The total variation of functions and applications in image analysis

Viktor Larsson

Supervised by Niels Christian Overgaard

May 9, 2012

Abstract

In this project the problem of image segmentation by minimizing energy functionals is considered. First the Rudin-Osher-Fatemi (ROF) model is introduced along with an example of an application in image denoising. Total Variation for a C^1 function is defined and then generalized to the non-differentiable case. Chambolle's algorithm for solving the ROF problem is then introduced. Finally the piecewise constant case for background/foreground image segmentation is considered by minimizing an energy functional. The energy functional approximates the well known segmentation model originally proposed by T. Chan and L. Vese. Two algorithms for performing the minimization is presented and compared.

1 Notation

Images can be considered functions $f : \Omega \to \mathbb{R}$ which maps a pixel $\boldsymbol{x} \in \Omega$ to the corresponding gray level $f(\boldsymbol{x})$. For functions in a continuous domain we write $f(\boldsymbol{x})$ where $\boldsymbol{x} = (x_1, x_2)$, while $f_{i,j}$ will be used when the domain is discrete.

Discrete images will be considered square $N \times N$ matrices. The set of discrete images, i.e. the set of square $N \times N$ matrices with real entries, is denoted by X. The set of vector fields in the discrete domain will be $Y = X \times X$.

For continuous domains we define the scalar products between functions u, v and between vector fields $\boldsymbol{p} = (p^1, p^2)$, $\boldsymbol{q} = (q^1, q^2)$ as

$$\langle u, v \rangle = \int_\Omega u(\boldsymbol{x})v(\boldsymbol{x})\; d\boldsymbol{x} \quad \text{and} \quad \langle \boldsymbol{p}, \boldsymbol{q} \rangle = \int_\Omega \boldsymbol{p} \cdot \boldsymbol{q}\; d\boldsymbol{x}$$

and for discrete domains

$$\langle u, v \rangle_X = \sum_{i,j} u_{i,j} v_{i,j} \quad \text{and} \quad \langle \boldsymbol{p}, \boldsymbol{q} \rangle_Y = \sum_{i,j} p^1_{i,j} q^1_{i,j} + p^2_{i,j} q^2_{i,j}$$

2 The Rudin-Osher-Fatemi (ROF) model

The Rudin-Osher-Fatemi model deals with the following problem: Given a function g, find a smooth function that approximates g in L_2 sense. To define what is meant by a function being smooth we define the Total Variation of a function $u \in C^1(\Omega, \mathbb{R})$ as

$$TV(u) = \int_\Omega |\nabla u(\boldsymbol{x})|\; d\boldsymbol{x} \tag{1}$$

where $|\cdot|$ is the usual Euclidean norm $|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2}$. The Total Variation is a semi-norm which measures how much the function varies on Ω.

To solve the problem Rudin, Osher and Fatemi [3] proposed minimizing the following functional

$$E_{ROF}(u) = TV(u) + \frac{1}{2\theta} ||u - g||_2^2$$

where $\theta > 0$ and $||\cdot||_2$ is the usual L_2 norm, $||f||_2^2 = \int_\Omega f(\boldsymbol{x})^2\; d\boldsymbol{x} = \langle f, f \rangle$.

The constant θ determines how important it is for u to approximate g well versus having low variation. When θ is large, the first term will dominate the expression and the minimizing u will have low variation. Similarly when θ is small the second term will dominate and then deviating from g in L_2 sense will be expensive.

2.1 Example: Image denoising

The ROF model can be used for removing noise from images. This particular application was studied by Rudin, Osher and Fatemi in the original paper [3].

Assume we are given an image g that is an observation of a *true* image u with zero-mean noise added. Assuming also that the *true* image has low variation then the image can be denoised by solving a ROF problem. In Figure 1 an example of an image being denoised by using this technique.

