

Published in final edited form as:

*Comput Med Imaging Graph.* 2009 June ; 33(4): 295–303. doi:10.1016/j.compmedimag.2009.01.005.

## FDG-PET Parametric Imaging by Total Variation Minimization

Hongbin Guo<sup>a,\*</sup>, Rosemary A Renaut<sup>a</sup>, Kewei Chen<sup>b</sup>, and Eric M Reiman<sup>b</sup>

Hongbin Guo: hb\_guo@asu.edu; Rosemary A Renaut: renaut@asu.edu; Kewei Chen: Kewei.Chen@bannerhealth.com; Eric M Reiman: Eric.Reiman@bannerhealth.com

<sup>a</sup> Arizona State University, Department of Mathematics and Statistics, Tempe, AZ 85287-1804

<sup>b</sup> Banner Alzheimer Institute and Banner Good Samaritan Positron Emission Tomography Center, Phoenix, AZ 85006

### Abstract

Parametric imaging of the cerebral metabolic rate for glucose (CMRGlc) using [<sup>18</sup>F]-fluorodeoxyglucose positron emission tomography is considered. Traditional imaging is hindered due to low signal to noise ratios at individual voxels. We propose to minimize the total variation of the tracer uptake rates while requiring good fit of traditional Patlak equations. This minimization guarantees spatial homogeneity within brain regions and good distinction between brain regions. Brain phantom simulations demonstrate significant improvement in quality of images by the proposed method as compared to Patlak images with post-filtering using Gaussian or median filters.

### Keywords

Total variation; graphical analysis; Patlak plot; PET quantification; Parametric imaging; FDG; Alzheimer's disease; uptake rate

## 1. Introduction

We focus on positron emission tomography (PET) parametric imaging for estimating the cerebral metabolic rate of glucose (CMRGlc) using the [<sup>18</sup>F]-fluorodeoxyglucose (FDG) tracer. The ability to derive accurate parameters depends upon the quality of data, the quantification method and the numerical algorithm. In this study, we refer to the time activity curve (TAC) from a given tissue location as the *output*, tissue TAC or TTAC, and the TAC from the blood pool (image-derived or arterial blood-sampled) as the *input*, plasma TAC or PTAC. Most existing quantification methods perform well for regions of interest (ROIs), but are not good for voxel level quantification due to the high level of noise. These include graphical methods, [20,18], linear least squares, the weighted integration method, [3], generalized linear least squares, [6,5], nonlinear least squares (NLS) and weighted NLS.

All the algorithms listed above perform the quantification at each voxel location separately; they do not consider the kinetic similarities among neighboring voxels within functionally-defined regions. Thus, voxel-by-voxel variation in a functionally-homogeneous region may be large because of noise in the data. But, by incorporating the spatial constraint that parameters in a functionally-homogeneous region should be similar, in any of the above methods, the quality of the resulting parametric image for the CMRGlc may be improved. Zhou et al, [25], for example, improved the parametric image quality by ridge regression with constraints on the rate constants. There, the estimation of parameters uses a linear components decomposition

\*Corresponding author. Tel: 1-480-965-8002, Fax: 1-480-965-4160. hguo1@asu.edu (Hongbin Guo).

of the kinetics in which each component represents a functional kinetic curve generated by clustering the TTACs, and the problem is solved voxel by voxel. Here we propose the use of a total variation (TV) penalty term which imposes spatial consistency between neighboring voxels.

The total variation penalty was first introduced in the context of image deblurring by Rudin *et al.*, [22]. TV can significantly suppress noise while recovering sharp edges because it does not penalize discontinuities. It has received much theoretical research attention and been utilized in many signal and image processing applications. While it was introduced for PET image reconstruction by Jonsson *et al.*, [13], and Kisilev *et al.*, [15], it has apparently not been applied for parametric PET imaging. Instead of calculating the uptake rate for each voxel by Patlak's method, we propose to minimize the TV of the uptake rate over the entire image while also maintaining a good least squares fit for the Patlak equations at all voxels. Thus the parameters of the whole image are spatially related by the TV and solved simultaneously. The resultant parametric image is expected to have spatial homogeneity over brain regions with similar kinetics and distinct edges between brain regions that have different kinetics. This is validated by phantom simulations.

In addition to proposing the new model with the TV penalty, we also pay careful attention to the computational complexity of the algorithm by taking advantage of implementations for large scale sparse matrix computations. In contrast to approximating the Hessian matrix, as is typical for quasi-Newton methods, our algorithm explicitly and accurately calculates both gradient and Hessian terms. The Hessian is efficiently recalculated at each iteration because of its sparsity. The Quasi-Minimal Residual (QMR) method, [7], is used to solve the resulting large-scale linear systems. With this efficient implementation, the procedure described here is computationally feasible.

The rest of the paper is organized as follows: The new algorithm with TV penalty is introduced in section 2, with relevant computational issues detailed in the appendix. The experimental data sets are described in section 3 and results reported in section 4. Issues relevant to the proposed TV-Patlak method and computational aspects are discussed in section 5. Conclusions are presented in section 6.

## 2. Methods

The Patlak plot has been developed for systems with irreversible trapping [20]. Most often it is applied for the analysis of FDG. The measured TTAC undergoes a transformation and is plotted against a *normalized time*. It is given by the expression

$$\frac{C_T(t)}{C_P(t)} \approx K \frac{\int_0^t C_P(s) ds}{C_P(t)} + V, \quad (1)$$

where  $C_T(t)$  is the measured TTAC, (in counts/min/g) and  $C_P$  is the PTAC (in counts/min/ml), i.e. the FDG concentration in plasma. For systems with irreversible compartments this plot yields a straight line after sufficient equilibration time. For the FDG tracer, the slope  $K$  represents the uptake rate which, together with lumped constant (LC) and glucose concentration in plasma ( $C_{pg}$ ) allows easy calculation of the  $CMRGlc = KC_{pg}/LC$ , (in mg/min/100g). The intercept  $V$  is given by  $V_0 + \nu B$  where  $V_0$  is the distribution volume of the reversible compartment and  $\nu B$  is the fractional blood volume.

