Multiscale TV flow with applications to fast denoising and registration

Prashant Athavale[a], Robert Xu[a], Perry Radau[b], Adrian Nachman[a] and Graham Wright[b]

[a]University of Toronto, Toronto, Canada;
[b]Sunnybrook Health Institute, Toronto, Canada

ABSTRACT

Medical images consist of image structures of varying scales, with different scales representing different components. For example, in cardiac images, left ventricle, myocardium and blood pool are the large scale structures, whereas infarct and noise are represented by relatively small scale structures. Thus, extracting different scales in an image i.e. multiscale image representation, is a valuable tool in medical image processing. There are various multiscale representation techniques based on different image decomposition algorithms and denoising methods. Gaussian blurring with varying standard deviation can be considered as a multiscale representation, but it diffuses the image isotropically, thereby diffusing main edges. On the other hand, inverse scale representations based on variational formulations preserve edges; but they tend to be time consuming and thus unsuitable for real-time applications.

In the present work, we propose a fast multiscale representation technique, motivated by successive decomposition of smooth parts based on total variation (TV) minimization. Thus, we smooth a given image at an increasing scale, producing a multiscale TV representation. As noise is a small scale component of an image, we can effectively use the proposed method for denoising . We also prove that the denoising speed, up to the time-step, is determined by the user, making the algorithm well-suited for *real-time* applications. The proposed method inherits edge preserving property from total variation flow. Using this property, we propose a novel multiscale image registration algorithm, where we register corresponding scales in images, thereby registering images efficiently and accurately.

1. INTRODUCTION

Decomposing an image into different scales is a natural way of studying medical images, as medical images consist of components with varying scales. Denoising of an image can be perceived as a decomposition of the given image f, into a denoised part u and a residual $f - u$. There are numerous methods proposed for denoising, based on partial differential equations, statistical analysis, wavelets, etc. We note that any image restoration technique leads to image decomposition. For example, an observed image f with additive noise is naturally decomposed into a denoised part u_λ, and the noise $f - u_\lambda$. Thus, denoising generates *a multiscale family*, $\{u_\lambda\}_{\lambda\in\Lambda}$ where λ denotes an algorithm-specific scale parameter, which belongs to some set Λ. This approach motivates multiscale image representation schemes.[1–4]

The goal of this paper is to propose a multiscale representation of a given image based on variational method. Mathematically, a grayscale digital image image can be considered as sampled version of a real valued function $f : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ denotes the domain of the image. Our approach is based on iterative (TV, L^2) decomposition of the function f into a TV-part, u, which is a function with bounded total variation, and a square integrable residual, $f - u$, i.e. the L^2-part. We successively apply the decomposition to the TV- parts obtained at increasing scales. We show that this approach gives us a differential equation similar to the total

Further author information: (Send correspondence to P.A.)
P.A. : E-mail: prashantva@gmail.com
R.X. : E-mail: robert.xu@sri.utoronto.ca
P.R. : E-mail: perry.radau@gmail.com
A.N. : E-mail: nachman@math.toronto.edu
G.W.: E-mail: graham.wright@sri.utoronto.ca

variation flow,[5–7] that yields a multiscale representation of the given image. We show that this approach is effective in fast denoising of medical images where the noise is of smaller scale than useful structures in the image. Furthermore, because of the multiscale nature of the algorithm, it can be used successfully for accurate registration of medical images. In this approach, we first register large scale images obtained using the multiscale TV flow and then use these registration parameters as initial guess for registering subsequent finer scale images.

1.1 Multiscale total variation flow

One of the earliest PDE-based methods for denoising[8] and multiscale representation[9] of a given image f is using the heat equation; which is equivalent to Gaussian smoothing. This yields a family of images, $\{u(\cdot,t)\}_{t\geq 0}$, which can be viewed as smoothed versions of f. In this linear set up, smoothing is implemented by a convolution with the two-dimensional Gaussian kernel. Hence, small scale details are removed due to the isotropic nature of Gaussian smoothing. We can say that $\{u(\cdot,t)\}_{t\geq 0}$ is a multiscale representation of f, as $u(\cdot,t)$ diffuses from the small scales in f into increasingly larger scales.