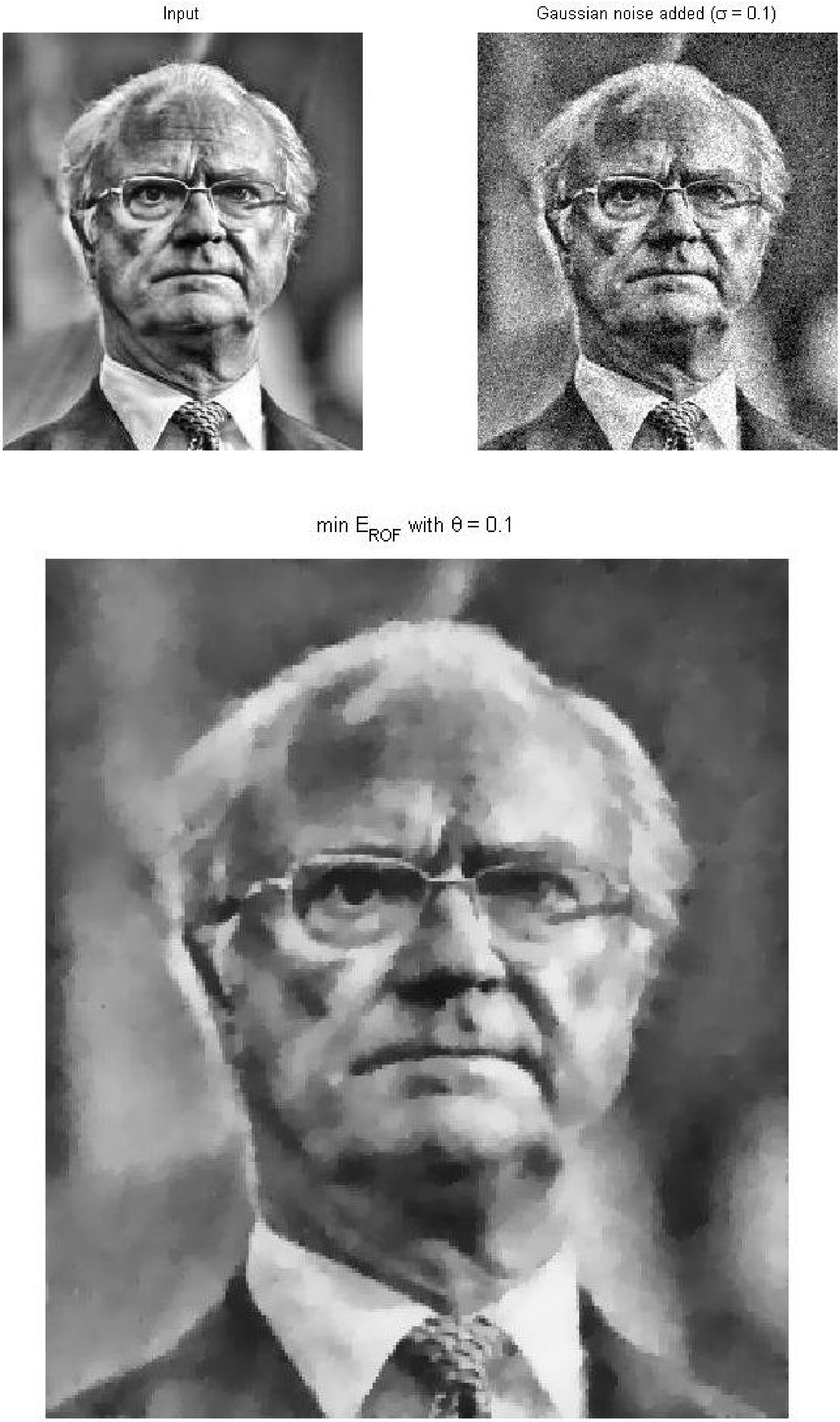

Figure 1: *Left:* Original image *Right:* Original image with added noise. *Bottom:* $\min_u E_{ROF}(u)$

2.2 An alternate definition of Total Variation

The definition of the Total Variation in (1) requires that the function is C^1. Since we might be interested in solutions that are not in C^1 we look at an alternate definition of the TV norm which does not have this requirement on u:

$$J(u) = \sup \left\{ \int_\Omega u(\boldsymbol{x}) \text{ div } \boldsymbol{p}(\boldsymbol{x}) \; d\boldsymbol{x} \; ; \; \boldsymbol{p} \in C_c^1(\Omega, \mathbb{R}^2), \; |\boldsymbol{p}(\boldsymbol{x})| \leq 1 \right\} \tag{2}$$

where $C_c^1(\Omega, \mathbb{R}^2)$ is the set of all continuously differentiable vector functions with compact support in Ω and $|\boldsymbol{p}(\boldsymbol{x})| \leq 1$ means that each individual vector $\boldsymbol{p}(\boldsymbol{x})$ has at most unit length.

This definition might seem odd at first but it can be shown that for C^1 functions the definitions are equivalent.

Let $u \in C^1(\Omega, \mathbb{R})$ and look at

$$\int_\Omega |\nabla u| \; d\boldsymbol{x} = \int_\Omega \nabla u \cdot \frac{\nabla u}{|\nabla u|} \; d\boldsymbol{x} = \sup_{|\boldsymbol{p}(\boldsymbol{x})| \leq 1} \int_\Omega \nabla u \cdot \boldsymbol{p} \; d\boldsymbol{x} \tag{3}$$

Note that in the last expression the supremum is taken over every vector function $\boldsymbol{p}$ (not necessarily in $C_c^1(\Omega, \mathbb{R}^2)$) that satisfies $|\boldsymbol{p}(\boldsymbol{x})| \leq 1$.

The last equality holds since the supremum is attained by letting $\boldsymbol{p}$ be of unit length and point in the direction of ∇u, i.e. $\boldsymbol{p} = \frac{\nabla u}{|\nabla u|}$. We will now show that the integral has the same value if we instead take supremum over the set $\{\boldsymbol{p} \in C_c^1(\Omega, \mathbb{R}^2) \; ; \; |\boldsymbol{p}(\boldsymbol{x})| \leq 1\}$.

Let

$$\boldsymbol{p}^* = \begin{cases} \frac{\nabla u}{|\nabla u|} & \text{if} \quad |\nabla u| \neq 0 \\ \boldsymbol{0} & \text{otherwise} \end{cases}$$

This choice of $\boldsymbol{p}$ attains the supremum but it is not in $C_c^1(\Omega, \mathbb{R}^2)$. It is however in $L_2(\Omega, \mathbb{R}^2)$, since it is a bounded function over a bounded domain.

It is known that $C_c^1(\Omega, \mathbb{R}^2)$ is a dense subset in $L_2(\Omega, \mathbb{R}^2)$. This means that for each $\epsilon > 0$ and for each $\boldsymbol{p} \in L_2(\Omega, \mathbb{R}^2)$ we can find a $\boldsymbol{q} \in C_c^1(\Omega, \mathbb{R}^2)$ such that $||\boldsymbol{p} - \boldsymbol{q}||_2 < \epsilon$. Furthermore it can be shown that this still holds if we add the constraint $|\boldsymbol{p}(\boldsymbol{x})| \leq 1$ to both sets.