The linear relationship (1) can be rewritten as

$$\left(\int_0^t C_p(s)ds\right)K + C_p(t)V \approx C_T(t), \quad (2)$$

and, assuming  $m$  dynamic frames over the period of equilibration, its discretized version is

$$\left(\int_0^{t_j} C_p(s)ds\right)K + C_p(t_j)V \approx C_T(t_j), \quad j=1, \dots, m. \quad (3)$$

In matrix format,

$$A \begin{pmatrix} K \\ V \end{pmatrix} \approx \mathbf{b}, \quad (4)$$

where  $A$  is a  $m$ -by-2 matrix,

$$A = \begin{pmatrix} \int_0^{t_1} C_p(s)ds, & C_p(t_1) \\ \int_0^{t_2} C_p(s)ds, & C_p(t_2) \\ \vdots & \vdots \\ \int_0^{t_m} C_p(s)ds, & C_p(t_m) \end{pmatrix},$$

and vector  $\mathbf{b} = (C_T(t_1), C_T(t_2), \dots, C_T(t_m))^T$ . If we were to solve (4) for each voxel independently we would obtain a parametric image lacking spatial homogeneity and with low signal-to-noise ratio (SNR). Image denoising techniques could then be applied as a separate task to improve the image quality. Instead, obtaining all voxel parameters as a result of a global optimization algorithm with a TV penalty for the entire image, the necessity for postprocessing should be eliminated.

Limiting the discussion here to 2D images (although our application of TV is 3D), we select the active voxels to be quantified by the application of a brain mask, yielding a total of  $N$  voxels. Equation (4) holds with common matrix  $A$  dependent on  $C_p(t)$  for each voxel  $i$ , but with  $K$ ,  $V$  and  $\mathbf{b}$  replaced by  $K^{(i)}$ ,  $V^{(i)}$ , and  $\mathbf{b}_i$ , respectively, where  $\mathbf{b}_i$  is obtained from the TTAC for voxel  $i$ . Collecting the unknowns of these  $N$  voxels in vectors of uptake rates and intercepts

$$\mathbf{x} = (K^{(1)}, K^{(2)}, \dots, K^{(N)})^T \quad \text{and} \quad \mathbf{y} = (V^{(1)}, V^{(2)}, \dots, V^{(N)})^T,$$

and requiring (4) in the least squares sense over all voxels, while maintaining minimal TV of the uptake rate for the selected image voxels, yields the global minimization problem

$$(\text{TV} - \text{Patlak}): \quad \min \Phi(\mathbf{x}; \mathbf{y}), \quad (5)$$

$$\Phi(\mathbf{x}; \mathbf{y}) = \|\mathbf{x}\|_{\text{TV}, \beta} + \alpha \sum_{i=1}^N \|W_i(A(x_i, y_i)^T - \mathbf{b}_i)\|_2^2. \quad (6)$$

Here the total variation norm is given by  $\|\mathbf{x}\|_{\text{TV}, \beta} = \sum_{i=1}^N \varphi_i(\mathbf{x})$  with

$\varphi_i(\mathbf{x}) = \sqrt{(x_i - x_{i_r})^2 + (x_i - x_{i_b})^2 + \beta^2}$ , and  $x_{i_r}$  and  $x_{i_b}$  are the values associated with voxels to the right and below voxel  $i$ . Theoretically the TV norm is  $\|\mathbf{x}\|_{\text{TV}, 0}$ , which is a seminorm on a space of bounded variation, [24]. The small constant  $\beta$  is used to avoid the numerical difficulty due to the lack of differentiability at the origin of  $\varphi_i$  for  $\beta = 0$ . The diagonal weight matrix for voxel  $i$  is given by

$$W_i = \text{diag} \left( \sqrt{\frac{\Delta t_j}{b_{ij} e^{\lambda t_j}}} \right), \quad j=1, \dots, m, \quad (7)$$

where  $\Delta t_j$  is the scan duration of the frame at time  $t_j$ ,  $\lambda$  is the tracer's decay constant and  $b_{ij}$  is the value of the  $i^{\text{th}}$  TTAC at frame  $j$ . This weighting is consistent with using a simulation with variance

$$\text{Var}(C_T(t_j)) = S c \frac{C_T(t_j) e^{\lambda t_j}}{\Delta t_j}, \quad (8)$$

[17].  $Sc$  is a common scale factor that need not be made explicit here because it is absorbed into the parameter  $\alpha$  in (6).

The objective function in (5) is convex, and can be solved using a standard Newton-type algorithm, [24] Chapter 8. To simplify the expressions we introduce the vector  $\mathbf{z} = [\mathbf{x}; \mathbf{y}]$ .

Further details on the calculation of gradient vector  $\nabla \Phi(\mathbf{z})$  and Hessian matrix  $\nabla^2 \Phi(\mathbf{z})$  are provided in the appendix. Some other aspects of the algorithm, including discussion about constants  $\alpha$  and  $\beta$ , here chosen to be 0.2 and  $10^{-8}$ , respectively, as well as other approaches for the solution of the TV problem are discussed in section 5.

### 3. Experimental Data

To validate the proposed parametric imaging method we performed experiments with simulated data. The MRI-based high-resolution Zubal head phantom is used to define the brain structures, [26]. Each voxel in the slice of  $256 \times 256$  voxels is of size  $1.5 \times 1.5 \text{ mm}^2$ . There are 128 slices for the entire head and 62 defined anatomical, neurological, and taxonomical structures. The 11 regions for slice 64, representing a total 13708 voxels, are given in table 1. For the purposes of the simulations, well-accepted values of the kinetic rate parameters of these structures for the two-tissue compartmental model of the FDG tracer [12], are assigned. Notice that in order to better model the true biochemical process we assume  $k_4 > 0$ . Because  $k_4$  is, however, relatively small, Patlak's method, for which it is assumed  $k_4 = 0$ , is still a suitable graphical method for quantifying the uptake rate.

The noise-free input function, with values given in kBq/ml, is given by the formulation introduced in [9],

$$u(t) = \begin{cases} 0 & t \in [0, 0.25] \\ 339.03(t - 0.25) & t \in [0.25, 0.4433] \\ -214.656t + 160.691 & t \in [0.4433, 0.65] \\ 21.165e^{-0.7501(t-0.65)^{0.2359}} & t > 0.65. \end{cases} \quad (10)$$

Although any reasonable input function, including clinical plasma samples, could be used for the simulation this formulation was validated as providing a good approximation to the plasma samples of a healthy subject. Given the input function, the **exact** phantom can then be generated using the rate constants for the structures detailed in table 1, [23]. The output time frames were generated assuming time frames with durations  $\Delta t_j, j = 1, \dots, m, m = 22$ , given in minutes, 0.2,  $8 \times 0.0333$ ,  $2 \times 0.1667$ , 0.2, 0.5,  $2 \times 1$ ,  $2 \times 1.5$ , 3.5,  $2 \times 5$ , 10 and 30.