Some variational methods also produce multiscale image representation,[1–3,8] while preserving edges. These techniques are based on minimizing a suitable energy. The variational approach for denoising as proposed by Rudin, Osher, Fatemi (ROF) can be viewed as the (TV, L^2) decomposition of the given image, f, into two parts, the clean image u_λ, which is forced to be in the space of functions with finite total variation ($\int_\Omega |\nabla u_\lambda| < \infty$), and the $L^2(\Omega)$ residual image, $v_\lambda = f - u_\lambda$, which contains texture, noise etc. The (TV, L^2) decomposition can be formulated in terms of the minimization problem as follows[10]

$$f = u_\lambda + v_\lambda, \quad u_\lambda := \underset{u}{\operatorname{arginf}} \Big\{ \int_\Omega |\nabla u| + \lambda \int_\Omega |f-u|^2 \Big\}, \tag{1}$$

where the minimizer u_λ is a denoised version[10,11] of f with the scale $\sim 1/\lambda$. The minimizer u_λ formally satisfies the following Euler-Lagrange differential equation

$$u_\lambda = f + \frac{1}{2\lambda}\operatorname{div}\left(\frac{\nabla u_\lambda}{|\nabla u_\lambda|}\right); \quad u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \quad \frac{\partial u}{\partial \mathbf{n}}\Big|_\Gamma = 0, \tag{2}$$

where Γ denotes the boundary of the image f. If we select the parameter $\lambda = \lambda_0$ in (1) to be very small, then we get an image u_{λ_0} with a large fidelity term, $\int_\Omega |f-u|^2$. On the other hand, if we choose the scaling parameter λ_0 to be large, then the image u_{λ_0} thus obtained will be closer to the given image, f.

Our idea is to decompose the original image, f, using a large parameter, λ_0, in the (TV, L^2) decomposition (1) to obtain the initial decomposition, $f = u_{\lambda_0} + v_{\lambda_0}$, and then decompose the image u_{λ_0} using the scaling parameter $\lambda_1 < \lambda_0$, to obtain the (TV, L^2) decomposition $u_{\lambda_0} = u_{\lambda_1} + v_{\lambda_1}$. We continue this process iteratively, each time decomposing the TV- part with $\lambda_{i+1} < \lambda_i$, thus producing the following nonlinear multiscale decomposition, which is also discussed in[1] with dyadic scale.

$$\begin{aligned} f &= u_{\lambda_0} + v_{\lambda_0} \\ &= u_{\lambda_1} + v_{\lambda_1} + v_{\lambda_0} \\ &= \dots \\ &= u_{\lambda_N} + \sum_{i=0}^{N} v_{\lambda_i}. \end{aligned}$$

Using the Euler-Lagrange differential equation (2) we get the following:

$$u_{\lambda_N} = f + \sum_{i=0}^{N} \frac{1}{2\lambda_i}\operatorname{div}\left(\frac{\nabla u_{\lambda_i}}{|\nabla u_{\lambda_i}|}\right). \tag{3}$$

We propose a similar decomposition, but with the sequence of parameters $\{\frac{\lambda_i}{\triangle\tau}\}$, where $\triangle\tau$ is a small intensity quanta, and $\lambda_{i+1} < \lambda_i$. We get the following summation equation after N steps:

$$u_{\lambda_N} = f + \sum_{i=0}^{N} \frac{1}{2\lambda_i}\operatorname{div}\left(\frac{\nabla u_{\lambda_i}}{|\nabla u_{\lambda_i}|}\right)\triangle\tau.$$

This equation motivates the following integro-differential equation:

$$u(x,t) = f(x) + \int_0^t \frac{1}{2\lambda(s)} \operatorname{div}\left(\frac{\nabla u(x,s)}{|\nabla u(x,s)|}\right) ds; \quad u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \quad \frac{\partial u}{\partial \mathbf{n}}\Big|_\Gamma = 0, \tag{4}$$