Therefore, for any $\epsilon > 0$ we can find a $\boldsymbol{q} \in C_c^1(\Omega, \mathbb{R}^2)$ with $|\boldsymbol{q}(\boldsymbol{x})| \leq 1$ such that $||\boldsymbol{p}^* - \boldsymbol{q}|| < \epsilon$. Look at the difference between the integrals

$$\left| \int_\Omega \nabla u \cdot \boldsymbol{p}^* \; d\boldsymbol{x} - \int_\Omega \nabla u \cdot \boldsymbol{q} \; d\boldsymbol{x} \right| = \left| \langle \nabla u, \boldsymbol{p}^* \rangle - \langle \nabla u, \boldsymbol{q} \rangle \right| =$$

$$\left| \langle \nabla u, \boldsymbol{p}^* - \boldsymbol{q} \rangle \right| \overset{Cauchy-Schwarz}{\leq} ||\boldsymbol{p}^* - \boldsymbol{q}||_2 ||\nabla u||_2 \; < \epsilon \; ||\nabla u||_2$$

Since $||\nabla u||_2$ is constant the difference between the integrals can be made arbitrarily small.
It follows that

$$\sup_{|\boldsymbol{p}(\boldsymbol{x})| \leq 1} \int_\Omega \nabla u \cdot \boldsymbol{p} \; d\boldsymbol{x} = \sup \left\{ \int_\Omega \nabla u \cdot \boldsymbol{p} \; d\boldsymbol{x} \; ; \; \boldsymbol{p} \in C_c^1(\Omega; \mathbb{R}^2), \; |\boldsymbol{p}(\boldsymbol{x})| \leq 1 \right\} \tag{4}$$

Now lets look at Gauss' theorem

$$\int_\Omega \text{div } \boldsymbol{F} = \int_{\partial\Omega} \boldsymbol{F} \cdot \boldsymbol{n} \; dS$$

Let $\boldsymbol{F} = u\boldsymbol{p}$. Since $\boldsymbol{p}$ has compact support it follows that

$$\int_\Omega \text{div } u\boldsymbol{p} \; d\boldsymbol{x} = \int_{\partial\Omega} u\boldsymbol{p} \cdot \boldsymbol{n} \; dS = 0$$

Using the product rule for divergence (div $\alpha\boldsymbol{\beta} = \nabla\alpha \cdot \boldsymbol{\beta} + \alpha$ div $\boldsymbol{\beta}$) we get

$$0 = \int_\Omega \text{div } u\boldsymbol{p} \; d\boldsymbol{x} = \int_\Omega \nabla u \cdot \boldsymbol{p} + u \text{ div } \boldsymbol{p} \; d\boldsymbol{x} \quad \Leftrightarrow$$

$$\int_\Omega \nabla u \cdot \boldsymbol{p} \; d\boldsymbol{x} = -\int_\Omega u \text{ div } \boldsymbol{p} \; d\boldsymbol{x} \tag{5}$$

Substituting (5) into (3) together with (4) gives us

$$\begin{aligned}
\int_\Omega |\nabla u| \; d\boldsymbol{x} &= \sup_{|\boldsymbol{p}(\boldsymbol{x})|\leq 1} \int_\Omega \nabla u \cdot \boldsymbol{p} \; d\boldsymbol{x} = \\
&= \sup\left\{ \int_\Omega \nabla u \cdot \boldsymbol{p} \; d\boldsymbol{x} \; ; \; \boldsymbol{p} \in C_c^1(\Omega;\mathbb{R}^2), \; |\boldsymbol{p}(\boldsymbol{x})| \leq 1 \right\} = \\
&= \sup\left\{ -\int_\Omega u \text{ div } \boldsymbol{p} \; d\boldsymbol{x} \; ; \; \boldsymbol{p} \in C_c^1(\Omega;\mathbb{R}^2), \; |\boldsymbol{p}(\boldsymbol{x})| \leq 1 \right\} = \\
&= \sup\left\{ \int_\Omega u \text{ div } \boldsymbol{p} \; d\boldsymbol{x} \; ; \; \boldsymbol{p} \in C_c^1(\Omega;\mathbb{R}^2), \; |\boldsymbol{p}(\boldsymbol{x})| \leq 1 \right\} = J(u)
\end{aligned}$$

The last equality follows from the fact that the integral is linear in $\boldsymbol{p}$ so we can simply move the minus sign into the divergence and then replace $(-\boldsymbol{p})$ with $\boldsymbol{p}$. $\square$

2.2.1 Example: Total Variation of a function not in C^1

With the alternate definition there are less constraints on the function so we can now calculate the Total Variation on a larger class of functions. Now lets look at an example with a function which is not C^1. Let D be a closed subset in Ω with a boundary ∂D that is piecewise C^1 and define the set indicator function

$$\mathbb{1}_D(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in D \\ 0 & \text{if } \boldsymbol{x} \notin D \end{cases}$$

This function is clearly not in C^1 but using the alternate definition we can calculate the Total Variation.

$$J(\mathbb{1}_D) = \sup_{|\boldsymbol{p}(\boldsymbol{x})|\leq 1} \int_\Omega \mathbb{1}_D \text{ div } \boldsymbol{p} \; d\boldsymbol{x} = \sup_{|\boldsymbol{p}(\boldsymbol{x})|\leq 1} \int_D \text{div } \boldsymbol{p} \; d\boldsymbol{x} \overset{Gauss}{=} \sup_{|\boldsymbol{p}(\boldsymbol{x})|\leq 1} \int_{\partial D} \boldsymbol{p} \cdot \boldsymbol{n} \; dS$$

By similar reasoning as above, the integral reaches its supremum when $\boldsymbol{p} = \boldsymbol{n}$ on ∂D. This gives us

$$\sup_{|\boldsymbol{p}(\boldsymbol{x})|\leq 1} \int_{\partial D} \boldsymbol{p} \cdot \boldsymbol{n} \; dS = \int_{\partial D} \boldsymbol{n} \cdot \boldsymbol{n} \; dS = \int_{\partial D} dS = \text{Per}(D)$$

where $\text{Per}(D)$ denotes the length of the boundary of the set D.

2.3 Discretization

Since we want to work with discrete images we need a discretization of the problem. To be able to define what is meant by variation in the discrete domain we will need to define discrete gradient and divergence operators.

