CONVEX RELAXATIONS FOR A GENERALIZATION OF MUMFORD-SHAH

ANDREW FISCHER-GARBUTT

1. Introduction

We can represent a grey scale image by a function $f : \Omega = [0,1]^n \to [0,1]$. In 1998 Mumford and Shah introduced a variational model for image segmentation. The model segments a given image f by minimizing the functional

$$MS(u) = \lambda \int_\Omega (u-f)^2 dx + \int_{\Omega \setminus S_u} |\nabla u|^2 dx + \mu \text{Length}(S_u). \tag{1}$$

Where S_u denotes the set of points where u has a jump discontinuity. For more about the Mumford Shah model see [3].

1.1. **The Convex Relaxation.**

In this section we will follow the argument in [1]. In order to construct our convex relaxation we start by considering a vector field

$$\phi = (\phi^x, \phi^t),$$

where $\phi^x : \mathbb{R}^{d+1} \to \mathbb{R}^d$ and $\phi^t : \mathbb{R}^{d+1} \to \mathbb{R}$. In addition, for a given function u, we also introduce the set

$$\Sigma_u = \{(x,t) : x \in \mathbb{R}^d, t \in \mathbb{R}, t \leq u(x)\} \tag{2}$$

and another set Γ_u denoting the boundary of Σ_u. Now we will start by considering the integral

$$\int_\Omega v \nabla \cdot \phi dx \tag{3}$$

where v is taken to be the characteristic function of Σ_u. This allows us to write

$$\int_\Omega v \nabla \cdot \phi dx = \int_{\Sigma_v} \nabla \cdot \phi dx. \tag{4}$$

Then we can apply the divergence theorem to the right hand side of the above equation to get

$$\int_\Omega v \nabla \cdot \phi dx = \int_{\Gamma_v} \nu_{\Gamma_v} \cdot \phi ds,$$

were ν_{Γ_u} denotes the outward normal to Γ_u. Now we can decompose this integral into parts to get

$$\int_\Omega v\nabla\cdot\phi dx = \int_{\Omega\setminus S_v} \frac{(\nabla v,-1)}{\sqrt{1+|\nabla v|^2}}\cdot(\phi_x,\phi_t)\sqrt{1+|\nabla v|^2}dx \tag{5}$$

$$+\int_{S_v}\int_{v_-(x)}^{v_+(x)}(\nu_{\Gamma_v},0)\cdot(\phi^x,\phi^t)dtds, \tag{6}$$

were v_+ and v_- denote the values of u at the top and bottom of the jump at x.

Now we can expand the dot products to get

$$\int_\Omega v\nabla\cdot\phi dx = \int_{\Omega\setminus S_v}(\nabla v)\cdot(\phi_x)-\phi_t dx+\int_{S_v}\int_{v_-(x)}^{v_+(x)}\nu_{\Gamma_u}\cdot\phi^x dtds. \tag{7}$$

We can use the right hand side of the above equation to write down a saddle point problem that is equivalent to minimizing the original Mumford Shah. The problem is

$$\inf_v \sup_\phi \int_\Omega v\nabla\cdot\phi dx \tag{8}$$

subject to the constraints:

(1) $v\in\{0,1\}$,
(2) $\phi^t \geq \frac{1}{4}|\phi_x|^2-\lambda(f-t)^2$,
(3) v is decreasing in the t-direction,
(4) $\left|\int_{t_1}^{t_2}\phi^x(x,s)ds\right| \leq \mu, \forall t_1,t_2$.

Now in order to make this problem convex we need to relax the constraint that $v\in\{0,1\}$. We do this by replacing it with the constraint that $0\leq v\leq 1$.

1.2. **Proposed Modification.** In our proposed modification we think of f and u as representing the orientations of the grains in a material, instead of the intensities in a grey scale image.

