Universität Münster

Fachbereich 10 Mathematik und Informatik

Master's Thesis

The Frank-Wolfe algorithm for inverse problems with anisotropic TV regularization

Lena Schmedt

Matr. 448820

Supervisor: Prof. Dr. Benedikt Wirth
Assisting Supervisor: Dr. Frank Wübbeling

Faculty of Mathematics and Computer Science, University of Münster, Germany

May 16, 2025

Contents

List of Figures

List of Algorithms

1. Introduction

The task to recover an underlying cause from measurements of its induced effects is an important task in many scientific disciplines. Typical problems can be found for example in medical imaging, where it is desired to reconstruct anatomic structures in the human body from MRI or X-ray measurements, or in the field of geophysics in order to map subsurface structures with help of seismic tomography. All these applications share a common goal: gaining insight into hidden phenomena through indirect measurements - or mathematically speaking to solve an underlying *inverse problem*.

Let X denote the space of possible sources and Y the space of measurements equipped with a mapping $K : X \to Y$ describing the measurement process. The goal of the inverse problem is then to reconstruct $u \in X$ from a measurement $f \in Y$ such that

$$Ku = f.$$

However, in most applications these inverse problems are often ill-posed in Hadamard´s sense [1, Def. 4]: Either a solution u of $Ku = f$ does not exist, a solution is not unique or small changes of f may lead to large changes of the solution u. In practice, it is not unusual for several of these conditions to be violated at once.

A common strategy to avoid these obstacles is to make use of *regularization*. Instead of attempting to solve the ill-posed original problem, which means to solve $u = K^{-1}f$ directly, one considers a variational formulation. For a given measurement $f \in Y$ one aims at finding

$$u \in \underset{v \in X}{\operatorname{argmin}} \frac{1}{\alpha} F_f(Kv) + G(v).$$

Here, $\alpha \in \mathbb{R}$, $F_f : Y \to (-\infty, +\infty]$ and $G : X \to (-\infty, +\infty]$ should be selected in a manner that F_f ensures closeness of the minimizer and the ground truth and G assures certain regularity and may incorporate a priori known properties of u. The regularization parameter α on the other hand regulates, which term gains how much impact on the solution.

This thesis will deal with the inverse problem of reconstructing a ground truth consisting of a linear combination of weighted indicator functions of rectangular sets from its truncated Fourier image. The *anisotropic total variation* is chosen as the regularization term. This setup relies on the problem formulated in [2]. Although this type of TV regularization does not produce a well-posed problem in Hadamard´s sense, the anisotropic total variation is well suited to produce images with sharp edges and is therefore a valid and promising choice as regularization term in our case. Moreover, note that since the considered ground truth has just horizontal and vertical edges, it can be parameterized by just finitely many values. This is a characteristic that is also true for images consisting solely of finitely many point masses, so one may expect similar behaviour when it comes to reconstruction. It was shown in [2] that this is true and we will briefly refer to these results in section 2.3.2.

The question we want to answer in this thesis is *how* to solve the reconstruction problem. In view of the above considerations it is natural to ponder if the two problems of point masses and weighted indicator functions may also be solved by similar algorithms. In [3] the classic *Frank-Wolfe algorithm* (FW) and a variation, the *sliding Frank-Wolfe algorithm* (SFW), were presented to reconstruct point masses, so it seems intuitive and appropriate to analyze whether the Frank-Wolfe algorithm and its sliding version are also suitable to restore a ground truth consisting of rectangles and how these algorithms can be applied.

This paper is organized as follows: In section 2 we will briefly discuss the role of regularization in inverse problems and recall the concepts of anisotropic total variation and the space of bounded variation. The anisotropic total variation will be introduced as regularization term in section 2.3. Here, the main problem is presented and the solvability properties from [2] are cited.

Section 3 is the main part of this thesis. Here, the classic Frank-Wolfe algorithm is introduced and adapted progressively to the total variation regularized problem. Therefore, concerning the minimization step, the *extreme points* of a total variation ball are determined and it is argued numerically, whether it is justified to assume that the optimal extreme points are rectangular shaped in our application. To analyze the stopping criterion of the Frank-Wolfe algorithm in the next part of the section, primal-dual optimality conditions are developed. The last part of this section is dedicated to the sliding aspect of the sliding Frank-Wolfe algorithm.

In the final section we test the classical Frank-Wolfe algorithm as well as the sliding Frank-Wolfe algorithm numerically and assess their ability to reduce reconstruction errors and to capture edges precisely.

2. Total variation regularization

In order to understand the problem discussed in this paper, we must first introduce its setup.
The goal is to reconstruct a ground truth composed of rectangular indicator functions from a slightly noisy measurement. To avoid the possible ill-posedness of this noisy inverse problem we approximate it by neighboring well-posed problems and solve them for an approximate solution. To do so, we are going to make use of regularization.

2.1. Regularization

Let X, Y be Banach spaces, $u^\dagger \in X$ the ground truth that should be reconstructed, $K : X \to Y$ a measurement operator and $Ku^\dagger + \omega = f \in Y$ the measurement with some noise $\omega \in Y$. Note, that f does not necessarily need to coincide with $f^\dagger = Ku^\dagger$. The inverse problem that is to be solved is of the form

$$\text{given } f \in Y, \text{ find } u \in X \text{ s.t. } Ku = f.$$

As explained in the introduction, in order to find an approximate solution of $Ku = f$ one can solve a regularized problem in form of the minimization problem below,

$$\min_{u \in X} J_\alpha^f(u), \quad J_\alpha^f(u) = \frac{1}{\alpha} F_f(Ku) + G(u) \tag{1}$$

Here, $F_f : Y \to \mathbb{R} \cup +\infty$ is the *fidelity term*, $G : X \to \mathbb{R} \cup +\infty$ the *regularization term* and $\alpha \geq 0$ the *regularization parameter*.
The fidelity term ensures that the solution of the minimization problem (1) (respectively its image under K) is close to the measured data f, whereas the regularization term makes the solution fit into the context of the given problem. This means, that G can compensate for lost information during the measurement with a priori known information of the ground truth such as sparsity or smoothness, or that it can remove artifacts from the reconstructed solution. The amount of the influence that G has on the solution of (1) is controlled by α. The larger α is, the more impact the regularization term gains, as described in [4].
As stated in [2], G and F_f are proper and convex functions, where the fidelity term satisfies $F_f(Ku) = 0 \Leftrightarrow Ku = f$ and $F_f(Ku) > 0$, otherwise. Therefore, $\alpha = 0$ is considered as the constraint forcing $F_f(Ku) = 0$, i.e. $Ku = f$. This means, for $\alpha = 0$ the objective functional may take the form $J_0^f(u) = \iota_{\{f\}}(Ku) + G(u)$, where

$$\iota_{\{f\}}(Ku) = \begin{cases} 0 & \text{if } f = Ku \\ +\infty & \text{if } f \neq Ku \end{cases}$$

This type of regularization with $\alpha = 0$ is especially used in a no-noise regime as for example in Theorem 2.12 that is cited from [2].

As stated in [4] common examples for the data-fidelity term F_f are L^1, L^2, Huber or Lorentzian estimators. The regularized problem that is considered in this paper indeed uses an ℓ^2 term, as pointed out in $P_\alpha(f)$.
One of the most used regularization terms G is the so called *Tikhonov regularization*. This is defined as $G(u) = \|\Gamma u\|_2^2$. Γ captures certain aspects of the ground truth and takes the form of e.g. the derivative, the Laplacian or just the identity. Thus, the regularizer limits the energy of the solution or enforces smoothness. As a consequence noisy pixels are reduced or sharp edges will be prevented.
This is not what is desired in our case. As described above, the considered ground truth is made up of rectangular indicator functions and thus a key characteristic of a solution should be sharp edges. So instead of Tikhonov regularization, we make use of the *total variation* (TV) of a function u as regularization term.

2.2. Total variation and the space of bounded variation

Definition 2.1 (Anisotropic total variation, [5, Def. 3.4])**.** Let Ω be a generic open set of $\mathbb{R}^n$ and $u \in L^1(\Omega)$. The *anisotropic total variation* of u in Ω is defined as

$$\mathrm{TV}(u;\Omega) = \sup\left\{ \int_\Omega u \,\mathrm{div}\,\phi \,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega,\mathbb{R}^n),\ \|\phi\|_{L^\infty(\Omega)} \le 1 \right\}, \tag{2}$$

where div is the divergence operator, $C_c^1(\Omega,\mathbb{R}^n)$ is the set of continuously differentiable functions from Ω to $\mathbb{R}^n$ with compact support in Ω and $\|\cdot\|_{L^\infty(\Omega)}$ is the supremum norm.

Remark 2.2. Note, that for continuously differentiable functions $u \in C^1(\Omega)$:

$$\mathrm{TV}(u;\Omega) = \int_\Omega |\nabla u| \,\mathrm{d}x = |\nabla u|(\Omega) \tag{3}$$

The first equality of (3) was stated in [5], whereas the second equality is a simple result for weighted measures. Let $\nu = f\mu$, where μ is a measure and f measurable. Then, by [1, Ex. 114.5], $\nu(\Omega) = \int_\Omega f \,\mathrm{d}\mu$. In our case, let μ be the Lebesgue measure and $|\nabla u|$ the measurable function, as it is continuous. We then get the desired $\int_\Omega |\nabla u| \,\mathrm{d}x = |\nabla u|(\Omega)$.

Equation (3) motivates the choice of $\mathrm{TV}(u;\Omega)$ as regularization term G in the regularization (1) in order to reconstruct indicator functions of rectangles. It indicates that this selection promotes solutions with sparse (distributional) gradients and sharp edges. Moreover, horizontal and vertical edges are preferred over diagonal ones as argued later in a slightly different context in A.1.

In any case, a minimizer u of (1) must satisfy $\mathrm{TV}(u;\Omega) < +\infty$. All functions fulfilling this condition are summed up in the space of functions of bounded variation $\mathrm{BV}(\Omega)$.

Definition 2.3 (Space of functions of bounded variation $\mathrm{BV}(\Omega)$)**.** Again, let Ω be a generic open set of $\mathbb{R}^n$. The *space of functions of bounded variation* is then defined as

$$\mathrm{BV}(\Omega) = \left\{ u \in L^1(\Omega) \,\middle|\, \mathrm{TV}(u;\Omega) < \infty \right\}. \tag{4}$$

Remark 2.4. In [5, Def. 3.1] the space $\mathrm{BV}(\Omega)$ is introduced in a slightly different way. This characterization will be needed in the proof of Proposition 2.9.

Proposition 2.5 (Equivalent definition of $\mathrm{BV}(\Omega)$, [5, Def. 3.1, Prop. 3.6])**.**

$$\begin{aligned}\widetilde{\mathrm{BV}}(\Omega) &= \left\{ u \in L^1(\Omega) \,\middle|\, \begin{array}{l}\textit{the distributional derivative of } u \textit{ is representable}\\ \textit{by a finite Radon measure in } \Omega\end{array} \right\}\\ &= \left\{ u \in L^1(\Omega) \,\middle|\, \begin{array}{c}\forall i \in \{1,\dots,n\} \textit{ there is a finite Radon measure } \mu_i \textit{ such that}\\ \displaystyle\int_\Omega u \frac{\partial \phi}{\partial x_i}\,\mathrm{d}x = -\int_\Omega \phi \,\mathrm{d}\mu_i \quad \forall \phi \in C_c^1(\Omega).\end{array} \right\}\end{aligned}$$

We write $Du = \mu = (\mu_1, \dots, \mu_n)$.
Indeed both characterizations $\mathrm{BV}(\Omega)$ *and* $\widetilde{\mathrm{BV}}(\Omega)$ *coincide: Let* $u \in L^1(\Omega)$*. Then,* u *belongs to* $\widetilde{\mathrm{BV}}(\Omega)$ *if and only if* $\mathrm{TV}(u;\Omega) < \infty$*. In addition,* $\mathrm{TV}(u;\Omega)$ *coincides with* $|Du|(\Omega)$ *for any* $u \in \mathrm{BV}(\Omega)$*.*

Proof. A proof of Proposition 2.5 can be obtained from [5, Prop. 3.6]. □

As $\mathrm{TV}(u;\Omega)$ and $\mathrm{BV}(\Omega)$ are crucial concepts in the course of this thesis, we will state some important properties in the following propositions.

Proposition 2.6 (Lower semicontinuity of TV, [6, Thm. 1.9])**.** *Let* $\Omega \subseteq \mathbb{R}^n$ *be an open set and* $\{u_i\}$ *a sequence of functions in* $\mathrm{BV}(\Omega)$ *which converges in* $L^1_{loc}(\Omega)$ *to a function* u*.*
Then

$$\mathrm{TV}(u;\Omega) \leq \liminf_{i\to\infty} \mathrm{TV}(u_i;\Omega) \tag{5}$$

Proof. The proof stems also from [6, Thm. 1.9]. Let $\phi \in C_c^1(\Omega)$ such that $\|\phi\|_\infty \leq 1$. Then:

$$\begin{aligned}\int_\Omega u \operatorname{div}\phi \,\mathrm{d}x &= \lim_{i\to\infty} \int_\Omega u_i \operatorname{div}\phi \,\mathrm{d}x\\ &\leq \liminf_{i\to\infty} \sup \left\{ \int_\Omega u_i \operatorname{div}\phi \,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega), \|\phi\|_\infty \leq 1 \right\}\\ &= \liminf_{i\to\infty} \mathrm{TV}(u_i;\Omega)\end{aligned}$$

Now, taking the supremum over all such ϕ on the left side leads to the desired statement. $\square$

Proposition 2.7 (Convexity of TV)**.** *The total variation is a convex functional, i.e. for every $u_1, u_2 \in BV(\Omega)$ and $0 \le t \le 1$:*

$$\mathrm{TV}(tu_1 + (1-t)u_2; \Omega) \le t\mathrm{TV}(u_1; \Omega) + (1-t)\mathrm{TV}(u_2; \Omega) \tag{6}$$

Proof. Let $u_1, u_2 \in \mathrm{BV}(\Omega)$ and $t \in [0,1]$. Then

$$\begin{aligned}
\mathrm{TV}(tu_1 + (1-t)u_2; \Omega) &= \sup\left\{ \int_\Omega (tu_1 + (1-t)u_2)\,\mathrm{div}\phi\,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \le 1 \right\} \\
&= \sup\left\{ \int_\Omega tu_1\,\mathrm{div}\phi\,\mathrm{d}x + \int_\Omega (1-t)u_2\,\mathrm{div}\phi\,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \le 1 \right\} \\
&\le t \sup\left\{ \int_\Omega u_1\,\mathrm{div}\phi\,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \le 1 \right\} \\
&\quad + (1-t)\sup\left\{ \int_\Omega u_2\,\mathrm{div}\psi\,\mathrm{d}x \,\middle|\, \psi \in C_c^1(\Omega, \mathbb{R}^n), \|\psi\|_{L^\infty(\Omega)} \le 1 \right\} \\
&= t\mathrm{TV}(u_1; \Omega) + (1-t)\mathrm{TV}(u_2; \Omega)
\end{aligned}$$

$\square$

Proposition 2.8 (Positive 1-homogeneity of TV)**.** *For any $u \in \mathrm{BV}(\Omega)$ and $t \in \mathbb{R}$ with $t > 0$ one gets*

$$\mathrm{TV}(tu; \Omega) = t \cdot \mathrm{TV}(u; \Omega).$$

Proof. Let $u \in \mathrm{BV}(\Omega)$ and $t > 0$. Then

$$\begin{aligned}
\mathrm{TV}(tu; \Omega) &= \sup\left\{ \int_\Omega tu \cdot \mathrm{div}\phi \,\middle|\, \phi \in C_c^\infty(\Omega, \mathbb{R}^2),\ \|\phi\|_\infty \le 1 \right\} \\
&\overset{t \geq 0}{=} t \cdot \sup\left\{ \int_\Omega u \cdot \mathrm{div}\phi \,\middle|\, \phi \in C_c^\infty(\Omega, \mathbb{R}^2),\ \|\phi\|_\infty \le 1 \right\} \\
&= t \cdot \mathrm{TV}(u; \Omega)
\end{aligned}$$

$\square$

Proposition 2.9 ($\mathrm{BV}(\Omega)$ as Banach space, [6, Rem. 1.12])**.** $\mathrm{BV}(\Omega)$ *becomes a Banach space, when endowed with the norm*

$$\begin{aligned}
\|u\|_{\mathrm{BV}} :&= \int_\Omega |u|\ \mathrm{d}x + |Du|(\Omega) \\
&= \|u\|_{L^1} + \mathrm{TV}(u; \Omega).
\end{aligned}$$

Proof. The proof is based on the argumentation in [6, Rem. 1.12].

