GENERALIZED NEWTON METHODS FOR ENERGY FORMULATIONS IN IMAGE PROCESSING

Leah Bar and Guillermo Sapiro

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, USA

ABSTRACT

Many problems in image processing are solved via the minimization of a cost functional. The most widely used optimization technique is the gradient descent, often used due to its simplicity and applicability where other optimization techniques, e.g., those coming from discrete optimization, can not be used. Yet, gradient descent suffers from a slow convergence, and often to just local minima which highly depends on the condition number of the functional Hessian. Newton-type methods, on the other hand, are known to have a rapid (quadratic) convergence. In its classical form, the Newton method relies on the L^2-type norm to define the descent direction. In this paper, we generalize and reformulate this very important optimization method by introducing a novel Newton method based on general norms. This generalization opens up new possibilities in the extraction of the Newton step, including benefits such as mathematical stability and smoothness constraints. We first present the derivation of the modified Newton step in the calculus of variation framework. Then we demonstrate the method with two common objective functionals: variational image deblurring and geodesic active contours. We show that in addition to the fast convergence, different selections norm yield different and superior results.

Index Terms— Newton method, Variational methods

1. INTRODUCTION

Optimization of a cost functional is a fundamental task in variational image analysis, where most of the optimization techniques are based on gradient flows. In the iterative gradient descent method, the step or search direction is the negative gradient. The definition of the gradient relies on an inner product structure, in most studies the L^2-type inner product is implicitly assumed.

Recently, generalized gradients approaches were introduced in image analysis by defining different inner product types[1]. Sundaramoorthi *et al.*, [1], reformulated the generic geometric active contour model by redefining the gradients with Sobolev-type inner products. As a result, significant improvement in region-based and edge-based segmentation was accomplished. Charpiat *et al.*, [2], derived the general gradient descent process associated with a symmetric positive linear operator which defines a new inner product. They demonstrated that the choice of the inner product can be seen as a prior on the deformation fields in shape warping and tracking applications. Later on, Eckstein *et al.*, [3], showed the importance of the norm selection in the context of shape matching.

The major weakness of the gradient descent methods is that despite its simplicity, the convergence rate can be very poor, many iterations are needed to achieve a local minima. Alternatively, it is well known in optimization theory that *Newton* methods are much faster, with a quadratic convergence [4]. The Newton step direction is basically the minimizer of the second-order approximation of the cost function. In the case of a real function $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, the step is denoted by $d_N = -\nabla^2 f(x) \nabla f(x)$. If the time step is $\Delta t = 1$, the method is a *pure* Newton, while *damped* Newton refers to the case where Δt is selected via line search process.

In the calculus of variation framework, the derivative is replaced by the first Gâteaux variation and the Hessian is replaced by the second Gâteaux variation. Few variational image processing studies had used the Newton method for optimization. Hintermüller and Ring, [5], solved the segmentation of grey scale image by the minimization of the Mumford-Shah functional. Zhang and Hancock, [6], developed an edge-preserving filter for smoothing images whose features reside on a curved manifold. Both works use standard L^2 norms.

In this paper we derive a generalized Newton method based on a general norm in the calculus of variations framework.[2] We begin by presenting the quadratic approximation of the functional, and continue with the mathematical derivation of the generalized Newton step. Numerical simulations demonstrate the performance of the algorithm for image segmentation and deblurring. We show that although the classical Newton method is very efficient, the results are quite poor. By using different norms in the generalized method, promising results are obtained with the advantage of high convergence rate. Furthermore, given a highly noisy image, the segmentation procedure turns to fail with the classical

Work partially supported by ONR, NSF, NGA, NIH, ARO, and DARA.

[1]There is also a rich optimization literature with generalized metrics and manifolds, here omitted due to space limitations.

[2]For mathematical details and additional examples see [7].

gradient descent method as pointed out in [1, 2]. Choosing an appropriate norm in the suggested algorithm alleviates the sensitivity problem and yields improved results at faster convergence rates.