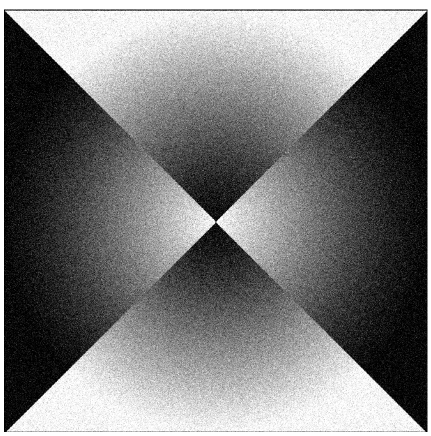
with $u(\cdot, 0) = f$ as the initial condition and $x \equiv (x_1, x_2) \in \Omega$. Here, the function $\lambda(t)$ is any real valued function that is monotone decreasing, and vanishing at infinity. Differentiating (4) in time, we get

$$\frac{\partial u}{\partial t} = \mu(t) \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right); \quad u : \Omega \times \mathbb{R}_+ \mapsto \mathbb{R}, \quad \frac{\partial u}{\partial \mathbf{n}}\Big|_\Gamma = 0, \tag{5}$$

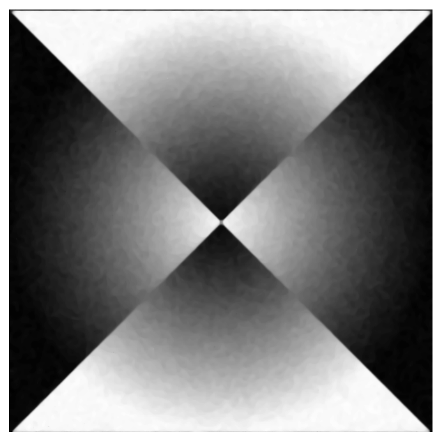
where we impose $u(t = 0) := f$ with Neumann boundary condition and $\mu(t) \equiv \frac{1}{2\lambda(t)}$ is the speed function. This equation produces a multiscale flow and it is closely related to the well-studied problem of total variation flow.[5–7] Thus we refer to (5) as multiscale TV flow. One of the important property is the speed of denoising is directly proportional to the function $\mu(t)$.

2. DENOISING WITH MULTISCALE TV FLOW

In this section, we demonstrate fast denoising applications with multiscale TV flow. The flow in (5) has many interesting properties like edge preserving smoothing,[6] and the dependence of the denoising speed on user-defined speed function $\mu(t)$, which will be explored in section 2.3.



(a) noisy synthetic image



(b) denoised image

Figure 1. Image denoising with Multiscale TV flow preserves edges.

2.1 Method

We can use the multiscale TV flow in (5) for denoising if we know an estimate of the noise variance or equivalently the signal to noise ratio. For denoising an image f, we start with $u(\cdot, t = 0) := f$ and solve (5) using finite difference scheme. At each time step we measure the variance of the residual, $f - u(\cdot, t)$, and the flow is terminated at $t = t_0$ when this variance is more than the estimated variance. The results are demonstrated in Figure 1.

2.2 Experimental results

We now demonstrate the denoising algorithm using multiscale TV flow. To this effect we added Gaussian noise with zero mean and variance, $\sigma^2 = 0.02$ to a synthetic image* to obtain a noisy image f shown in Figure

*size of the synthetic image is 663×663 pixels.

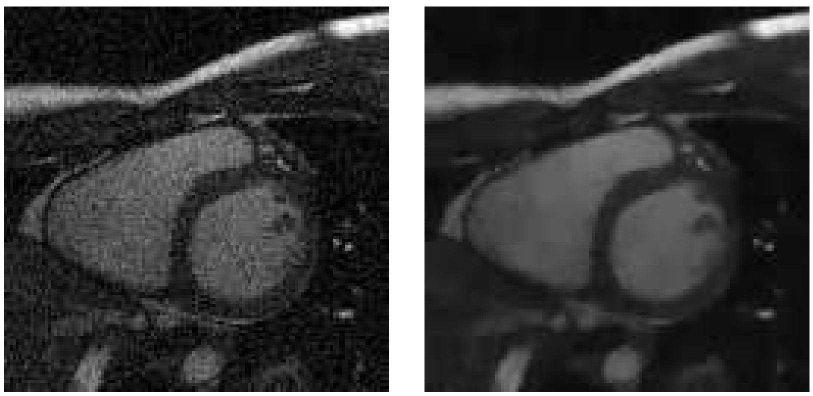
1(a). The image intensity is scaled in the range from 0 to 1 before adding the noise. We used this image as the initial condition for (5) and we terminated the TV flow when the variance of the residual $f - u(\cdot, t_0)$ became more than σ^2. The denoised image $u(\cdot, t_0)$ is shown in Figure 1(b). (In practice, we could estimate the noise variance empirically or statistically,[12] which is beyond the scope of this paper.) We observe that the image is denoised, without diffusing the main edges.