- $\mathrm{BV}(\Omega)$ is a vector space of $L^1(\Omega)$: Clearly, $\mathrm{BV}(\Omega)$ is a subset of $L^1(\Omega)$. It remains to show its linearity properties. Therefore, let $u, v \in \mathrm{BV}(\Omega)$ and $c \in \mathbb{R}$.
Then:

$$\begin{aligned}\int_\Omega (u + cv) \operatorname{div} \phi \, \mathrm{d}x &= \int_\Omega u \operatorname{div} \phi \, \mathrm{d}x + c \int_\Omega v \operatorname{div} \phi \, \mathrm{d}x \\ &\overset{u,v \in \mathrm{BV}(\Omega), \text{ Prop. } 2.5}{=} - \int_\Omega \phi \, \mathrm{d}Du - c \int_\Omega \phi \, \mathrm{d}Dv \\ &= - \int_\Omega \phi \, \mathrm{d}(Du + cDv) \qquad \forall \phi \in C_c^1(\Omega)\end{aligned}$$

This implies, again with Proposition 2.5, $u + cv \in \mathrm{BV}(\Omega)$ as its variational derivative is a finite Radon measure.

- $\|\cdot\|_{\mathrm{BV}}$ satisfies the norm properties: Let $u, v \in \mathrm{BV}(\Omega)$. Then

$$\begin{aligned}\|u + v\|_{\mathrm{BV}} &= \|u + v\|_{L^1} + \mathrm{TV}(u + v; \Omega) \\ &= \|u + v\|_{L^1} + \sup \left\{ \int_\Omega u + v \operatorname{div}\phi \, \mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \leq 1 \right\} \\ &\leq \|u\|_{L^1} + \|v\|_{L^1} + \sup \left\{ \int_\Omega u \operatorname{div}\phi \, \mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \leq 1 \right\} \\ &\quad + \sup \left\{ \int_\Omega v \operatorname{div}\phi \, \mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \leq 1 \right\} \\ &= \|u\|_{L^1} + \mathrm{TV}(u; \Omega) + \|v\|_{L^1} + \mathrm{TV}(v; \Omega) = \|u\|_{\mathrm{BV}} + \|v\|_{\mathrm{BV}}\end{aligned}$$

Moreover for $u \in \mathrm{BV}(\Omega)$ and $c \in \mathbb{R}$:

$$\begin{aligned}\|cu\|_{\mathrm{BV}} &= \|cu\|_{L^1} + \mathrm{TV}(cu; \Omega) = \|cu\|_{L^1} + |D(cu)|(\Omega) \\ &= |c| \, \|u\|_{L^1} + |c| \, |Du|(\Omega) = |c| \, \|u\|_{\mathrm{BV}}\end{aligned}$$

Now, let $u \in \mathrm{BV}(\Omega)$ such that $\|u\|_{\mathrm{BV}} = 0$, i.e. $\overbrace{\|u\|_{L^1}}^{\geq 0} + \overbrace{|Du|(\Omega)}^{\geq 0} = 0$, so both summands need to be 0. From $|Du|(\Omega) = 0$ follows that u must be a constant function. Under this condition with $\|u\|_{L^1} = 0$ we get $u = 0$.
Altogether, $\|\cdot\|_{\mathrm{BV}}$ satisfies all norm conditions and thus is a norm on $\mathrm{BV}(\Omega)$.

- $\mathrm{BV}(\Omega)$ is complete with respect to $\|\cdot\|_{\mathrm{BV}}$: Suppose $\{u_i\}$ is a Cauchy sequence in $\mathrm{BV}(\Omega)$. Then, by definition of the norm, $\{u_i\}$ is also Cauchy in the complete space $L^1(\Omega)$. Therefore, there exists a function $u \in L^1(\Omega)$ such that $u_i \to u$ in $L^1(\Omega)$. As u_i is Cauchy in $\mathrm{BV}(\Omega)$ its norm $\|u_i\|_{\mathrm{BV}}$ is bounded. Thus, $|Du_i| = \mathrm{TV}(u_i; \Omega) < \infty$ for $i \to \infty$ and by lower semicontinuity of the total variation (Prop. 2.6) $\mathrm{TV}(u; \Omega) < \infty$, i.e. $u \in \mathrm{BV}(\Omega)$. It remains to show $u_i \to u$ in $\mathrm{BV}(\Omega)$, but $u_i \to u$ in $L^1(\Omega)$ is already done, so just show $\mathrm{TV}(u_i - u; \Omega) \to 0$ for $i \to \infty$.

Let $\epsilon > 0$. Then, there is $N \in \mathbb{N}$ s.t. for all $i, n \geq N$: $\|u_i - u_m\|_{\mathrm{BV}} \leq \epsilon$ i.e. $\mathrm{TV}(u_i - u_n; \Omega) \leq \epsilon$. As $u_n \to u$ in $L^1(\Omega)$, also $(u_i - u_n) \to (u_i - u)$ in $L^1(\Omega)$. By lower semicontinuity again:

$$\mathrm{TV}(u_i - u; \Omega) \leq \liminf_{n\to\infty} \mathrm{TV}(u_i - u_n; \Omega) \leq \epsilon$$

$\Rightarrow \|u_i - u\|_{\mathrm{BV}} \to 0$. This proves completeness of $\mathrm{BV}(\Omega)$ and it becomes a Banach space when equipped with the norm as defined above.

□

Remark 2.10. With Proposition 2.9 it is valid to optimize over the space of $\mathrm{BV}(\Omega)$-functions in Equation (1).

2.3. The TV regularized problem

In this section we formally introduce the inverse problem and its associated regularized problem, that we are going to analyze in this thesis. The problem as well as the contents of this section are taken from [2].

2.3.1. Setup of the problem

Let $\Omega = (\mathbb{R}/\mathbb{Z})^2 = \mathbb{T}^2$ the two-dimensional flat torus, that can be identified with $[0,1]^2$ with periodic boundary conditions. Let the ground truth image be a piecewise constant real-valued function of the form

$$u^\dagger = \sum_{m=1}^{M} \sum_{n=1}^{N} u^\dagger_{mn} \mathbb{1}_{[x_m, x_{m+1}[\times[y_n, y_{n+1}[} \in \mathrm{BV}(\Omega) \tag{7}$$

again, with periodic boundary conditions. $\mathbb{1}_A$ is the characteristic function of a set A, i.e.

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

Further, let $x_1 < \cdots < x_M$ and $y_1 < \cdots < y_M$ be points in $\mathbb{R}/\mathbb{Z}$, where we identify $x_{M+1} = x_1$ and analogue $y_{N+1} = y_1$ due to periodicity of Ω. Moreover, we consider the interval $[a, b[$ with $b < a$ as the interval $[a, b+1[\subset \mathbb{R}$ projected onto $\mathbb{R}/\mathbb{Z}$. The minimal distance of all points $\{x_i \mid i = 1, \ldots, M\}$ and $\{y_j \mid j = 1, \ldots, N\}$ is denoted by

$$\Delta = \min\left\{\min\{\mathrm{dist}(x_m, x_{m+1}) \mid m = 1, \ldots, M\},\ \min\{\mathrm{dist}(y_n, y_{n+1}) \mid n = 1, \ldots, N\}\right\}.$$

Here, dist is the Euclidean distance between two points in Ω defined via

$$\mathrm{dist}((x_1, y_1), (x_2, y_2)) = \min\{|(x_1 - x_2 + k, y_1 - y_2 + l)|_2 \mid k, l \in \mathbb{Z}\}$$

for all $(x_1, y_1), (x_2, y_2) \in \Omega$.

Further, let $K : \mathrm{BV}(\Omega) \to Y$ be the linear bounded measurement operator from the space of functions of bounded variations introduced in Definition 2.3 into the finite dimensional image space Y equipped with the Euclidean norm.
In our case, K is chosen to be the truncated Fourier transform, i.e. Ku is of the form

$$\begin{aligned} Ku &= (\hat{u}_k)_{k \in \mathbb{Z}^2, |k|_\infty \leq \Phi} \\ &= \left(\int_\Omega e^{-2\pi i (x,y)\cdot(k_1,k_2)} u(x,y) \,\mathrm{d}(x,y) \right)_{k_1,k_2 \in \mathbb{Z}, |k_1|, |k_2| \leq \Phi} \end{aligned} \tag{8}$$

with cut-off frequency $\Phi \geq 1$.
The concrete regularized problem in form of Equation (1) reads as follows:

$$\min_{u \in \mathrm{BV}(\Omega)} J_\alpha^f(u), \quad J_\alpha^f(u) = \frac{1}{2\alpha} |Ku - f|_2^2 + \mathrm{TV}(u; \Omega) \tag{$P_\alpha(f)$}$$

where $|\cdot|_2$ is the ℓ^2-norm on the finite-dimensional vectorspace Y.

Remark 2.11. In the analysis of $(P_\alpha(f))$, we are going to need the preadjoint measurement operator of K as well as the related predual spaces. Therefore, we will introduce the associated notation.
Let the predual space of $\mathrm{BV}(\Omega)$ be denoted by $\mathrm{BV}(\Omega)^\#$ and let $Y^\#$ be the predual space of Y. Note, that Y can be identified with $Y^\#$ as it is finite dimensional by assumption. The preadjoint measurement operator $K^\# : Y^\# \to \mathrm{BV}(\Omega)^\#$ is defined by

$$\langle K^\# w, u \rangle = \langle w, Ku \rangle \qquad \text{for all } u \in \mathrm{BV}(\Omega),\ w \in Y^\# = Y$$

where $\langle \cdot, \cdot \rangle$ describes the dual pairing between an element from a topological vectorspace and an element from its dual.
Let $w \in Y$. Then the preadjoint $K^\#$ of the truncated Fourier transform is of the form

$$K^\# w = \sum_{|k|_\infty \leq \Phi} w_k \cdot e^{2\pi i (x,y)\cdot(k_1,k_2)}. \tag{9}$$

2.3.2. Solvability of the problem

Before dealing with the question of how to solve $P_\alpha(f)$, we need to discuss the conditions under which it can be solved at all. To do so, we will cite the main results from the article [2]. The assumption, we need to make, regards the distributional gradient of the ground truth and is presented below:

Assumption 1 (Consistent gradient direction [2, Assumption 1])**.** *On a vertical line, the (distributional) x-derivative of $u^\dagger$ does not change sign, and equivalently, on a horizontal* *line, the (distributional) y-derivative of* $u^\dagger$ *does not chance sign as well. Mathematically speaking, for each m the restriction* $D_x u^\dagger \mid_{(\{x_m\}\times\mathbb{R}/\mathbb{Z})}$ *of the measure* $D_x u^\dagger$ *to the vertical line* $\{x_m\} \times \mathbb{R}/\mathbb{Z}$ *is either a nonnegative or nonpositive measure. Accordingly,* $D_y u^\dagger \mid_{(\mathbb{R}/\mathbb{Z}\times\{y_n\})}$ *is nonnegative or nonpositive for every n.*

Now consider two different cases. On the one hand, the exact reconstruction of $u^\dagger$ from $Ku^\dagger = f^\dagger$ in a no-noise regime and on the other hand the case of a noisy environment, i.e. the attempt to reconstruct $u^\dagger$ from a noisy measurement f^δ with noise level δ, i.e. f^δ satisfies $\frac{1}{2}|f^\delta - f^\dagger|_2^2 < \delta$.
Indeed, for a noise free environment, one can expect exact reconstruction of the ground truth, as presented in the theorem below:

Theorem 2.12 (Exact reconstruction [2, Thm. 2])**.** *Let Assumption 1 about a consistent gradient direction hold and let K be the truncated Fourier transform defined in (8). There exists a constant* $C > 0$ *such that, if* $\Delta > \frac{C}{\Phi}$*, then* $u^\dagger$ *is the unique minimizer of*

$$J_0^{f^\dagger}(u) = \mathrm{TV}(u) + \iota_{\{f^\dagger\}}(Ku).$$

In a noisy regime with noise level δ, one can not expect exact nor unique reconstruction, but a bound on the reconstruction error depending on δ.

Theorem 2.13 (Convergence for vanishing noise [2, Thm. 3])**.** *Let Assumption 1 (consistent gradient direction) hold and let K be the truncated Fourier transform from (8). There exists a constant* $C > 0$ *such that, if* $\Delta > \frac{C}{\Phi}$*, then any minimizer* u^δ *of*

$$J_\alpha^{f^\delta}(u) = \frac{1}{2\alpha}|Ku - f^\delta|_2^2 + \mathrm{TV}(u;\Omega)$$

for the choice of $\alpha = \sqrt{\delta}$ *satisfies*

$$\left\| u^\delta - u^\dagger \right\|_1 \leq C\delta^{\frac{1}{4}}.$$

Remark 2.14. The estimation of the convergence rate in Theorem 2.13 may be suboptimal as numerical results from [2] imply. Those results rather suggest that an upper bound of $C\delta^{\frac{1}{2}}$ can be achieved, although this was not proved yet.
Moreover, it is important to note, that the measurement operator K can not be chosen completely arbitrarily, as it needs to satisfy certain source conditions that are required to prove Theorem 2.12 and Theorem 2.13. These source conditions are of the form that there exists a dual variable $w \in Y^\#$ such that $K^\# w$ satisfies particular regularity conditions. Indeed, as shown in [2], the choice of K being the truncated Fourier transform is suitable as the required source conditions are fulfilled in that case.

3. Frank-Wolfe algorithm

In the previous section, we have seen, that under certain conditions, the TV regularized optimization problem $P_\alpha(f)$ yields a solution that either coincides with the ground truth $u^\dagger$ or converges to it in the $L^1(\Omega)$-norm for decreasing noise level δ. Nevertheless, we still need to answer the question of *how* to determine solutions u of $P_\alpha(f)$ and which algorithm is suitable for this task.

There are various approaches for algorithms to numerically solve $P_\alpha(f)$. As outlined in [7] these algorithms can be categorized according to whether they require to solve a system of linear equations or not. Those that do not, are in general based on the dual formulation of the minimization problem. Thus, they are computationally efficient but may be limited, when the measurement operator K becomes non-trivial and complex. Alternatively, one can consider algorithms that directly operate on the primal formulation of $P_\alpha(f)$ and usually use a smooth approximation of $\mathrm{TV}(u;\Omega)$ for optimization. Plenty of them are based on the Euler-Lagrange equations and therefore a system of linear equations, often involving second-order differential operators, needs to be solved. Another class of methods, which avoids both the need for second-order differential approximations and the explicit dual formulation of $P_\alpha(f)$, is based on first-order optimization techniques. One such algorithm to solve $P_\alpha(f)$ and subject of this paper is the *Frank-Wolfe algorithm* (FW), also known as conditional gradient (CG) method. In the past it was not as popular due to sub-optimal convergence rates but the algorithm has become more favoured, as it is suitable to incorporate complicated constraints of the given minimization problem and each iteration of the algorithm has just low complexity. This is the case, since the FW does not need difficult and expensive projections but just solves a simpler linear subproblem in each iteration. This is very useful for current topics such as large-scale machine learning problems or image processing [cf. 8].
The FW was first introduced in [9] but we are going to present the version that was showcased in [3].

The Frank-Wolfe algorithm aims to solve a minimization problem of the form

$$\min_{u \in C} \quad J(u). \tag{10}$$

Here, C is a weakly compact convex set of a Banach space, therefore bounded and J is supposed to be a differentiable convex function.

Remark 3.1. Note that there is no need for a Hilbert structure as C just needs to be a subset of a Banach space.
Moreover, the convexity of the functional J has just an impact on the convergence rate of the algorithm and not on the applicability of the algorithm in general. According to [10], in case of non-convexity of the objective functional J the convergence rate of the Frank-Wolfe algorithm reduces from $\mathcal{O}(\frac{1}{k})$ to $\mathcal{O}(\frac{1}{\sqrt{k}})$, but the problem can still be solved

by the FW.
Furthermore, the weak compactness of C is often demanded in order to ensure well-definedness of the minimization step in line 2. Nevertheless, this is not a necessary condition for well-definedness as we will show later in Lemma 3.8.

