Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2020 February ; 11313: . doi:10.1117/12.2549625.

Simultaneously spatial and temporal Higher-Order Total Variations for noise suppression and motion reduction in DCE and IVIM

Renjie He^a, Yao Ding^a, Abdallah S. R. Mohamed^a, Sweet Ping Ng^a, Rachel B. Ger^a, Hesham Elhalawani^a, Baher A. Elgohari^a, Kristina H. Young^{b,c}, Katherine A. Hutcheson^a, Clifton D. Fuller^a, Stephen Y. Lai^{*,a}

^aUniversity of Texas, MD Anderson Cancer Center, Houston, TX 77030

^bEarle A. Chiles Research Institute, Providence Cancer Center, Portland, OR 97213

^cRadiation Oncology Division, The Oregon Clinic, Portland, OR 97213

Abstract

In many applications based on kinetic evaluation analysis and model fitting, quantitative mapping retrieved on data series from modalities such as MRI is completed on a voxel-by-voxel basis, where motion and low signal to noise ratio (SNR) would considerably degenerate the reliability of estimations. The coherence of image series in space and time can be used as prior knowledge to mitigate this occurrence. In this study, spatial and temporal higher-order total variations (HOTVs) are applied on a data series of MRI signal (e.g. dynamic contrast-enhanced (DCE) MRI and intravoxel incoherent motion (IVIM) MRI) to exploit the coherence of signal in space and time to minimize the variabilities caused by motion as well as improving quality of images with low SNR while retaining the physical details of original data properly. Simultaneously applying spatial and temporal HOTVs on images is non-trivial in implementation since it is a non-smooth optimization problem with multiple regularizers. Therefore, we use the proximal gradient method as well as a primal-dual split proximal mechanism to address the problem properly. In addition to increase the reliability of quantitative parametric map estimations, this preprocessing procedure can be included into many existing map estimation algorithms and pipelines effortlessly. We demonstrate our method on the parametric maps estimation for DCE MRI and IVIM MRI.

Keywords

Higher-order total variations; DCE; IVIM; Regularization method; Non-smoothness; Proximal gradient; Primal-dual splitting method

1. INTRODUCTION

Degeneration in the estimation of parametric maps such as those derived from dynamic contrast-enhanced (DCE) MRI can result from intensity deviations caused by motion or

* SYLai@mdanderson.org; phone 713-792-6528.

from images with low signal to noise ratio (SNR), therefore, motion reduction and noise suppression are essential for reliable DCE parametric mapping in clinical studies, especially in certain regions such as head and neck where motion is difficult to control and SNR is intrinsically low. This is especially true when data fitting models, such as those employed in DCE MRI and intravoxel incoherent motion (IVIM) MRI, are implemented in a voxel-wise manner, where voxels in regions affected by motion show sudden intensity deviation in their time-course curves and spatial deformation in positions. Thus, motion occurring in a time frame series not only disrupts DCE/IVIM data fitting, but also makes the determination of a proper arterial input function (AIF) difficult in DCE MRI. MR noise causes spatial and time-curve irregularities, which causes additional uncertainties in many aspects of the estimation procedures.

To solve the problem, a common method is using deformable image registration (DIR) to suppress motion in time frame series. However, aside from the burden in the computation of DIR, the image in motion is usually blurred. Therefore, it's necessary that separate noise suppression has to be considered in addition to motion reduction by DIR.

Previously, researchers tried to embed noise suppression inside each mapping method (model based). However, this was inefficient since researchers have to develop algorithms for each of the models. For example, in DCE analysis, various pharmacokinetic (PK) models [1] (or tracer kinetic models) have been designed to describe the behavior of administered contrast agent (CA), quantify the CA kinetic parameters, and interpret their physiological relevancies in DCE-MRI experiments. Under various physiological conditions, there are compartmental PK models and spatially distributed kinetic models. The general Two-Compartment Exchange Model (2CXM), in which contrast agent is distributed into two separate compartments (vascular and extracellular-extravascular spaces), reduces to the simpler extended Tofts Model (ETM), the three-parameter compartmental tissue uptake model (CTUM) and Tofts Model (TM). Unlike compartmental models, spatially distributed models such the distributed parameter (DP) model and tissue homogeneity (TH) model, account for both spatial and temporal variations of an administered CA and correspond more closely to reality. Moreover, they potentially increase modeling accuracy and provide additional information. Even with the simplified compartmental PK models such as TM, the parameters could be composed of fast and slow components, which is not unusual in certain applications. Based on these facts, applying the motion reduction and noise suppression to each model becomes tedious and impractical, and requires considerable modifications of existing workflows.

To meet the demand of all these scenarios we need an unified approach. To this end, we use the mechanism of total variation to develop a spatial and temporal higher-order total variations (HOTVs [2, 3]) methods for both noise suppression as well as motion reduction in the preprocessing of DCE and IVIM. The previously described limitations on the in-vivo parameter mapping can be addressed by jointly exploiting the intrinsic signal's correlation (cohesion or coherence) in space domain and temporal dimension by means of HOTVs. Mathematically the reconstruction problem is equivalent to convex optimization with non-smooth regularizers [4] of HOTVs. With an iterative approach using convex optimization

with non-smooth regularizations, HOTVs will allow us to approximate the intrinsic signal from deteriorated data collection.

In our implementation, convex optimization with non-smooth regularizers is fulfilled in a flexible scheme using proximal mapping and adjoint operations where multiple HOTVs can be included as constraints, thereby facilitating the regularizations on spatial and temporal domains simultaneously. This is in contrast to existing DCE and IVIM analysis methods that do not provide the ability to apply motion reduction and noise suppression simultaneously. Moreover, our method has minimized modifications of conventional algorithms, physiological models, and processing workflows.

The major contribution of this paper is on the practical implementation of the non-smooth optimization problem with multiple regularizers. This method is universal for the estimation of parametric maps from data series acquisitions (e.g. DCE/ IVIM, etc.). The applications of our method on DCE and IVIM mapping show recognizable improvements when compared to current solutions.