2. MATHEMATICAL DERIVATION

The second order Taylor expansion of the common cost functional ($\Omega \in \mathbb{R}^n, f \in C^1(\Omega)$)

$$\mathcal{F}(f) := \int_\Omega I(f(x), \nabla f(x))\, dx,$$

motivates the Newton's method. Let $\hat{f}$ be the minimizer of this functional. The quadratic approximation to the variation $\mathcal{F}(\hat{f}+\psi)$ is given by

$$\mathcal{Q}(\psi) := \mathcal{F}(\hat{f}) + < \nabla\mathcal{F}(\hat{f}) \mid \psi > + \frac{1}{2} < \mathcal{H}_{\hat{f}}\, \psi \mid \psi >, \quad (1)$$

where $\mathcal{H}_{\hat{f}}$ denotes the Hessian of the functional at $\hat{f}$, and $\psi : \Omega \to \mathbb{R}$ stands for the variation. The first variation is expressed as

$$< \nabla\mathcal{F}(\hat{f}) \mid \psi >= \int_\Omega \left(I_f \psi + \sum_{i=1}^{n} I_{f_{x_i}} \psi_{x_i} \right) dx,$$

and the second variation is

$$< \mathcal{H}_f\, \psi \mid \psi >= \int_\Omega \left(\sum_{i,j=1}^{n} I_{f_{x_i x_j}} \psi_{x_i} \psi_{x_j} + 2 \sum_{i=1}^{n} I_{f f_{x_i}} \psi \psi_{x_i} + I_{ff} \psi^2 \right) dx. \quad (2)$$

In the case of $n = 2\,(x_1 = x, x_2 = y)$, the integrand of (2) can be expressed in a quadratic form

$$\begin{pmatrix} \psi & \psi_x & \psi_y \end{pmatrix} \begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix} \begin{pmatrix} \psi \\ \psi_x \\ \psi_y \end{pmatrix} \quad (3)$$

where $a = I_{ff}$, $b = I_{f_x f_x}$, $c = I_{f_y f_y}$, $d = I_{f f_x}$, $e = I_{f f_y}$, and $f = I_{f_x f_y}$.

Consider an abstract infinite dimensional Euclidean space - a vector space endowed with an inner product (symmetric bilinear positive-definite form) such that [2] (see also [1])

$$< u \mid v >_{\mathcal{L}} = < \mathcal{L}u \mid v >,$$

where $\mathcal{L} : L^2 \to L^2$ is a symmetric positive definite linear operator. Thus, (1) can be rewritten as

$$m(\psi) := \mathcal{F}(\hat{f}) + < \nabla\mathcal{F}(\hat{f}) \mid \psi >_{\mathcal{L}} + < \mathcal{H}_{\hat{f}}\, \psi \mid \psi >_{\mathcal{L}}\,. \quad (4)$$

Minimization of $m(\psi)$ with respect to ψ is carried out using the first Gâteaux derivative i.e.

$$\frac{\partial}{\partial \varepsilon}\, m(\psi + \varepsilon\eta)\bigg|_{\varepsilon=0} = 0,$$

which yields the following integral equation

$$\begin{aligned} < \nabla\mathcal{F}(\hat{f}) \mid \eta >_{\mathcal{L}} + \frac{1}{2} < \mathcal{H}_{\hat{f}}\, \psi \mid \eta >_{\mathcal{L}} \\ + \frac{1}{2} < \psi \mid \mathcal{H}_{\hat{f}}\, \eta >_{\mathcal{L}} = 0. \end{aligned} \quad (5)$$