The Frank-Wolfe algorithm is then defined as follows:

Algorithm 1 Classic Frank-Wolfe Algorithm (general case)

1: **for** $k = 0, \dots, n$ **do**
2: Minimize: $s^{[k]} \in \operatorname{argmin}_{s \in C} J(u^{[k]}) + \mathrm{d}J(u^{[k]})[s - u^{[k]}]$.
3: **if** $\mathrm{d}J(u^{[k]})[s^{[k]} - u^{[k]}] = 0$ **then**
4: $u^{[k]}$ solution of (10). Stop.
5: **else**:
6: Step research: $\gamma^{[k]} \leftarrow \frac{2}{k+2}$ or $\gamma^{[k]} \in \operatorname{argmin}_{\gamma \in [0,1]} J(u^{[k]} + \gamma(s^{[k]} - u^{[k]}))$.
7: Update: $u^{[k+1]} \leftarrow u^{[k]} + \gamma^{[k]}(s^{[k]} - u^{[k]})$.
8: **end if**
9: **end for**

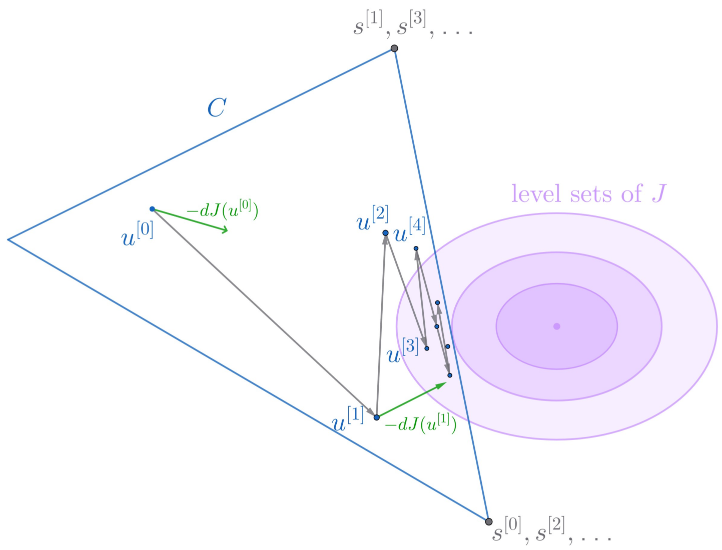


Figure 1: Illustration of how the classic Frank-Wolfe algorithm (Algorithm 1) works in the general setting introduced in (10).

Note, that Figure 1 illustrates that any minimizer $s^{[k]}$ that is determined in line 2 of Algorithm 1 is an *extreme point* of the convex set C. The definition of extreme points can be found in Definition 3.2 and the corresponding proposition explaining this behaviour following from the Bauer Maximum Principle is presented in Proposition 3.3.

Definition 3.2 (Extreme points, [11, Def. 7.61])**.** A point x in a set C is said to be an *extreme point* of C if it cannot be written as a strict convex combination of distinct points in C. This means, for any convex combination with $x_1, x_2 \in C$, $0 < \lambda < 1$ and

$$x = \lambda x_1 + (1-\lambda)x_2$$

follows that $x_1 = x_2 = x$.

Proposition 3.3 ([11, Cor. 7.70])**.** *If C is a nonempty compact convex subset of a locally convex Hausdorff space, then every continuous linear functional has a maximizer and a minimizer that are extreme points of C.*

3.1. Frank-Wolfe algorithm for the TV regularized problem

It is easy to see that the TV regularized problem $P_\alpha(f)$ does not satisfy the conditions presented above that are required to apply the Frank-Wolfe algorithm directly.
On the one hand, in $P_\alpha(f)$ we optimize over the whole space of functions of bounded variation, i.e. $C = \mathrm{BV}(\Omega)$. Note that this is in particular an unbounded set and therefore, it is not a weakly compact convex set of a Banach space as the Frank-Wolfe algorithm demands. Moreover, the objective functional $J_\alpha^f(u) = \frac{1}{2\alpha}|Ku - f|_2^2 + \mathrm{TV}(u;\Omega)$ is not differentiable, as the regularizer $\mathrm{TV}(u;\Omega)$ is not differentiable [cf. 12, Ch. 3].

In the proceeding we want to modify the original problem $P_\alpha(f)$ in such a way that we can apply the Frank-Wolfe algorithm and still restore minimizers of the original objective.
As a first step, following the idea from [3], we are going to reformulate the original problem $P_\alpha(f)$, such that the considered objective becomes differentiable.

Lemma 3.4. *Let $C = \{(t,u) \in \mathbb{R}_+ \times \mathrm{BV}(\Omega) \mid \mathrm{TV}(u;\Omega) \le t \le M\}$ with $M = \frac{|f|_2^2}{2\alpha}$. Then, the original TV regularized problem*

$$\min_{u \in \mathrm{BV}(\Omega)} J_\alpha^f(u), \quad J_\alpha^f(u) = \frac{1}{2\alpha}|Ku - f|_2^2 + \mathrm{TV}(u;\Omega) \tag{$P_\alpha(f)$}$$

is equivalent to the problem

$$\min_{(t,u) \in C} \widetilde{J}_\alpha^f(t,u), \quad \widetilde{J}_\alpha^f(t,u) = \frac{1}{2}|Ku - f|_2^2 + \alpha t \tag{$\widetilde{P}_\alpha(f)$}$$

in the sense that u is a solution to $P_\alpha(f)$ if and only if (t,u) is a solution to $\widetilde{P}_\alpha(f)$ for some feasible $t \ge 0$.

Remark 3.5. Note, that trivially a minimizer u^* of $P_\alpha(f)$= $\frac{1}{2\alpha}|Ku - f|_2^2 + \mathrm{TV}(u;\Omega)$ is also a minimizer of the functional $\frac{1}{2}|Ku - f|_2^2 + \alpha \mathrm{TV}(u;\Omega)$ and vice versa. With abuse of notation denote the latter functional also as $P_\alpha(f)$.

Proof of Lemma 3.4.

- Show that any solution u^* of $P_\alpha(f)$ is feasible for $\widetilde{P}_\alpha(f)$ and vice versa:
  Let $u^* \in \mathrm{BV}(\Omega)$ be a minimizer of J_α^f. Then:

$$J_\alpha^f(u^*) = \frac{1}{2\alpha}\underbrace{|Ku^* - f|_2^2}_{\geq 0} + \mathrm{TV}(u^*;\Omega) \overset{\text{optimality}}{\leq} J_\alpha^f(0) = \frac{1}{2\alpha}|f|_2^2 := M$$

$$\Rightarrow \mathrm{TV}(u^*;\Omega) \leq \frac{|f|_2^2}{2\alpha} = M$$

  Thus, the restriction of the optimization problem $\widetilde{P}_\alpha(f)$ to the bounded set C is valid since one can only consider the $\mathrm{BV}(\Omega)$-functions u satisfying $\mathrm{TV}(u;\Omega) \leq M$ without excluding possible solutions.
  Notice that trivially, any feasible function u from a pair $(t, u) \in C$, in particular including the minimizing function $\tilde{u}^*$, is a $\mathrm{BV}(\Omega)$-function and thus feasible for $P_\alpha(f)$.

- Solution of $P_\alpha(f)$ is a solution of $\widetilde{P}_\alpha(f)$ for some admissible t:
  Let u^* be a solution of $P_\alpha(f)$. For any pair (t, u) minimizing $\widetilde{P}_\alpha(f)$ one can observe $t = \mathrm{TV}(u;\Omega)$, as this minimizes the second part of the functional $\widetilde{J}_\alpha^f$. Having this, both problems read the same and thus by assumption, $(t, u) = (\mathrm{TV}(u^*;\Omega), u^*)$ is a minimizer of $\widetilde{P}_\alpha(f)$.

- Solution of $\widetilde{P}_\alpha(f)$ yields a solution of $P_\alpha(f)$:
  Let $(\tilde{t}^*, \tilde{u}^*)$ be a minimizer of $\widetilde{P}_\alpha(f)$. Then, by above considerations, $\tilde{t}^* = \mathrm{TV}(\tilde{u}^*;\Omega)$ and both problems coincide again. We can conclude that $\tilde{u}^*$ is also minimizer of $P_\alpha(f)$ and both problems are equivalent.

□

As shown above in Lemma 3.4, the constructed auxiliary problem is equivalent to the one we aim to solve. With the above adjustments we have reached differentiability of the objective $\widetilde{J}_\alpha^f$, as stated in the lemma below.

Lemma 3.6 (Differentiability of $\widetilde{J}_\alpha^f$)**.** *The modified objective* $\widetilde{J}_\alpha^f(t, u) = \frac{1}{2}|Ku - f|_2^2 + \alpha t$ *is differentiable.*

Proof. Compute the total derivative $D\widetilde{J}_\alpha^f$ at position $(t, u) \in C$ and evaluate it in direction $(s, v) \in C$. We get

$$\begin{aligned}
D\widetilde{J}_\alpha^f(t,u)[(s,v)] &= D_t\widetilde{J}_\alpha^f(t,u)[s] + D_u\widetilde{J}_\alpha^f(t,u)[v] \\
&= D_t\left(\frac{1}{2}|Ku-f|_2^2+\alpha t\right)[s] + D_u\left(\frac{1}{2}\langle Ku-f, Ku-f\rangle_2+\alpha t\right)[v] \\
&= \alpha s + \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{2}\langle K(u+rv)-f, K(u+rv)-f\rangle_2+\alpha t\right)\Bigg|_{r=0} \\
&= \alpha s + \frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{1}{2}\langle Ku-f, Ku-f\rangle_2 + \frac{r^2}{2}\langle Kv, Kv\rangle_2 + r\langle Ku-f, Kv\rangle_2+\alpha t\right)\Bigg|_{r=0} \\
&= \alpha s + \langle Ku-f, Kv\rangle_2 = \alpha s + \langle K^{\#}(Ku-f), v\rangle_2 \\
&= \alpha s + \int_\Omega K^{\#}(Ku-f)\cdot v\;\mathrm{d}x
\end{aligned} \tag{11}$$

and with this the differentiability of $\widetilde{J}_\alpha^f$. Note, that this derivative is linear in s and v. □

What still needs to be discussed is the question whether the minimization step in line 2 of Algorithm 1, to find

$$(s^{[k]}, v^{[k]}) \in \underset{(s,v)\in C}{\operatorname{argmin}}\; D\widetilde{J}_\alpha^f(t,u)[s,v],$$

is well defined.

In the current setting the space we minimize above, $C = \{(t,u) \in \mathbb{R}_+ \times \mathrm{BV}(\Omega) \mid \mathrm{TV}(u;\Omega) \le t \le M\}$, lacks in compactness.
On the one hand C is convex but not weakly compact, as by [13, Thm. 3.31] this would mean that $\mathrm{BV}(\Omega)$ was reflexive, but it is not as stated in [14, Ch. 2].
On the other hand, we can also not achieve weak*-compactness by interpreting $\mathrm{BV}(\Omega)$ as the dual space of its predual $\mathrm{BV}(\Omega)^{\#}$. This would have sufficed, as the differential from Equation (11) can be represented by $(\alpha, K^{\#}(Ku-f)) \in \mathbb{R}\times\mathrm{BV}(\Omega)^{\#}$ so with weak*-compactness any sequence produced in line 2 of the algorithm would have had a convergent subsequence attaining its limit in C. However, the Banach-Alaoglu Theorem [15, Thm. 3.16] that would have given the desired weak*-compactness can not be applied in the current case, as the considered ball $\{u \in \mathrm{BV}(\Omega) \mid \mathrm{TV}(u;\Omega) \le M\}$ is not closed with respect to the Banach space norm $\|\cdot\|_{\mathrm{BV}} := \|\cdot\|_{L^1} + \mathrm{TV}(\cdot;\Omega)$ on $\mathrm{BV}(\Omega)$ but just with respect to the seminorm $\mathrm{TV}(\cdot;\Omega)$, which is not sufficient.
To rectify this issue, there is the need to adjust the set C in order to achieve closedness of the ball in $\|\cdot\|_{\mathrm{BV}}$. The property that prevents closedness is that the total variation

is invariant under the addition of constants. This can be seen as for every $c \in \mathbb{R}$

$$\begin{aligned}\mathrm{TV}(u+c;\Omega) &= \sup\left\{\int_\Omega (u+c)\,\mathrm{div}\phi\,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega,\mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \le 1\right\}\\ &= \sup\{\int_\Omega u\,\mathrm{div}\phi\,\mathrm{d}x + \underbrace{\int_\Omega c\,\mathrm{div}\phi\,\mathrm{d}x}_{=\int_{\partial\Omega} c\,\phi\cdot\nu\,\mathrm{d}A=0} \mid \phi \in C_c^1(\Omega,\mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \le 1\}\\ &= \sup\left\{\int_\Omega u\,\mathrm{div}\phi\,\mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega,\mathbb{R}^n), \|\phi\|_{L^\infty(\Omega)} \le 1\right\}\\ &= \mathrm{TV}(u;\Omega).\end{aligned}$$

In order to prohibit these vertical shifts, we force the functions $u \in \mathrm{BV}(\Omega)$ to satisfy $\int_\Omega u\ \mathrm{d}x = 0$. With the Poincaré-type inequality from [5, Thm. 3.44], stating that for any $u \in \mathrm{BV}(\Omega)$ with vanishing integral on a bounded connected domain:

$$\|u\|_{L^1(\Omega)} = \int_\Omega |u|\ \mathrm{d}x \le D \cdot \mathrm{TV}(u;\Omega), \qquad D \in \mathbb{R},\ \text{depending only on } \Omega,$$

we can conclude that this additional restriction also limits the L^1-norm of u.
Therefore, every $u \in C^0$ with

$$C^0 = \left\{(t,u) \in \mathbb{R}_+ \times \mathrm{BV}(\Omega) \mid \mathrm{TV}(u;\Omega) \le t \le M, \int_\Omega u\ \mathrm{d}x = 0\right\} \tag{12}$$

also satisfies

$$\|u\|_{\mathrm{BV}} = \mathrm{TV}(u;\Omega) + \|u\|_{L^1(\Omega)} \le M + D \cdot \mathrm{TV}(u;\Omega) \le (1+D)\cdot M.$$

Thus, the created set C^0 is closed in norm, as any pair $(t,u) \in C^0$ satisfies $|t| \le M < \infty$ and $\|u\|_{\mathrm{BV}} \le (1 + D \cdot M) < \infty$ and the Banach-Alaoglu Theorem can be applied in order to ensure weak*-compactness.

Remark 3.7 (Modified Problem). The new setting in which we want to solve the TV regularized problem $P_\alpha(f)$ reads as follows:
Let

$$\mathrm{BV}^0(\Omega) := \left\{u \in \mathrm{BV}(\Omega) \mid \int_\Omega u\ \mathrm{d}x = 0\right\}. \tag{13}$$

This still forms a Banach space as it is well-defined under addition and scalar multiplication and inherits the Banach space properties from $\mathrm{BV}(\Omega)$. Moreover define the adjusted measurement operator and its predual

$$K^0 := K|_{\mathrm{BV}^0(\Omega)} : \mathrm{BV}(\Omega)^0 \to Y \quad \text{and} \quad (K^0)^\# : Y \to \mathrm{BV}^0(\Omega)^\#. \tag{14}$$

With this the initial optimization problem $(P_\alpha(f))$ restricts to

$$\min_{u\in \mathrm{BV}^0(\Omega)} J_\alpha^{f,0}(t,u), \quad J_\alpha^{f,0}(t,u) = \frac{1}{2\alpha}|K^0 u - f|_2^2 + \mathrm{TV}(u;\Omega) \qquad (P_\alpha^{\mathrm{BV}^0}(f))$$

respectively

$$\min_{(t,u)\in C^0} \widetilde{J}_\alpha^{f,0}(t,u), \quad \widetilde{J}_\alpha^{f,0}(t,u) = \frac{1}{2}|K^0 u - f|_2^2 + \alpha t \qquad (\widetilde{P}_\alpha^{\mathrm{BV}^0}(f))$$

with $C^0 := \Big\{(t,u) \in \mathbb{R}_+ \times \mathrm{BV}^0(\Omega) \;\big|\; \mathrm{TV}(u;\Omega) \le t \le M\Big\}$.