2. METHODS

2.1 Total variations

Regularization is important in reconstructing signal from incomplete and noised data. Conventionally it's represented by minimization of the following energy function [5, 6, 7, 8]

$$u^* = \arg \min_u \{F(u) + \lambda G(u)\}; \quad (1)$$

where F is the fidelity term measuring how reconstruction is fitted to the observation data, and G is a regularization term representing prior knowledge through a convex function. $\lambda (> 0)$ is used to balance F and G . Proper regularization and fidelity terms are dependent on the problem at hand.

Generally in image reconstruction, the fidelity term can be chosen as ℓ^2 distance based on the assumption that the noise will be Gaussian distributed [9]. There are many options for the regularization term, the popular choices being the total variation (TV) method [10, 8, 11] and ℓ^1 regularization based on wavelet or tight-frame technique [12, 13].

However, TV models may suffer from staircase effects (i.e. false edges) in flat regions [14, 15]. The staircase effects are caused by the fact that the TV minimization forces the flat regions to be piecewise constant.

To mollify the staircases of TV, higher-order total variations (HOTVs) have been proposed and studied comprehensively [16, 17, 18, 19], especially the second-order total variation (SOTV).

The SOTV was first introduced in [19] to suppress the staircase effects of TV. Theoretically the SOTV restrain a restored image toward a 'smooth' one from the observed images, and analyses [20, 21] show that SOTV is superior to TV in reconstructions or restoration in smooth regions.

The non-smoothness of HOTVs (and SOTV) make gradient based algorithms such as the Newton method unfeasible (2). To address these difficulties, several algorithms have been proposed recently, including the split Bregman method [22, 23], primal-dual (PD) algorithm [24, 25, 26], alternative direction method of multiplier [27] and fixed-point algorithm [28, 29].

2.2 Higher order methods

For a given datum $u_0 \in L^s(\Omega)$, $\Omega \subset \mathbb{R}^2$, $s = 1, 2$, non-negative regularisation parameters α, β , convex functions $f: \mathbb{R}^2 \rightarrow \mathbb{R}^+$, $g: \mathbb{R}^4 \rightarrow \mathbb{R}^+$ with at most linear growth at infinity, and a suitable linear operator T , the general form of a combined first and second order non-smooth regularization procedure for the reconstruction of u is solving [30]

$$\min_u \{ H(u) = \frac{1}{s} \int_{\Omega} |u_0 - Tu|^s dx + \alpha \int_{\Omega} f(\nabla u) dx + \beta \int_{\Omega} g(\nabla^2 u) dx \} \quad (2);$$

The bounded Hessian $BH(\Omega)$, which consists of all functions $u \in W^{1,1}(\Omega)$ such that ∇u is a function of bounded variation, is the appropriate space for minimization of this function. With a fairly large weight α in the first order term, which aims to preserve sharp details, and adopting a smaller weight β for the second order term, which aims to mollify the staircase effect induced by the first order regulariser without introducing any noticeable smear, the image can be better reconstructed from observed data.

Besides HOTVs, there are efforts towards the reduction of staircase effects in applying TV. The recently proposed total generalized variation (TGV) model [31] produced the currently best higher-order reconstruction results in the field. However the numerical solution for TGV regularization is in general about ten times slower than available methods for HOTVs [30]. On the other hand, the classical setting of the non-smoothness problem in the space $W^{2,1}(\Omega)$ is by means of relaxation to “round off” the total variation term. The generality of this approach includes both the classical variational formulation in $W^{2,1}$ (such as the Huber regularized version of the total variation) as well as the non-smooth norm minimization setting in $BH(\Omega)$, which constitutes the relaxed version of (2), and is minimized by means of an operator splitting technique for the cases $f(x) = |x|$ and $g(x) = |x|$ as elucidated in [30]. However, the Huber regularized version can only reduce these staircase artifacts to a certain extent (e.g. eliminate the staircasing effect only in areas with small gradient) [30]. The modification in [30] by employing the split Bregman method is feasible, but we are aiming to apply multiple higher order regularizers on spatial and temporal dimensions, which requires more efficient approaches as we will state below. From a different view point, Chambolle and Lions [32] suggested a higher order solution by means of an infimal-convolution of two convex regularizers, and a modified infimal-convolution approach has been recently proposed [33, 34]. Other higher-order methods including higher-order diffusion filters and Euler’s elastica are discussed in [30], however these are beyond the scope of our concerns. General discussion on modern regularization methods for inverse problems can be referred to [34], and demystified and comprehensive numerical implementations can be referred to [35].

2.3 Primal-Dual Method

The PD algorithm was first introduced in [37], and investigated thoroughly in [38, 39]. The PD algorithm is a common approach for various models in image processing. The mechanism of the PD algorithm [25] to solve the model of (3) is described as follow. For a primal problem

$$\min_x \{F(Kx) + G(x)\}; \quad (3)$$

where $x \in \mathbb{R}^N$, K is a linear operator, G and F are convex functions, the PD algorithm can be expressed as [25, 7] :

Algorithm 1 (PD algorithm for problem (3)).

1. Initialization: Choose $\tau, \sigma > 0$ such that $\tau\sigma \|K\|_2^2 \leq 1$, $\theta \in [0, 1)$, $x_0 = 0$, $y_0 = 0$, $\bar{x}_0 = x_0$, $n = 0$;
2. Iterations ($n = 1$): Update $x_n, y_n, \bar{x}_n$ as follows

$$\begin{cases} y_{n+1} = \text{prox}_{\sigma} [F^*](y_n + \sigma K \bar{x}_n) \\ x_{n+1} = \text{prox}_{\tau} [G](x_n - \tau K^T y_{n+1}) \\ \bar{x}_{n+1} = x_{n+1} + \theta(x_{n+1} - x_n) \end{cases}$$

In PD algorithm, to solve the proposed higher order model, we should obtain the conjugate function of SOTV (or HOTV) and the corresponding proximal mapping (see details in [35]).