Using the quadratic form (3) yields

$$\begin{aligned} &< \mathcal{H}_{\hat{f}}\, \psi \mid \eta >_{\mathcal{L}} = < \mathcal{L}\left(\mathcal{H}_{\hat{f}}\, \psi \right) \mid \eta > \\ &= \int_\Omega \mathcal{L}\left(a\psi + d\psi_x + e\psi_y\right)\eta + \mathcal{L}\left(d\psi + b\psi_x + f\psi_y\right)\eta_x \\ &+ \mathcal{L}\left(e\psi + f\psi_x + c\psi_y\right)\eta_y\, dx. \end{aligned} \quad (6)$$

In the same fashion,

$$\begin{aligned} &< \psi \mid \mathcal{H}_{\hat{f}}\, \eta >_{\mathcal{L}} = < \mathcal{L}(\psi) \mid \mathcal{H}_{\hat{f}} \psi > \\ &= \int \left(a\eta + d\eta_x + e\eta_y\right)\mathcal{L}(\psi) + \left(d\eta + b\eta_x + f\eta_y\right)\mathcal{L}(\psi_x) \\ &+ \left(e\eta + f\eta_x + c\eta_y\right)\mathcal{L}(\psi_y)\, dx. \end{aligned} \quad (7)$$

By substituting (6) and (7) into (5), and using integration by parts and the fundamental lemma of calculus of variations, we end up with the following partial differential equation

$$\begin{aligned} &\mathcal{L}(a\psi + d\psi_x + e\psi_y) - \partial_x\left[\mathcal{L}(d\psi + b\psi_x + f\psi_y)\right] \\ &- \partial_y\left[\mathcal{L}(e\psi + f\psi_x + c\psi_y)\right] \\ &+ a\mathcal{L}(\psi) + d\mathcal{L}(\psi_x) + e\mathcal{L}(\psi_y) \\ &- \partial_x\left[d\mathcal{L}(\psi) + b\mathcal{L}(\psi_x) + f\mathcal{L}(\psi_y)\right] \\ &- \partial_y\left[e\mathcal{L}(\psi) + f\mathcal{L}(\psi_x) + c\mathcal{L}(\psi_y)\right] \\ &= -\mathcal{L}\left(I_f - \partial_x(I_{f_x}) - \partial_y(I_{f_y})\right). \end{aligned} \quad (8)$$

The iterative generalized damped-Newton algorithm is thereby:

1. `Initialize` $f = f_1$
2. `Solve` (8) `and get the direction` ψ_n
3. `Choose` Δt `by standard backtracking line search.`
4. `Set` $f_{n+1} = f_n - \Delta t \psi_n$
5. `If` $(||f^{n+1} - f^n||_{L_2} < \varepsilon ||f^n||_{L_2})$ `stop.`

3. GEODESIC ACTIVE CONTOUR

As a first example, we address the classical geodesic active contour problem. The segmenting contour is the minimizer of a cost functional which consists of a region-based fidelity term, [8], and edge-based geodesic active contour [9].

Let u denotes the observed image. The contour is implicitly represented by the zero level set of a function ϕ. Then

$$\begin{aligned} \mathcal{F}_1(\phi, c_1, c_2) := \int_\Omega (u - c_1)^2 H(\phi) + (u - c_2)^2 (1 - H(\phi)) \\ + g\left(|\nabla u|\right)\, \delta(\phi)\, |\nabla(\phi)|\, dx, \end{aligned} \quad (9)$$

where $g(|\nabla u|) = \frac{1}{1+|\nabla u|^2/\lambda}$, $\lambda \in \mathbb{R}^+$, $c_1, c_2 \in \mathbb{R}$, and $H(\cdot)$ is the heaviside function. Following Chan and Vese, [8], the heaviside function is approximated as[3]

$$H_\epsilon(x) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan\left(\frac{x}{\epsilon}\right)\right),$$

and

$$\delta_\epsilon(x) = \partial_x H_\epsilon(x) = \frac{1}{\pi}\frac{\epsilon}{\epsilon^2 + x^2}.$$