In our experiments, in order to control the computational overhead of the reconstructions, we reduced the image size of the phantom from  $256 \times 256$  to  $128 \times 128$ . To do this we needed to relabel the structure assigned to a given voxel. This was done by averaging the quantity  $K_1 k_3 / (k_2 + k_3)$  over the 2-by-2 neighboring voxels at the finer resolution. Then the voxel at the coarser level was labeled as belonging to the structure for which this average is closest to the structure value. The kinetic values were then assigned, the TTAC output values calculated, and then projected to the sinogram space yielding projected noise-free sinogram data  $P$ . For simplification, the instrumental and physical effects, including attenuation, Compton scattering, decay and random coincidences, were not simulated. Poisson noise was then added to the projected data using  $S = \text{poissrnd}(P)$  where the Matlab®, [19] function `poissrnd` uses vector  $P$  as the means of Poisson densities to generate the noisy sinogram data  $S$ . Based on this noisy sinogram data, concentration images are reconstructed using the Expectation-Maximization (EM) algorithm, [14].

Several data sets were generated: To investigate errors introduced by violation of the irreversibility assumption,  $k_4 = 0$ , we tested both  $k_4 = 0$  and  $k_4 > 0$ . Similarly, we tested both with and without noise in the sinograms to investigate the effects of the proposed method due to noise in the sinograms. Gaussian noise with noise levels 0%, 5%, 10%, 15% and 20% was added to the input function, i.e.

$$C_p(t_j) = u(t_j)(1 + CV\eta_j), \quad (11)$$

where  $\eta_j$  is selected from a standard normal distribution ( $G(0, 1)$ ), and  $CV = 0, 0.05, 0.10, 0.15$  and  $0.20$ . 100 random realizations are tested in each case.

We summarize the test data sets in table 2, using a character triple to classify each test. The first character of the triple indicates whether  $k_4 = 0$  or  $k_4 > 0$ , 0 or +, respectively. The second character indicates the noise level on the input function, 0 or +, for noise-free, or with added noise, respectively. To indicate the noise level + is replaced by 1, 2, 3 and 4 to indicate noise levels 5%, 10%, 15% and 20% as necessary. The third character indicates whether noise is added to the sinogram, again 0 or +, respectively. Therefore, for example, +3+ represents the test case for  $k_4 > 0$ , 15% noise in  $u(t)$  and Poisson noise added to the sinogram.

## 4. Results

To evaluate the simulations quantitatively we define the relative error of the  $k^{\text{th}}$  realization for voxel  $i$ :

$$r_{ik} = \frac{\widehat{K}_k^{(i)} - K_{\text{true}}^{(i)}}{K_{\text{true}}^{(i)}}, \quad i=1, \dots, N, \quad k=1, \dots, 100,$$

where  $\widehat{K}_k^{(i)}$  is the estimated value of the true value of the uptake rate  $K_{\text{true}}^{(i)}$  at the  $k^{\text{th}}$  realization for voxel  $i$ . Over 100 random simulations, and  $N$  voxels, we calculate the bias, i.e. the mean of the relative errors, and the deviation from the mean:

$$\bar{r} = \frac{\sum_{i=1}^N \sum_{k=1}^{100} r_{ik}}{100N}, \quad d = \frac{\sum_{i=1}^N \sum_{k=1}^{100} |r_{ik} - \bar{r}|}{100N}. \quad (12)$$

Absolute relative errors  $|r_{ik}|$  and associated mean  $\bar{R} = \sum_{i=1}^N \sum_{k=1}^{100} |r_{ik}| / 100N$ , and deviation,  $D = \sum_{i=1}^N \sum_{k=1}^{100} ||r_{ik}| - \bar{R}| / N$ , are also calculated. Note the deviations are  $l_1$  measurements which do not overweight outlier and large error samples, as is the case for the  $l_2$ -based measurements such as the root mean squared error.

In the images shown in the figures we illustrate the calculated uptake rates  $K$  of the FDG. Images for the CMRGlc can be obtained by directly scaling  $K$ . In figure 1 we compare the result of using Patlak and TV-Patlak for estimating the uptake rates with respect to no noise, 20% noise in the input function, Poisson noise in the sinogram, and finally with respect to the case in which the irreversibility assumption is violated but without noise in the sinogram or input data. In each case the histogram of the relative errors is given on the left, the Patlak image in the middle and the TV-Patlak on the right. The different scales in the histograms are due to the total number of results illustrated. When there is no noise (triples 000 and +00) the histogram illustrates results over all voxels but only one simulation, while for the noisy simulations the results are for all voxels over all 100 realizations of the noise. The TV-Patlak images are more homogeneous in all cases and the relative errors are smaller. The figures clearly show the improvements of employing the TV-Patlak method as compared to using Patlak independently for each voxel. This is confirmed in figure 2 in which images with noise in the sinogram, positive  $k_4$  and different noise levels in the input function are shown.

Quantitative measurements, confirming the illustrations, are presented in table 3. There we also present the results for conventional Patlak's method with post-smoothing by two standard filters:

1. a Gaussian filter (Patlak-GF), size 3-by-3 with standard deviation 0.5, generated by Matlab functions `filter2(fspecial('gaussian', 3, 0.5), img)` and
2. a median filter (Patlak-MF) generated by Matlab function `medfilt2(img)` for a 3-by-3 neighborhood.

Consistent with the observation in [21,16], we find that violation of the Patlak assumption,  $k_4 = 0$ , introduces about 10% bias;  $\bar{r} \approx 0$  when  $k_4 = 0$  but  $\bar{r} \approx 10\%$  for  $k_4 > 0$ . The rows  $\overline{TV}$  (std) and “# 10% (#15%)” provide complementary supporting information, indicating that the TV is minimized by TV-Patlak; as compared to Patlak, Patlak-GF and Patlak-MF the number of

voxels with larger error is reduced. In particular, we emphasize that **TV-Patlak provides a better noise removal mechanism than popular post-filtering approaches.**

In figures 3 and 4 we illustrate the uptake rates and relative error in the uptake rates, respectively, calculated by Patlak, TV-Patlak, Patlak-GF and Patlak-MF for one simulated data case +4+, i.e.  $k_4 > 0$ , 20% noise in the input function and Poisson noise in the sinograms. The uptake rate image generated by Patlak-MF is visually smoother than that by TV-Patlak, but the equivalent histograms show that the relative error is higher for Patlak-MF than for TV-Patlak; the Patlak-MF image is over-smoothed.

In figure 5 we illustrate the relative error of noisy case +4+ for gray matter regions in the phantom, including frontal lobes, occipital lobes, insula cortex, temporal lobes and globus pallidus, which are of interest for Alzheimer's disease research. The error bars show that all estimation methods for gray matter regions have negative bias and there are fewer cases with high error by TV-Patlak.