2.4 First-Order Primal-Dual Splitting Method for HOTVs

The optimization of the sum of two convex functions takes the form of

$$\min_{x \in \mathbb{R}^n} f(x) + h(Bx), \quad (4)$$

The functions f and h in (4) may represent the data fidelity term and the regularization term. Denoting $\Gamma_0(\mathbb{R}^n)$ the set of proper, lower semi-continuous convex functions from $\mathbb{R}^n$ to $(-\infty, +\infty]$, here $f \in \Gamma_0(\mathbb{R}^n)$, $h \in \Gamma_0(\mathbb{R}^m)$, and $B: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear transformation matrix.

Under the assumption that the functions f and h have a closed-form solution of proximity operator, a general primal-dual method is introduced in Chambolle and Pock [25] to solve problem (4), with proof for the convergence of the algorithms. Subsequently the primal-dual method of Chambolle and Pock [25] is shown to be equivalent to the well-known proximity point algorithm (PPA) in He and Yuan [38], which provides the convenience in convergence analysis. Pock and Chambolle proposed a preconditioning technique in [39] to accelerate the primal-dual method, with the convergence of preconditioning technique followed straightly after the PPA method. The primal-dual method of Chambolle and Pock [25] is investigated in the applications in medical image reconstruction [7, 40, 41].

Chambolle and Pock [9] did not use the gradient operator of the function in (4), when the objective function in (4) is differentiable with Lipschitz continuous gradient, Condat [42], Vu [43], and Combettes and Pesquet [44] introduced a primal-dual split method. For the problem of sum of three convex functions, under the condition of a smooth function of Lipschitz continuous gradient, a non-smooth proximable function, and linear composite functions, the objective function is presented as:

$$\min_{x \in R^n} f(x) + g(x) + h(Bx), \quad (5)$$

where $f \in \Gamma_0(R^n)$, $g \in \Gamma_0(R^n)$, $h \in \Gamma_0(R^m)$, $f: R^n \rightarrow R$ is differentiable, and its gradient ∇f is Lipschitz continuous with Lipschitz constant $\beta > 0$; $B: R^n \rightarrow R^m$ is a linear transformation matrix. To solve problem (5), Condat [42] proposed a primal-dual splitting method, and proved the existence of convergence in an infinite-dimensional Hilbert space using the fixed point theory of non-expansive mappings.

The basic principle of primal-dual splitting method was introduced by Condat [42], Vu [43] simultaneously, and subsequently by Combettes and Pesquet [44, 45], and preconditioning technique is also used for primal-dual splitting method [46]. Denote H be a real Hilbert space with its inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|^{1/2}$. $\Gamma_0(H)$ is defined as above, when $f \in \Gamma_0(H)$, its Fenchel conjugate $f^* \in \Gamma_0(H)$ is defined by $f^*(u) = \sup_x \langle x, u \rangle - f(x)$ and its proximity operator is defined by $\text{prox}_{\lambda f}(u) = \arg \min_x \{ (1/2)\|x - u\|^2 + \lambda f(x) \}$; $\lambda > 0$ is a positive constant. The subdifferential of f is defined as the set-valued operator $\partial f(u) = \{ v \in H: f(u') - f(u) + \langle v, u' - u \rangle, \forall u' \in H \}$. If f is differential at μ , then $\partial f(u) = \{ \nabla f(u) \}$. H_1 and H_2 are two real Hilbert spaces with $B: H_1 \rightarrow H_2$ being a bounded linear operator of adjoint B^* and induced norm

$$\| B \| = \sup_{\| x \| \leq 1} \{ \| Bx \|, x \in H_1 \} < +\infty. \quad (6)$$

Assuming $f: H_1 \rightarrow R$ is convex, differentiable on H_1 , with β -Lipschitz continuous gradient ∇f for some $\beta > 0$; satisfies $\| \nabla f(x) - \nabla f(y) \| \leq \beta \| x - y \|$, $\forall x, y \in H_1$. When proximity operator of $g \in \Gamma_0(H_1)$ and $h \in \Gamma_0(H_2)$ have a closed-form solution, and $B: H_1 \rightarrow H_2$ is a bounded linear operator with adjoint B^* , Condat [1] considered the solution of following optimization problem:

$$\text{Find } \hat{x} \in \arg \min_x f(x) + g(x) + h(Bx), \quad (7)$$

The dual formulation of the primal problem (7) is

$$\text{Find } \hat{y} \in \arg \min_y (f + g)^*(-B^*y) + h^*(y). \quad (8)$$

The corresponding saddle point of primal problem (7) and dual problem (9) is constructed as:

$$\text{Find } (\hat{x}, \hat{y}) \in \arg \min_{x,y} \max \{ f(x) + g(x) + \langle Bx, y \rangle - h^*(y) \}. \quad (9)$$

Monotone variational inclusion can be used to find the pair $(\hat{x}, \hat{y})$ by way of the following:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial g(\hat{x}) + B^* \hat{y} + \nabla f(\hat{x}) \\ -B \hat{x} + \partial h^*(\hat{y}) \end{pmatrix}, \quad (10)$$

where g and h^* represent the subdifferential corresponding g and h^* . The iterative algorithm to solve problem (7) is proposed by Condat [1] as the following.

Algorithm 2 (primal-dual splitting method for problem (7)). Choose the proximal parameters $\tau > 0$, $\sigma > 0$ and the relaxation parameters $\{\rho_n\}$. Give an initial value $(x_0, y_0) \in H_1 \times H_2$. For $n = 0, 1, 2, \dots$, the iterative $\{x_n\}$ and $\{y_n\}$ are updated by

1. $\tilde{x}_{n+1} = \text{prox}_{\tau g}(x_n - \tau(\nabla f(x_n) + e_{f,n}) - \tau B^* y_n) + e_{g,n}$;
2. $\tilde{y}_{n+1} = \text{prox}_{\sigma h^*}(y_n + \sigma B(2\tilde{x}_{n+1} - x_n)) + e_{h,n}$;
3. $(x_{n+1}, y_{n+1}) = \rho_n(\tilde{x}_{n+1}, \tilde{y}_{n+1}) + (1 - \rho_n)(x_n, y_n)$, where $e_{f,n} \in H_1$, $e_{g,n} \in H_1$, and $e_{h,n} \in H_2$ model errors from the approximate computation of the operators ∇f , $\text{prox}_{\tau g}$, and $\text{prox}_{\sigma h^*}$ respectively. The algorithm stops when stopping criteria are encountered.