Our proposed modification is to change the Mumford-Shah by replacing the term that penalizes the length of the singular set with a term that instead penalizes a sub-additive function g of the jump height. When we do this the Mumford-Shah changes into

$$\lambda\int_\Omega(u-f)^2dx+\int_{\Omega\setminus S_u}|\nabla u|^2dx+\mu\int_{S_u}g(u_+(x)-u_-(x))ds.$$

were $u_+(x)$ denotes the value of u at the top of the jump and $u_-(x)$ denotes the value of u at the bottom of the jump at x.

The goal of this proposed modification is to help with the problem of determining the boundaries of grains in a material. For this reason we will consider the function $h : [0, \infty) \to \mathbb{R}$ given by

$$h(\theta) = \begin{cases} \frac{\theta}{\theta^*}(1 - \log(\frac{\theta}{\theta^*})) & \theta \le \theta^* \\ 1 & \theta \ge \theta^* \end{cases}$$

where θ^* is a parameter that is known to be between $10°$ and $30°$. We will take g to be a periodic extension of h. We can define g by taking $g(x) = h(|x|)$ on some interval $[-L, L]$ and then extending g periodically to a function defined on the entire real line.

1.3. **Sub-Additivity.** In order for a unique minimizer to exist g must be sub-additive. To prove that g is sub-additive we'll start by proving that h is sub-additive. We will do this using several cases. Firstly if $\theta_1, \theta_2 \le \theta^*$ and $\theta_1 + \theta_2 \le \theta^*$, then we have that

$$h(\theta_1 + \theta_2) = \frac{\theta_1 + \theta_2}{\theta^*}(1 - \log(\frac{\theta_1 + \theta_2}{\theta^*})) \tag{9}$$

$$= \frac{\theta_1}{\theta^*}(1 - \log(\frac{\theta_1 + \theta_2}{\theta})) + \frac{\theta_2}{\theta^*}(1 - \log(\frac{\theta_1 + \theta_2}{\theta^*})). \tag{10}$$

Now since $-\log(x)$ is a decreasing function we have that

$$h(\theta_1 + \theta_2) = \frac{\theta_1}{\theta^*}(1 - \log(\frac{\theta_1 + \theta_2}{\theta})) + \frac{\theta_2}{\theta^*}(1 - \log(\frac{\theta_1 + \theta_2}{\theta^*})) \tag{11}$$

$$\le \frac{\theta_1}{\theta^*}(1 - \log(\frac{\theta_1}{\theta})) + \frac{\theta_2}{\theta^*}(1 - \log(\frac{\theta_2}{\theta^*})) \tag{12}$$

$$= h(\theta_1) + h(\theta_2). \tag{13}$$

If $\theta_1 + \theta_2 \ge \theta^*$ but θ_1 and θ_2 are both less than or equal to θ^*, then we have that

$$h(\theta_1 + \theta_2) = 1$$

and

$$h(\theta_1) + h(\theta_2) = \frac{\theta_1}{\theta^*}(1 - \log(\frac{\theta_1}{\theta^*})) + \frac{\theta_2}{\theta^*}(1 - \log(\frac{\theta_2}{\theta^*})).$$

Now since θ_1 and θ_2 are both less than or equal to θ^* we have that $-\log(\frac{\theta_1}{\theta^*}) \ge 0$ and $-\log(\frac{\theta_2}{\theta^*}) \ge 0$. So therefore it follows that

$$h(\theta_1) + h(\theta_2) = \frac{\theta_1}{\theta^*}(1 - \log(\frac{\theta_1}{\theta^*})) + \frac{\theta_2}{\theta^*}(1 - \log(\frac{\theta_2}{\theta^*})) \tag{14}$$

$$\ge \frac{\theta_1}{\theta^*} + \frac{\theta_2}{\theta^2}. \tag{15}$$

Therefore since $\theta_1 + \theta_2 \ge \theta_2$ we have that $h(\theta_1) + h(\theta_2) \ge 1 = h(\theta_1 + \theta_2)$.