Finally, we note that the computational cost for the TV-Patlak Algorithm 1 is about 6.5 seconds while for Patlak, Patlak-GF and Patlak-MF they are 0.48, 0.55 and 0.58 seconds respectively for a 2D image on a PC with 1GHz CPU and Matlab code.

## 5. Discussion

In this section we discuss issues relevant to the proposed TV-Patlak method.

1. *Regularization constant  $\alpha$ :* There are many approaches for determining the choice of an appropriate regularization constant  $\alpha$ . A good reference would be [24] in which the methods of unbiased predictive risk-estimation, generalized cross validation, and the L-curve are described. In general, the choice of  $\alpha$  depends on the noise level of the problem. Here  $\alpha$  balances the homogeneity of the uptake rate against the residual of the traditional Patlak least squares data fit. For the PET imaging application, noise in the TTAC data depends on the scanner, the reconstruction method, the tracer dosage and even the kinetics of individuals. It is therefore possible to make a standard parameter setting for commonly-used environments. The most convenient method for the selection of  $\alpha$  is the so-called L-curve, [10,11], which plots total variation against  $\sum_{i=1}^N \|W_i(A(x_i, y_i)^T - \mathbf{b}_i)\|_2^2$  for all tested  $\alpha$ . The L-curve clearly displays the compromise between of the homogeneity and the residual of the Patlak fitting equations. The  $\alpha$  corresponding to the left lower corner of the L-curve is considered as the optimal choice. One representative L-curve of our simulations is illustrated in figure 6. We found for our simulations that a suitable choice is  $\alpha = 0.2$ , but certainly it will in general depend on the reconstruction algorithm. For example, the simple EM method and filtered backprojection algorithms introduce more noise than the maximum a posteriori (MAP) algorithm, [8,1]. For real data not only are there additional sources of noise but the choice for  $\alpha$  will also depend on the specific tracer. However, once an appropriate  $\alpha$  is found by L-curve for a specific imaging environment, it can be fixed for future imaging calculations.
2. *Constant  $\beta$ :* We use Newton's method for the case  $\beta \neq 0$  to solve the optimization problem (5). Other methods for solving the TV problem with  $\beta \neq 0$  are discussed in [24], which includes the primal-dual method [4]. A good choice of  $\beta$  avoids numerical difficulties for small derivatives and provides a good approximation of the TV. For our tests we found that the results are not sensitive to the choice of  $\beta$  and thus suggest the use of  $\beta = 10^{-8}$  as a good choice for FDG-PET brain imaging. If we wish to avoid



the choice for  $\beta$  the TV term in (5) can be reformulated as a set of linear constraints, and other algorithms are possible, [2].

3. *Boundary values:* At each iteration of the algorithm, after updating  $\mathbf{x}$  and before doing any function evaluation or other calculation, the *boundary* voxels need to be updated. The value of each boundary voxel is set to the average of its active neighbors in four directions.
4. *Computational efficiency:* For computational efficiency and to minimize memory usage, it is important to not only use sparse storage strategies for the relevant matrices but also to use appropriate algorithms for the solution of the large scale sparse linear systems given by (9). Here we use the QMR iteration, [7]. Moreover, if a general Newton's or quasi-Newton method with approximated Hessian were used, the cost in time and memory would be much more expensive because the approximated Hessian matrix is generally dense. In addition to achieving sparsity, the Hessian matrix is calculated accurately; and the convergence should be faster. For our simulations convergence is achieved in eleven iteration steps on average. The Matlab code of the TV-Patlak method can be downloaded from <http://math.asu.edu/~hongbin>.

## 6. Conclusions

A qualitative improvement in imaging of PET uptake can be achieved by using a global model, with the total variation as a penalty term, to obtain the voxel uptake rates. The resultant uptake images have spatial homogeneity over brain regions with similar kinetics and distinct edges between brain regions that have different kinetics. It is statistically validated that the TV-Patlak significantly reduces the relative errors of the calculated parameters as compared with those generated by Patlak's graphical method, and post-smoothing by Gaussian and median filters.

## Acknowledgments

This work was supported by grants from the state of Arizona, the NIH, R01 MH057899 and P30 AG19610 and the NSF DMS 0652833 and DMS 0513214. The authors thank Chi-Chuan Chen for providing the transition matrix, generated by the Monte Carlo method, which is used in the PET image reconstructions and Drs Bouman and Musfata for offering their MAP reconstruction code, which provides comparisons between reconstruction methods. We also acknowledge Dr. Laszlo Balkay and Dr. Jeffrey Fessler for providing PET simulator and PET reconstruction software, respectively.

## References

1. Bouman CA, Sauer K. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Tr Im Proc* March;1996 5 (3):480–492.
2. Boyd, S.; Vandenberghe, L. *Convex Optimization*. Cambridge University Press; 2004.
3. Carson R, Huang S, Green M. Weighted integration method for local cerebral blood flow measurements with positron emission tomography. *J Cereb Blood Flow Metab* 1986;6 (2):245–58. [PubMed: 3485644]
4. Chan TF, Golub GH, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J Sci Comput* 1999;20 (6):1964–1977.
5. Chen K, Lawson M, Reiman E, Cooper A, Feng D, Huang S, Bandy D, Ho D, Yen L, Palant A. Generalized linear least squares method for fast generation of myocardial blood flow parametric images with N-13 ammonia PET. *Med Imag* 1998;7 (3):236–243.
6. Feng D, Huang S. An unbiased parametric imaging algorithm for nonuniformly sampled biomedical system parameter estimation. *IEEE Trans Med Imag* 1996;15 (4):512–518.
7. Freund RW, Nachtigal NM. QMR: A quasi-minimal residual method for non-Hermitian linear systems. *Numerische Mathematik* 1991;60:315–340.



