Projected Gradient Flows for BV / Level Set Relaxation

Martin Burger[*1] and **Michael Hintermüller** [**2]

[1] Institut für Industriemathematik, Johannes Kepler Universität, Altenbergerstr. 69, A 4040 Linz, Austria.
[2] Institut für Mathematik, Karl-Franzens Universität, Heinrichstr. 36, A-8010 Graz, Austria.

This paper introduces a new level set method based on projected gradient flows for problems that can be solved by a recently introduced relaxation approach. For the class of problems the relaxation is exact, it can be shown that the solution of the flow converges to a solution of the relaxed problem for large time, and the level sets of the limit are solutions of the original problem. We introduce a simple computational scheme based on explicit time discretization and apply the method to imaging examples.

1 Introduction

The solution of optimization problems with unknown topology is a challenging task appearing in various applications ranging from image processing (cf. e.g. [12] for an overview) over inverse problems (cf. e.g. [4]) to structural optimization (cf. e.g. [2]). In the last years, significant progress in the development of computational methods for these problems has been made, using the level set approach and shape sensitivity analysis as basic ingredients (cf. [4, 9, 12] and the references therein).

Level set based optimization methods can handle rather general structures and even changes of topology as opposed to classical shape design approaches. However, already from their construction it is clear that these techniques are of a local nature and might converge to local rather than global minimizers for the usually highly nonconvex problem. An example of such a local minimum is well-known in applications to image segmentation, where standard level set approaches fail to obtain inner contours in most cases. In general the computation of global minimizers is open and might even be impossible unless special assumptions on the problem are made. As shown in [6] (see also [5] for a related problem) such a global minimization is possible for an interesting special class, namely optimization problems of the form

$$\int_\Omega g\; u \;\mathrm{d}x + J(u) \quad \rightarrow \quad \min_{u \in BV(\Omega;\{0,1\})}, \tag{1}$$

where $g \in L^1(\Omega)$, $BV(\Omega;\{0,1\})$ denotes the subset of functions u of bounded variation such that $u(x) \in \{0,1\}$ for almost every $x \in \Omega$, and the functional J denotes the total variation seminorm, formally given as $\int_\Omega |\nabla u|\; dx$. The main idea of the global minimization is to consider the convex relaxation

$$J_g(u) := \int_\Omega g\; u \;\mathrm{d}x + J(u) \quad \rightarrow \quad \min_{u \in BV(\Omega;[0,1])}, \tag{2}$$

$BV(\Omega;[0,1])$ denotes the subset of functions u of bounded variation such that $0 \leq u(x) \leq 1$ for almost every $x \in \Omega$. Roughly speaking, the main reason for the exactness of the relaxation is that the non-coercive objective functional will not prevent bang-bang type controls caused by the constraint $0 \leq u \leq 1$. Besides the exactness of the relaxation one can also show that almost every level set of a minimizer of (2) is a solution of the original topology optimization problem. The proof is based on the co-area formula for functions of bounded variation and will be detailed in Section 2.

Once the exactness of the relaxation and the relation of minimizers for (1) and (2) is understood, the remaining open problem amounts to solve the convex problem (2). In [6], the minimization was carried out via a gradient flow for a penalized unconstrained functional, which gave reasonable results but still enforces the tuning of penalization parameters. In this paper we introduce a projected gradient flow, whose convergence to the solution of (2) can be shown and which is straight-forward to implement. The method is then applied to (suitable reformulations of) some imaging problems.

2 BV / Level Set Relaxation

In this section, we briefly recall the basic results on the exactness of the convex relaxation (2) following the approach in [11, 6]. For this sake assume that $\overline{u}$ is a minimizer of (1) and $\hat{u}$ is a minimizer of (2). For $0 < t < 1$ define (in an almost everywhere sense)

$$u^t(x) = \begin{cases} 1 & \text{if } \hat{u}(x) > t \\ 0 & \text{else} \end{cases}$$

* e-mail: martin.burger@jku.at, Phone: +43 732 2468 9224, Fax: +43 732 2468 9224
** e-mail: michael.hintermueller@uni-graz.at, Phone: +43 316 380 8546, Fax: +43 316 380 9815

Then, with a simple transformation of variables we obtain

$$\int_\Omega g\ \hat{u}\ dx = \int_\Omega \int_0^{\hat{u}(x)} g(x)\ dt\ dx = \int_0^1 \int_\Omega g\ u^t\ dx\ dt.$$