Finally if $\theta_1 \ge \theta^*$ or $\theta_2 \ge \theta^*$ we have that $h(\theta_1 + \theta_2) = 1$, and since one of θ_1 and θ_2 is greater than or equal to θ^* and because $h(\theta) \ge 0$, we have $h(\theta_1) + h(\theta_2) \ge 1$. This is sufficient to show that h is sub-additive.

Now we need to show that g is sub-additive. Without loss of generality we can consider $\theta_1, \theta_2 \in [-L, L]$, then we have that

$$g(\theta_1 + \theta_2) \le h(|\theta_1 + \theta_2|). \tag{16}$$

Using the triangle inequality and the fact that h is an increasing function, we have that

$$g(\theta_1 + \theta_2) \leq h(|\theta_1| + |\theta_2|). \tag{17}$$

Finally by sub-additivity of h we have that

$$\begin{aligned} g(\theta_1 + \theta_2) &\leq h(|\theta_1|) + h(|\theta_2|) && (18)\\ &= g(\theta_1) + g(\theta_2). && (19) \end{aligned}$$

So therefore g is sub-additive as desired.

2. Computing the Projection

2.1. **The Strategy.** We can think of the projecting onto our constraints as a projection onto an intersection of convex sets. To project onto the intersection of convex sets we can use the Boyle Dykstra aligorithm. All that we need know to use this algorithm is how to project onto each of the convex sets. For a complete description of this algorithm see [3].

2.2. **First Constraint for 2D image.** The first constraint we will deal with is the constraint that

$$\left| \int_{t_1}^{t_2} \phi^x(x, s) ds \right| \leq \mu, \forall t_1, t_2.$$

Because we are working on a discretized grid we can represent this problem by projecting a point

$$(X_1, X_2, \cdots, X_n, Y_1, \cdots, Y_n) \notin K_1$$

onto the set

$$K_1 = \{(x_1, \cdots, x_n, y_1, \cdots, y_n) \colon (x_1 + x_2 + \cdots + x_n)^2 + (y_1 + y_2 + \cdots y_n)^2 \leq \mu\}.$$

We can recast this problem as an optimization problem

$$\min_{(x_1+x_2+\cdots+x_n)^2+(y_1+y_2+\cdots+y_n)^2=\mu} \sum_{j=1}^{n} (x_j - X_j)^2 + (y_j - Y_j)^2.$$

Now using Lagrange multipliers we can write down the system of nonlinear equations

$$x_j - X_j = \lambda(x_1 + \cdots + x_n) \tag{20}$$

$$y_j - Y_j = \lambda(y_1 + \cdots + y_n) \tag{21}$$

$$(x_1 + x_2 + \cdots + x_n)^2 + (y_1 + y_2 + \cdots y_n)^2 = \mu. \tag{22}$$

Equations (1) and (2) imply that

$$x_1 + x_2 + \cdots + x_n - X_1 - X_2 - \cdots - X_n = n\lambda(x_1 + x_2 + \cdots + x_n) \tag{23}$$

$$y_1 + y_2 + \cdots + y_n - Y_1 - Y_2 - \cdots - Y_n = n\lambda(y_1 + y_2 + \cdots + y_n) \tag{24}$$

This implies that

$$x_1 + x_2 + \cdots + x_n = -\frac{\alpha\mu}{n\lambda - 1}$$

and

$$y_1 + y_2 + \cdots + y_n = -\frac{\beta\mu}{n\lambda - 1}.$$

Now this allows us to rewrite (3) as

$$\frac{\alpha^2\mu^2}{(n\lambda - 1)^2} + \frac{\beta^2\mu^2}{(n\lambda - 1)^2} = \mu$$

Then we can solve this equation for λ to get

$$\lambda = \frac{\pm\sqrt{\mu(\alpha^2 + \beta^2)} + 1}{n}.$$

Now if we plug in $x_1 + x_2 + \cdots + x_n$ and $y_1 + y_2 + \cdots + y_n$ then we have

$$x_j = -\frac{\lambda\alpha\mu}{n\lambda - 1} + X_j$$
$$y_j = -\frac{\lambda\beta\mu}{n\lambda - 1} + Y_j.$$