8. Geman S, McClure DE. Bayesian image analysis: An application to single photon emission tomography. *Proc Amer Statist Assoc Statistical Computing Section* 1985:12–18.
9. Guo H, Renaut R, Chen K. An input function estimation method for FDG-PET human brain studies. *Nuclear Medicine and Biology* 2007;34 (5):483–492. [PubMed: 17591548]
10. Hansen PC. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review* 1992;34:561–580.
11. Hansen PC, O’Leary DP. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing* 1993;14:1487–1503.
12. Huang SC, Phelps ME, Hoffman EJ, Sideris K, Selin CJ, Kuhl DE. Noninvasive determination of local cerebral metabolic rate of glucose in man. *Am J Physiol* 1980;238 (E):69–82.
13. Jonsson E, Huang SC, Chan T. Total variation regularization in positron emission tomography. *Technical Reports on Image Processing CAM 98-48*, UCLA. 1998
14. Kaufman L. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans Med Imag Mar*;1987 6 (1):37–51.
15. Kisilev P, Zibulevsky M, Zeevi YY. Wavelet representation and total variation regularization in emission tomography. *ICIP* 2001;(1):702–705.
16. Lammertsma A, Brooks D, Frackowiak R, Beaney R, Herold S, Heather J, Palmer A, Jones T. Measurement of glucose utilisation with [18f]2-fluoro-2-deoxy-d-glucose: a comparison of different analytical methods. *J Cereb Blood Flow Metab* 1987;7 (2):161–72. [PubMed: 3558499]
17. Logan J, Fowler J, Volkow N, Ding Y, Wang G, Alexoff D. A strategy for removing the bias in the graphical analysis method. *J Cereb Blood Flow Metab* 2001;21 (3):307–20. [PubMed: 11295885]
18. Logan J, Fowler JS, Volkow ND, Wolf AP, Dewey SL, Schlyer DJ, MacGregor RR, Hitzemann R, Bendriem B, Gatley SJ. Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects. *J Cereb Blood Flow Metab* 1990;10:740–747. [PubMed: 2384545]
19. Matlab. Matlab is a registered trademark of MathWorks, Inc. 2008. URL <http://www.mathworks.com>
20. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J Cereb Blood Flow Metab* 1983;3 (1):1–7. [PubMed: 6822610]
21. Phelps ME, Huang SC, Hoffman EJ, Selin CE, Kuhl D. Tomographic measurement of local cerebral glucose metabolic rate in man with (<sup>18</sup>F) fluorodeoxyglucose: Validation of method. *Ann Neurol* 1979;6:371–388. [PubMed: 117743]
22. Rudin L, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D* 1992;60:259–268.
23. Sokoloff L, Reivich M, Kennedy C, Rosiers MHD, Patlak CS, Pettigrew KD, Sakurada M, Shinohara M. The [<sup>14</sup>C] deoxyglucose method for the measurement of local cerebral glucose metabolism: theory procedures and normal values in the conscious and anesthetized albino rat. *J Neurochem* 1977;28:897–916. [PubMed: 864466]
24. Vogel, CR. *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics; Philadelphia, PA, USA: 2002.
25. Zhou Y, Endres CJ, Brasic JR, Huang SC, Wong DF. Linear regression with spatial constraint to generate parametric images of ligand-receptor dynamic PET studies with a simplified reference tissue model. *NeuroImage* 2003;18 (4):975–989. [PubMed: 12725772]
26. Zubal IG, Harrell CR, Smith EO, Rattner Z, Gindi G, Hoffer PB. Computerized three-dimensional segmented human anatomy. *Medical Physics* Feb;1994 21:299–302. [PubMed: 8177164]

## A. Appendix

To simplify notation, we assume that the weighting matrix in (5) has been absorbed into  $A$  and  $\mathbf{b}_i$ . To simplify the expressions in the objective function (6) we introduce the vector function  $\mathbf{f}(\mathbf{z}) = (f_1(\mathbf{z}), f_2(\mathbf{z}), \dots, f_{2N}(\mathbf{z}))^T$ , where  $f_i(\mathbf{z}) = \varphi_k(\mathbf{x})^2$  for  $i = 1 \dots N$  and  $f_i(\mathbf{z}) = \|A[x_{i-N}; y_{i-N}] - \mathbf{b}_{i-N}\|^2$  for  $i = N + 1 \dots 2N$ .

## Gradient calculation

To find the gradient of  $\Phi(\mathbf{z})$ , we first derive the Jacobian of  $\mathbf{f}(\mathbf{z})$ .

1. For  $i = 1, \dots, N$  and  $l = 1, \dots, 2N$

$$\frac{\partial f_i}{\partial z_l} = \begin{cases} 4x_i - 2x_{i_b} - 2x_{i_r}, & \text{if } l=i \\ 2x_{i_b} - 2x_i, & \text{if } l=i_b \\ 2x_{i_r} - 2x_i, & \text{if } l=i_r \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

2. For  $i = N + 1, \dots, 2N$  and  $l = 1, \dots, 2N$

$$\frac{\partial f_i}{\partial z_l} = \begin{cases} p_{i1}, & \text{if } l=i-N \\ p_{i2}, & \text{if } l=i \\ 0, & \text{otherwise,} \end{cases} \quad (14)$$

where

$$\begin{pmatrix} p_{i1} \\ p_{i2} \end{pmatrix} = 2A^T A \begin{pmatrix} x_{i-N} \\ y_{i-N} \end{pmatrix} - 2A^T \mathbf{b}_{i-N}.$$

Because  $\nabla_{\mathbf{x}} \varphi_i = 1/(2\varphi_i) \nabla_{\mathbf{x}} f_i$  for  $i = 1, \dots, N$ ,  $\nabla_{\mathbf{z}} \Phi = (\nabla \mathbf{f}(\mathbf{z}))^T \mathbf{g}$ , where

$$\mathbf{g} = [1/(2\varphi_1), 1/(2\varphi_2), \dots, 1/(2\varphi_N), \alpha, \alpha, \dots, \alpha]^T.$$

## Hessian matrix

To find the Hessian matrix for  $\Phi(\mathbf{z})$ , we first derive the Hessian matrix for each function  $f_i(\mathbf{z})$ .

1. For  $i = 1, \dots, N$ ,  $\nabla_{\mathbf{xx}}^2 f_i$  is sparse

$$\nabla_{\mathbf{xx}}^2 f_i = \begin{pmatrix} i & i_b & i_r \\ 4 & -2 & -2 \\ -2 & 2 & 0 \\ -2 & 0 & 2 \end{pmatrix} \begin{matrix} i \\ i_b \\ i_r \end{matrix}. \quad (15)$$

2. For  $i = N + 1, \dots, 2N$ , the Hessian matrix is again sparse.

$$\Delta_{\mathbf{zz}}^2 f_i = \begin{pmatrix} i-N & i \\ q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \begin{matrix} i-N \\ i \end{matrix} \quad \text{where} \quad \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = 2A^T A. \quad (16)$$