Moreover, the co-area formula for functions of bounded variation implies

$$J(\hat{u}) = \int_0^1 \text{Per}(\{u > t\})\ dt = \int_0^1 J(u^t)\ dt.$$

By combining the two identities we conclude

$$J_g(\hat{u}) = \int_0^1 J_g(u^t)\ dt. \tag{3}$$

Since $u^t \in BV(\Omega; \{0,1\})$, we clearly have $J_g(u^t) \geq J(\overline{u})$ for all t. On the other hand, the inequality $J_g(\overline{u}) \geq J_g(\hat{u})$ holds due to the relaxation. Hence, we have

$$J_g(\hat{u}) = \int_0^1 J_g(u^t)\ dt \geq \int_0^1 J_g(\overline{u})\ dt = J_g(\overline{u}) \geq J_g(\hat{u}). \tag{4}$$

From (4) we can conclude that $J_g(\hat{u}) = J_g(\overline{u}) = J_g(u^t)$, i.e., the relaxation is exact and for each minimizer $\hat{u}$ of the relaxed problem (2) we can compute minimizers of (1) by taking level sets (respectively the indicator functions u^t of the level sets of $\hat{u}$).

The benefit of this BV / level set relaxation is the possibility to solve a convex minimization problem with bound constraints instead of a problem with non-convex constraints, which corresponds to an infinite-dimensional version of a combinatorial optimization problem, and to determine minimizers of the original problem by a simple procedure. Due to the convexity of the relaxed problem, one can also be sure that each minimizer is a global one, so we compute global minimizers of the original non-convex problem.

3 A Projected Gradient Flow

In the following we construct a simple projected gradient flow for the solution of the relaxed problem (2). The gradient flow for the functional J_g is a *total variation flow* with source g, given (in a formal way) as

$$\frac{\partial u}{\partial t} = \text{div}\left(\frac{\nabla u}{|\nabla u|}\right) - g.$$

Since we are looking for a solution with $u \in [0,1]$, we introduce a formal version of the projected gradient flow

$$\frac{\partial u}{\partial t} = \begin{cases} \max\left\{\text{div}\left(\frac{\nabla u}{|\nabla u|}\right) - g, 0\right\} & \text{if } u = 0, \\ \min\left\{\text{div}\left(\frac{\nabla u}{|\nabla u|}\right) - g, 0\right\} & \text{if } u = 1, \\ \text{div}\left(\frac{\nabla u}{|\nabla u|}\right) - g & \text{else.} \end{cases} \tag{5}$$

In this way, we have that $\frac{\partial u}{\partial t} \leq 0$ if $u = 1$ and $\frac{\partial u}{\partial t} \geq 0$ if $u = 0$, and hence, the solution must satisfy $0 \leq u \leq 1$ for all $t \geq 0$.

A mathematically sound definition of a solution can be given by a variational inequality similar to the one derived for the total variation flow by Feng and Prohl [7]. We look for a solution in

$$u \in K := \{\varphi \in L^1(0,T;BV) \cap C(0,T;L^2) \mid 0 \leq \varphi \leq 1 \text{ a.e. }\},$$

such that

$$\int_0^s \int_\Omega \frac{\partial v}{\partial t}(v-u)\ dx\ dt + \int_0^s \left[J_g(v) - J_g(u)\right]\ dt \geq \frac{1}{2}\int_\Omega (u-v)^2\ dx\Big|_0^s \qquad \forall\, v \in K \cap H^1(0,T;BV). \tag{6}$$

The existence of such a solution $u \in K$ can be proved in a standard way by considering the time-discrete version on a time grid $\{t_k = k\tau\}_{j=0,\dots,N}$ with $\tau = \frac{t}{N}$. An approximate solution can be defined as

$$u^N(t) = u^N(t_{k+1}) \qquad \text{for } t \in (t_k, t_{k+1}],$$

where $u^N(t_{k+1})$ is defined iteratively via

$$u^N(t_{k+1}) = \arg\min_{u \in BV(\Omega;[0,1])} \left[J_g(u) + \frac{1}{2\tau} \int_\Omega (u - u^N(t_k))^2 \, dx \right]. \tag{7}$$

This sequential optimization problems provide a-priori bounds for $u(t)$, by standard weak convergence and lower semiconuity techniques one can construct a convergent subsequence of u^N and verify that the variational inequality holds for its limit. Thus, we obtain (cf. [1, Chapter 3] for a detailed proof in a more general setup)

Proposition 3.1 *There exists a weak solution $u \in K$ of the projected gradient flow according to the definition* (6).

Besides the existence of a solution to the flow, we are interested in its long time behaviour, since the motivation of constructing the flow is to obtain convergence to a solution of the relaxed problem:

Theorem 3.2 *Let $u \in K$ be a weak solution of* (6)*. Then there exists a subsequence $u(t_k)$ converging in the weak-* topology of BV, and each accumulation point (in the weak-* topology) of $(u(t))_{t \in \mathbb{R}_+}$ is a solution of* (2).