The following theorem ensured the convergence of Algorithm 2.

Theorem 1 (see [1]). Assume $\tau > 0$, $\sigma > 0$, and $\{\rho_n\}$, $\{e_{f,n}\}$, $\{e_{g,n}\}$, and $\{e_{h,n}\}$ be the parameters sequences of Algorithm 2. Let β be the Lipschitz constant of ∇f . Suppose that $\beta > 0$ and the following conditions encountered:

- i. $1/\tau - \sigma \|B\|^2 > \beta/2$;
- ii. $\{\rho_n\} \subset (0, \delta)$, where $\delta = 2 - (\beta/2)(1/\tau - \sigma \|B\|^2)^{-1}$;
- iii. $\sum(\delta - \rho_n) = +\infty$;
- iv. $\sum \rho_n \|e_{f,n}\| < +\infty$, $\sum \rho_n \|e_{g,n}\| < +\infty$, and $\sum \rho_n \|e_{h,n}\| < +\infty$.

Then the iterated solution of $\{x_n\}$ and $\{y_n\}$ will converge weakly to $\hat{x}$ of (7) and $\hat{y}$ of (8), respectively.

By using Moreau's identity $\text{prox}_{\sigma h^*}(\mu) = \mu - \sigma \text{prox}_{(1/\sigma)h}((1/\sigma)\mu)$, the proximal operator $\text{prox}_{\sigma h^*}$ of adjoint can be retrieved from proximal operator $\text{prox}_{(1/\sigma)h}$. It's obvious that the role of primal variable x and dual variable y in Algorithm 2 can be swapped, provides a new iterative algorithm with the convergence ensured by Theorem 1. In practice, the performance of the two iterative algorithms is nearly the same, providing additional flexibility in implementation.

2.5 Accelerated gradient methods

Suppose that we want to solve the following convex optimization problem:

$$\min_{x \in \mathbb{R}^n} g(x) + h(x)$$

where $g(x)$ is assumed to be convex and differentiable, and $h(x)$ is convex. Then the generalized gradient descent can be formulated as follows:

Step 0: choose initial $x_0 \in \mathbb{R}^n$

Step k : $k \geq 1$: $\mathbf{x}^{(k)} = \text{prox}_h \left(x^{(k-1)} - t_k \nabla g(x^{(k-1)}), t_k \right)$

where a proximal operator is defined as $\text{prox}_h(y, t) = \arg \min_{x \in \mathbb{R}^n} h(x) + \frac{1}{2t} \|y - x\|^2$

It is known that if $\nabla g(x)$ is Lipschitz continuous and proximal operator can be evaluated, the convergence rate of will be $O(1/k)$. This result can be accelerated to achieve $O(1/k^2)$. The original acceleration idea for smooth functions is proposed by Nesterov in [47], another acceleration idea for smooth functions is introduced in [48], in [49], smoothing techniques for non-smooth functions, coupled with original acceleration idea is advised, and in [50], acceleration idea for composite functions is included. Beck and Teboulle [51, 52] made extension of Nesterov [48] to composite functions, arrived at the famous FISTA (Fast iterative-thresholding algorithms). Tseng [53] unified analysis of acceleration techniques. The idea of acceleration still remains an active topic of research (for non-smooth, composite functions, etc.).

Roughly speaking, acceleration is achieved by introducing one more sequence of numbers y_k constructed as a specific linear combination of x_k and x_{k-1} ; proximal function operates then on y_k instead of x_k . In case of FISTA we have:

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$y_{k+1} = x_k + \frac{t_k - 1}{t_{k+1}}(x_k - x_{k-1})$$

Note, that the sequence t_k satisfies $t_k^2 = t_{k+1}^2 - t_{k+1}$; this is justified in the proof of the convergence for FISTA.

2.6 HOTVs for spatial and temporal constraints

With the combination of first-order and second-order total-variations (TVs) as regularizers in reconstruction, spatial and temporal constraints are applied on (dynamic) MRI signal to minimize the variabilities caused by motion or low SNR while preserving the details of true signal or image, and the reconstructed signal is used in the estimation of MR parametric maps. To overcome staircase effects of first-order total variations (TV) [54], second-order TVs are included. While this constrained convex optimization problem with multiple non-smooth regularizers is hard in numerical completion, we initially transform the constrained optimization into an unconstrained optimization problem by including an indication function

on the defined domain of solution and then using the proximal gradient method on the unconstrained optimization problem. Since the proximal mapping of the indicator function is a simple projection operator, a flexible fulfillment of convex optimization with non-smooth regularizers can be easily developed to fit various reconstruction problems based on the primal-dual proximal mechanism. Besides flexibility, the most important advantage of our algorithm is that it can be easily repeated and extended to cases need to apply multiple TVs.

The constrained optimization problem is $\min_x f(x)$, where $f(x)$ is the cost function to be minimized, x belongs to the domain of definition (e.g. the value range of MRI data, $0 \leq x \leq \alpha$), the constrained optimization can be transformed into an unconstrained optimization problem $\min_x f(x) + I(x)$, where $I(x)$ is an indication function on the defined domain of solution. Using the proximal gradient method on the unconstrained optimization problem, then we have an iterative solution to the non-constrained optimization problem as

$$x(k) = \text{PROX}_{t(k)I} (x(k) - t(k) \text{grad} (f(x(k)))) ,$$

where $\text{PROX}_{t(k)I}$ is the proximal mapping of the indicator function $I(x)$, $t(k)$ is the step length in iterations k . Since the proximal mapping of the indicator function is a simple projection operator, and the adjoint operator of TVs are available, a flexible fulfillment of the convex optimization with non-smooth regularizers can be easily developed to fit various reconstruction problems based on the primal-dual splitting proximal mechanism [42]. In sum, this method will find