Since there are many ways to define these we must decide which properties we want them to fulfill. Drawing inspiration from (5) we decide that we want $\langle \nabla u, \boldsymbol{p} \rangle_Y = \langle u, -\text{div } \boldsymbol{p} \rangle_X$ to hold for every $u \in X$ and $\boldsymbol{p} \in Y$. This later turns out to give some nice properties when defining the discrete Total Variation.

Define the gradient operator $\nabla : X \to Y$ as $\nabla u = ((\nabla u)^1, (\nabla u)^2)$ where

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & 1 \leq i < N \\ u_{1,j} - u_{i,j} & i = N \end{cases} \quad \text{and} \quad (\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & 1 \leq j < N \\ u_{i,1} - u_{i,j} & j = N \end{cases}$$

and the divergence operator div $: Y \to X$ as

$$(\text{div }\boldsymbol{p}\ )_{i,j} = \begin{cases} p^1_{i,j} - p^1_{i-1,j} & 1 < i \le N \\ p^1_{i,j} - p^1_{N,j} & i = 1 \end{cases} \quad + \quad \begin{cases} p^2_{i,j} - p^2_{i,j-1} & 1 < j \le N \\ p^2_{i,j} - p^2_{i,N} & j = 1 \end{cases}$$

Note that these definitions of the operators are cyclic. This means that they wrap around when they reach the borders of the images. To simplify the calculations we let $u_{N+1,j}$ denote $u_{1,j}$ and similarly for the other borders.

A quick check verifies that the desired property holds

$$\langle \nabla u, \boldsymbol{p} \rangle_Y = \sum_{i,j} (\nabla u)_{i,j} \cdot \boldsymbol{p}_{i,j} = \sum_{i,j} (u_{i+1,j} - u_{i,j}) p^1_{i,j} + (u_{i,j+1} - u_{i,j}) p^2_{i,j} =$$

The summation can be reordered so that each term contains a single $u_{i,j}$

$$= \sum_{i,j} (p^1_{i-1,j} - p^1_{i,j} + p^2_{i,j-1} - p^2_{i,j})\ u_{i,j} = \sum_{i,j} -(p^1_{i,j} - p^1_{i-1,j} + p^2_{i,j} - p^2_{i,j-1})\ u_{i,j} = \langle u, -\text{div }\boldsymbol{p} \rangle_X$$

□

Discrete versions of the two definitions of Total Variation can then be defined as

$$TV(u) = \sum_{i,j} |\nabla u| \quad \text{and} \quad J(u) = \max_{|\boldsymbol{p}_{i,j}| \le 1} \sum_{i,j} u_{i,j}\ (\text{div }\boldsymbol{p})_{i,j} = \max_{|\boldsymbol{p}_{i,j}| \le 1} \langle u, \text{div }\boldsymbol{p} \rangle$$

Using definition of the discrete gradient and divergence we can show that for the discrete versions of the Total Variation norms are equivalent for all $u \in X$.

$$TV(u) = \sum_{i,j} |\nabla u| = \max_{|\boldsymbol{p}_{i,j}| \le 1} \sum_{i,j} (\nabla u)_{i,j} \cdot \boldsymbol{p}_{i,j} = \max_{|\boldsymbol{p}_{i,j}| \le 1} \langle \nabla u, \boldsymbol{p} \rangle_Y$$

$$= \max_{|\boldsymbol{p}_{i,j}| \le 1} \langle u, -\text{div }\boldsymbol{p} \rangle_X = \max_{|\boldsymbol{p}_{i,j}| \le 1} \langle u, \text{div }\boldsymbol{p} \rangle_X = J(u)$$

□

2.4 Chambolle's algorithm

In [1] Chambolle proposes an algorithm for solving the ROF problem

$$\min_u \left\{ J(u) + \frac{1}{2\theta} ||u - g||^2 \right\} = \min_u \max_{|\boldsymbol{p}_{i,j}| \le 1} \left\{ \langle u, \text{div }\boldsymbol{p} \rangle_X + \frac{1}{2\theta} ||u - g||^2 \right\}$$

for a given function $g \in X$ and $\theta > 0$.

We look at the expression

$$\langle u, \text{div }\boldsymbol{p} \rangle_X + \frac{1}{2\theta} ||u - g||^2$$

and simplify it by completing the square

$$\begin{aligned}
&\langle u, \text{div }\boldsymbol{p} \rangle_X + \frac{1}{2\theta} \left(||u||^2 - 2\langle u, g \rangle_X + ||g||^2 \right) = \\
&= \frac{1}{2\theta} \left[2\theta\ \langle u, \text{div }\boldsymbol{p} \rangle_X + ||u||^2 - 2\langle u, g \rangle_X + ||g||^2 \right] = \\
&= \frac{1}{2\theta} \left[(-2) \left(\langle u, g \rangle_X - \langle u, \theta \text{ div }\boldsymbol{p} \rangle_X \right) + ||u||^2 + ||g||^2 \right] = \\
&= \frac{1}{2\theta} \left[||u||^2 - 2\langle u, g - \theta \text{ div }\boldsymbol{p} \rangle_X + ||g||^2 \right] = \\
&= \frac{1}{2\theta} \left[||u - (g - \theta \text{ div }\boldsymbol{p})||^2 - ||\theta \text{ div }\boldsymbol{p} - g||^2 + ||g||^2 \right]
\end{aligned}$$

Now assume that we can reorder the min and max.

$$\max_{|\boldsymbol{p}_{i,j}|\leq 1} \min_u \frac{1}{2\theta}\left[||u - (g - \theta \text{ div } \boldsymbol{p})||^2 - ||\theta \text{ div } \boldsymbol{p} - g||^2 + ||g||^2 \right]$$