2.3 Denoising speed

We define the star-norm which measures the noise and oscillation content in an image, as the dual of the TV-seminorm with respect to the L^2-inner product.[11] Using standard techniques[3,11,13] we can show that the star-norm of the speed of the flow, i.e.

$$\left\| \frac{\partial u}{\partial t} \right\|_* = \mu(t). \tag{6}$$

Thus, we can control the denoising speed. We can exploit this property for real-time denoising of images, by appropriately choosing the function $\mu(t)$.



Noisy image

Denoised image

Figure 2. Denoising of cardiac image with multiscale TV flow.

Table 1. Denoising speed can be controlled by changing the speed function $\mu(t)$.

Speed function $\mu(t)$	computational time (ms)
1	402
2^t	137
4^t	98
8^t	70

To illustrate this control over the denoising speed, we performed a set of denoising experiments with a cardiac image with Rician noise $R(\nu, \sigma)$ with $\nu = 0, \sigma^2 = 0.005$. We terminated the TV flow when the SNR[†] of the denoised image became less than 14 dB. We performed this experiments for various speed functions $\mu(t)$. The denoising results are depicted in Figure 2. The computational times[‡] for various speed functions $\mu(t)$ are delineated in Table 1. We observe that if we choose a fast growing speed function, the speed of denoising increases as predicted by equation (6).

[†] $SNR(f, u) := 20 \log_{10} \frac{\|u\|_{L^2}}{\|u-f\|_{L^2}}$dB where f is the original image and u is the denoised image.

[‡] The code was written in Matlab, and run on 2.3 GHz Intel Core i5, Mac OS X. The image size is 130×130 pixels.

3. HIERARCHICAL REGISTRATION USING MULTISCALE TV FLOW

Image registration is one of the important problems in medical imaging; where given two images and which may be taken at different times, for different subjects or using different modalities, we need to find a suitable transformation $y : \Omega \to \mathbb{R}^2$ such that transformed version of the template image f is "similar" to the reference image g. This can be formulated as an optimization problem where we minimize the following functional

$$\mathcal{D}ist(f[y], g) + \alpha\mathcal{S}(y),$$

where '$\mathcal{D}ist$' denotes a suitable distance measure, $f[y]$ denotes the template image which is transformed through the transformation y, and the regularization term, $\mathcal{S}$, measures the reasonability of the transformation. One of the issues with image registration is that the registration may be time consuming, or can get stuck in local minima. To avoid this, multilevel or multiscale[14,15] approaches are employed. In this section, an application of the multiscale TV flow is demonstrated in the context of hierarchical image registration, where we register larger scales in the images to be registered, and then use the registration parameters thus obtained as an initial guess for the finer scale images, thus obtaining a robust and efficient registration. For this demonstration we use rigid registration, i.e. the transformation $y(x) \equiv (y_1(x), y_2(x))$ for $x \equiv (x_1, x_2)$ is given by

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \end{bmatrix},$$

where θ is the angle of rotation, d_1 and d_2 represent translations in x_1 and x_2 directions respectively. We denote the parameters for the rigid registration as a vector $\omega \equiv \langle\theta, d_1, d_2\rangle$. We used mutual information[16,17] as the similarity measure. We did not use any explicit regularizer for the experiments.