The first variation is

$$< \nabla\mathcal{F}_1 \mid \psi >= \int_\Omega \delta(\phi)\left((u-c_1)^2 - (u-c_2)^2 - \nabla\cdot\left(g\frac{\nabla\phi}{|\nabla\phi|}\right)\right)\psi\, dx, \tag{10}$$

and the second variation matrix of the quadratic form is

$$\begin{pmatrix} \delta''(\phi)(\mathcal{P} + g|\nabla\phi|) & \frac{g\delta'(\phi)\phi_x}{|\nabla\phi|} & \frac{g\delta'(\phi)\phi_y}{|\nabla\phi|} \\ \frac{g\delta'(\phi)\phi_x}{|\nabla\phi|} & \frac{g\delta(\phi)\phi_y^2}{|\nabla\phi|^{3/2}} & \frac{-g\delta(\phi)\phi_x\phi_y}{|\nabla\phi|^{3/2}} \\ \frac{g\delta'(\phi)\phi_y}{|\nabla\phi|} & \frac{-g\delta(\phi)\phi_x\phi_y}{|\nabla\phi|^{3/2}} & \frac{g\delta(\phi)\phi_x^2}{|\nabla\phi|^{3/2}} \end{pmatrix}, \tag{11}$$

where $\mathcal{P} = (u-c_1)^2 + (u-c_2)^2$.

In this example we tested the case of a highly noisy image (Fig. 1), where Gaussian noise of 18.2 dB SNR was added to the original image. Using the standard L^2 inner product of the first and second variations results in a noisy level set function in both the classical gradient descent method (first row) and the Newton method (second row). This is due to the high (noisy) gradients caused by the noise, and the fact that the geodesic active contour functional is minimized along prominent gradients. Better results are obtained using the Sobolev gradient descent [1] (third row). Convergence with $\int |H(\phi_n) - H(\phi_{n+1}|)\, dx < 10$ was achieved after 59 iterations. Using the suggested generalized Newton method, the new inner product was selected such that the symmetric positive definite operator $\mathcal{L}_1$ is a convolution with a Gaussian kernel G_σ,

$$< u \mid v >_{\mathcal{L}_1} = < G_\sigma * u \mid v > .$$

This operation smoothes the incremental level set function in the generalized Newton method and eliminates high perturbations (fourth row). Here we used the truncated conjugate gradient method to solve Eq. (8), with 10 iterations and one single Newton step. The result looks better even when compared to the Sobolev gradient descent method (note the right leg and the left arm), and has the additional advantage of computational efficiency (3 times faster, see [7] for details).

[3] In the current experiment ϵ was set to 0.01.

