Brix, G., Griebel, J., et al.: Tracer Kinetic Modelling of Tumour Angiogenesis Based on Dynamic Contrast-Enhanced CT and MRI Measurements. *Eur. J. Nucl. Med. Mol. Imaging* 37(suppl. 1), S30–S51 (2010)
6. Raue, A., Kreutz, C., et al.: Addressing Parameter Identifiability by Model-Based Experimentation. *IET Syst. Biol.* 5(2), 120–130 (2011)
7. Evgeniou, T., Pontil, M.: Regularized Multi-Task Learning. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2004*, pp. 109–117. ACM, New York (2004)
8. Forsgren, A., Gill, P., et al.: Interior Methods for Nonlinear Optimization. *SIAM Review* 44(4), 525–597 (2002)
9. Tofts, P.S., Brix, G., et al.: Estimating Kinetic Parameters from Dynamic Contrast-Enhanced T1-Weighted MRI of a Diffusible Tracer: Standardized Quantities and Symbols. *Journal of Magn. Reson. Imaging* 10(3), 223–232 (1999)
10. Golub, G.H., Heath, M.T., et al.: Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* 21, 215–223 (1979)
11. Romain, B., Letort, V., Lucidarme, O., d’Alché-Buc, F., Rouet, L.: Registration of Free-Breathing Abdominal 3D Contrast-Enhanced CT. In: Yoshida, H., Hawkes, D., Vannier, M.W. (eds.) *MMCP 2012 and CCAAI 2012*. LNCS, vol. 7601, pp. 274–282. Springer, Heidelberg (2012)