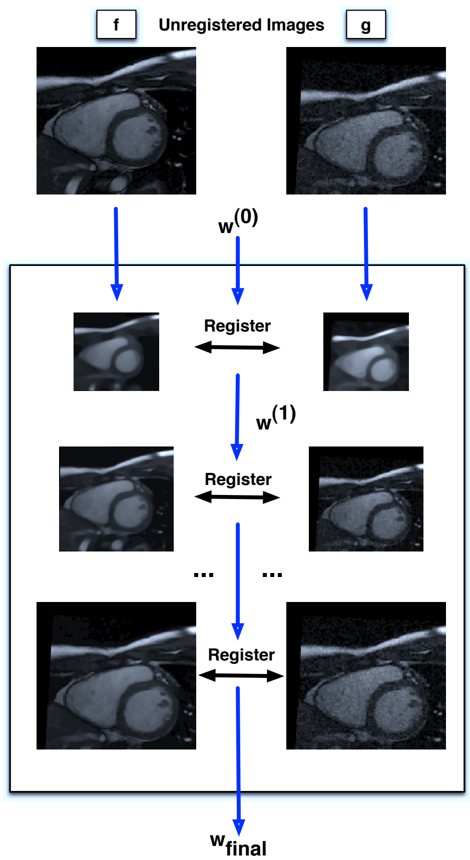


Figure 3. Multiscale TV Flow Registration

3.1 Method

The registration method is illustrated in Figure 3, where images f and g and the initial guess of the registration parameter vector $\omega^{(0)} = \langle \theta^{(0)}, d_1^{(0)}, d_2^{(0)} \rangle$ are supplied as inputs to the algorithm.

We first register the large scale features. To this effect, TV flow is applied to both input images to produce coarse scale images $u^f(\cdot, t_0)$ and $u^g(\cdot, t_0)$ respectively, using a large value for t_0. This is followed by an integer downsampling[§] operator D, for some integer N_0. Then registration between the downsampled images $D(u^f(\cdot, t_0), N_0)$ and $D(u^g(\cdot, t_0), N_0)$ is performed to obtain the registration parameters $\langle \theta^{(1)}, d_1^{(1)}, d_2^{(1)} \rangle$.

This registration and downsampling procedure is then iterated with an updated initial guess for the registration parameters for each iteration. For the $i + 1^{st}$ iteration step we choose $t_{i+1} < t_i$, as smaller stopping times correspond to finer scales. The integer factors for downsampling are chosen so that $N_{i+1} < N_i$. We choose dyadic downsampling i.e. $N_{i+1} = N_i/2$. Under this scheme if the optimal transformation for the i^{th} step is $\omega^{(i)} = \langle \theta^{(i)}, d_1^{(i)}, d_2^{(i)} \rangle$, then the initial guess for the $i + 1^{st}$ step is chosen as $\left\langle \theta^{(i)}, 2d_1^{(i)}, 2d_2^{(i)} \right\rangle$. Finally, the registration parameter vector ω_{final}, obtained with the full scale images (i.e. with no downsampling), is the estimated optimal transformation.

3.2 Experimental results

In this study, registration experiments were performed with a short axis MR image of a human heart. To test the robustness of the multiscale TV flow algorithm, we take an MR image f, and add Rician noise $R \sim \text{Rice}(\nu, \sigma)$ to produce a noisy image g. Then, we apply a known rigid transformation: $\hat{\omega} = \langle \theta, d_1, d_2 \rangle$ to the image g, where θ is the angle of rotation and $(d_1,\, d_2)$ represent translations in x_1-and x_2-directions respectively. We define the ground truth transformation to be $\hat{\omega} = \langle 3, 10, 15 \rangle$. Then we perform multiscale TVregistration between f and g as described in section 3, using the similarity metric mutual information (MI)[17] and the simplex optimization algorithm.[18] We define the registration to be successful if $\|\omega_{final} - \hat{\omega}\|_{l^\infty} < 1$.

Registration results for these experiments are shown in Table 2. The first column is the parameter σ in the Rician distribution, which represents the amount of added noise. The second and the third columns show the number of successful registrations out of 10 trials under varying noise parameter σ with and without denoising using multiscale TV flow. Finally, the difference in the number of successful registrations is shown in the last column. The advantage of the proposed method is demonstrated, as we observed 32% more success with hierarchical registration using multiscale TV flow, compared to registration without the TV flow.