P r o o f. Let $\hat{u}$ be a minimizer of (2), then $v = \hat{u}$ is an admissible test function in (6), since $\frac{\partial v}{\partial t} = 0$. Hence, we obtain

$$\int_0^s [J_g(\hat{u}) - J_g(u)] \, dt \geq \frac{1}{2} \int_\Omega (u - \hat{u})^2 \, dx \Big|_0^s ,$$

which implies in particular that $\int_0^s [J_g(u) - J_g(\hat{u})] \, dt$ and $\int_\Omega (u - \hat{u})^2 \, dx$ are uniformly bounded from above with respect to s. Moreover, we have $J_g(u) \geq J_g(\hat{u})$ due to the definition of the minimizer and hence, there exists a sequence $t_k \to \infty$ such that $J_g(u(t_k)) \to J_g(\hat{u})$. On the other one can show by standard means that $J_g(u(t))$ is nonincreasing with respect to t and hence, $J_g(u(t)) \to J_g(\hat{u})$ From the uniform boundedness we can extract a weak-* convergent subsequence (again denoted by $u(t_k)$). From the lower semicontinuity of J_g we deduce that for each for each limit $\tilde{u} \in BV(\Omega;[0,1])$, the inequality $J_g(\tilde{u}) \leq J_g(\hat{u})$ holds, and thus, $\tilde{u}$ is a minimizer of (2). □

In order to construct numerical approximations to the projected gradient flow, one could either use an implicit time discretization based on the sequential optimization problems, or an explicit time discretization based on the formal version (5). An explicit time discretization is rather easy to perform. Assume that $D_{\Delta x}(u)$ is a discretization of div $\left(\frac{\nabla u}{|\nabla u|}\right)$ on a grid of size Δx, then we compute a time step (of size Δt) as

$$v(t + \Delta t) = u(t) + \Delta t \, D_{\Delta x}(u(t)) - g, \qquad u(t + \Delta t) = \mathcal{P}_{[0,1]}(v(t + \Delta t)),$$

where $\mathcal{P}_{[0,1]}$ is the pointwise projection operator to the interval $[0,1]$. As usual for explicit time discretizations of second order parabolic equations, we have to choose the time step as $\Delta t = \mathcal{O}((\Delta x)^2)$ in order to obtain stability.

Note that the projected gradient flow can be interpreted as a level set method, the right-hand side of the flow corresponds to the mean curvature of level sets of u. Moreover, the level sets of long time limits are solution of the original problem and it therefore seems reasonable to expect some kind of convergence of the level sets of $u(t)$, which will be confirmed by the numerical tests below.

4 Applications

In the following we discuss the application of the approach to two prominent model problems in image processing, namely the denoising of binary images via the ROF-model and the segmentation of images via the Mumford-Shah model.

4.1 ROF-Model for Binary Images

The classical ROF-model (cf. [10]) for denoising blocky images f is given by the minimization of the functional

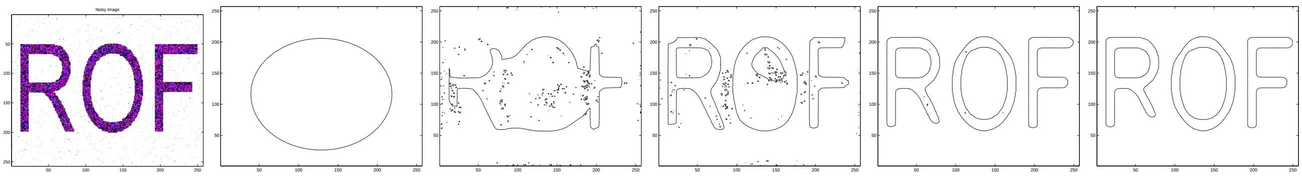
$$J_\lambda(u) = \lambda \int_\Omega (u - f)^2 \, dx + J(u)$$

for some positive parameter λ to be chosen in dependence on the noise in the image. If one looks for binary images, one minimizes J_λ over $u \in BV(\Omega;\{0,1\})$. Since $u \in \{0,1\}$ almost everywhere, we obtain

$$(u - f)^2 = u^2 - 2uf + f^2 = u - 2uf + f^2.$$

For the minimization we can ignore the constant term, and hence the ROF-model for binary images is equivalent to (1) with $g = \lambda(1 - 2f)$. Hence, we can directly apply the relaxation and the corresponding gradient flow for the ROF model.