Now substituting in our expression for λ we get

$$x_j = \mp\alpha\mu\left(\frac{\pm\sqrt{\mu(\alpha^2 + \beta^2)} + 1}{n}\right)\frac{1}{\sqrt{\mu(\alpha^2 + \beta^2)}} + X_j \tag{25}$$

$$= \mp\frac{\alpha\mu}{n}\left(\pm 1 + \frac{1}{\sqrt{\mu(\alpha^2 + \beta^2)}}\right) + X_j \tag{26}$$

$$y_j = \mp\frac{\beta\mu}{n}\left(\pm 1 + \frac{1}{\sqrt{\mu(\alpha^2 + \beta^2)}}\right) + Y_j. \tag{27}$$

Now we can use these expressions for x_j and y_j to evaluate the energy to get

$$\frac{\mu^2(\alpha^2 + \beta^2)}{n}\left(\pm 1 + \frac{1}{\sqrt{\mu(\alpha^2 + \beta^2)}}\right)^2.$$

Since we are trying to minimize the energy this suggests that we should use the minus branch. So we have that

$$\lambda = \frac{1 - \sqrt{\mu(\alpha^2 + \beta^2)}}{n}.$$

2.3. **First Constraint for 1D image.** In the case of a 1 dimensional image we can also pose the problem of projecting a given point $(X_1, X_2, \cdots, X_n)$ onto our constrait as an optimization problem. However in this case our optimization problem takes the simplier form

$$\min_{x_1 + x_2 + \cdots + x_n = \sqrt{\mu}} \sum_{j=1}^{n} (x_j - X_j)^2.$$

Now we can write our the Lagrange multipliers for this problem as

$$2(x_j - X_j) = \lambda \tag{28}$$

$$x_1 + x_2 + \cdots + x_n = \pm\sqrt{\mu}. \tag{29}$$

Now if we add the constraints together and then apply our constraint we end up with

$$\sqrt{\mu} - X_1 - X_2 - \cdots - X_n = \frac{n\lambda}{2}.$$

So we have that

$$\lambda = \frac{2}{n}(\pm\sqrt{\mu} - \alpha)$$

where $\alpha = X_1 + X_2 + \cdots + X_n$. Now this allows us to solve for x_j to get

$$x_j = \frac{1}{n}(\pm\sqrt{\mu} - \alpha) + X_j.$$

If we put this into our formula for the energy we get

$$\frac{1}{n}(\pm\sqrt{\mu} - \alpha)^2$$

we see the energy is minimized when we choose $\pm\sqrt{\mu}$ to have the opposite sign as α.

2.4. **Second Constraint.** We also need to project onto the constraint

$$\phi^t \geq \frac{1}{4}|\phi_x|^2 - \lambda(f-t)^2.$$

Since in practice we work with a discrete grid projecting onto this constraint is equivalent to projecting each of our discrete points onto a parabola.

Now to project onto the second constraint we need to project onto a parabola in the case of a one dimensional image or a quadratic cone in the case of a two dimensional image. In particular the 1-dimensional case reduces to the problem of projecting a point (X,T) onto the curve

$$t = \frac{1}{4}x^2 - \lambda(f-s)^2. \tag{30}$$

Now the normal to our parabola at a point (x,t) is

$$N = (\frac{1}{2}x, -1)$$

Then we can express our problem of finding the projection in a geometric way as the equation

$$(x, \frac{1}{4}x^2 - \lambda(f-s)^2) + \kappa(\frac{1}{2}x, -1) = (X,T).$$

Now this gives us the system

$$x + \frac{1}{2}\kappa x = X \tag{31}$$

$$\frac{1}{4}x^2 - \lambda(f-s)^2 - \kappa = T. \tag{32}$$

Now we can solve the second equation for κ and then plug this into the first equation to get

$$x + \frac{1}{2}x(\frac{1}{4}x^2 - \lambda(f-s)^2) - T) = X.$$

This simplifies to the depressed cubic

$$x^3 - 4(\lambda(f-s)^2 + T)x - 8X = 0. \tag{33}$$

3. Modified Mumford Shah constraints

If we instead want to minimize our modified version of Mumford Shah only one constraint changes.