Table 2. Registration Results

Rician Noise (σ^2)	Registration Success with TV flow	Registration Success without TV flow	Improvement
0.001	9/10	8/10	1/10
0.002	6/10	3/10	3/10
0.003	9/10	5/10	4/10
0.004	7/10	4/10	3/10
0.005	7/10	1/10	6/10
0.010	4/10	2/10	2/10
Mean	7/10	3.8/10	3.2/10

4. CONCLUSION

In this work, we proposed a novel multiscale TV flow (5), which can be used for fast image denoising. The speed of the denoising can be controlled by selecting an apropriate speed functions. This multiscale denoising is useful for images with small-scale noise.

[§]The integer downsampling of image h with integer N is defined as $D(h(x), N) \stackrel{\text{def}}{=} h(Nx)$.

The TV flow is also known to have edge preserving property like the standard TV flow.[5–7] We can use this feature to generate multiscale images without blurring main edges. This proves to be of great value for hierarchical image registration of noisy images. We see in section 3 that hierarchical registration using different scales is an efficient and robust registration method. This approach produces better registration results compared to registration without the TV flow. The algorithm was tested for rigid registration and its use for other types of registrations is subject to further investigation.

REFERENCES

[1] Tadmor, E., Nezzar, S., and Vese, L., "A multiscale image representation using hierarchical (BV, L^2) decompositions," *Multiscale Model. Simul.* **2**(4), 554–579 (2004).

[2] Tadmor, E., Nezzar, S., and Vese, L., "Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation," *Commun. Math. Sci.* **6**(2), 1–26 (2008).

[3] Athavale, P. and Tadmor, E., "Integro-differential equations based on (BV, L^1) image decomposition," *SIAM J. Imaging Sci.* **4**(1), 300–312 (2011).

[4] Tadmor, E. and Athavale, P., "Multiscale image representation using integro-differential equations," *Inverse problems and imaging* **3**, 693–710 (November 2009).

[5] Andreu, F., Ballester, C., Caselles, V., and Mazón, J., "Minimizing total variational flow," *Differential and integral equations* **14**, 321–360 (March 2001).

[6] Andreu, F., Caselles, V., Diaz, J., and Mazón, J., "Some qualitative properties for the total variational flow," *Journal of functional analysis* **188**, 516–547 (2002).

[7] Andreu-Vaillo, F., Caselles, V., and Mazón, J., [*Parabolic quasilinear equations minimizing linear growth Functionals*], vol. 223 of *Progress in mathematics*, Birkhäuser, Switzerland (2004).

[8] Tikhonov, A. and Arsenin, V., [*Solutions of ill-posed problems*], V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York (1977). Translated from the Russian, Preface by translation editor Fritz John, Scripta Series in Mathematics.

[9] Koenderink, J., "Structure of images," *Biological Cybernetics* **50**, 363–370 (1984).

[10] Rudin, L., Osher, S., and Fatemi, E., "Nonlinear total variation based noise removal algorithms," *Phys. D* **60**, 259–268 (1992).

[11] Meyer, Y., [*Oscillating patterns in image processing and nonlinear evolution equations*], vol. 22 of *University Lecture Series*, American Mathematical Society, Providence, RI (2001). The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[12] Trees, H. L. V., [*Detection, Estimation, and Modulation Theory, Volume 1*], vol. 1, John Wiley & Sons, New York (2001).

[13] Attouch, H., Buttazzo, G., and Michaille, G., [*Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization*], MPS-SIAM series on optimization, SIAM, Philadelphia, PA (2005).

[14] Modersitzki, J., [*Numerical Methods for Image Registration*], Oxford University Press, New York (2004).

[15] Paquin, D., Levy, D., Schreibmann, E., and Xing, L., "Multiscale image registration," *Mathematical biosciences and engineering* **3**, 389–418 (April 2006).

[16] Collignon, A., Vandermeulen, P., Suetens, P., and Marchal, G., "Automated multi-modality medical image registration based on information theory," *Computational Imaging and Vision* **3**, 263–274 (1995).

[17] Viola, P. and Wells, W., "Alignment by maximization of mutual information," *Int. J. Comput. Vis.* **24**(2), 137–154 (1997).

[18] Nelder, J. and Mead, R., "A simplex method for function minimization," *The Computer Journal* **7**, 308–313 (1965).