Then for a fix $\boldsymbol{p}$ we should find the u that minimizes the expression. Since u is only present in one term this is easy and u is chosen as

$$u^* = g - \theta \text{ div } \boldsymbol{p}$$

Now to find $\boldsymbol{p}$ we need to maximize the rest of the expression.

$$\begin{aligned}
\boldsymbol{p}^* &= \arg\max_{|\boldsymbol{p}_{i,j}|\leq 1} \left\{ \frac{1}{2\theta}||g||^2 - \frac{1}{2\theta}||\theta \text{ div } \boldsymbol{p} - g||^2 \right\} = \\
&= \arg\max_{|\boldsymbol{p}_{i,j}|\leq 1} \left\{ -\frac{1}{2\theta}||\theta \text{ div } \boldsymbol{p} - g||^2 \right\} = \\
&= \arg\min_{|\boldsymbol{p}_{i,j}|\leq 1} \left\{ ||\theta \text{ div } \boldsymbol{p} - g||^2 \right\}
\end{aligned}$$

If we define the set $\theta K = \{\theta \text{ div } \boldsymbol{p} \ ; \ \boldsymbol{p} \in Y, \ |\boldsymbol{p}_{i,j}| \leq 1\}$, which is a closed and convex set, then θ div $\boldsymbol{p}^*$ is the projection $\pi_{\theta K}(g)$ (see the appendix for definition of projection onto closed convex sets). So the solution to the original problem is

$$u^* = g - \pi_{\theta K}(g)$$

For finding the projection $\pi_{\theta K}(g)$ Chambolle proposes the following iterative scheme

$$\begin{aligned}
\boldsymbol{p}^0_{i,j} &= 0 \\
\boldsymbol{p}^{n+1}_{i,j} &= \frac{\boldsymbol{p}^n_{i,j} + \tau(\nabla(\text{div } \boldsymbol{p}^n - g/\theta\ ))_{i,j}}{1 + \tau|(\nabla(\text{div } \boldsymbol{p}^n - g/\theta\ ))_{i,j}|}
\end{aligned}$$

where $\tau > 0$ is the step length, controlling how far we go each step. In [1] Chambolle proves the following theorem.

Theorem: Let $\tau \leq 1/8$ and define $\boldsymbol{p}^n$ as above. Then

$$\theta \text{ div } \boldsymbol{p}^n \to \pi_{\theta K}(g) \quad \text{when} \quad n \to \infty$$

In the implementation the stopping critereon used for the iteration is

$$||\boldsymbol{p}^{n+1} - \boldsymbol{p}^n||_{\infty,2} = \max_{\boldsymbol{x}} |\boldsymbol{p}^{n+1}(\boldsymbol{x}) - \boldsymbol{p}^n(\boldsymbol{x})| < \epsilon$$

This stopping critereon was also used by Chambolle in [1].

3 Image segmentation

Image segmentation is the problem of dividing an image f into foreground and background. In this section we look at the special case where both the foreground and background have almost constant gray levels. For this problem T. Chan and L. Vese [4] proposed minimizing the following functional

$$E_{CV}(D, c_1, c_2) = \text{Per}(D) + \lambda_1 \int_D (f(\boldsymbol{x}) - c_1)^2 \ d\boldsymbol{x} + \lambda_2 \int_{\Omega/D} (f(\boldsymbol{x}) - c_2)^2 \ d\boldsymbol{x}$$

where $D \subset \Omega$ is the segmentation. The first term will keep the boundary short and the two following terms will make sure that f is approximately c_1 inside D and c_2 outside. This functional can be simplified by writing it in terms of set indicator functions $\mathbb{1}_D$. Letting $\lambda = \lambda_1 = \lambda_2$ and remembering that $\text{Per}(D) = TV(\mathbb{1}_D)$ (section 2.2.1) we get

$$E^2_{CV}(\mathbb{1}_D, c_1, c_2) = TV(\mathbb{1}_D) + \lambda \int_\Omega (f(\boldsymbol{x}) - c_1)^2 \mathbb{1}_D + (f(\boldsymbol{x}) - c_2)^2(1 - \mathbb{1}_D) \ d\boldsymbol{x} =$$

$$= TV(\mathbb{1}_D) + \lambda \int_\Omega ((f(\boldsymbol{x}) - c_1)^2 - (f(\boldsymbol{x}) - c_2)^2)\mathbb{1}_D \; d\boldsymbol{x} + \lambda \int_\Omega (f(\boldsymbol{x}) - c_2)^2 \; d\boldsymbol{x}$$

Since the last term is constant we can disregard it when minimizing. The problem then becomes

$$\min_u \left\{ TV(u) + \lambda \int_\Omega ((f(\boldsymbol{x}) - c_1)^2 - (f(\boldsymbol{x}) - c_2)^2) u(\boldsymbol{x}) \; d\boldsymbol{x} \; ; \; u = \mathbb{1}_D, \; D \subset \Omega \right\} \tag{6}$$

where the functional is minimized over every set indicator function $\mathbb{1}_D$ such that $D \subset \Omega$.
To solve (6) Chan et al. [5] proposed looking at the relaxed problem of minimizing

$$E(u, c_1, c_2) = \underbrace{\int_\Omega |\nabla u(\boldsymbol{x})| \; d\boldsymbol{x}}_{TV(u)} \; + \lambda \int_\Omega \underbrace{((f(\boldsymbol{x}) - c_1)^2 - (f(\boldsymbol{x}) - c_2)^2)}_{r(\boldsymbol{x}, c_1, c_2)} u(\boldsymbol{x}) \; d\boldsymbol{x}$$

over all u that satisfy $0 \le u(\boldsymbol{x}) \le 1$. Note that the resulting u after minimizing the functional is not necessarily a binary function (since we're minimizing over every u) but in [5] Chan et al. shows that if u minimizes $E(\cdot)$ then $\mathbb{1}_{\{u(\boldsymbol{x}) > \mu\}}$ is a global minimizer to $E(\cdot)$ for almost every $\mu \in [0, 1]$.