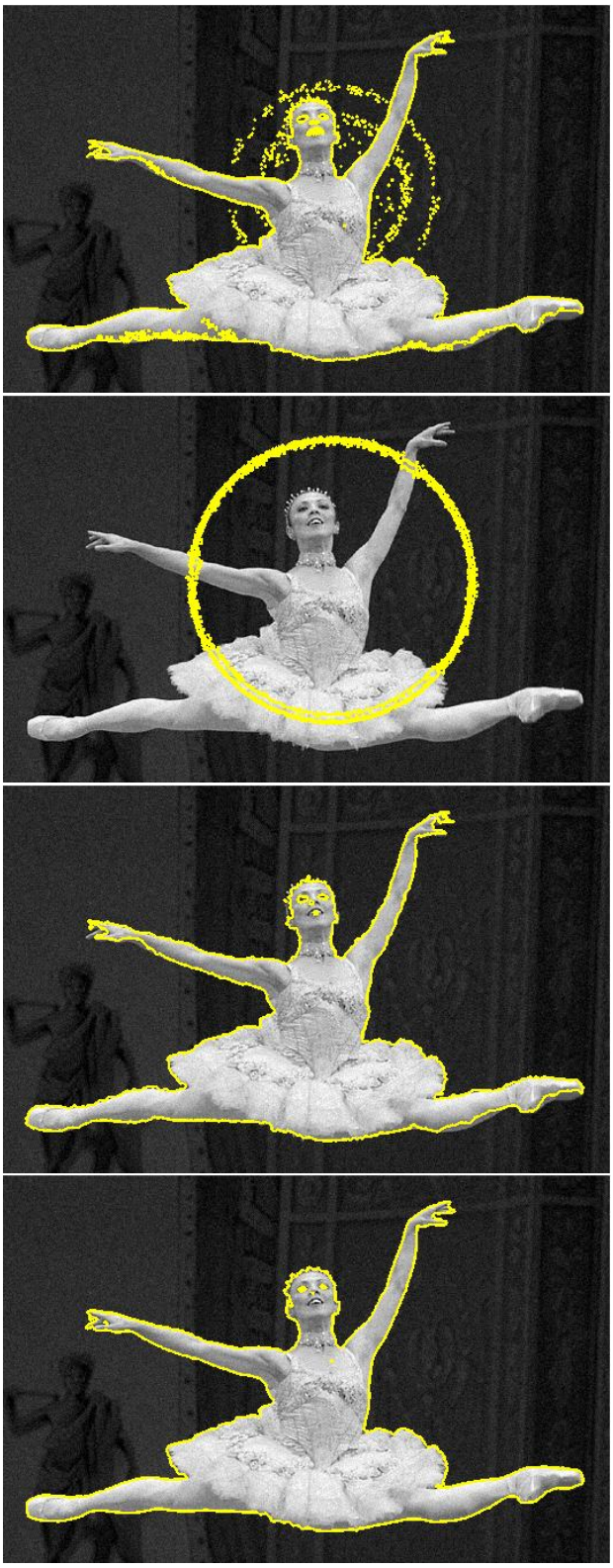

Fig. 1. Segmentation by geodesic active contour. *First row*: gradient descent. *Second row*: standard Newton method. *Third row*: Sobolev gradient descent [1]. *Fourth row*: generalized Newton with a smoothing operator.

4. IMAGE DEBLURRING

In the second example we look at the total variation deblurring method [10]. The observed (blurred) image is denoted by g, h is the (known) blur kernel, and f is the (unknown) clean image:

$$\mathcal{F}_2(f) := \frac{1}{2}\int_\Omega (h * f - g)^2 dx + \alpha \int_\Omega |\nabla f|\, dx. \qquad (12)$$

The first variation is given by

$$< \mathcal{F}_2 \mid \psi >= \int_\Omega \left[(h * f - g) * h(-x) - \alpha \nabla \cdot \left(\frac{\nabla f}{|\nabla f|} \right) \right] \psi \, dx,$$

and the second variation matrix is

$$\begin{pmatrix} h(x) * h(-x) * & 0 & 0 \\ 0 & \alpha \frac{f_y^2}{|\nabla f|^{3/2}} & \alpha \frac{-f_x f_y}{|\nabla f|^{3/2}} \\ 0 & \alpha \frac{-f_x f_y}{|\nabla f|^{3/2}} & \alpha \frac{f_x^2}{|\nabla f|^{3/2}} \end{pmatrix}. \qquad (13)$$

In this experiment the original *Einstein* image was blurred by a pill-box kernel of radius 2. The blurred image is shown in Fig. 2, top left. The recovered image using the classical Newton method is shown top right. The artifacts due to the sensitive inverse operation (the collar and left shoulder) can be easily seen. The bottom left figure is the outcome of the generalized Newton method with the $\mathcal{L}_1$ smoothing operator. While this operator gave good results in the previous active contour example, poor results are obtained in the deblurring case. It can be explained by the fact that smoothing the incremental image f prevents the desired sharpening operation. Better results are accomplished using the Sobolev operator such that [1, 2]

$$< u \mid v >_{H^1} = \int_\Omega u(x) \cdot v(x) dx + \lambda \int_\Omega D_x u(x) \cdot D_x v(x) dx,$$

and $\mathcal{L}_{H^1}(u) = u - \lambda \nabla^2 u$. The suggested recovered image is shown bottom right, where **18** internal loops in the conjugated gradients stage and two Newton loops with tolerance of $\varepsilon = 0.006$ were used. Similar results are obtained using the classical gradient descent with **70** iterations.