Lemma 3.8 (Well-posedness of $\widetilde{P}_\alpha^{\mathrm{BV}^0}(f)$)**.** *The optimization problem $\widetilde{P}_\alpha^{\mathrm{BV}^0}(f)$ is well-posed to be solved by the Frank-Wolfe algorithm.*

Proof. The differentiability of the objective $\widetilde{J}_\alpha^{f,0}(t,u)$ can be obtained exactly as in the proof of Lemma 3.6 with derivative

$$D\widetilde{J}_\alpha^{f}(t,u)[(s,v)] = \alpha s + \int_\Omega (K^0)^\#(K^0 u - f)\cdot v \, \mathrm{d}x. \tag{15}$$

The well-definedness of the minimization step now follows immediately from the considerations before, as we have now assured that C^0 is closed and the Banach-Alaoglu Theorem [15, Thm. 3.16] can be applied. □

3.2. Minimization step

The core of the Frank-Wolfe algorithm is the minimization step in line 2 of Algorithm 1, in our application finding

$$(s^{[k]}, v^{[k]}) \in \operatorname*{argmin}_{(s,v)\in C^0} D\widetilde{J}_\alpha^{f,0}(t,u)[s,v].$$

To characterize solutions of this optimization problem more precisely, we make use of Proposition 3.3. As proved in the previous section, C^0 is nonempty since $(t,u) = (t,0) \in C^0$ for some $t \le M$, weak*-compact and convex. Moreover $D\widetilde{J}_\alpha^{f,0}(t,u)$ is linear and weak*-continuous since it is an element from $\mathrm{BV}^0(\Omega)^\#$ and thus weak*-continuous by definition of the weak*-topology. This shows that the requirements of Proposition 3.3 are satisfied and thus the minimizer in each iteration can be found among the extreme points of the convex set C^0.

The aim of the following section is therefore to characterize the extreme points of $C^0 = \Big\{(t,u) \in \mathbb{R}_+ \times \mathrm{BV}^0(\Omega) \;\big|\; \mathrm{TV}(u;\Omega) \le t \le M\Big\}$ in order to determine the possible minimizers that are created in line 2 of Algorithm 1 by minimizing the linear functional $D\widetilde{J}_\alpha^{f}$.

Note that extreme points of C^0 are of the form

$$(0,0) \quad \text{or} \quad (M, u^{\text{ext}})$$

with u^{ext} an extreme point of the ball $\mathcal{B}^{M,0} := \left\{u \in \mathrm{BV}^0(\Omega) \mid \mathrm{TV}(u;\Omega) \leq M\right\}$. Thus the next step is to analyze the structure of the extreme points of $\mathcal{B}^{M,0}$.

3.2.1. Extreme points of the TV ball

This passage is dedicated to characterize the extreme points of $\mathcal{B}^{M,0}$ and relies in large parts on the work done in [16] and [17].
In the beginning we need to establish some theoretical background.

Definition 3.9 (Perimeter, [5, Def. 3.35])**.** Let $E \subset \Omega$ be a measurable set and Ω an open set in $\mathbb{R}^n$. The *perimeter* of E in Ω is defined as the total variation of the indicator function of the set E, i.e.

$$\mathrm{Per}(E;\Omega) = \mathrm{TV}(\mathbb{1}_E;\Omega) = \sup\left\{\int_E \mathrm{div}\phi \ \mathrm{d}x \mid \phi \in C_c^1(\Omega;\mathbb{R}^n), \|\phi\|_\infty \leq 1\right\}.$$

A set E has *finite perimeter* in Ω if $\mathrm{Per}(E;\Omega) < \infty$ is finite.

Lemma 3.10 ([5, Prop. 3.38])**.** *With the notation from Definition 3.9 one gets the equality*

$$\mathrm{Per}(E;\Omega) = \mathrm{Per}(\Omega \setminus E;\Omega).$$

The next theorem needed in the following section is the so-called *coarea formula.* This gives another characterization of the perimeter of a function.

Theorem 3.11 (Coarea formula, [5, Thm. 3.40])**.** *Let $\Omega \subset \mathbb{R}^n$ be an open set and $u \in \mathrm{BV}(\Omega)$. Define the sets*

$$U^{(t)} = \{x \in \Omega \mid u(x) > t\}.$$

Then, the following equality holds:

$$\mathrm{TV}(u;\Omega) = \int_{-\infty}^{\infty} \mathrm{Per}(U^{(t)};\Omega) \ \mathrm{d}t.$$

Remark 3.12. From Lemma 3.10 one can deduce that the coarea formula also holds for level sets defined as

$$\widetilde{U}^{(t)} = \{x \in \Omega \mid u(x) \leq t\}.$$

Definition 3.13 (Decomposable set, [16, Def. 4.4])**.** Let $E \subset \Omega$ be a set of finite perimeter. E is called *decomposable* if there exists a partition of E into two sets A and B i.e. $A \cup B = E$ and $A \cap B = \emptyset$, such that $|A| > 0$, $|B| > 0$ and $\mathrm{Per}(E;\Omega) =$

$\mathrm{Per}(A;\Omega) + \mathrm{Per}(B;\Omega)$.
A set of finite perimeter is called *indecomposable* if it is not decomposable.

This directly leads to the family of simple sets, that are going to play a crucial role in the characterization of the extreme points of $\mathcal{B}^{M,0}$.

Definition 3.14 (Simple set, [16, Def. 4.5])**.** A set of finite perimeter $E \subset \Omega$ is called *simple* if both E and $\Omega \setminus E$ are indecomposable.

Definition 3.15 (Measure theoretic interior and exterior, [16])**.** Let us further define the *measure theoretic interior* of E as

$$E^1 := \left\{ x \in \mathbb{R}^d | \lim_{r\to 0} \frac{|E \cap B_r(x)|}{|B_r(x)|} = 1 \right\},$$

where $B_r(x)$ denotes the ball of radius r around the point $x \in \mathbb{R}^d$. Analogously, define the *measure theoretic exterior* as

$$E^0 := \left\{ x \in \mathbb{R}^d \mid \lim_{r\to 0} \frac{|E \cap B_r(x)|}{|B_r(x)|} = 0 \right\}.$$

Moreover, the *essential boundary* of E is defined as $\partial^* E = \mathbb{R}^d \setminus (E^0 \cup E^1)$.

Theorem 3.16 (Constancy Theorem, [16, Lem. 4.6])**.** *Let $u \in \mathrm{BV}(\Omega)$ and $E \subset \Omega$ an indecomposable set such that*

$$\mathrm{TV}(u; E^1) = 0.$$

Then there exists $c \in \mathbb{R}$ such that $u(x) = c$ almost everywhere in E.

Proof. A proof of the above theorem can be found in [18, Prop. 2.15]. □

The first step to analyze the extreme points of the target set $\mathcal{B}^{M,0}$ is to understand the structure of the extreme points of the related set $\mathcal{B}^M_{\mathcal{N}}$ introduced below in equation (16).

Let $\mathcal{N}$ be the nullspace of the total variation, i.e.

$$\mathcal{N} := \{u \in \mathrm{BV}(\Omega) \mid \mathrm{TV}(u;\Omega) = 0\}.$$

Remark 3.17. Constant functions $u \equiv c$ for some $c \in \mathbb{R}$ clearly belong to the nullspace of the total variation, in other words $\mathbb{R} \subseteq \mathcal{N}$. Moreover, elements of $\mathcal{N}$ need to be constant on Ω by definition of the total variation. As Ω is a connected set, those functions u are solely allowed to take a single value on the whole domain, so we get equality, which means $\mathcal{N} = \mathbb{R}$.

The nullspace is a closed subspace of the space of functions of bounded variation and thus it is valid to consider the quotient space

$$\mathrm{BV}(\Omega)_{\mathcal{N}} := \mathrm{BV}(\Omega)/\mathcal{N} = \{u_{\mathcal{N}} = u + \mathcal{N} \mid u \in \mathrm{BV}(\Omega)\}$$

endowed with the quotient topology. Furthermore, let

$$TV_{\mathcal{N}} : \mathrm{BV}(\Omega)_{\mathcal{N}} \to [0, +\infty) \quad \text{with} \quad \mathrm{TV}_{\mathcal{N}}(u_{\mathcal{N}}; \Omega) := \mathrm{TV}(u; \Omega).$$

This describes a well-defined map on the quotientspace $\mathrm{BV}(\Omega)_{\mathcal{N}}$, as the total variation is constant on each equivalence class $u_{\mathcal{N}}$. Having this, one can consider the ball of radius M in the quotientspace $\mathrm{BV}(\Omega)_{\mathcal{N}}$, defined by

$$\mathcal{B}_{\mathcal{N}}^{M} := \{u_{\mathcal{N}} \in \mathrm{BV}(\Omega)_{\mathcal{N}} \mid \mathrm{TV}_{\mathcal{N}}(u_{\mathcal{N}}; \Omega) \leq M\}. \tag{16}$$

Remark 3.18. Notice that the set $\mathcal{B}^{M,0}$ can be regarded as a canonical choice of representatives for the equivalence classes of $\mathcal{B}_{\mathcal{N}}^{M}$, by selecting for each class the unique element with vanishing mean, which is obtained by subtracting the constant $\frac{1}{|\Omega|}\int_\Omega u \, \mathrm{d}x \in \mathcal{N}$ from u.
Thus, it is natural to consider the ball in the quotient space $\mathcal{B}_{\mathcal{N}}^{M}$ and determine extreme points there in order to deduce from these results the extreme points of the desired set $\mathcal{B}^{M,0}$.

To determine the extreme points of $\mathcal{B}_{\mathcal{N}}^{M}$, denoted by $\mathrm{Ext}(\mathcal{B}_{\mathcal{N}}^{M})$, we base our argument on Theorem 4.7 from [16], providing a more detailed version of the proof than presented in the original source. In addition, we adapt both the statement and the proof to the case where the ball has radius M.

Theorem 3.19 (Extreme points of $\mathcal{B}_{\mathcal{N}}^{M}$, [16, Thm. 4.7])**.** *We have*

$$Ext(\mathcal{B}_{\mathcal{N}}^{M}) = \left\{ \pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E; \Omega)} + \mathcal{N} \mid E \text{ simple} \right\}.$$

Proof. We proceed to prove the equality of sets above by proving each inclusion separately.

Step 1 ("$\subseteq$"):

Firstly, note that for any $u_{\mathcal{N}} \in \mathrm{Ext}(\mathcal{B}_{\mathcal{N}}^{M})$ must hold $u_{\mathcal{N}} \neq \mathcal{N}$. If this was not the case, i.e. if $u_{\mathcal{N}} = \mathcal{N}$, then we could write with any E being a connected simple set $u_{\mathcal{N}} = \frac{1}{2}(\frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + \mathcal{N}) + \frac{1}{2}(-\frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + \mathcal{N})$, which is a non-trivial convex combination contradicting the extremality of $u_{\mathcal{N}} = \mathcal{N}$.
Therefore, let $u_{\mathcal{N}} \in Ext(\mathcal{B}_{\mathcal{N}}^{M})$ and choose a suitable $u \in \mathrm{BV}(\Omega)$ such that $u + \mathcal{N} = u_{\mathcal{N}}$ and $u \notin \mathcal{N}$. Then $\mathrm{TV}(u; \Omega) \neq 0$. With $0 \neq \mathrm{TV}(u; \Omega) = \mathrm{TV}_{\mathcal{N}}(u_{\mathcal{N}}, \Omega) \in \{0, M\}$, it is obvious that $\mathrm{TV}(u; \Omega) = M$.

The next goal is to show that u attains only two values almost everywhere in Ω. Therefore, define the functional $F : [-\infty, +\infty] \to \mathbb{R}$ by

$$F(s) = \int_{-\infty}^{s} \mathrm{Per}(\{u(x) \le t\}; \Omega)\ \mathrm{d}t.$$

This satisfies $F(-\infty) = 0$ and with the Coarea Formula 3.11 and Remark 3.12 follows

$$F(+\infty) = \int_{-\infty}^{+\infty} \mathrm{Per}(\{u(x) \le t\}; \Omega)\ \mathrm{d}t = \mathrm{TV}(u; \Omega) = M.$$

Moreover, as the function $t \mapsto \mathrm{Per}(\{u(x) \le t\}; \Omega)$ is integrable on $\mathbb{R}$, F becomes a continuous functional. Thus, the intermediate value theorem ([e.g. 19]) can be applied and yields the existence of an $s \in \mathbb{R}$ such that $F(s) = \frac{M}{2}$. Set

$$u_1 := 2\min(u, s), \quad u_2 := 2\max(u - s, 0).$$

Then

$$\frac{1}{2}u_1 + \frac{1}{2}u_2 = \min(u, s) + \max(u - s, 0) = \left\{ \begin{array}{ll} u + 0 = u & \text{, if } u < s \\ u + 0 = s + 0 & \text{, if } u = s \\ s + u - s = u & \text{, if } u > s \end{array} \right\} = u.$$

Furthermore, $\mathrm{TV}(u_1; \Omega) = \mathrm{TV}(u_2; \Omega) = M$, due to the following calculation.

$$\begin{aligned} \frac{M}{2} &= F(s) = \int_{-\infty}^{s} \mathrm{Per}(\{u(x) \le t\}; \Omega)\ \mathrm{d}t = \int_{-\infty}^{s} \mathrm{Per}(\{\frac{1}{2}u_1(x) \le t\}; \Omega)\ \mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \mathrm{Per}(\{\frac{1}{2}u_1(x) \le t\}; \Omega)\ \mathrm{d}t - \int_{s}^{+\infty} \underbrace{\mathrm{Per}(\{\frac{1}{2}u_1(x) \le t\}; \Omega)}_{\overset{\text{Lem. 3.10}}{=} \mathrm{Per}(\{\frac{1}{2}u_1(x) > t\}; \Omega) \overset{\frac{1}{2}u_1 \le s}{=} \mathrm{Per}(\emptyset; \Omega) = 0} \mathrm{d}t \\ &= \int_{-\infty}^{+\infty} \mathrm{Per}(\{\frac{1}{2}u_1(x) \le t\}; \Omega)\ \mathrm{d}t \overset{\text{Thm. 3.11}}{=} \mathrm{TV}(\frac{1}{2}u_1; \Omega) = \frac{1}{2}\mathrm{TV}(u_1; \Omega) \\ &\Rightarrow \mathrm{TV}(u_1; \Omega) = M \end{aligned}$$

A similar calculation yields $\mathrm{TV}(u_2; \Omega) = M$.
We have chosen $u_{\mathcal{N}}$ to be an extreme point of $\mathcal{B}_{\mathcal{N}}^{M}$, so $u_{\mathcal{N}} = (u_1)_{\mathcal{N}} = (u_2)_{\mathcal{N}}$ as the convex combination from above would not be possible otherwise. Due to Remark 3.17, we can conclude that there exist $c_1, c_2 \in \mathbb{R}$ such that $u = u_1 + c_1 = u_2 + c_2$.
Let $x \in \Omega$ such that $u(x) \ge s$. Then,

$$u(x) = u_1(x) + c_1 = 2 \cdot \min(u(x), s) + c_1 = 2s + c_1.$$

Similarly, we get for $x \in \Omega$ with $u(x) < s$ that

$$u(x) = u_2 + c_2 = 2 \cdot \max(u(x) - s, 0) + c_2 = c_2.$$

This observation shows, that u takes at most two values everywhere, namely $2s + c_1$ or c_2. From the property that $\mathrm{TV}(u;\Omega) = M \neq 0$, it becomes clear that u cannot be a constant function and therefore needs to take at least two values, so altogether, u takes exactly two values. Moreover, from $2s + c_1 \geq s$ and $c_2 < s$ it follows $2s + c_1 > c_2$.

As the representatives $\tilde{u}$ of the equivalence class $u_{\mathcal{N}}$ are of the form $\tilde{u} = u + c$, where $c \in \mathbb{R}$ can be chosen arbitrarily, we could pick without loss of generality the representative with $c = -c_2$ or the one with $c = -2s - c_1$. Then, $\tilde{u}$ takes either the values $c_2 - c_2 = 0$ and $2s + c_1 - c_2 =: a > 0$ or the values $2s + c_1 - 2s - c_1 = 0$ and $c_2 - 2s - c_1 =: a < 0$. For simplicity, $\tilde{u}$ will still be denoted by u. Thus, we can suppose that $u(x) \in \{0, a\}$ almost everywhere.
Now, define the set $E := \{x \in \Omega \mid u(x) = a\}$. Having this, it is clear, that u needs to be a scaled version of the characteristic function $\pm\mathbb{1}_E$, as u vanishes outside of E and is constant on the inside. To determine the scaling d, we exploit $\mathrm{TV}(u;\Omega) = M$. Thus,

$$\begin{aligned} M = \mathrm{TV}(u;\Omega) =& \mathrm{TV}(d \cdot \pm\mathbb{1}_E;\Omega) = \pm d \cdot \mathrm{TV}(\mathbb{1}_E;\Omega) = \pm d \cdot \mathrm{Per}(E;\Omega) \\ &\Rightarrow d = \pm\frac{M}{\mathrm{Per}(E;\Omega)} \end{aligned}$$

and u is of the form $u = \pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)}$.