3.1 The algorithm of Bresson et al.

In [2, Section 3] Bresson et al. propose an algorithm for dealing with the problem of minimizing E under the constraint $0 \le u(\boldsymbol{x}) \le 1$. They define the functional

$$E_B(u, v) = TV(u) + \frac{1}{2\theta} ||u - v||_2^2 + \langle \lambda r, v \rangle$$

where $\theta > 0$ and look at the problem

$$\min_{\substack{u \\ 0 \le v \le 1}} \; E_B(u, v)$$

When θ is small the second term will keep $u \approx v$ and if $u = v$ we get that $E_B(u, v) = E(u)$. The functional $E_B(u, v)$ can be minimized by alternating minimizing in u and v.

When minimizing in u we consider v to be fixed and then the last term in E_B is constant and can be disregarded. So the problem we are interested in is then

$$\min_u \; TV(u) + \frac{1}{2\theta} ||u - v||_2^2$$

This is a ROF problem and can be solved using Chambolle's algorithm.

When minimizing in v the first term is constant and can be disregarded. This gives the following problem

$$\min_{0 \le v \le 1} \; \frac{1}{2\theta} ||u - v||_2^2 + \langle \lambda r, v \rangle$$

Define the set $B = \{v \; ; \; 0 \le v(\boldsymbol{x}) \le 1\}$ which is a closed and convex set. Then the problem of finding v^* becomes

$$\begin{aligned}
v^* =& \arg\min_{v \in B} \left\{ \frac{1}{2\theta} ||u - v||_2^2 + \langle \lambda r, v \rangle \right\} = \\
=& \arg\min_{v \in B} \left\{ \frac{1}{2\theta} \left(||u||_2^2 - 2\langle u, v \rangle + ||v||_2^2 + 2\theta \langle \lambda r, v \rangle \right) \right\} = \\
=& \arg\min_{v \in B} \left\{ ||v||_2^2 - 2\langle u - \theta\lambda r, v \rangle \right\} = \\
=& \arg\min_{v \in B} \left\{ ||v - (u - \theta\lambda r)||_2^2 - ||u - \theta\lambda r||_2^2 \right\} = \\
=& \arg\min_{v \in B} \left\{ ||v - (u - \theta\lambda r)||_2^2 \right\} = \pi_B(u - \theta\lambda r)
\end{aligned}$$

The projection $\pi_B(u - \theta\lambda r_1)$ can be calculated by simply cutting $u(\boldsymbol{x}) - \theta\lambda r(\boldsymbol{x}, c_1, c_2)$ to the interval $[0, 1]$, i.e.

$$v^*(\boldsymbol{x}) = \max\{\min\{u(\boldsymbol{x}) - \theta\lambda r(\boldsymbol{x}, c_1, c_2),\ 1\}, 0\}$$

So to perform the segmentation we iterate minimizing u and v as shown above and then when the iteration converges we select $\{u(\boldsymbol{x}) > 0.5\}$ as the segmentation.

3.2 Thresholding method

In this section another method for performing the image segmentation by minimizing the functional E is presented. It is a special case of a method proposed by Chambolle in [6] for solving a more general problem.

It can be shown that if w^* is the solution to the ROF problem

$$\min_w TV(w) + \frac{1}{2}||w - g||_2^2$$

and we construct u by thresholding w^* around zero

$$u(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad w^*(\boldsymbol{x}) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

then u is the solution to the problem

$$\min_{0 \leq u \leq 1} TV(u) - \langle g, u \rangle$$

Now if we let $g = -\lambda r$ we get

$$TV(u) - \langle g, u \rangle = TV(u) + \langle \lambda r, u \rangle = E(u)$$

Which is the functional we are interested in minimizing.

4 Results and Discussion

In Figure 2 Bresson's algorithm has been applied to a satellite picture. The segmentation has been performed several times with varying λ. When λ is small the term $TV(u)$ will dominate the expression and the perimeter of the segmentation will be smoother.

A comparison of the two presented segmentation methods can be found in Figure 3. Note that although both algorithms converge to similar results, the algorithm proposed by Bresson et al. requires significantly more processing time. This is since Bresson's method (for fix c_1 and c_2) requires multiple ROF problems to be solved while the Thresholding method only requires one.

When testing the algorithms it was noticed that the Thresholding method required a lower ϵ for the stopping critereon when solving the ROF problem.