$$x = \arg \min_x f(x) + g(x) + \sum_{m=1}^M h_m(L_m x),$$

where the first term ($f(x)$) can be data consistence constraints that measure the distance between the original signal and the reconstructed signal, in general, $f(x) = \|O - x\|$, Where O is the original signal, x is reconstruction, and $\|\cdot\|$ is distance measure of the vectors (such as L_2 -norm). The second term ($g(x)$) will be an indicator function which changes original domain-constrained problem into a non-constrained problem, and the multiple regularizers ($h_m(\cdot), m = 1, \dots, M$ with various order of total variations (TVs, $L_m x$) on spatial and temporal contents [2, 3]) can be flexibly included in the third term. The solution is an iterative algorithm which aims at minimizing a sum of functions by successive evaluations of their gradients or proximity operators using a proximal algorithm, in which only first-order information of the functions is exploited. Fenchel–Rockafellar conjugate is involved for the proximal operator, which satisfies the useful Moreau identity.

Our algorithm is presented as below, where FISTA [51, 52] (in the updating of t with t_{new}) is introduced to accelerate the convergence of iteration: Choose the parameters $\tau > 0, \sigma > 0, \rho > 0$, initial $t=1$ and the initial estimates $x^{(0)}, u_1^{(0)}, \dots, u_M^{(0)}$, then iterate, for $i = 0, 1, \dots$

$$\begin{cases}
 \text{For } m = 1, \dots, M \\
 \tilde{u}_m^{(i+1)} = \text{prox}_{\sigma h_m^*}(u_m^{(i)} + \sigma L_m x^{(i)}) \\
 u_m^{(i+1)} = \rho \tilde{u}_m^{(i+1)} + (1 - \rho) u_m^{(i)} \\
 \tilde{x}^{(i+1)} = \text{prox}_{\tau g}\left(x^{(i)} - \tau \Delta f(x^{(i)}) - \tau \sum_{m=1}^M L_m^*(2\tilde{u}_m^{(i+1)} - u_m^{(i)})\right) \\
 x^{(i+1)} = \rho \tilde{x}^{(i+1)} + (1 - \rho) x^{(i)} \\
 t_{new} = (1 + \sqrt{1 + 4 * t^2})/2; \quad \theta = (t - 1)/(t_{new}); \quad t = t_{new}; \\
 x^{(i+1)} = x^{(i+1)} + \theta * (x^{(i+1)} - x^{(i)}); \quad u_m^{(i+1)} = u_m^{(i+1)} + \theta * (u_m^{(i+1)} - u_m^{(i)});
 \end{cases}$$

Where L_m^* is the adjoint operator of L_m , and $\text{prox}_{\sigma h_m^*}(u) = u - \sigma \text{prox}_{h_m/\sigma}(u/\sigma)$. The numerical implementations of proximal operator of HOTVs can be referred to [35].

With proper adjustment of weighting factors of regularizers, the optimization would converge with around 20 iterations in practice for DCE and IVIM data.

3. RESULTS

The DCE/IVIM maps are retrieved on the reconstructed signals with HOTVs. We implement a linear DCE solver to significantly boost the speed of parameter estimation while preserving the consistency and repeatability against initial estimations [55, 56, 57]. A Segmented-Unconstrained algorithm, which is likely the most frequently used algorithm in IVIM analysis [58, 59], is adopted in our demonstration section below.

3.1 DCE-MRI Imaging

DCE-MRI scans were obtained using GE 3.0T Discovery MR750 scanners with a 6-channel Flex phased-array coil (GE Healthcare Technology, Milwaukee, WI). Prior to DCE-MRI, T1 mapping was performed using six variable flip angles (FA) 3D spoiled gradient recalled echo (SPGR) sequence (FA = 2°, 5°, 10°, 15°, 20°, and 25°; TR/TE = 5.5/2.1 ms, FOV = 25.6 cm, slice number = 30, voxel size = $2 \times 2 \times 4 \text{ mm}^3$). The DCE-MRI was acquired using a multi-phase 3D Fast SPGR sequence to gain sufficient signal-to-noise ratio (SNR), contrast, and temporal resolution (FA = 15°, TR/TE = 3.6/1 ms, voxel size = $2 \times 2 \times 4 \text{ mm}^3$, temporal resolution = 5.5s, number of repetitions = 56, pixel bandwidth = 326 Hz, ASSET acceleration = 2). Gadobutrol (Gadovist, Bayer Healthcare, Germany) was administered at a dose of 0.1 mmol/kg of body weight at 3 ml/s followed by the same amount of saline at 3 ml/s, via a power injector (Spectris MR Injector, MedRad, Pittsburgh Pa).

3.2 Computation of the kinetic model

Post-processing of the DCE-MRI images was performed at a workstation running an in-house Matlab based pipeline (Matlab, MathWorks, MA, USA). Before quantitative analysis, motion correction and noise suppression were applied using HOTVs as described above. To quantify the physiological parameters using DCE-MRI, the arterial input function (AIF) of the contrast agent (CA) entering the tissue was pre-determined. T1 map was calculated to convert the signal intensity into concentration time course. Extended Tofts model [1] assumes that the CA resides in and exchanges between two compartments in the tissue: the

vascular space and extracellular extravascular space (EES). When this model is used, the differential equation describing the kinetic behavior of the CA in the tissue of interest is given by:

$$\frac{dC_{TOI}(t)}{dt} = (K^{trans} + v_p \cdot k_{ep}) \cdot C_p(t) - k_{ep} \cdot C_{TOI}(t) + v_p \cdot \frac{dC_p(t)}{dt}$$

where $C_{TOI}(t)$ and $C_p(t)$ are the concentrations of the CA at time t in the tissue of interest (TOI) and blood plasma, respectively, K^{trans} are the transfer (CA permeability rate) constants between blood plasma and the EES of the TOI. $k_{ep} = K^{trans}/V_e$ is the transfer (CA permeability rate) constant (min^{-1}) from the TOI back to the blood plasma, where V_e is the distribution volume fractions of CA in the EES per unit volume of tissue. When the kinetic model includes a vascular term, V_p that is the capillary plasma volume fractions per unit volume of tissue. Otherwise, by ignoring vascular term, the extended Tofts model is reduced to Tofts model [1] as:

$$\frac{dC_{TOI}}{dt} = K^{trans} \cdot C_p(t) - k_{ep} \cdot C_{TOI}$$

The AIF for the pharmacokinetic (PK) analysis was derived from selected arteries. The relevance of several analytical AIF models in DCE-MRI have been previously extensively investigated [60]. Recently computer simulations were performed to evaluate and compare a population of AIF models with the Parker model [61]. The results demonstrated that a six-parameter linear function plus bi-exponential function AIF model was almost equivalent to the Parker AIF. It should be noted the former is computationally faster and more reliable in functional fitting when compared to the Parker AIF. However, predetermining the arrival time (AT) and time to peak (TTP) of upslope for each AIF time series is usually not accurate in using the above six-parameter model. Therefore, we extended the six-parameter model to a bi-exponential and bi-linear function where the AT and TTP of the upslope are included as parameters to be estimated in AIF fitting.

The new AIF model function (AIFM) was defined as:

$$\text{AIFM} = \min \left(\overbrace{\max \left(p1, p1 + (p2 - p1) \cdot \frac{t - \min(p4, p5)}{\text{abs}(p5 - p4)} + (p3 - p1) \cdot \frac{t - \min(p4, p5)}{\text{abs}(p5 - p4)} \right)}^{\text{part 1}}, \right. \\ \left. \overbrace{\min(\text{uplimit}, p2 \cdot \exp(-p6 \cdot (t - \max(p5, p4))) + p3 \cdot \exp(-p7 \cdot (t - \max(p5, p4))))}^{\text{part 2}} \right)$$

where $p1$ to $p7$ are parameters to be determined by functional fitting²³, $\min(p4, p5)$ is AT, and $\max(p4, p5)$ is TTP; t is time points, $\min$ is minimal value operator, $\max$ is the maximal value operator, and abs is absolute value operator. “*uplimit*” is a constant that is estimated by the maximal possible value of data to be fitted. The fitting cost function is implemented by optimization on $\| \text{DATA} - \text{AIFM} \|$, where DATA could be DCE concentration or DCE signal.

By using this extended AIF model function, the AIF fitting can be completed in a more precise and reliable manner. Figure 1 shows the fitting process as well as the fitting results where a 56 time point AIF time series is presented. The PK modelling was done on a pixel-by-pixel basis using a linearization equation of the models used (i.e. Tofts and extended Tofts) as described by Kenya Murase [55]. Subsequently, we implemented the linear least-squares method to acquire the PK parameters.[56, 57]. This method is preferred to the conventional nonlinear least-squares method [1], because it is faster and it does not require initial estimation, and has no local optima problems.

3.3 Analysis of DCE-MRI

Figure 2 shows the time courses of DCE MRI intensity before and after noise suppression with HOTVs, (a) and (b) are voxels site in Arterial Input Function (AIF) region, (c) and (d) are voxels site in region of tissue of interest, (a) and (c) are before filtering, (b) and (d) are after filtering. It is shown that the details of AIF and tissue time course are preserved properly.

In Figure 3, we demonstrate consecutive DCE MRI frame time series (from left to right and from up to down) before and after filtering with HOTVs, where both noise suppression, as well as motion reduction, are compared side-by-side visually, with the original image in lead, followed by the reconstructed image. In this case, the motion is moderate and noticeable, which becomes a challenge to conventional DCE mapping algorithms. With the improved reconstruction of MRI frame time series, it's possible for conventional DCE mapping algorithms to be applied effectively.

In Figure 4, we demonstrated another consecutive DCE MRI frame time series (from left to right and from up to down) before and after filtering with HOTVs, where both noise suppression, as well as motion reduction, are compared side-by-side visually. In this case, the motion is severe and data recovered by HOTVs is compromised. However, with the original data, it's impossible to generate acceptable parametric maps by conventional DCE mapping algorithms. While with the improved reconstruction of MRI frame time series and using standardized AIF [57], it's still possible for conventional DCE mapping algorithms to be able to generate parametric maps with reasonable quality.

Figure 5 (a), (b), (c) and (d) show the motion reduction by the HOTVs filtering on data acquired with optimized MR sequence for head and neck, each row represents one of four consecutive time frames, the first column is imaged before filtering, the second column is imaged after filtering, and the third column is the difference between the first and second column showing the motion reduction caused by the filtering on optimized data acquisition. Figure 6 is the DCE maps corresponding to Figure 5, they are generated with the Tofts model, (a) and (b) are K_{trans} , and (c) and (d) are V_e . (a) and (c) are before filtering, (b) and (d) are after filtering. It's noticeable that after HOTVs filtering, the DCE maps reveal more details than before. From our current studies on head and neck DCE MRI, the spatial and temporal higher-order total variations can help in suppressing noise while reducing the motion influence on the estimation of DCE parameters. The HOTVs filtering is fulfilled on image slices, with parallel computations implemented on multi-cores.

We also investigated HOTVs to exploit spatial-temporal coherence in IVIM. Reconstruction of IVIM signal by HOTVs considerably increased SNR while retaining the details of images; preliminary acquisitions using this approach are shown *in vivo* in a pelvis case, with improved D^* and f (Figure 7). We plan to perform a similar investigation for head and neck.

4. CONCLUSION

The initial applications on DCE MRI and IVIM MRI mapping with HOTVs based reconstruction show promising improvements on both normal and degenerated data acquisition. However, comparisons are only carried out by visual inspection, deliberated quantitative comparisons are needed to be completed in future studies to further validate the qualification of this new approach.