Thus

$$\nabla^2 \Phi(\mathbf{z}) = \sum_{i=1}^N \begin{pmatrix} \nabla_{\mathbf{xx}}^2 \varphi_i & 0_{N \times N} \\ 0_{N \times N} & 0_{N \times N} \end{pmatrix} + \alpha \sum_{i=N+1}^{2N} \nabla^2 f_i,$$

where

$$\nabla_{\mathbf{x}\mathbf{x}}^2 \varphi_i = -\frac{1}{4\varphi_i^3} \nabla_{\mathbf{x}} f_i (\nabla_{\mathbf{x}} f_i)^T + \frac{1}{2\varphi_i} \nabla_{\mathbf{x}\mathbf{x}}^2 f_i,$$

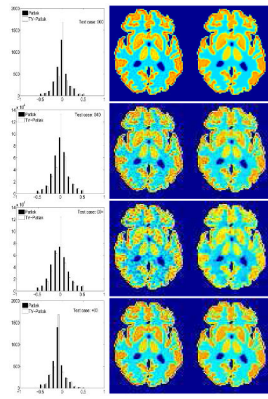
## Biographies

**Hongbin Guo** received his Ph.D. degree in computational mathematics from the Department of Mathematics, Fudan University, in 2000. He is currently an Assistant Research Professor with the Department of Mathematics and Statistics at Arizona State University. His main research interest is computational mathematics with a focus on the numerical linear algebra and its applications in medical imaging, particularly for the quantification of brain function with positron emission tomography and magnetic resonance imaging, co-registration, functional clustering, image restoration and classification.

**Rosemary A. Renaut** received the Ph.D. degree in applied mathematics from the University of Cambridge, U.K., in 1985. Since 1987, she has been with the Department of Mathematics, Arizona State University, where she is now a Full Professor. Dr. Renaut is a Fellow of the Institute for Mathematics and its Applications, and a Chartered Mathematician. Her research interests are broad and include the design and evaluation of computational methods for the solution of partial differential equations, with specific emphasis on high-order and spectral methods, as well as the development of novel algorithms for solving inverse problems, specifically as applied to medical image reconstruction and restoration.

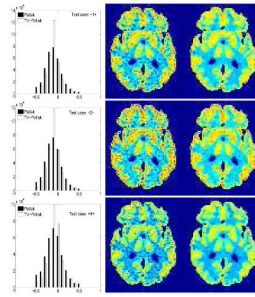
**Kewei Chen** got his master degree from Beijing Normal University in 1986 and his Ph.D. Degree from UCLA in 1993. He is currently the director and a senior biomathematician of the Computational Image Analysis Program, Banner Alzheimer's Institute. He has been using and developing neuroimaging analytic techniques in PET, magnetic resonance imaging (MRI) and functional MRI. His main methodological research interests include the tracer kinetic modeling in PET, measuring local and global volume changes using various MRI techniques, brain functional connectivity, and multi-modal data integration. One of his primary research interests is the use of neuroimaging techniques in the study of Alzheimer's disease.

**Eric M Reiman** is Executive Director of the Banner Alzheimer's Institute, Clinical Director of the Neurogenomics Division at the Translational Genomics Research Institute (TGen), Professor and Associate Head of Psychiatry at the University of Arizona, and Director of the Arizona Alzheimer's Consortium. He received his undergraduate and medical degrees and most of psychiatry residency training at Duke University. He completed his residency training, became an Assistant Professor of Psychiatry and developed a leadership role in positron emission tomography research at Washington University in St. Louis, before moving to Arizona. His research interests include brain imaging, genomics, and their use in the unusually early detection and tracking of Alzheimer's disease and the rigorous and rapid evaluation of promising Alzheimer's disease-slowing, risk-reducing and prevention therapies.



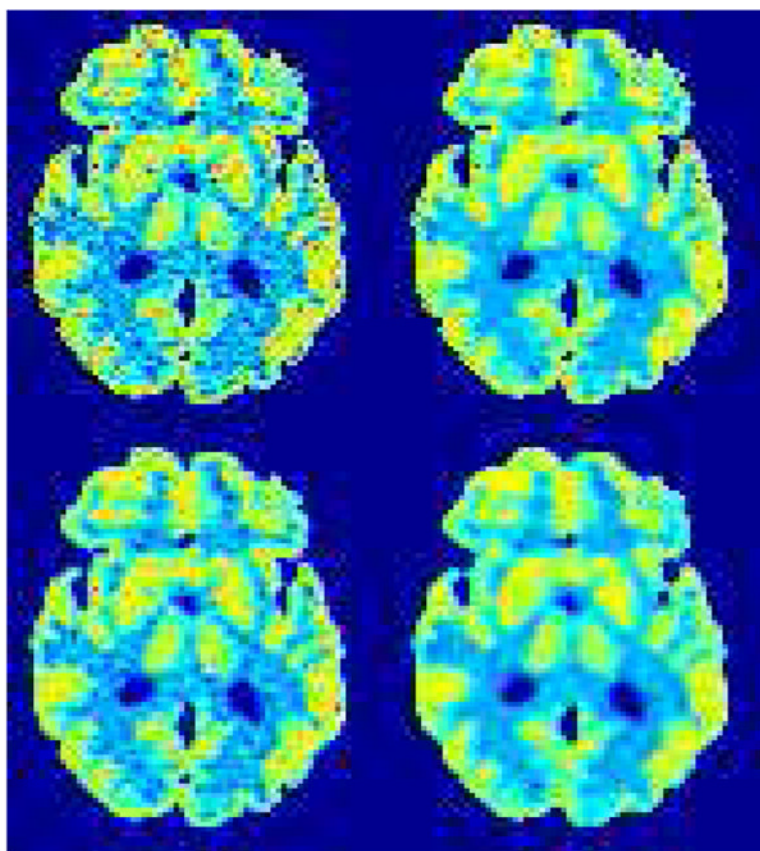
**Figure 1.**

Comparison of Patlak and TV-Patlak for imaging the uptake rate  $K$ . In each case on the left is the histogram for the relative error as compared to the exact simulated value, in the middle the Patlak image and on the right the TV-Patlak image. The first row compares the estimation by Patlak and TV-Patlak with no noise added. The second, third and fourth rows provide the comparison with 20% noise added in the input, Poisson noise added to the sinogram, and positive  $k_4$ , respectively.