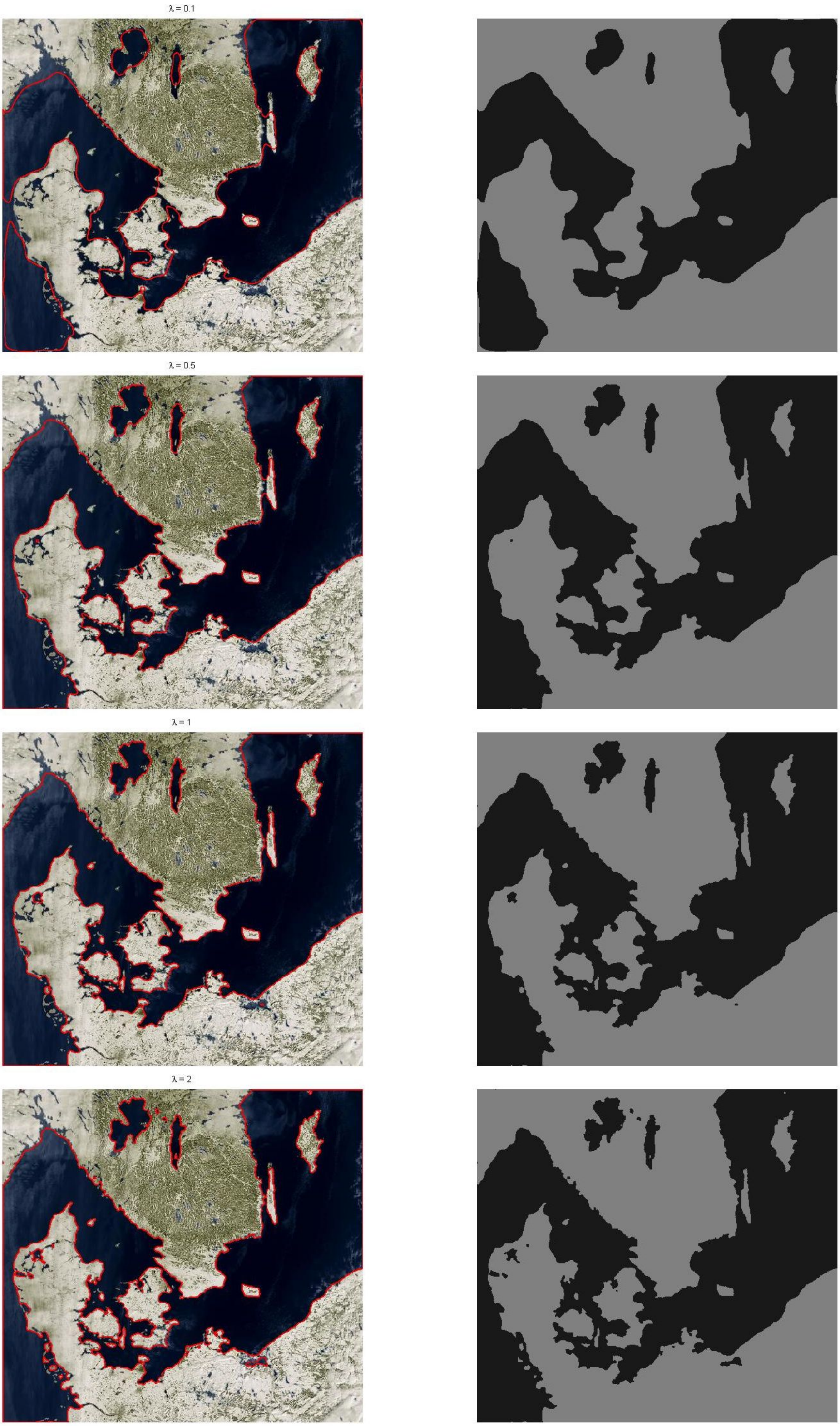

Figure 2: Segmentation using Bresson's algorithm with varying λ

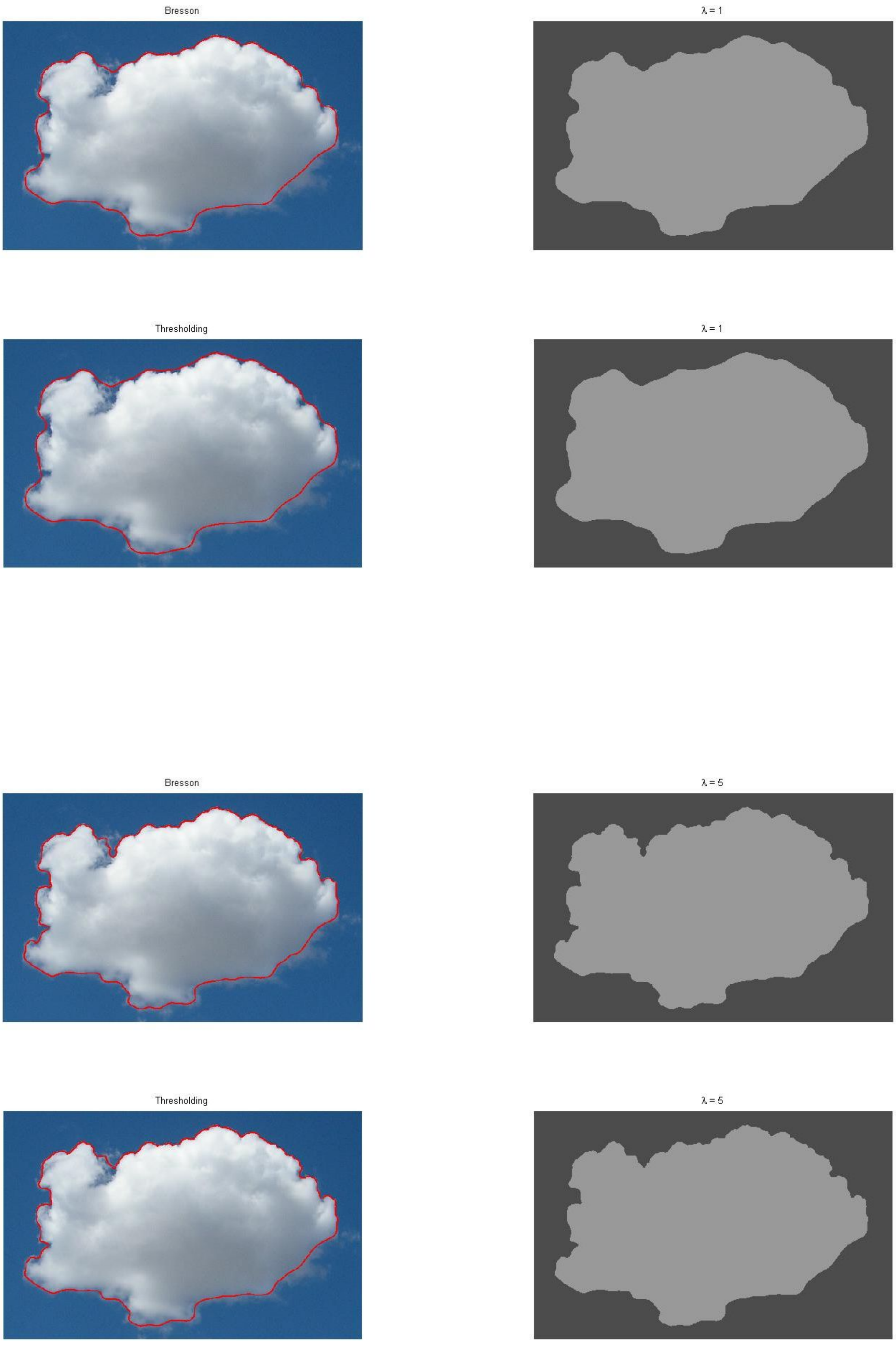

Figure 3: Comparison of the two algorithms for image segmentation.

References

[1] A. Chambolle, *An Algorithm for Total Variation Minimization and Applications*, Journal of Mathematical Imaging and Vision, 2004

[2] X. Bresson, S. Esedoglu, P. Vandergheynst, J. P. Thiran, S. Osher, *Fast Global Minimization of the Active Contour/Snake Model*, Journal of Mathematical Imaging and Vision, 2005

[3] L. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 1992

[4] T. Chan, L. A. Vese, *Active Contours Withour Edges*, IEEE Transactions on image processing, 2001

[5] T. Chan, S. Esedoglu, M. Nikolova, *Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models*, SIAM J. on Applied Mathematics, 2004

[6] A. Chambolle, *Total variation minimization and a class o binary MRF models*, Energy Minimization Methods in Computer Vision and Pattern Recognition, 2005

Appendix: Convexity

Definition: The set $C \subset \mathbb{R}^n$ is called *convex* if

$$x, y \in C \ , \ \lambda \in [0,1] \Rightarrow \lambda x + (1-\lambda)y \in C$$

Lets look at an example of a convex set. Consider the closed unit ball in $\mathbb{R}^n$.

$$B = \{x \in \mathbb{R}^n \ ; \ ||x|| \leq 1\}$$

This is a convex set since if $x, y \in B$ and $\lambda \in [0,1]$ then

$$||\lambda x + (1-\lambda)y|| \leq ||\lambda x|| + ||(1-\lambda)y|| = \lambda ||x|| + (1-\lambda)||y|| \leq \lambda + (1-\lambda) = 1$$

so $\lambda x + (1-\lambda)y \in B$. $\square$

Another example of a convex set is

$$K = \{\text{div } \boldsymbol{p} \ ; \ \boldsymbol{p} \in Y, \ |\boldsymbol{p}_{i,j}| \leq 1\}$$

Consider two elements div $\boldsymbol{p}$ and div $\boldsymbol{q}$ and look at

$$\lambda \text{div } \boldsymbol{p} + (1-\lambda)\text{div } \boldsymbol{q} = \text{div } \Bigg(\lambda \boldsymbol{p} + (1-\lambda)\boldsymbol{q}\Bigg) \in K$$