REFERENCES

- [1]. Sourbron SP, Buckley DL, Classic models for dynamic contrast-enhanced MRI. *NMR Biomed.* 8; 26(8):1004–27, 2013. [PubMed: 23674304]
- [2]. Chan TF, Shen J, Image Processing and Analysis - Variational, PDE, Wavelet, and Stochastic Methods, SIAM, ISBN 0–89871-589-X, 2005.
- [3]. Caselles V, Chambolle A, and Novaga M., The discontinuity set of solutions of the TV denoising problem and some extensions, *SIAM, Multiscale Modeling and Simulation*, vol. 6 n. 3, 2007.
- [4]. Yang T, Mahdavi M, Jin R, and Zhu S. An efficient primal dual prox method for non-smooth optimization. *Machine Learning*, Volume 98, Issue 3, pp 369–406, 3 2015
- [5]. Herman GT and Davidi R, Image reconstruction from a small number of projections, *Inverse Problems*, 24, 045011, 2008.
- [6]. Pan X, Sidky EY and Vannier M, Why do commercial CT scanners still employ traditional, filtered backprojection for image reconstruction? *Inverse Problems*, 25, 2009.
- [7]. Sidky EY, Jorgensen JH and Pan X, Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm, *Phys. Med. Biol.* 57, 3065–3091, 2012. [PubMed: 22538474]
- [8]. Sidky EY, Kao CM and Pan X, Accurate image reconstruction from few views and limited-angle data in divergent-beam CT, *J. X-ray Sci. Technol.* 14, 119–139, 2006.
- [9]. Herman GT, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Springer, 2nd edition, 2009.
- [10]. Rudin LI, Osher S and Fatemi E, Nonlinear total variation based noise removal algorithms, *Physica D*, 60, 259–268, 1992.
- [11]. Strong D and Chan T, Edge-preserving and scale-dependent properties of total variation regularization, *Inverse Problems*, 19, 165–187, 2003.
- [12]. Zhao B, Gao H, Ding H and Molloy S, Tight-frame based iterative image reconstruction for spectral breast CT, *Med. Phys.* 40, 031905, 2013. [PubMed: 23464320]
- [13]. Zhou W, Cai J and Gao H, Adaptive tight frame based medical image reconstruction: a proof-of-concept study for computed tomography, *Inverse Problems*, 29, 125006, 2013.
- [14]. Buades A, Coll B and Morel JM, The staircasing effect in neighborhood filters and its solution, *IEEE Trans. Image Process.* 15, 1499–1505, 2006. [PubMed: 16764274]
- [15]. Chambolle A and Lions PL, Image recovery via total variation minimization and related problems, *Numer. Math.* 76, 167–188, 1997.
- [16]. Bredies K, Kunisch K and Pock T, Total generalized variation, *SIAM Journal on Image Sciences*, 3, 492–526, 2010.
- [17]. Chan T, Esedoglu S and Park FE, A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In *Proceedings of IEEE International Conference on Image Processing*, p. 4137–4140, Hong Kong, 2005.

- [18]. Chen H, Song J and Tai X, A dual algorithm for minimization of the LLT model, *Adv. Comput. Math*, 31, 115–130, 2009.
- [19]. Lysaker M, Lundervold A and Tai X, Noise removal using fourthorder partial differential equation with applications to medical magnetic resonance images in space and time, *IEEE Trans. Image Process*, 12, 1579–1590, 2003. [PubMed: 18244712]
- [20]. Greer JB and Bertozzi AL, H1 solutions of a class of fourth order nonlinear equations for image processing, *Discrete Contin. Dyn. Syst*, 1–2, 349–366, 2004.
- [21]. Weickert J, *Anisotropic Diffusion in Image Processing*, B.G. Teubner, Stuttgart, 1998.
- [22]. Chen Y, Hager W, Yashtini M, Ye X and Zhang H, Bregman operator splitting with variable stepsize for total variation image reconstruction, *Comput. Optim. Appl*, 54, 317–342, 2013.
- [23]. Goldstein T and Osher S, The split Bregman algorithm for L1 regularized problems, *SIAM J. Imaging Sci*, 2, 323–343, 2009.
- [24]. Chambolle A, An algorithm for total variation minimization and applications, *J. Math. Imaging Vision*, 20 (2004) 89–97.
- [25]. Chambolle A and Pock T, A first-order primal-dual algorithm for convex problem with applications to imaging, *J. Math. Imaging Vision*, 40, 120–145, 2011.
- [26]. Zhu M and Chan T, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, *UCLA CAM Report*, 34, 2008.
- [27]. Ouyang Y, Chen Y and Lan G, An accelerated linearized alternating direction method of multipliers, *SIAM J. Imaging Sci*, 8, 644–681, 2015
- [28]. Li Q, Micchelli CA, Shen L and Xu Y, A proximity algorithm accelerated by Gauss-Seidel iterations for L1/TV denoising models, *Inverse Problems*, 28, 095003, 2012.
- [29]. Micchelli CA, Shen L and Xu Y, Proximity algorithms for image models: denoising, *Inverse Problems*, 27, 045009, 2011.
- [30]. Papafitsoros K, Schönlieb CB, A Combined First and Second Order Variational Approach for Image Reconstruction, *Journal of Mathematical Imaging and Vision*, Volume 48, Issue 2, pp 308–338, 2 2014
- [31]. Chambolle A and Lions P Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.
- [32]. Setzer S and Steidl G Variational methods with higher order derivatives in image processing. *Approximation XII*, pages 360–386, 2008.
- [33]. Setzer S, Steidl G, and Teuber T Infimal convolution regularizations with discrete l1-type functionals. *Comm. Math. Sci*, 9:797–872, 2011.
- [34]. Benning M, & Burger M, Modern regularization methods for inverse problems. *Acta Numerica*, 27, 1–111, 2018.
- [35]. Lu W, Duan J, Qiu Z, Pan Z, Wen R, Bai L, Implementation of high-order variational models made easy for image processing, *Mathematical Methods in the Applied Sciences*, Volume39, Issue14, Pages 4208–4233, 30 9 2016
- [36]. Arrow K, Hurwicz L and Uzawa H, *Studies in Linear and Non-linear Programming* Stanford Mathematical Studies in the Social Sciences. Stanford University Press, California, 1958.
- [37]. Chen Y, Lan G and Ouyang Y, Optimal primal-dual methods for a class of saddle point problems, *SIAM J. Optim*, 24, 1779–1814, 2013.
- [38]. He B and Yuan X, Convergence analysis of primal-dual algorithms for a saddle point problem: From contraction perspective, *SIAM J. Imaging Sci*, 5, 119–149, 2012.
- [39]. Pock T and Chambolle A, “Diagonal preconditioning for first order primal-dual algorithms in convex optimization,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV ‘11)*, pp. 1762–1769, Barcelona, Spain, November 2011.
- [40]. Tang Y, Peng J, Yue S, and Xu J, “A primal dual proximal point method of chambolle-pock algorithms for ℓ_1 -TV minimization problems in image reconstruction,” in *Proceedings of the 5th International Conference on Biomedical Engineering and Informatics (BMEI ‘12)*, pp. 12–16, IEEE, Chongqing, China, October 2012.
- [41]. Tang Y, Cai Y, Wang X, and Peng J, “A primal dual proximal point method of Chambolle-Pock algorithm for total variation image reconstruction,” in *Proceedings of the IEEE International*