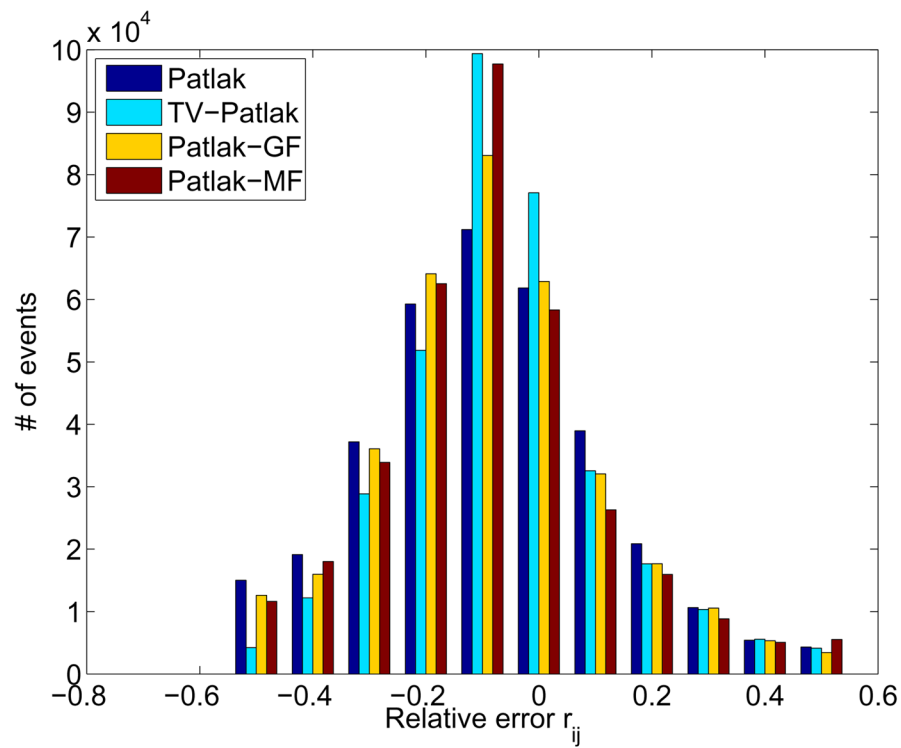


**Figure 2.**

Comparison of Patlak and TV-Patlak for imaging the uptake rate  $K$ . In each case on the left is the histogram for the relative error as compared to the exact simulated value, in the middle the Patlak image and on the right the TV-Patlak image. In each case Poisson noise is added to the sinogram and  $k_4 > 0$ . The level of noise for 100 realizations added to the input is 5%, 10% and 20% respectively, for each row from top to bottom.



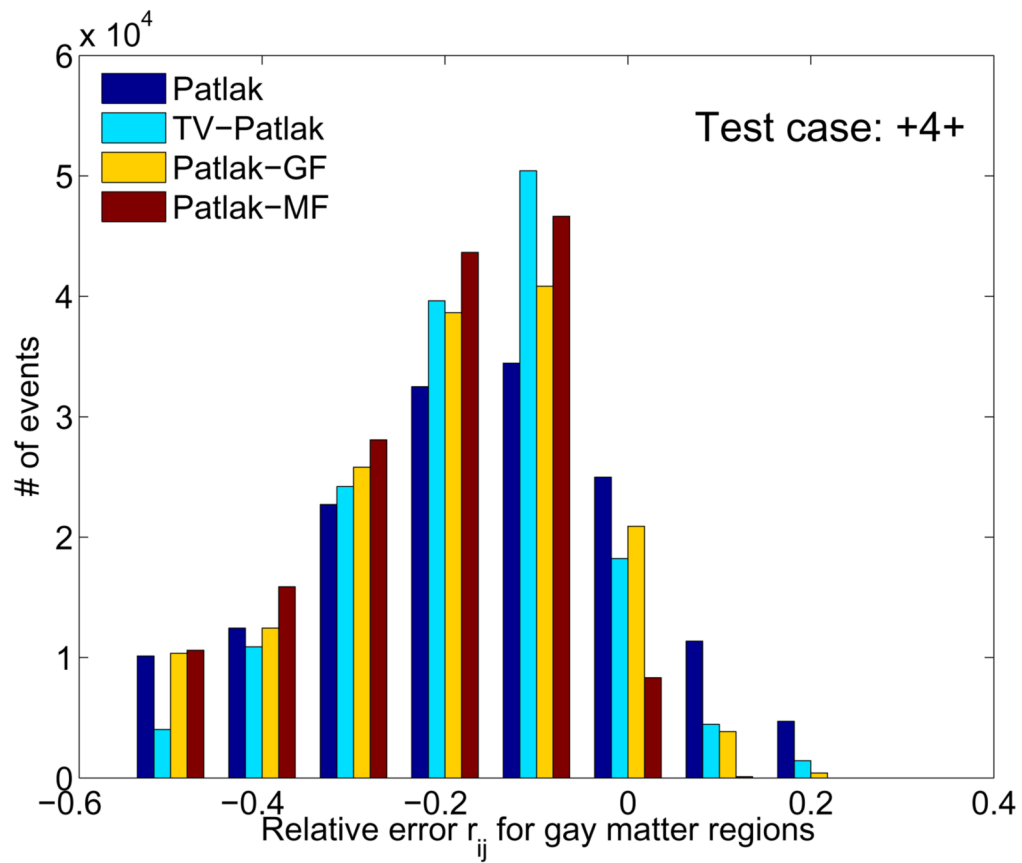
**Figure 3.** Comparison of Patlak (upper-left), TV-Patlak (upper-right), Patlak-GF (bottom-left) and Patlak-MF (bottom-right) for imaging the uptake rate  $K$  for one simulated data case +4+, i.e.  $k_4 > 0$ , 20% noise in the input function and Poisson noise in the sinogram.



**Figure 4.**

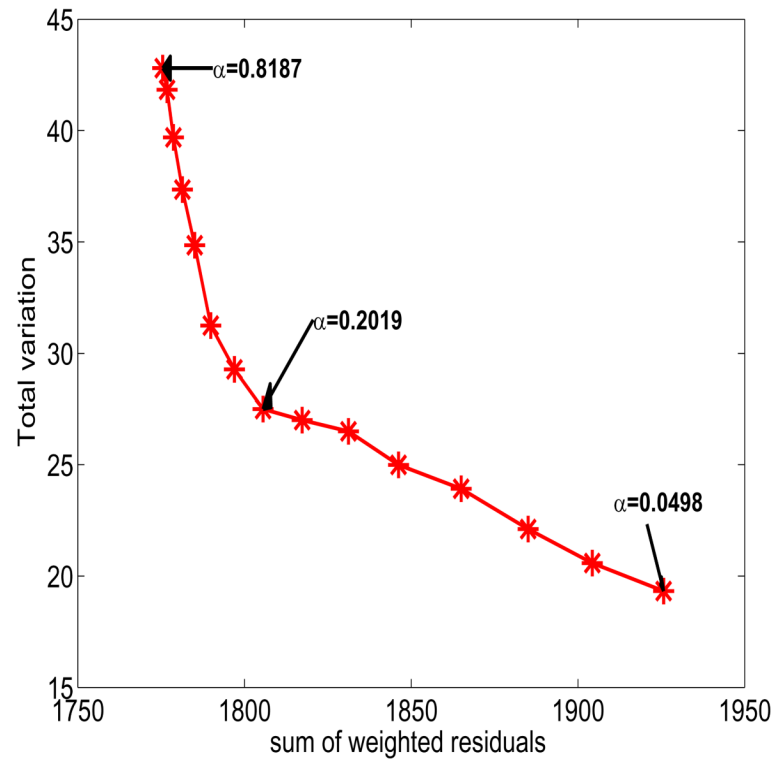
Histograms for the relative error (12) in the uptake rate calculated by Patlak, TV-Patlak, Patlak-GF and Patlak-MF, for 100 realizations of the data case +4+, i.e.  $k_4 > 0$ , 20% noise in the input function and Poisson noise in the sinogram.