It is still left to show, that the constructed set E is a simple set in Ω. Therefore it is sufficient to show E and $\Omega \setminus E$ are not decomposable.

E is not decomposable:
Let by contradiction E be decomposable. Then, there exists a partition of E in sets A and B of finite perimeter satisfying $\mathrm{Per}(A;\Omega) > 0$, $\mathrm{Per}(B;\Omega) > 0$ and $\mathrm{Per}(E;\Omega) = \mathrm{Per}(A;\Omega) + \mathrm{Per}(B;\Omega)$. Define

$$u_1 = \pm M \cdot \frac{\mathbb{1}_A}{\mathrm{Per}(A;\Omega)} \quad \text{and} \quad u_2 = \pm M \cdot \frac{\mathbb{1}_B}{\mathrm{Per}(B;\Omega)}.$$

With this we get

$$\begin{aligned} u =& \pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} \\ =& \pm M \cdot \frac{\mathbb{1}_A + \mathbb{1}_B}{\mathrm{Per}(E;\Omega)} = \pm M \left(\frac{\mathbb{1}_A}{\mathrm{Per}(A;\Omega)} \cdot \frac{\mathrm{Per}(A;\Omega)}{\mathrm{Per}(E;\Omega)} + \frac{\mathbb{1}_B}{\mathrm{Per}(B;\Omega)} \cdot \frac{\mathrm{Per}(B;\Omega)}{\mathrm{Per}(E;\Omega)} \right) \\ =& \frac{\mathrm{Per}(A;\Omega)}{\mathrm{Per}(E;\Omega)} u_1 + \frac{\mathrm{Per}(B;\Omega)}{\mathrm{Per}(E;\Omega)} u_2 \end{aligned}$$

and have found a non-trivial convex combination of u, which forms a contradiction to u being an extreme point. Therefore, E is an indecomposable set.

$\Omega \setminus E$ is not decomposable:
Again, suppose by contradiction that $\Omega \setminus E$ is decomposable. Let $\Omega \setminus E = A \cup B$ be a

suitable decomposition. Define, similar to the above computations,

$$u_1 = \mp M \cdot \frac{\mathbb{1}_A}{\mathrm{Per}(A;\Omega)} \quad \text{and} \quad u_2 = \pm M \cdot \frac{1 - \mathbb{1}_B}{\mathrm{Per}(B:\Omega)}.$$

With this we get

$$\begin{aligned} u &= \pm\, M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} = \pm M \cdot \frac{1 - \mathbb{1}_{\Omega\setminus E}}{\mathrm{Per}(E;\Omega)} = \pm M \cdot \frac{1 - \mathbb{1}_A - \mathbb{1}_B}{\mathrm{Per}(E;\Omega)} \\ &= \pm M \cdot \frac{1 - \mathbb{1}_B}{\mathrm{Per}(B;\Omega)} \cdot \frac{\mathrm{Per}(B;\Omega)}{\mathrm{Per}(E;\Omega)} \mp M \cdot \frac{\mathbb{1}_A}{\mathrm{Per}(A;\Omega)} \cdot \frac{\mathrm{Per}(A;\Omega}{\mathrm{Per}(E;\Omega)} \\ &= \frac{\mathrm{Per}(B;\Omega)}{\mathrm{Per}(E;\Omega)} u_2 + \frac{\mathrm{Per}(A;\Omega)}{\mathrm{Per}(E;\Omega)} u_1 \end{aligned}$$

With Lemma 3.10,

$$\frac{\mathrm{Per}(A;\Omega)}{\mathrm{Per}(E;\Omega)} + \frac{\mathrm{Per}(B;\Omega)}{\mathrm{Per}(E;\Omega)} = \frac{\mathrm{Per}(A;\Omega) + \mathrm{Per}(B;\Omega)}{\mathrm{Per}(\Omega \setminus E;\Omega)} = \frac{\mathrm{Per}(\Omega \setminus E;\Omega)}{\mathrm{Per}(\Omega \setminus E;\Omega)} = 1.$$

Therefore, the above is again an non-trivial convex combination of u and forms a contradiction for u being an extreme point. This shows that $\Omega \setminus E$ is not decomposable and together with the part above this proves that E is a simple set. Thus, the first inclusion $Ext(\mathcal{B}_{\mathcal{N}}^M) \subseteq \left\{ \pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + \mathcal{N} \mid E \textit{ simple} \right\}$ is proven.

Step 2 ("$\supseteq$"):

Now consider the opposite inclusion. Let $E \subset \Omega$ be a given simple set. Suppose by contradiction that $\pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + \mathcal{N}$ is no extreme point of $Ext(\mathcal{B}_{\mathcal{N}}^M)$, i.e. there exist $u_1, u_2 \in \mathrm{BV}(\Omega)$ with $\mathrm{TV}(u_1;\Omega) \leq M$, $\mathrm{TV}(u_2;\Omega) \leq M$ and

$$\pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + \mathcal{N} = \lambda(u_1 + \mathcal{N}) + (1 - \lambda)(u_2 + \mathcal{N}),$$

with $\lambda \in (0,1)$ and $u_1 + \mathcal{N} \neq u_2 + \mathcal{N}$.
In the next step we can choose representatives of the classes $u_i + \mathcal{N}$ with $i \in \{1,2\}$ and a representative of $\pm M \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + \mathcal{N}$, which means to choose arbitrary $c_0, c_1, c_2 \in \mathbb{R}(= \mathcal{N})$. Then:

$$\begin{aligned} \pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + c_0 &= \lambda(u_1 + c_1) + (1 - \lambda)(u_2 + c_2) \\ \Leftrightarrow \pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + c &= \lambda u_1 + (1 - \lambda) u_2 \end{aligned} \tag{17}$$

for some $c \in \mathbb{R}$, depending on c_0, c_1 and c_2. By building the derivative, this gives the equation

$$\pm M \cdot \frac{D\mathbb{1}_E}{\operatorname{Per}(E;\Omega)} = \lambda Du_1 + (1-\lambda)Du_2. \tag{18}$$

Moreover, for any measurable $A \subset \Omega$ one gets the equality

$$M \cdot \frac{|D\mathbb{1}_E|(A)}{\operatorname{Per}(E;\Omega)} = \lambda|Du_1|(A) + (1-\lambda)|Du_2|(A). \tag{19}$$

To show this, it is a quite obvious observation that due to the triangle inequality we have " $\leq$ " in (19), so it remains to show, that " $<$ " is not possible for any $A \subset \Omega$.

Proof of (19): Again, show the claim by contradiction. Assume that there exists a measurable set $A \subset \Omega$ such that $M \cdot \frac{|D\mathbb{1}_E|(A)}{\operatorname{Per}(E;\Omega)} < \lambda|Du_1|(A) + (1-\lambda)|Du_2|(A)$. Then, as $|D\mathbb{1}_E|(\Omega) = \operatorname{TV}(\mathbb{1}_E;\Omega) = \operatorname{Per}(E;\Omega)$:

$$\begin{aligned}
M = M \cdot \frac{|D\mathbb{1}_E|(\Omega)}{\operatorname{Per}(E;\Omega)} &\overset{\substack{A\cap A^c=\emptyset,\\ A\cup A^c=\Omega}}{=} M \cdot \left(\frac{|D\mathbb{1}_E|(A)}{\operatorname{Per}(E;\Omega)} + \frac{|D\mathbb{1}_E|(A^c)}{\operatorname{Per}(E;\Omega)}\right) \\
&< \lambda|Du_1|(A) + (1-\lambda)|Du_2|(A) + M \cdot \frac{|D\mathbb{1}_E|(A^c)}{\operatorname{Per}(E;\Omega)} \\
&< \lambda|Du_1|(A) + (1-\lambda)|Du_2|(A) + \lambda|Du_1|(A^c) + (1-\lambda)|Du_2|(A^c) \\
&\overset{\substack{A\cap A^c=\emptyset\\ A\cup A^c=\Omega}}{=} \lambda \underbrace{|Du_1|(\Omega)}_{\leq M} + (1-\lambda)\underbrace{|Du_2|(\Omega)}_{\leq M} \\
&\leq M \qquad \Rightarrow M < M
\end{aligned}$$

This shows that the assumption leads to a contradiction and thus proves (19).

Clearly, $|D\mathbb{1}_E|$ is just supported on the essential boundary $\partial^* E$ (Def. 3.15), as it is the derivative of the characteristic function of a set of finite perimeter. Therefore, $|D\mathbb{1}_E|(E^1) = |D\mathbb{1}_E|(E^0) = 0$ and with (19), as E^1 and E^0 are measurable sets, we get

$$\begin{aligned}
&|Du_1|(E^0) = |Du_1|(E^1) = |Du_2|(E^0) = |Du_2|(E^1) = 0 \\
\overset{\text{Prop. 2.5}}{\Leftrightarrow}\ &\operatorname{TV}(u_1;E^0) = \operatorname{TV}(u_1;E^1) = \operatorname{TV}(u_2;E^0) = \operatorname{TV}(u_2;E^1) = 0.
\end{aligned}$$

This shows that the requirements of the constancy theorem (Thm. 3.16) are satisfied for the indecomposable sets E and $\Omega \setminus E$. Note, that when considering $\Omega \setminus E$ as an indecomposable set, $(\Omega \setminus E)^1 = E^0$. Applying Thm. 3.16, one gets that u_1 and u_2 are both constant on E and on $\Omega \setminus E$, i.e. there are constants $c_1, c_2, d_1, d_2 \in \mathbb{R}$ such that

$$u_1 = d_1 \cdot \mathbb{1}_E + c_1 \quad \text{and} \quad u_2 = d_2 \cdot \mathbb{1}_E + c_2. \tag{20}$$

Furthermore, from $M = \lambda \cdot \operatorname{TV}(u_1;\Omega) + (1-\lambda)\cdot\operatorname{TV}(u_2;\Omega)$ due to Equation (19) and with $\operatorname{TV}(u_1;\Omega) \leq M$ as well as $\operatorname{TV}(u_2;\Omega) \leq M$ by assumption, it follows that $\operatorname{TV}(u_1;\Omega) =$

$\mathrm{TV}(u_2;\Omega) = M$. Combining this with the the explicit formulas from above yields

$$M = \mathrm{TV}(u_i;\Omega) = \mathrm{TV}(d_i \cdot \mathbb{1}_E + c_i;\Omega) = |d_i| \cdot \mathrm{TV}(\mathbb{1}_E;\Omega) = |d_i| \cdot \mathrm{Per}(E;\Omega)$$
$$|d_i| = \frac{M}{\mathrm{Per}(E;\Omega)} > 0 \qquad \text{for } i \in \{1,2\}.$$

The next step is to determine the signs of d_1 and d_2.

It can be seen that d_1 and d_2 cannot have an opposite sign. Assume, they had. From (18) and with $Du_i = d_i \cdot D\mathbb{1}_E$ for $i \in \{1,2\}$ it follows

$$\begin{aligned}
&M \cdot \underbrace{\frac{|D\mathbb{1}_E|(\Omega)}{\mathrm{Per}(E;\Omega)}}_{=1} = |\lambda D u_1 + (1-\lambda) D u_2|(\Omega) \\
\Leftrightarrow\ &M = |\lambda d_1 \cdot D\mathbb{1}_E + (1-\lambda) d_2 \cdot D\mathbb{1}_E|(\Omega) = |\lambda d_1 + (1-\lambda) d_2| \cdot |D\mathbb{1}_E|(\Omega) \\
\Leftrightarrow\ &M = |\lambda d_1 + (1-\lambda) d_2| \cdot \mathrm{Per}(E;\Omega) \\
&\overset{\substack{\text{assumption of} \\ \text{opposite sign}}}{<} (\lambda |d_1| + (1-\lambda)|d_2|) \cdot \mathrm{Per}(E;\Omega) \\
&= \left(\lambda \frac{M}{\mathrm{Per}(E;\Omega)} + (1-\lambda) \frac{M}{\mathrm{Per}(E;\Omega)}\right) \cdot \mathrm{Per}(E;\Omega) = M
\end{aligned}$$

This forms a contradiction, so sign d_1 = sign d_2.

Moreover the signs of d_1 and d_2 need to coincide with the sign of $\pm M \cdot (\mathrm{Per}(E;\Omega))^{-1}$. This is due to (17):

$$\begin{aligned}
&\pm M \cdot \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} + c = \lambda u_1 + (1-\lambda) u_2 \\
\Leftrightarrow\ &\pm M \cdot (\mathrm{Per}(E;\Omega))^{-1} \cdot \mathbb{1}_E + c = \lambda(d_1 \mathbb{1}_E + c_1) + (1-\lambda)(d_2 \mathbb{1}_E + c_2) \\
\Leftrightarrow\ &\pm M \cdot (\mathrm{Per}(E;\Omega))^{-1} \cdot \mathbb{1}_E + c = (\lambda d_1 + (1-\lambda) d_2)\mathbb{1}_E + (\lambda c_1 + (1-\lambda) c_2)
\end{aligned}$$

If the prefactor of the characteristic function on the left hand side $\pm M \cdot (\mathrm{Per}(E;\Omega))^{-1}$ is larger than zero, also the prefactor on the right side $\lambda d_1 + (1-\lambda) d_2$ needs to be larger than zero to ensure a positive bulge over the set E. This is only possible, if at least one of d_1 and d_2 is positive, and with the above considerations also the other one has to have a positive sign. Similarly, if the left hand prefactor is negative also d_1 and d_2 have a negative sign.
Hence, $d_1 = d_2 = \pm \frac{M}{\mathrm{Per}(E;\Omega)}$. Now, one can put this into (20) and receive

$$u_1 = \pm \frac{M}{\mathrm{Per}(E;\Omega)} \mathbb{1}_E + c_1 \quad \text{and} \quad u_2 = \pm \frac{M}{\mathrm{Per}(E;\Omega)} \mathbb{1}_E + c_2$$

with $c_1, c_2 \in \mathbb{R}(=\mathcal{N})$, which means

$$u_1 + \mathcal{N} = \pm \frac{M}{\mathrm{Per}(E;\Omega)} \mathbb{1}_E + \mathcal{N} = u_2 + \mathcal{N}.$$

This contradicts the assumption that $\pm\frac{M}{\mathrm{Per}(E;\Omega)}\mathbb{1}_E + \mathcal{N}$ is no extreme point of $Ext(\mathcal{B}^M_\mathcal{N})$ and thus proves the second inclusion of the claim and finishes the proof. □

The previous theorem has shown, that there is a good characterization of the extreme points of $\mathcal{B}^M_\mathcal{N}$. To derive the extreme points of $\mathcal{B}^{M,0}$, we transfer the calculated extreme points of $\mathcal{B}^M_\mathcal{N}$ into extreme points of $\mathcal{B}^{M,0}$. To do so, the next theorem can be applied.

Theorem 3.20 ([20, Theorem 9.2.3])**.** *Let X and Y be linear spaces, K a convex subset of X, and A an affine map of X into Y. If A is injective, then*

$$\mathrm{ext}A(K) = A(\mathrm{ext}K).$$

Proof. First of all recall from the definition of affinity [e.g. 20, Definition 9.1.4] that a map $A : X \to Y$ is called affine, if $A(tx + (1-t)y) = tAx + (1-t)Ay$ for all $x, y \in X$ and $t \in [0, 1]$.

- $\mathrm{ext}A(K) \subset A(\mathrm{ext}K)$
  Let Ax be an extreme point of $A(K)$. Show that x is an extreme point of K. Therefore, let $x = ty+(1-t)z$ with $y, z \in K$ and $t \in (0, 1)$ be a convex combination of x. Now, apply A on both sides of the convex combination and yield

$$Ax = A(ty + (1-t)z) \overset{A \text{ affine}}{=} tAy + (1-t)Az.$$

  As Ax was supposed to be an extreme point of $A(K)$, one can conclude $Ax = Ay = Az$. As A is injective, also $x = y = z$, which proves, that x is an extreme point of K.
- $\mathrm{ext}A(K) \supset A(\mathrm{ext}K)$
  Now, let x be an extreme point of K and show that Ax is an extreme point of $A(K)$. Again, let $Ax = tAy + (1-t)Az$ be a convex combination of Ax, i.e. $Ay, Az \in A(K), t \in (0, 1)$. By affinity of A, one gets $Ax = A(ty + (1-t)z)$ as above and with injectivity $x = ty + (1-t)z$. Due to the choice of x as an extreme point of K, we can conclude $x = y = z$, and with this $Ax = Ay = Az$, which shows that Ax is an extreme point of $A(K)$.