It is in K since $\lambda \boldsymbol{p} + (1-\lambda)\boldsymbol{q} \in Y$ and

$$|\lambda \boldsymbol{p}_{i,j} + (1-\lambda)\boldsymbol{q}_{i,j}| \leq \lambda |\boldsymbol{p}_{i,j}| + (1-\lambda)|\boldsymbol{q}_{i,j}| \leq \lambda + (1-\lambda) = 1$$

$\square$

Definition: Let $C \subset \mathbb{R}^n$ be a closed convex set, $g \in \mathbb{R}^n$ and define

$$\pi_C(g) = \arg\min_{z \in C} ||z - g||$$

then $\pi_C(g)$ is called the *projection* of x onto C.

Theorem: The projection $\pi_C(g)$ always exists and is unique.
Proof: We first prove the existence. Let a be any point in C. Then minimizing $||x-g||$ over every $x \in C$ is the same as minimizing it over the closed and bounded set

$$\{x \in C \ ; \ ||x - g|| \leq ||a - g||\}$$

and since $x \to ||x - g||$ is a continuous function the existence follows from Weierstrass theorem.

For the uniqueness, assume there are two elements x_1 and x_2 which both minimizes $||x-g||$ and let $\tau = ||x_1 - g|| = ||x_2 - g||$. From $x_1 \neq x_2$ it follows that

$$0 < ||x_1 - x_2||^2 = ||(x_1 - g) - (x_2 - g)||^2 = \underbrace{||x_1 - g||^2}_{\tau^2} - 2\langle x_1 - g, x_2 - g\rangle + \underbrace{||x_2 - g||^2}_{\tau^2} =$$

Using the identity $2\langle \alpha, \beta \rangle = ||\alpha + \beta||^2 - ||\alpha||^2 - ||\beta||^2$ the expression becomes

$$= 2\tau^2 - \Bigg(||x_1 - g + x_2 - g||^2 - \underbrace{||x_1 - g||^2}_{\tau^2} - \underbrace{||x_2 - g||^2}_{\tau^2}\Bigg) =$$

$$= 4\tau^2 - ||x_1 + x_2 - 2g||^2 = 4\tau^2 - 4||\frac{x_1 + x_2}{2} - g||^2 \quad \Leftrightarrow$$

$$||\frac{x_1 + x_2}{2} - g||^2 < \tau^2$$

Let $\hat{x} = \frac{x_1 + x_2}{2}$.

$$||\hat{x} - g||^2 < \tau^2 \Leftrightarrow ||\hat{x} - g|| < \tau$$

So $\hat{x}$ is closer to g than x_1 and x_2. This is a contradiction since $\hat{x} \in C$ due to convexity. $\square$