**Figure 5.**

Histograms for the relative error (12) in the uptake rate of gray matter regions calculated by Patlak, TV-Patlak, Patlak-GF and Patlak-MF, for 100 realizations of the data case +4+, i.e.  $k_4 > 0$ , 20% noise in the input function and Poisson noise in the sinogram.



**Figure 6.**

L-curve for the simulation case '+4+'. Total variation against the sum of weighted residuals,

i.e.  $\sum_{i=1}^N \|W_i(A(x_i, y_i)^T - \mathbf{b}_i)\|_2^2$ , for  $\alpha$  from 0.0498 to 0.8187 is plotted.  $\alpha = 0.2$  is associated with the left lower "corner".

**Table 1**

Brain regions and rate constants[12] for slice 64 of the Zubal head phantom [26].

Brain Regions	$K_1$ ml/min/g	$k_2$ 1/min	$k_3$ 1/min	$k_4$ 1/min
frontal lobes occipital lobes insula cortex temporal lobes globus pallidus	0.102	0.130	0.062	0.0068
thalamus	0.082	0.105	0.060	0.0068
putamen caudate nucleus	0.070	0.070	0.054	0.0068
internal capsule corpus collosum other white matter	0.054	0.109	0.045	0.0058

**Table 2**

Summary of the test cases. Here 0 in columns two and three indicates the noise-free case, while + indicates noise was added. In column one the 0 indicates  $k_4 = 0$ . The same comments apply, but with the irreversibility assumption violated for all triples starting with +, indicating  $k_4 > 0$ .

$k_4$	$u(t)$	Sinogram	Comments
0	0	0	Errors only caused by reconstruction
	0	+	Errors in TTACs caused by reconstruction plus Poisson noise
	+	0	Errors in $C_p(t)$
	+	+	General case for irreversible compartmental model

Table 3

Results of phantom simulation: comparison of relative errors between the Patlak (P), TV-Patlak (TV-P), Patlak-GF (P-GF) and Patlak-MF (P-MF) methods.  $\overline{TV}$  is the mean of the total variation over 100 simulations in each case, std denotes its standard deviation. The rows “#10% (#15%)” report the average number of voxels with high errors,  $|r_{ik}| > 10\%$  ( $|r_{ik}| > 15\%$ ) per image. The first row indicates the specific simulation.

Data case	000	040	00+	+00	+1+	+2+	+4+
P	$\bar{r}$	-0.009	0.005	-0.009	-0.105	-0.103	-0.088
	$d$	0.116	0.136	0.157	0.106	0.151	0.165
	$R$	0.116	0.136	0.157	0.147	0.177	0.182
	$D$	0.088	0.094	0.101	0.081	0.104	0.113
	$\overline{TV}$	19.488	22.816	24.811	17.246	22.901	23.526
	std	0.000	0.193	0.235	0.000	0.219	0.225
	#10%	1429	1718	2020	2099	2272	2265
	#15%	995	1166	1453	1285	1733	1732
	$\bar{r}$	-0.005	0.020	0.001	-0.101	-0.094	-0.063
	$d$	0.104	0.120	0.128	0.097	0.121	0.123
TV/P	$R$	0.105	0.121	0.128	0.145	0.157	0.155
	$D$	0.095	0.096	0.099	0.078	0.089	0.091
	$\overline{TV}$	15.663	17.624	15.484	13.650	13.607	13.988
	std	0.000	0.156	0.220	0.000	0.241	0.254
	#10%	1345	1462	1522	1905	2164	2097
	#15%	974	1046	1102	1163	1441	1415
	$\bar{r}$	-0.012	0.002	-0.012	-0.108	-0.106	-0.091
	$d$	0.124	0.137	0.151	0.115	0.142	0.143
	$R$	0.125	0.137	0.152	0.162	0.175	0.175
	$D$	0.106	0.106	0.104	0.088	0.101	0.102
P/GF	$\overline{TV}$	19.052	20.635	22.254	16.881	20.231	20.502
	std	0.000	0.126	0.171	0.000	0.147	0.155
	#10%	1492	1589	1874	2240	2297	2279
							2205
							21.653
							0.106
							0.173
							0.151
							0.251
							15.344

Data case	000	040	00+	+00	+1+	+2+	+4+
	#15%	1183	1191	1342	1328	1722	1713
PMF	$\bar{f}$	-0.017	-0.002	-0.021	-0.112	-0.116	-0.113
	$d$	0.125	0.138	0.148	0.117	0.139	0.140
	$R$	0.128	0.138	0.150	0.169	0.182	0.181
	$D$	0.113	0.115	0.114	0.096	0.103	0.104
	$\overline{TV}$	15.938	16.295	16.795	14.104	15.026	15.121
	std	0.000	0.119	0.130	0.000	0.137	0.143
	#10%	1499	1522	1709	2342	2396	2362
	#15%	1118	1152	1251	1335	1730	1710
							1615

**Algorithm 1**

Given initial guess  $\mathbf{z} = [\mathbf{x}; \mathbf{y}]$  and tolerance  $\varepsilon > 0$

---

**Repeat**

1 **Solve** for  $\Delta \mathbf{z}$  in

$$\nabla^2 \Phi(\mathbf{z}) \Delta \mathbf{z} = -\nabla \Phi(\mathbf{z}). \quad (9)$$

2 **Stop** if  $|\nabla \Phi^T \Delta \mathbf{z}| \leq \varepsilon$ .

3 **Line search:** Choose step size  $s$ .

4 **Update:**  $\mathbf{z} = \mathbf{z} + s \Delta \mathbf{z}$ .

---