In order to illustrate the behaviour for the ROF model we perform the gradient flow on a binary image with 256×256 pixels, corrupted by Gaussian noise with mean zero and variance 10%, with parameter $\lambda = 20$. The figure below shows the noisy image (left), and the level sets $\partial\{u > 0.5\}$ for the initial guess and the time steps 100, 200, 1000, and 1500 (from left to right). One observes that the noise is eliminated successfully, with the payoff of losing a small part of the letter R, but the letters are still clearly recognizable.



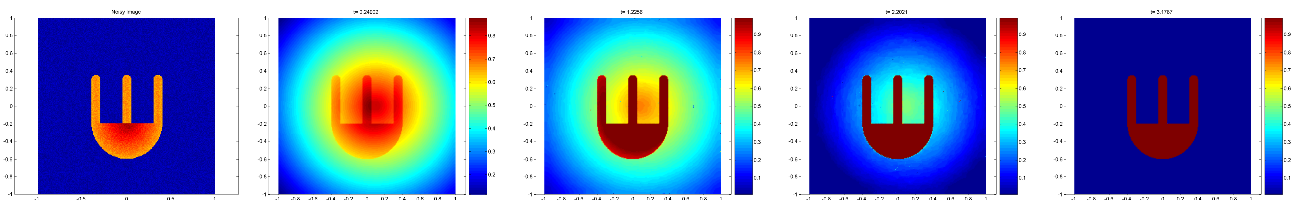
4.2 Mumford-Shah Segmentation

The most frequently used approach for image segmentation is the so-called Mumford-Shah model (cf. [8]), which adds penalties on the smoothness of the image in the segmented regions and on the length of interfaces between segmented regions. In the case of segmentation to binary images, the Mumford-Shah model can be written as

$$J_\lambda(u) = \lambda \int_\Omega \left[u(c_1 - f)^2 + (1-u)(c_2 - f)^2 \right] \, dx + J(u)$$

to be minimized over $u \in BV(\Omega; \{0,1\})$ and $(c_1, c_2) \in \mathbb{R}^2$. Note that for fixed u the minimizer $(c_1(u), c_2(u))$ can be computed explicitely. On the other hand, for fixed (c_1, c_2), the minimization over u is equivalent to (1) with $g = \lambda(c_1 - f)^2 - \lambda(c_2 - f)^2$. Hence, it seems natural to perform an alternate minimization with respect to (c_1, c_2) and with respect to u. Instead of an exact minimization of u in each step, we only perform a few time steps of the projected gradient flow, which yields a computationally efficient method.

As an example for the application of the method we investigated the segmentation of an image with 256×256 pixels. For the segmentation we use a noisy image (not binary, but with a significant discontinuity) perturbed by additive random noise with mean zero and variance of 10%. The initial value for the projected gradient flow was the indicator function of a large ball, the parameter λ was chosen as 10^3. The figure below shows the noisy image (first on the left) and subsequently the reconstructed level set functions u at different time steps of the flow. One observes that intermediate values are encountered during the flow, but the image becomes again binary for large time.



Acknowledgements The work of M.B. has been supported by the Austrian national science foundation FWF through project SFB F 013 / 08.

References

[1] L.Ambrosio, N.Gigli, G.Savare, *Gradient Flows* (Birkhäuser, Basel, 2005).

[2] M.P.Bendsøe, O.Sigmund, *Topology Optimization* (Springer, Berlin, 2002).

[3] M.Burger, M.Hintermüller, *Relaxation of shape and topology optimization problems*, Preprint (2005).

[4] M.Burger, S.Osher, *A survey on level set methods for inverse problems and optimal design* , European J. Appl. Math. (2005), to appear

[5] T.F.Chan, S.Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math. (2005), to appear.

[6] T.F.Chan, S.Esedoglu, M.Nikolova, *Algorithms for finding global minimizers of denoising and segmentation models*, CAM Report 04-54 (UCLA, 2004)

[7] X.Feng, A.Prohl, *Analysis of total variation flow and its finite element approximations*, ESAIM: Math. Mod. Numer. Anal. **37** (2003), 533-556.

[8] D.Mumford, J.Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Commun. Pure Appl. Math. **42** (1989), 577-685.

[9] S.J.Osher, R.P.Fedkiw, *The Level Set Method and Dynamic Implicit Surfaces* (Springer, New York, 2002).

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

[11] G.Strang, *Maximal flow through a domain*, Math. Program. **26** (1983), 123-143.

[12] R.Tsai, S.Osher, *Level set methods and their applications in image science*, Comm. Math. Sciences **1** (2003), 623 - 656.