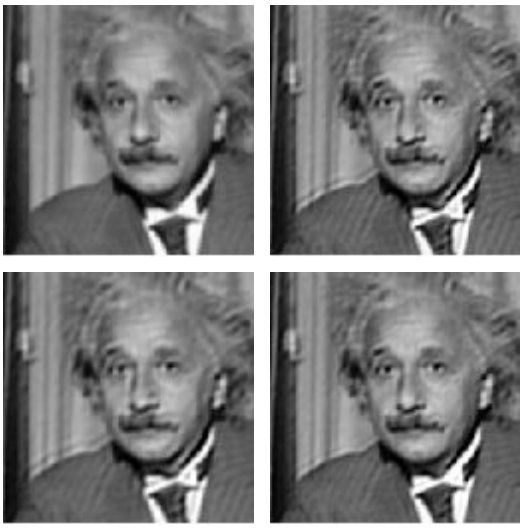

Fig. 2. Total Variation deblurring. *Top-left*: blurred image. *Top-right*: recovered image using the classical Newton method. *Bottom-left*: recovered image using generalized Newton method with $\mathcal{L}_1$ smoothing operator. *Bottom-right*: recovered image using generalized Newton method with $\mathcal{L}_{H^1}$ Sobolev operator.

5. CONCLUSIONS

We have extended the Newton method by using different inner products in the variational framework, following [1, 2]. Experimental results show the advantage of the method in computational efficiency and noisy data performance. Our future research includes incorporating advanced optimization tools to compensate for non-convex cases, e.g., trust region [7, 11].

6. REFERENCES

[1] G. Sundaramoorthi, A. Yezzi, and A. C. Mennucci, "Sobolev active contours," *Int. J. Comput. Vision*, vol. 73, no. 3, pp. 345–366, 2007.

[2] G. Charpiat, P. Maurel, J.-P. Pons, R. Keriven, and O. Faugeras, "Generalized gradients: Priors on minimization flows," *Int. J. Comput. Vision*, vol. 73, no. 3, pp. 325–344, 2007.

[3] I. Eckstein, J.-P. Pons, Y. Tong, C.-C. J. Kuo, and M. Desbrun, "Generalized surface flows for mesh processing," in *Proc. of Symposium on Geometry Processing*, 2007, pp. 183–192.

[4] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

[5] M. Hintermüller and W. Ring ., "An inexact Newton-CG-type active contour approach for the minimization of the Mumford-Shah functional," *Journal of Mathematical Imaging and Vision*, vol. 20, pp. 19–42, 2004.

[6] F. Zhang and E. R. Hancock, "A Riemannian weighted filter for edge-sensitive image smoothing," in *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, Washington, DC, USA, 2006, pp. 590–593, IEEE Computer Society.

[7] L. Bar and G. Sapiro, "Generalized Newton method for energy formulations in image processing," *IMA Preprint Series 2195, University of Minnesota*, http://www.ima.umn.edu/preprints/apr2008/2195.pdf.

[8] T. F. Chan and L. Vese, "Active contour without edges," *IEEE Trans. Image Processing*, vol. 10, no. 2, pp. 266–277, February 2001.

[9] G. Sapiro, *Geometric partial differential equations and image analysis*, Cambridge University Press, Cambridge, United Kingdom, 2001.

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1-4, pp. 259–268, 1992.

[11] P.-A. Absil, C. G. Baker, and K. A. Gallivan, "Trust-region methods on Riemannian manifolds," *Found. Comput. Math.*, vol. 7, no. 3, pp. 303–330, July 2007.