□

As explained before and as done in [17, Prop. 3.1], we will use Theorem 3.19 and Theorem 3.20 to describe the extreme points of $\mathcal{B}^{M,0}$:

Theorem 3.21 (Extreme points of $\mathcal{B}^{M,0}$, [17, Prop. 3.1])**.** *We have*

$$Ext(\mathcal{B}^{M,0}) = \left\{ \pm \frac{M}{\operatorname{Per}(E;\Omega)} \mathring{\mathbb{1}}_E \mid E \,\text{simple} \right\}$$

with $\mathring{u} := u - \frac{1}{|\Omega|}\int_\Omega u(x)\,\mathrm{d}x$ *for any* $u \in \mathrm{BV}(\Omega)$.

Proof. Consider the setup and notation of Theorem 3.19. We have already shown in that theorem, that the extreme points of $\mathcal{B}^M_{\mathcal{N}}$ are of the form

$$Ext(\mathcal{B}^M_{\mathcal{N}}) = \left\{ \left(\pm \frac{M}{\operatorname{Per}(E;\Omega)} \cdot \mathbb{1}_E \right)_{\mathcal{N}} \mid E \,\text{simple} \right\}.$$

According to Theorem 3.20 we aim at finding an affine and injective map $A : \mathrm{BV}(\Omega)_{\mathcal{N}} \to \mathrm{BV}^0(\Omega)$. Define this map via

$$A(u_{\mathcal{N}}) := u - \frac{1}{|\Omega|}\int_\Omega u(x)\,\mathrm{d}x = \mathring{u}.$$

Then:

- A is well-defined: Let $u_{\mathcal{N}} = v_{\mathcal{N}}$, which means $u = v + c$ for some $c \in \mathbb{R}$. Then

$$\begin{aligned} A(u_{\mathcal{N}}) &= u - \frac{1}{|\Omega|}\int_\Omega u(x)\,\mathrm{d}x \\ &= v + c - \frac{1}{|\Omega|}\int_\Omega (v+c)(x)\,\mathrm{d}x \\ &= \underbrace{c - \frac{1}{|\Omega|}|\Omega| c}_{=0} + v - \frac{1}{|\Omega|}\int_\Omega v(x)\,\mathrm{d}x = A(v_{\mathcal{N}}), \end{aligned}$$

  so A is independent of the representative of the equivalence class and thus well-defined on $\mathrm{BV}(\Omega)_{\mathcal{N}}$.

- A maps into $\mathrm{BV}^0(\Omega)$: Let $u_{\mathcal{N}} \in \mathrm{BV}(\Omega)_{\mathcal{N}}$. Then, $A(u_{\mathcal{N}})$ clearly is a $\mathrm{BV}(\Omega)$-function, as $u \in \mathrm{BV}(\Omega)$ and the total variation does not change when a constant $c \in \mathbb{R}$ is added. Furthermore,

$$\begin{aligned} \int_\Omega A(u_{\mathcal{N}}) &= \int_\Omega \left(u(y) - \frac{1}{|\Omega|}\int_\Omega u(x)\,\mathrm{d}x \right) \mathrm{d}y \\ &= -\frac{|\Omega|}{|\Omega|}\int_\Omega u(x)\,\mathrm{d}x + \int_\Omega u(y)\,\mathrm{d}y = 0, \end{aligned} \tag{21}$$

  so $A(u_{\mathcal{N}}) \in \mathrm{BV}^0(\Omega)$.

- A is affine: Let $u_{\mathcal{N}}, v_{\mathcal{N}} \in \mathrm{BV}(\Omega)_{\mathcal{N}}$ and $t \in [0,1]$.

$$\begin{aligned} A(\underbrace{tu_{\mathcal{N}} + (1-t)v_{\mathcal{N}}}_{=(tu+(1-t)v)_{\mathcal{N}}}) &= tu + (1-t)v - \frac{1}{|\Omega|}\int_\Omega (tu+(1-t)v)(x)\ \mathrm{d}x \\ &= t\left(u - \frac{1}{|\Omega|}\int_\Omega u(x)\ \mathrm{d}x\right) + (1-t)\left(v - \frac{1}{|\Omega|}\int_\Omega v(x)\ \mathrm{d}x\right) \\ &= tA(u_{\mathcal{N}}) + (1-t)A(v_{\mathcal{N}}) \end{aligned}$$

- A is injective: Let $A(u_{\mathcal{N}}) = A(v_{\mathcal{N}})$. Then:

$$\begin{aligned} & u - \frac{1}{|\Omega|}\int_\Omega u(x)\ \mathrm{d}x = v - \frac{1}{|\Omega|}\int_\Omega v(x)\ \mathrm{d}x \\ \Leftrightarrow\quad & u - v = \frac{1}{|\Omega|}\int_\Omega (u-v)(x)\ \mathrm{d}x \quad \in \mathbb{R} = \mathcal{N} \\ \Rightarrow & (u_{\mathcal{N}}) = (v_{\mathcal{N}}) \end{aligned}$$

Having this, the conditions of Theorem 3.20 are satisfied. Considering the convex set $\mathcal{B}^M_{\mathcal{N}}$, the theorem gives us

$$\mathrm{ext}A(\mathcal{B}^M_{\mathcal{N}}) = A(\mathrm{ext}(\mathcal{B}^M_{\mathcal{N}})).$$

Notice, that $A(\mathcal{B}^M_{\mathcal{N}}) = \mathcal{B}^{M,0}$. For the first inclusion let $u_{\mathcal{N}} \in \mathcal{B}^M_{\mathcal{N}}$. Then,

$$\mathrm{TV}(A(u_{\mathcal{N}});\Omega) = \mathrm{TV}(u - \frac{1}{|\Omega|}\int_\Omega u(x)\ \mathrm{d}x;\Omega) = \mathrm{TV}(u;\Omega) \le M.$$

With the above considerations in (21) also $\int_\Omega A(u_{\mathcal{N}}) = 0$, so $A(u_{\mathcal{N}}) \in \mathcal{B}^{M,0}$.
On the other hand, let $u \in \mathcal{B}^{M,0}$. Then, one can write $u = u - \frac{1}{|\Omega|}\int_\Omega u(x)\ \mathrm{d}x \in A(u_{\mathcal{N}})$, as the integral of u over Ω is zero. Obviously, $u_{\mathcal{N}} \in \mathcal{B}^M_{\mathcal{N}}$, as $\mathrm{TV}_{\mathcal{N}}(u_{\mathcal{N}};\Omega) = \mathrm{TV}(u;\Omega) \le M$, so $u \in A(\mathcal{B}^M_{\mathcal{N}})$. This yields the desired equality $A(\mathcal{B}^M_{\mathcal{N}}) = \mathcal{B}^{M,0}$.

Putting everything together, one gets

$$\begin{aligned} \mathrm{ext}(\mathcal{B}^{M,0}) &= A(\mathrm{ext}(\mathcal{B}^M_{\mathcal{N}})) \\ &= A\left(\left\{\left(\pm\frac{M}{\mathrm{Per}(E;\Omega)}\cdot \mathbb{1}_E\right)_{\mathcal{N}} \mid E\,\text{simple}\right\}\right) \\ &= \left\{\pm\frac{M}{\mathrm{Per}(E;\Omega)}\cdot\left(\mathbb{1}_E - \frac{1}{|\Omega|}\int_\Omega \mathbb{1}_E(x)\ \mathrm{d}x\right) \mid E\,\text{simple}\right\} \\ &= \left\{\pm\frac{M}{\mathrm{Per}(E;\Omega)}\cdot \mathring{\mathbb{1}}_E \mid E\,\text{simple}\right\} \end{aligned}$$

□

3.2.2. Optimal simple sets under the truncated Fourier transform

Above we have seen that the minimizing pair $(s, v) \in C^0$ that is produced in line 2 of the Frank-Wolfe algorithm is of the form $(s, v) = (0, 0)$ or $(s, v) = (M, \pm\frac{M}{\mathrm{Per}(E;\Omega)} \cdot \mathring{\mathbb{1}}_E)$ with E being a simple set. The minimization problem in line 2 thus reduces to

$$\begin{aligned}
&\underset{(s,v)\in C^0}{\operatorname{argmin}}\ \alpha s + \int_\Omega (K^0)^\#(K^0u - f) \cdot v \ \mathrm{d}x \\
=&\underset{\substack{\text{E simple}\\ \epsilon\in\{-1,1\}}}{\operatorname{argmin}}\ \alpha M + \int_\Omega (K^0)^\#(K^0u - f) \cdot \frac{\epsilon M}{\mathrm{Per}(E;\Omega)} \mathring{\mathbb{1}}_E \ \mathrm{d}x \\
=&\underset{\substack{\text{E simple}\\ \epsilon\in\{-1,1\}}}{\operatorname{argmin}}\ M \cdot \left(\alpha + \frac{\epsilon}{\mathrm{Per}(E;\Omega)} \int_\Omega (K^0)^\#(K^0u - f)\Big(\mathbb{1}_E - \frac{1}{|\Omega|} \underbrace{\int_\Omega \mathbb{1}_E \ \mathrm{d}y}_{=|E|}\Big) \mathrm{d}x\right) \\
=&\underset{\substack{\text{E simple}\\ \epsilon\in\{-1,1\}}}{\operatorname{argmin}}\ M \cdot \left(\alpha + \frac{\epsilon}{\mathrm{Per}(E;\Omega)} \left(\int_E (K^0)^\#(K^0u - f) \ \mathrm{d}x - \frac{|E|}{|\Omega|} \int_\Omega (K^0)^\#(K^0u - f) \ \mathrm{d}x\right)\right).
\end{aligned}$$

Recall that the considered setting requires that $(K^0)^\# : Y^\# = Y \to \mathrm{BV}^0(\Omega)^\#$, where the latter is the predual space of $\mathrm{BV}^0(\Omega)$ and can be characterized as $\mathrm{BV}^0(\Omega)^\# = \left\{v \in \mathrm{BV}(\Omega)^\# \middle| \int_\Omega v \ \mathrm{d}x = 0\right\}$. Therefore, $\int_\Omega (K^0)^\#(K^0u - f) \ \mathrm{d}x = 0$ and the problem reduces further to

$$\underset{\substack{\text{E simple}\\ \epsilon\in\{-1,1\}}}{\operatorname{argmin}}\ M \cdot \left(\alpha + \frac{\epsilon}{\mathrm{Per}(E;\Omega)} \int_E (K^0)^\#(K^0u - f) \ \mathrm{d}x\right).$$

This is minimized for $\epsilon = -\mathrm{sign}(\int_E (K^0)^\#(K^0u - f))$ and the optimal simple sets are minimizers of the problem

$$\underset{\text{E simple}}{\operatorname{argmin}} \underbrace{-\frac{1}{\mathrm{Per}(E;\Omega)} \left|\int_E (K^0)^\# \underbrace{(K^0u - f)}_{=:w} \ \mathrm{d}x\right|}_{=:J}. \tag{22}$$

Here $(K^0)^\# w$ is of the form $\sum_{|k|_\infty \le \Phi} w_k \cdot e^{2\pi i(x,y)\cdot(k_1,k_2)} (= K^\# w)$ as given in (9) with the additional condition that $\int_\Omega (K^0)^\# w \ \mathrm{d}x = 0$. As introduced in section 2.3.1 we identify Ω with $[0, 1]^2$.

Remark 3.22. The condition $\int_\Omega (K^0)^\# w \ \mathrm{d}x = 0$ can be achieved by forcing $w_{0,0} = 0$.

This can be seen as

$$\int_\Omega (K^0)^\# w \, \mathrm{d}x = \int_\Omega \sum_{|k|_\infty \le \Phi} w_k \cdot e^{2\pi i (x,y)\cdot(k_1,k_2)} \, \mathrm{d}x$$
$$= \sum_{|k|_\infty \le \Phi} w_k \cdot \int_0^1 e^{2\pi i x k_1} \, \mathrm{d}x \cdot \int_0^1 e^{2\pi i y k_2} \, \mathrm{d}y.$$

with

$$\int_0^1 e^{2\pi i z k_i} \, \mathrm{d}z = \begin{cases} 1 & \text{if } k_i = 0 \\ \frac{e^{2\pi i k_i} - 1}{2\pi i k_i} = 0 & \text{if } k_i \neq 0. \end{cases}$$

This can be plugged in above and yields

$$\int_\Omega (K^0)^\# w \, \mathrm{d}x = \sum_{|k|_\infty \le \Phi} w_k \cdot \left\{ \begin{matrix} 1 & \text{if } k_1 = k_2 = 0 \\ 0 & \text{else} \end{matrix} \right\} = w_{0,0}.$$

$(K^0)^\# w$ is the combination of trigonometric polynomials and looks as follows:

$$\begin{aligned} (K^0)^\# w = &\sum_{|k|_\infty \le \Phi} w_k \cdot e^{2\pi i (x,y)\cdot(k_1,k_2)} \\ = &\sum_{0 \le k_1, k_2 \le \Phi} a_{k_1,k_2} \cos(2\pi k_1 x)\cos(2\pi k_2 y) + b_{k_1,k_2} \sin(2\pi k_1 x)\sin(2\pi k_2 y) \\ &\quad + c_{k_1,k_2} \sin(2\pi k_1 x)\cos(2\pi k_2 y) + d_{k_1,k_2} \cos(2\pi k_1 x)\sin(2\pi k_2 y), \end{aligned}$$

where the coefficients $a_{k_1,k_2}, b_{k_1,k_2}, c_{k_1,k_2}$ and d_{k_1,k_2} depend on w and are in such a way that the integral over Ω vanishes.

In the following, we want to provide evidence for the assumption that the optimal simple sets solving (22) may be assumed to be rectangular. The idea, that cutting the possibly round boundaries of a simple set E may improve the functional J, stems from the following considerations. Therefore, we look at the individual components of J and analyze how they react on a cut-off of δ on one side of the set E as illustrated in Figure 2. Here, E is supposed to be the circular area, where $(K^0)^\# w$ is negative.

- $\mathrm{Per}(E;\Omega)$: Notice that the anisotropic perimeter of an arbitrary connected convex and bounded simple set E equals the perimeter of the surrounding axis-aligned rectangle with same heights and same width as E. A proof of this can be found in the appendix in Lemma A.1. Then,

  $$\Delta \mathrm{Per}(E;\Omega) = -2\delta$$

  as the upper and lower boundary is shortened by δ each.
- $|\int_E (K^0)^\# w|$: The area of the domain cut away is approximately of magnitude

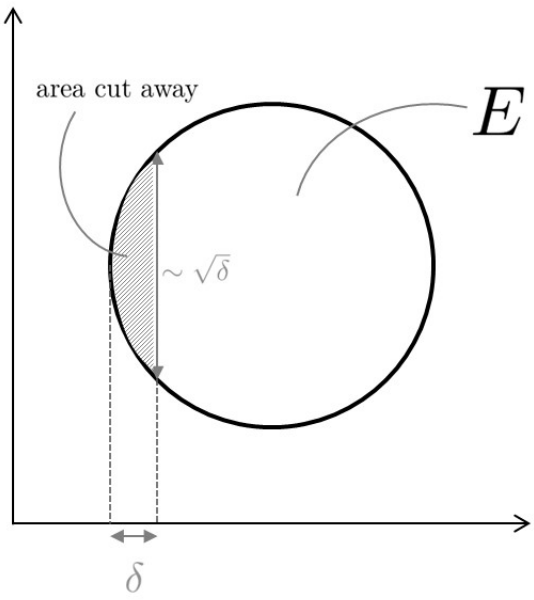


Figure 2: Illustration of the cut-away of the left edge of the simple set E by δ.