$$\int_{S_\theta} (\int_{\theta^-(s)}^{\theta^+(s)} (v_{\Gamma_\theta}, 0) \cdot (\phi^x, \phi^t) dt) ds \tag{34}$$

$$= \mu \int_{S_\theta} g(\theta^+(s) - \theta^-(s)) ds \tag{35}$$

$$\Rightarrow | \int_{t_1}^{t_2} \phi^x(x,s) ds| \leq \mu g(t_2 - t_1), \forall t_1, t_2. \tag{36}$$

Now we need to show that this constraint is convex. Suppose ϕ_1^x, ϕ_2^x satisfy the constraint, and let $\kappa \in [0,1]$

$$\left| \int_{t_1}^{t_2} \kappa \phi_1^x(x,s) + (1-\kappa)\phi_2^x(x,s) ds \right| \leq \mu g(t_2 - t_1) \tag{37}$$

We have that $\forall t_1, t_2$

$$\left| \int_{t_1}^{t_2} \kappa \phi_1^x(x,s) + (1-\kappa)\phi_2^x(x,s) ds \right| \tag{38}$$

$$\leq |1-\kappa| \left| \int_{t_1}^{t_2} \phi_2^x(x,s) ds \right| + |\kappa| \left| \int_{t_1}^{t_2} \phi_1^x(x,s) ds \right| \tag{39}$$

$$\leq (1-\kappa)\mu g(t_2 - t_1) + \kappa g(t_2 - t_1) \tag{40}$$

$$= \mu g(t_2 - t_1). \tag{41}$$

4. Conclusion

In conclusion we can find a solution to our modified version of Mumford Shah by solving the saddle point problem

$$\inf_v \sup_\phi \int_\Omega v \nabla \cdot \phi dx \tag{42}$$

subject to the constraints:

(1) $0 \leq v \leq 1$,
(2) $\phi^t \geq \frac{1}{4}|\phi^x| - \lambda(f-t)^2$,
(3) v is decreasing in the t-direction,
(4) $\left| \int_{t_1}^{t_2} \phi^x(x,s) ds \right| \leq \mu g(t_2 - t_1), \forall t_1, t_2$.

In practical terms this means using a scheme that iteratively updates a discretized version of v and ϕ by performing a gradient descent for v, projecting v onto the constraints, preforming a gradient ascent for ϕ then projecting ϕ onto the relevant constraints. Projecting onto the various constraints can be accomplished using the Boyle and Dykstra algorithm. For a description of this algorithm see [2]. For information about algorithms for minimizing the Mumford-Shah functional also see [5] and [4].

So far we have written code to find a minimizer for Mumford Shah. In future research we would like to modify our code to find a minimizer for our modified version of Mumford Shah. Doing this should be possible with only a slight modification to the code we currently have.

References

1. Giovanni Alberti, Guy Bouchitté, and Gianni Dal Maso, *The calibration method for the Mumford-Shah functional and free-discontinuity problems*, Calc. Var. Partial Differential Equations **16** (2003), no. 3, 299–333. MR 2001706
2. James P. Boyle and Richard L. Dykstra, *A method for finding projections onto the intersection of convex sets in Hilbert spaces*, Advances in order restricted statistical inference (Iowa City, Iowa, 1985), Lect. Notes Stat., vol. 37, Springer, Berlin, 1986, pp. 28–47. MR 875647
3. David Mumford and Jayant Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math. **42** (1989), no. 5, 577–685. MR 997568
4. Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle, *An algorithm for minimizing the mumford-shah functional*, 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 1133–1140.
5. Evgeny Strekalovskiy, Antonin Chambolle, and Daniel Cremers, *A convex representation for the vectorial mumford-shah functional*, 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1712–1719.