- Conference on Medical Imaging Physics and Engineering (ICMIPE '13), pp. 6–10, IEEE, October 2013.
- [42]. Condat L, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.
 - [43]. Vũ BC, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.
 - [44]. Combettes PL and Pesquet J-C, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued and Variational Analysis*, vol. 20, no. 2, pp. 307–330, 2012.
 - [45]. Combettes PL, Condat L, Pesquet J-C, and Vu BC, A Forward-Backward View of Some Primal-Dual Optimization Methods in Image Recovery, 2014 IEEE International Conference on Image Processing, ICIP 2014
 - [46]. Wen M, Peng J, Tang Y, Zhu C, and Yue S, A Preconditioning Technique for First-Order Primal-Dual Splitting Method in Convex Optimization, *Mathematical Problems in Engineering*, Volume 2017, Article ID 3694525, 11 pages
 - [47]. Yu Nesterov., A method for solving a convex programming problem with convergence rate $O(1/k^2)$, *Soviet Mathematics Doklady*, vol. 27, no. 2, 1983
 - [48]. Yu Nesterov., On an approach to the construction of optimal methods of minimization of smooth convex function, *Ekonom. i. Mat. Metody* (In Russian), 24, 509–517 (1988).
 - [49]. Yu Nesterov., Smooth minimization of non-smooth functions. *Mathematical programming, Serie A*, 103, 127–152, 2005.
 - [50]. Yu Nesterov., Gradient methods for minimizing composite objective function. *CORE Discussion Paper*, 76, 2007
 - [51]. Beck A and Teboulle M, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. on Imaging Sciences*, 2(1), 183–202, 2009
 - [52]. Beck A and Teboulle M, Gradient-based algorithms with applications to signal recovery, in: Eldar Y and Palomar D (Eds.), *Convex Optimization in Signal Processing and Communications*, 2009
 - [53]. Tseng P, On accelerated proximal gradient methods for convex-concave optimization, (unpublished) 2008
 - [54]. Jalalzai K Some Remarks on the Staircasing Phenomenon in Total Variation-Based Image Denoising. *Journal of Mathematical Imaging and Vision*, Volume 54, Issue 2, pp 256–268, 2016
 - [55]. Murase K Efficient method for calculating kinetic parameters using T1-weighted dynamic contrast-enhanced magnetic resonance imaging. *Magn Reson Med*; 51:858–62, 2004. [PubMed: 15065262]
 - [56]. Ke C-C et al. COMPARISON OF THE KINETIC PARAMETERS ESTIMATED WITH DIFFERENT NUMERICAL METHODS IN DCE-MRI, *ISMRM* 2011
 - [57]. Jones KM, Pagel MD, Cardenas-Rodriguez J, Linearization improves the repeatability of quantitative dynamic contrast-enhanced MRI. *Magnetic resonance imaging*; 47:16–24, 2018 [PubMed: 29155024]
 - [58]. Gurney-Champion OJ et al. Comparison of six fit algorithms for the intra-voxel incoherent motion model of diffusion-weighted magnetic resonance imaging data of pancreatic cancer patients, *PLoS One.*; 13(4): e0194590, 2018 [PubMed: 29617445]
 - [59]. Barbieri S et al., Impact of the calculation algorithm on biexponential fitting of diffusion-weighted MRI in upper abdominal organs, *MRM*, Volume 75, Issue 5, Pages 2175–2184, 5 2016 [PubMed: 26059232]
 - [60]. Balvay D, Ponvianne Y, Claudon M, Cuenod CA, Arterial input function: Relevance of eleven analytical models in DCE-MRI studies. Paper presented at: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro; 14–17 May 2008, 2008.
 - [61]. He D, Xu L, Qian W, Clarke J, Fan X A simulation study comparing nine mathematical models of arterial input function for dynamic contrast enhanced MRI to the Parker model. *Australas Phys Eng Sci Med*;41(2):507–518, 2018. [PubMed: 29569210]

Figure 1.
The fitting process and fitting results where a 56 time point AIF time series (in Cyan) is presented.

Figure 2.
The time courses of DCE before (a and c) and after (b and d) noise suppression

Figure 3. Consecutive DCE MRI frame time series (from left to right and from up to down) before (column 1 and 3) and after (column 2 and 4) filtering with HOTVs, the motion is moderate and noticeable, images are compared side-by-side.

Figure 4. Consecutive DCE MRI frame time series (from left to right and from up to down) before (odd) and after (even) filtering with HOTVs, the motion is severe and hard to be recovered, the same time frames are compared one after another.

Figure 5.
The motion reduction by the HOTVs filtering on optimized data acquisition

Figure 6.
The DCE maps generated with the Tofts model before and after HOTV's filtering

Figure 7.

Higher-order total variation (HOTV) application in a pelvic case example, showing uncorrected (left panels) and HOTV-correction (right panels). From top left: IVIM with low SNR at $b=800$ vs HOTVs filtering, IVIM with SNR at $b=400$ vs HOTVs filtering; estimation of S_0 (signal) maps with and without HOTV correction. From top right: D, D^* , and f maps, with and without HOTV correction.