$\delta \cdot \sqrt{\delta} = \delta^{\frac{3}{2}}$. Moreover, the integrand $(K^0)^{\#}w$ is supposed to be negative on the set E, and of order δ. Altogether, this yields that the absolute value of the integral will change by approximately

$$\Delta(|\int_E (K^0)^{\#}w)| \approx -\delta \cdot \delta \cdot \sqrt{\delta} = -\delta^{\frac{5}{2}}.$$

The overall change of J that is induced by the curtailment of E is thus given by

$$\begin{aligned}\Delta J &= -\frac{1}{\mathrm{Per}(E;\Omega)} \cdot \Delta\left(\left|\int_E (K^0)^{\#}w)\right|\right) - \Delta\left(\frac{1}{\mathrm{Per}(E;\Omega)}\right) \cdot \left|\int_E (K^0)^{\#}w\right| \\ &= -\frac{1}{\mathrm{Per}(E;\Omega)} \cdot \Delta\left(\left|\int_E (K^0)^{\#}w)\right|\right) + \frac{1}{\mathrm{Per}(E;\Omega)^2} \cdot \Delta\mathrm{Per}(E;\Omega) \cdot \left|\int_E (K^0)^{\#}w\right| \\ &\approx \frac{1}{\mathrm{Per}(E;\Omega)} \cdot \left(\delta^{\frac{5}{2}} - 2\delta \cdot \frac{1}{\mathrm{Per}(E;\Omega)} \cdot \left|\int_E (K^0)^{\#}w\right|\right).\end{aligned}$$

Hence, depending on the integral of $(K^0)^{\#}w$ over E and the perimeter of the set E, it is possible that a straightening of E by a small δ may improve the objective functional J. In the proceeding we will determine the optimal simple sets for two exemplary cases numerically and analyze, whether those numerical examples indicate that the optimal simple sets can be assumed to be rectangular, when considering the truncated Fourier transform as measurement operator.

Example 1 $((K^0)^\# w)_1 = \cos(2\pi x) + \cos(2\pi y) - \cos(2\pi x)\cos(2\pi y)$

Note, that $\int_\Omega ((K^0)^\# w)_1 \,\mathrm{d}x = 0$, so it is indeed a permitted measurement. To identify the optimal simple set E, we try three different approaches.

The first approach is to characterize the level sets of $((K^0)^\# w)_1$ and search for the level set minimizing J. The second is to start with the outer edges of the level sets and form rectangles with different heights and widths around them. Lastly, we combine both approaches, so we start with the level sets and proceedingly cut the edges, such that the outer edges are horizontal and vertical. Note that there possibly remain level set-shaped parts connecting the vertical and horizontal segments. The programming code can be found under `https://github.com/lenasme/SlidingFrankWolfe_Thesis_Schmedt`.

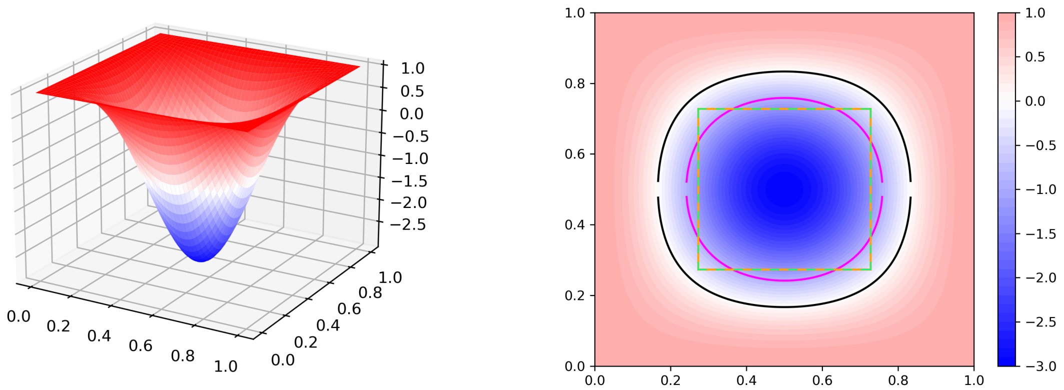


Figure 3: left: Illustration of $((K^0)^\# w)_1$ over the domain $\Omega = [0,1]^2$;
right: Top view of the optimal simple sets resulting from different approaches.
—— level set with value 0: $J \approx -0.17397$
—— optimal level set with value -0.2028: $J \approx -0.19379$
—— optimal set when considering rectangles: $J \approx -0.21213$
—— optimal set when considering the level set approach with horizontal and vertical cuts: $J \approx -0.21213$

Note, that the sets resulting from the combined and the rectangular approach are in the exact same place.

Figure 3 illustrates the findings. The objective functional J becomes smallest, when considering the rectangular or the combined approach. Note that both approaches lead to the same set E that is indeed rectangular.

Example 2 $((K^0)^\# w)_2 = -\cos(2\pi x)\cos(2\pi y) + \frac{1}{2}\cos(4\pi x)\cos(4\pi y)$

Again, $\int_\Omega ((K^0)^\# w)_2 \,\mathrm{d}x = 0$ and with this $((K^0)^\# w)_2$ forms an admissible measurement in our sense. Similar to the first example, we consider the level set, a rectangular and a combined approach. The numerical results are summed up in Figure 4.

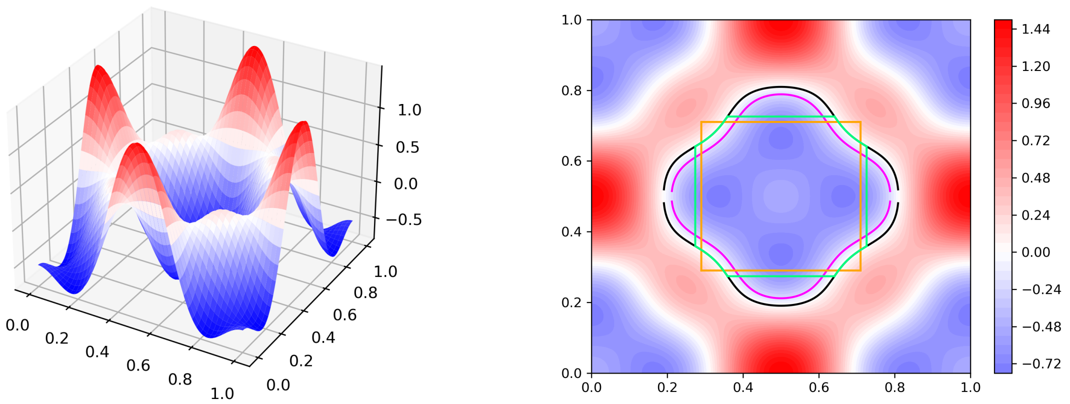


Figure 4: left: Illustration of $((K^0)^{\#}w)_2$ over the domain $\Omega = [0,1]^2$;
right: Top view of the optimal simple sets resulting from different approaches.
— level set with value 0: $J \approx -0.04791$
— optimal level set with value -0.2028: $J \approx -0.04947$
— optimal set when considering rectangles: $J \approx -0.05483$
— optimal set when considering the level set approach with horizontal and vertical cuts: $J \approx -0.05573$

It can be seen, that the pure level set approach yields the worst outcome for this example, while the combined approach leads to the best result, as the objective J takes here the smallest value. Nevertheless, restricting to rectangles does still a good job and J differs just by approximately 9×10^{-4} from the best solution, which is a lot better in comparison to what the best level set does, which produces an error of magnitude 6.3×10^{-3}.

In conclusion, these two examples indeed indicate that restricting the simple sets to be rectangular is justified. Even though small mistakes are done as pointed out in the second example, rectangles yield good approximations of the optimal simple sets. Therefore, assume in the following that the minimizing functions $v^{[k]}$ that are searched for in line 2 of the Frank-Wolfe algorithm are of the form

$$v^{[k]} \in \left\{ \pm \frac{M}{\mathrm{Per}(E;\Omega)} \mathring{\mathbb{1}}_E \middle| E \text{ is rectangular} \right\}. \tag{23}$$

These findings can be exploited in the formulation of the Frank-Wolfe algorithm. Thus, the updated algorithm is presented in Algorithm 2, including the results presented in Theorem 3.21 and Equation (22).

Algorithm 2 Classic Frank-Wolfe Algorithm (minimization step adapted for $P_\alpha^{\mathrm{BV}^0}(f)$)

1: **for** $k = 0, \dots, n$ **do**
2: Minimize: $E_* \in \underset{E \text{ rectangular}}{\operatorname{argmin}} \; -\frac{1}{\operatorname{Per}(E;\Omega)} \left| \int_E (K^0)^\#(K^0 u^{[k]} - f) \, \mathrm{d}x \right|$.
3: **if** $\mathrm{d}J(u^{[k]})[s^{[k]} - u^{[k]}] = 0$ **then**
4: $u^{[k]}$ solution of $(P_\alpha^{\mathrm{BV}^0}(f))$. Stop.
5: **else**:
6: Step research: $\gamma^{[k]} \leftarrow \frac{2}{k+2}$ or $\gamma^{[k]} \in \operatorname{argmin}_{\gamma \in [0,1]} J((1-\gamma)u^{[k]} + \gamma \frac{M\epsilon}{\operatorname{Per}(E_*;\Omega)} \mathring{\mathbb{1}}_{E_*})$.
7: Update: $u^{[k+1]} \leftarrow (1 - \gamma^{[k]})u^{[k]} + \gamma^{[k]}(\frac{M\epsilon}{\operatorname{Per}(E_*;\Omega)} \mathring{\mathbb{1}}_{E_*})$.
8: **end if**
9: **end for**

3.3. Stopping criterion

Having analyzed the minimization step of the Frank-Wolfe algorithm, it is still necessary to find a practical stopping criterion that can be used in line 3 of Algorithm 2.

Therefore, we want to make use of the *primal-dual optimality conditions* introduced in Corollary 3.27. The theorems and definitions that are required accordingly are presented below.

Definition 3.23 (Subdifferential, [1, Def. 98])**.** Let X be a Banach space and X^* its dual. The *subdifferential* of a convex function $f : X \to \mathbb{R} \cup \{+\infty\}$ in $x \in X$ is defined by

$$\partial f(x) := \left\{ x^* \in X^* \middle| f(y) - f(x) \geq \langle x^*, y - x \rangle \quad \text{for all } y \in X \right\}. \tag{24}$$

Definition 3.24 (Legendre-Fenchel conjugate, [1, Def. 101])**.** Let X be a Banach space and X^* its dual space. The *Legendre-Fenchel conjugate* of a convex function $f : X \to \mathbb{R} \cup +\infty$ is the function $f^* : X^* \to (-\infty, +\infty]$ defined by

$$f^*(x^*) = \sup_{x \in X} \{\langle x^*, x \rangle - f(x)\}. \tag{25}$$

Theorem 3.25 (Fenchel inequality, [1, Thm. 104])**.** *Let f be proper convex, $x \in X, x^* \in X^*$. Then*

$$\langle x^*, x \rangle = f(x) + f^*(x^*) \Leftrightarrow x^* \in \partial f(x) \Leftrightarrow x \in \partial f^*(x^*).$$

Theorem 3.26 (Fenchel duality, [21, Thm. 4.4.3])**.** *Let X and Y be Banach spaces, $F : X \to \mathbb{R} \cup \{+\infty\}$ and $G : Y \to \mathbb{R} \cup \{+\infty\}$ be convex functions and $K : X \to Y$ a* *linear bounded operator. Assume further that*

$$K\ domF \cap contG \neq \emptyset.$$

Then

$$\underbrace{\inf_{x\in X}\{F(x)+G(Kx)\}}_{\text{primal}} = \underbrace{\sup_{y^*\in Y^*}\{-F^*(K^*y^*)-G^*(-y^*)\}}_{\text{dual}}, \tag{26}$$

i.e. strong duality holds, and the supremum of the dual problem is attained if finite.

Proposition 3.27 (Primal-dual optimality conditions, [1, Cor. 106])**.** *Let strong duality hold.*

$$x \in X \text{ solves the primal and } y^* \in Y^* \text{ the dual problem } \Leftrightarrow \begin{cases} Kx & \in \partial G^*(y^*) \\ -K^*y^* & \in \partial F(x) \end{cases}$$

The obstacle that arises when we consider the theory above is that there is the need to know about the dual of the Banach space X. In our application $X = \mathrm{BV}(\Omega)$, but as remarked for example in Remark 3.12 of [5], just very little can be said on the dual space of $\mathrm{BV}(\Omega)$. To avoid these difficulties, we consider our optimization problem $P_\alpha^{\mathrm{BV}^0}(f)$ no longer over the space $\mathrm{BV}^0(\Omega)$ but over the easier to handle space $L^{2,0}(\Omega)$, defined as

$$L^{2,0}(\Omega) := \{u \in L^2(\Omega) \mid \int_\Omega u(x)\ \mathrm{d}x = 0\}.$$

Due to the Embedding Theorem stated in [5, Cor. 3.49] we obtain in our case with $\Omega = (\mathbb{R}/\mathbb{Z})^2$ that $\mathrm{BV}(\Omega)$ embeds continuously into $L^2(\Omega)$. The same is true for the restriction on vanishing integrals, which means $\mathrm{BV}^0(\Omega) \hookrightarrow L^{2,0}(\Omega)$ and in particular $\mathrm{BV}^0(\Omega) \subset L^{2,0}(\Omega)$.

Remark 3.28 (Modified Setting). The optimization problem, no longer viewed over $\mathrm{BV}^0(\Omega)$ but over the space $L^{2,0}(\Omega)$, reads as follows:

$$\min_{u\in L^{2,0}(\Omega)} J_\alpha^{f,0}(u) \qquad J_\alpha^{f,0}(u) = \frac{1}{2\alpha}|K^0u - f|_2^2 + \mathrm{TV}(u;\Omega). \tag{$P_\alpha^{L^{2,0}}(f)$}$$

or equivalently

$$\min_{(t,u)\in C^{L^{2,0}}} \widetilde{J}_\alpha^{f,0}(t,u) \qquad \widetilde{J}_\alpha^{f,0}(t,u) = \frac{1}{2}|K^0u - f|_2^2 + \alpha t \tag{$\widetilde{P}_\alpha^{L^{2,0}}(f)$}$$

with $C^{L^{2,0}} := \Big\{(t,u) \in \mathbb{R}_+ \times L^{2,0}(\Omega) \ \mid\ \mathrm{TV}(u;\Omega) \le t \le M = \frac{|f|_2^2}{2\alpha}\Big\}$.

According to Proposition 2 of [22], as the proof works the same when restricting to the space $L^{2,0}(\Omega)$, this problem is well-defined and there exists a minimizer $u \in L^{2,0}(\Omega)$ of the above problem.

Moreover, the solutions u^{BV^0} of $P_\alpha^{\mathrm{BV}^0}(f)$ and $u^{L^{2,0}}$ of $P_\alpha^{L^{2,0}}(f)$ will coincide since by definition of the objective functional, the solution $u^{L^{2,0}}$ has to satisfy $u^{L^{2,0}} \in L^{2,0}(\Omega) \cap \mathrm{BV}^0(\Omega)$, as $\mathrm{TV}(u^{L^{2,0}};\Omega) = \infty$, otherwise. Therefore, both problems will have the same solutions and $L^{2,0}(\Omega)\cap\mathrm{BV}^0(\Omega)$ is the natural space to minimize the objective functional.

3.3.1. Optimality conditions

In order to be able to apply Proposition 3.27, we must first verify that strong duality holds. To do so, consider Theorem 3.26. To obtain strong duality, we only need to show that $K^0 \operatorname{dom}F \cap \operatorname{cont}G \neq \emptyset$, as the other requirements are satisfied since in our application $F = \mathrm{TV}(\cdot, \Omega)$ and $G = \frac{1}{2\alpha}|\cdot - f|_2^2$ fulfill convexity and the truncated Fourier transform K^0 is linear and bounded.

Existence of $K^0 v \in K^0 \operatorname{dom}F \cap \operatorname{cont}G$:

Let $v \equiv 0 \in L^{2,0}(\Omega)$. Then, $F(v) = \mathrm{TV}(v;\Omega) = 0 < +\infty$ and thus $v \in \operatorname{dom}F$. We need to show that $K^0 v \in \operatorname{cont}G$. This is indeed the case, as G maps $w \mapsto \frac{1}{2\alpha}|w - f|_2^2$ which is continuous in every $w \in Y$, so in particular in $K^0 v \in Y$. Therefore, $K^0 v \in K^0 \operatorname{dom}F \cap \operatorname{cont}G \neq \emptyset$ and Theorem 3.26 can be applied.

Having this, the strong duality required in Proposition 3.27 is satisfied. Therefore, from the primal-dual optimality condition we can deduce

$$\left\{\begin{array}{ll} Kx & \in \partial F^*(y) \\ -K^*y & \in \partial G(x) \end{array}\right\} \Rightarrow x \in X \text{ solves the primal } P_\alpha^{L^{2,0}}(f).$$

Note that with the Fenchel inequality Theorem 3.25 this condition can be reformulated in order to get rid of the Legendre-Fenchel conjugate and takes following equivalent form, when adapted to our application:

$$\left\{\begin{array}{ll} y & \in \partial F(Ku) = \partial\left(\frac{1}{2\alpha}|K^0u - f|_2^2\right) \\ -(K^0)^{\#}y & \in \partial G(u) = \partial \mathrm{TV}(u;\Omega) \end{array}\right\} \Rightarrow u \in L^{2,0}(\Omega) \text{ solves the primal } P_\alpha^{L^{2,0}}(f). \tag{27}$$

Remark 3.29. Since $L^{2,0}(\Omega)$ is a Hilbert space, the Riesz representation theorem yields $(K^0)^* = (K^0)^{\#}$ and we can identify the adjoint of K^0 with the preadjoint.
Moreover, the first condition $y \in \partial\left(\frac{1}{2\alpha}|K^0u - f|_2^2\right)$ is equivalent to $y = \frac{1}{\alpha}(K^0u - f)$, as the subdifferential $\partial\left(\frac{1}{2\alpha}|\cdot - f|_2^2\right)$ is given by its derivative $\frac{1}{\alpha}(\cdot - f)$.

What remains in order to be able to formulate a specific optimality condition, is to understand the structure and appearance of the subdifferential $\partial\mathrm{TV}(\cdot;\Omega)$. This is of particular interest, as this condition is going to be the core of the stopping criterion, that will be used in line 3 of the Frank-Wolfe algorithm.

By Definition 3.23, we get

$$\begin{aligned}\partial \mathrm{TV}(0;\Omega) &= \left\{\eta \in (L^{2,0})^*(\Omega) \middle| \mathrm{TV}(u;\Omega) - \underbrace{\mathrm{TV}(0;\Omega)}_{=0} \geq \langle \eta, u-0\rangle, \quad \forall u \in L^{2,0}(\Omega)\right\} \\ &= \left\{\eta \in L^{2,0}(\Omega) \mid \underbrace{\mathrm{TV}(u;\Omega)}_{\geq 0} \geq \int_\Omega \eta u \,\mathrm{d}x, \quad \forall u \in L^{2,0}(\Omega)\right\} \\ &\overset{\text{holds for all } u\in L^{2,0}(\Omega)}{=} \left\{\eta \in L^{2,0}(\Omega) \mid \mathrm{TV}(u;\Omega) \geq \left|\int_\Omega \eta u \,\mathrm{d}x\right|, \quad \forall u \in L^{2,0}(\Omega)\right\}.\end{aligned} \tag{28}$$

Following the approach presented in Chapter 2.2 of [23], our goal is to express the subdifferential in terms of sets such that a resulting stopping criterion is consistent with the representation of solutions of the minimization step in line 2 of the algorithm. In order to obtain a set-based formulation of the subdifferential some additional tools are required. On the one hand the Coarea formula which was already introduced in Theorem 3.11 and on the other hand the so-called Layer-Cake formula introduced below.

Proposition 3.30 (Layer-cake formula, [24, Thm. 1.13])**.** *Let u be a measurable and nonnegative function on Ω. For $x \in \Omega$ one has*

$$u(x) = \int_0^{+\infty} \mathbb{1}_{\underbrace{\{x\in\Omega \mid u(x) > t\}}_{=:\{u>t\}}}(x) \,\mathrm{d}t.$$

Remark 3.31. What we are going to need is the applied case of u being nonnegative and $\eta \in L^{2,0}(\Omega)$. Then the Layer-cake formula gives

$$\begin{aligned}\int_\Omega \eta(x)u(x)\,\mathrm{d}x &= \int_\Omega \eta(x) \int_0^\infty \mathbb{1}_{\{u>t\}} \,\mathrm{d}t \,\mathrm{d}x \\ &\overset{\text{Fubini}}{=} \int_0^\infty \int_\Omega \mathbb{1}_{\{u>t\}} \eta(x) \,\mathrm{d}x \,\mathrm{d}t \\ &= \int_0^\infty \int_{\{u>t\}} \eta(x) \,\mathrm{d}x \,\mathrm{d}t.\end{aligned} \tag{29}$$

Lemma 3.32 (Subdifferential of $\mathrm{TV}(0;\Omega)$, [23, Chapter 2.2])**.** *The subdifferential of the total variation at 0 given in (28) can equivalently be expressed via the following set*

$$\begin{aligned}\partial \mathrm{TV}(0;\Omega) = \Bigg\{\eta \in L^{2,0}(\Omega) \Bigg| &\forall E \subset \Omega \text{ with } 0 < |E| < +\infty \text{ and } \mathrm{Per}(E;\Omega) < +\infty : \\ &\underbrace{\left|\int_\Omega \eta \frac{\mathbb{1}_E}{\mathrm{Per}(E;\Omega)} \,\mathrm{d}x\right| \leq 1}_{\Leftrightarrow \left|\int_E \eta \,\mathrm{d}x\right| \leq \mathrm{Per}(E;\Omega)}\Bigg\}.\end{aligned} \tag{30}$$

Proof. To prove the desired equality of sets, we are going to prove both inclusions separately.

- (28) $\subset$ (30): Let $\eta \in$ (28), i.e. for every $u \in L^{2,0}(\Omega)$ we have the inequality $|\int_\Omega \eta u \, \mathrm{d}x| \leq \mathrm{TV}(u;\Omega)$. Let $u = \mathring{\mathbb{1}}_E \in L^{2,0}(\Omega)$, with $E \subset \Omega$ such that $0 < |E| < +\infty$ and $\mathrm{Per}(E;\Omega) < +\infty$. Then

$$\left| \int_\Omega \eta \mathring{\mathbb{1}}_E \, \mathrm{d}x \right| \leq \mathrm{TV}(\mathring{\mathbb{1}}_E;\Omega).$$

  Note that

$$\left| \int_\Omega \eta \mathring{\mathbb{1}}_E \, \mathrm{d}x \right| = \left| \int_\Omega \eta \cdot \left(\mathbb{1}_E - \frac{1}{|\Omega|} \int_\Omega \mathbb{1}_E \, \mathrm{d}y \right) \mathrm{d}x \right| = \left| \int_E \eta \, \mathrm{d}x - |E| \underbrace{\int_\Omega \eta \, \mathrm{d}x}_{=0} \right| = \left| \int_E \eta \, \mathrm{d}x \right|.$$

  Moreover, $\mathrm{TV}(\mathring{\mathbb{1}}_E;\Omega) = \mathrm{TV}(\mathbb{1}_E;\Omega) = \mathrm{Per}(E;\Omega)$, as $\mathring{\mathbb{1}}_E$ and $\mathbb{1}_E$ just differ by a constant, which has no effect on the total variation as argued before. Thus, one gets the inequality

$$\left| \int_E \eta \, \mathrm{d}x \right| \leq \mathrm{Per}(E;\Omega)$$

  and can conclude $\eta \in$ (30).

- (30) $\subset$ (28): Let $u \in L^{2,0}(\Omega)$ and let $\eta \in$ (30), i.e. for every $E \subset \Omega$ with $0 < |E| < +\infty$ and $\mathrm{Per}(E;\Omega) < +\infty$ the inequality $|\int_E \eta| \leq \mathrm{Per}(E;\Omega)$ holds. As the Layer-cake formula is solely defined for nonnegative functionals, decompose u in its positive and its negative part, i.e. write $u = u^+ - u^-$ with

$$u^+(x) := \max\{u(x), 0\} \geq 0 \quad \text{and} \quad u^-(x) := \max\{-u(x), 0\} \geq 0.$$

  With the Coarea formula one gets

$$\begin{aligned} \mathrm{TV}(u^+;\Omega) &= \int_{-\infty}^{+\infty} \mathrm{Per}(\{u^+ > t\};\Omega) \, \mathrm{d}t = \int_{-\infty}^{\infty} \mathrm{Per}(\underbrace{\{u^+ \leq t\}}_{=\emptyset \ \forall t<0};\Omega) \, \mathrm{d}t \\ &= \int_0^{+\infty} \mathrm{Per}(\{u^+ \leq t\};\Omega) \, \mathrm{d}t = \int_0^{+\infty} \mathrm{Per}(\{u^+ > t\};\Omega) \, \mathrm{d}t \end{aligned} \tag{31}$$

  and similarly

$$\mathrm{TV}(u^-;\Omega) = \int_{-\infty}^{+\infty} \mathrm{Per}(\{u^- > t\};\Omega) \, \mathrm{d}t = \int_0^{+\infty} \mathrm{Per}(\{u^- > t\};\Omega) \, \mathrm{d}t. \tag{32}$$

  Moreover $\mathrm{TV}(u;\Omega) = \mathrm{TV}(u^+;\Omega) + \mathrm{TV}(u^-;\Omega)$. This can be seen, as

$$\mathrm{TV}(u;\Omega) = \sup \left\{ \int_\Omega u^+ \mathrm{div}\phi \, \mathrm{d}x - \int_\Omega u^- \mathrm{div}\phi \, \mathrm{d}x \,\middle|\, \phi \in C_c^1(\Omega, \mathbb{R}^2), \ \|\phi\|_\infty \leq 1 \right\}.$$

Note that the support of u^+ and u^- is disjoint, so the optimal ϕ in the above equation can be chosen in such a way that $\phi = \phi_1 + \phi_2$, where ϕ_1 and ϕ_2 also have disjoint support and

$$\begin{aligned}
\mathrm{TV}(u;\Omega) &= \sup\left\{\int_\Omega u^+ \mathrm{div}\phi_1 \ \mathrm{d}x - \int_\Omega u^- \mathrm{div}\phi_2 \ \mathrm{d}x \,\middle|\, \phi_1,\phi_2 \in C_c^1(\Omega,\mathbb{R}^2),\ \|\phi_{1,2}\|_\infty \le 1\right\} \\
&\overset{\text{minus taken by } \phi_2}{=} \sup\left\{\int_\Omega u^+ \mathrm{div}\phi_1 \ \mathrm{d}x \,\middle|\, \phi_1 \in C_c^1(\Omega,\mathbb{R}^2),\ \|\phi_1\|_\infty \le 1\right\} \\
&\qquad + \sup\left\{\int_\Omega u^- \mathrm{div}\phi_2 \ \mathrm{d}x \,\middle|\, \phi_2 \in C_c^1(\Omega,\mathbb{R}^2),\ \|\phi_2\|_\infty \le 1\right\} \\
&= \mathrm{TV}(u^+;\Omega) + \mathrm{TV}(u^-;\Omega).
\end{aligned}$$

Let $E_1 := \{x \in \Omega \mid u^+(x) > t\}$ and $E_2 := \{x \in \Omega \mid u^-(x) > t\}$. Then, $E_{1,2}$ satisfy the conditions of (30) and we get by assumption on η

$$\left|\int_{\{x\in\Omega | u^+(x)>t\}} \eta(x) \ \mathrm{d}x\right| \le \mathrm{Per}(\{x \in \Omega \mid u^+(x) > t\};\Omega)$$

and

$$\left|\int_{\{x\in\Omega | u^-(x)>t\}} \eta(x) \ \mathrm{d}x\right| \le \mathrm{Per}(\{x \in \Omega \mid u^-(x) > t\};\Omega).$$

Now integrate both sides over t from 0 to $+\infty$ and obtain

$$\int_0^{+\infty} \left|\int_{\{u^+>t\}} \eta(x) \ \mathrm{d}x\right| \mathrm{d}t \le \int_0^{+\infty} \mathrm{Per}(\{u^+ > t\};\Omega) \ \mathrm{d}t \overset{(31)}{=} \mathrm{TV}(u^+;\Omega) \tag{33}$$

as well as

$$\int_0^{+\infty} \left|\int_{\{u^->t\}} \eta(x) \ \mathrm{d}x\right| \mathrm{d}t \le \int_0^{+\infty} \mathrm{Per}(\{u^- > t\};\Omega) \ \mathrm{d}t \overset{(32)}{=} \mathrm{TV}(u^-;\Omega). \tag{34}$$

In the end, the aim is to estimate $|\int_\Omega \eta u \ \mathrm{d}x| \le \mathrm{TV}(u;\Omega)$. To verify this, start with the left hand side.

$$\begin{aligned}\left|\int_\Omega \eta u \;\mathrm{d}x\right| &= \left|\int_\Omega \eta\left(u^+ - u^-\right)\;\mathrm{d}x\right| = \left|\int_\Omega \eta u^+ \;\mathrm{d}x - \int_\Omega \eta u^- \;\mathrm{d}x\right| \\ &\leq \left|\int_\Omega \eta u^+ \;\mathrm{d}x\right| + \left|\int_\Omega \eta u^- \;\mathrm{d}x\right| \\ &\overset{\text{Layer-cake (29)}}{=} \left|\int_0^{+\infty}\int_{\{u^+>t\}} \eta \;\mathrm{d}x \;\mathrm{d}t\right| + \left|\int_0^{\infty}\int_{\{u^->t\}} \eta \;\mathrm{d}x \;\mathrm{d}t\right| \\ &\leq \int_0^{+\infty}\left|\int_{\{u^+>t\}} \eta \;\mathrm{d}x\right| \;\mathrm{d}t + \int_0^{\infty}\left|\int_{\{u^->t\}} \eta \;\mathrm{d}x\right| \;\mathrm{d}t \\ &\overset{(33),(34)}{\leq} \mathrm{TV}(u^+;\Omega) + \mathrm{TV}(u^-;\Omega) = \mathrm{TV}(u;\Omega)\end{aligned}$$

This proves that η satisfies the conditions of (28), as $u \in L^{2,0}(\Omega)$ was chosen arbitrarily.

In the above considerations both inclusions were shown, so the desired equality of sets holds. □

We have seen, that $\partial\mathrm{TV}(0;\Omega)$ is given by equation (30). Nevertheless, what is needed for the primal-dual optimality conditions described in equation (27) based on Proposition 3.27 is $\partial\mathrm{TV}(u;\Omega)$, where $u \neq 0$ is allowed. To understand the structure of $\partial\mathrm{TV}(u;\Omega)$, we make use of the following lemma.

Lemma 3.33 ([22, Lem. 10])**.** *Let X be a Hilbert space and $F : X \to \mathbb{R} \cup \{+\infty\}$ a positively 1-homogeneous convex functional. Then, for each $x \in X$ we have*

$$\partial F(x) = \left\{\xi \in \partial F(0) \middle| \langle \xi, x\rangle = F(x)\right\}. \tag{35}$$

As $L^{2,0}(\Omega)$ is a Hilbert space and from Propositions 2.8 and 2.7 we have 1-homogeneity and convexity of the total variation, the above lemma is applicable and leads directly to the following characterization of the subdifferential of the total variation for an arbitrary $u \in L^{2,0}(\Omega)$.

Lemma 3.34 (Subdifferential of $\mathrm{TV}(u;\Omega)$)**.** *The subdifferential of the total variation of an arbitrary $u \in L^{2,0}(\Omega)$ is given by*

$$\partial\mathrm{TV}(u;\Omega) = \left\{\eta \in \partial\mathrm{TV}(0;\Omega) \,\middle|\, \int_\Omega \eta u \;\mathrm{d}x = \mathrm{TV}(u;\Omega)\right\}. \tag{36}$$

3.3.2. Formulation of a stopping criterion

The next proposition gives a stopping criterion that is applicable in our case.

Proposition 3.35 (Stopping Criterion, [23, Rem. 4])**.** *A valid stopping criterion is given by*

$$\sup_E \frac{|\int_E \eta^{[k]}|}{\operatorname{Per}(E;\Omega)} \leq 1 \quad \text{with} \quad \eta^{[k]} = -\frac{1}{\alpha}(K^0)^\#(K^0 u^{[k]} - f)$$

together with the modified update step introduced in line 8 of Algorithm 3.
By Lemma 3.32, $\eta^{[k]} \in \partial \mathrm{TV}(0;\Omega)$ and due to optimality of $a^{[k]}$ in line 8 of Algorithm 3 we have $\int_\Omega \eta^{[k]} u^{[k]} = \mathrm{TV}(u^{[k]};\Omega)$, so indeed $\eta^{[k]} \in \partial \mathrm{TV}(u^{[k]};\Omega)$. This ensures that the primal-dual optimality conditions from equation (27) are satisfied and $u^{[k]}$ is a solution of $P_\alpha^{L^{2,0}}(f)$.

This stopping criterion as well as the new update step that is required in the above proposition to ensure optimality are included in Algorithm 3. Note that the new update step will decrease the objective more than the standard update step of forming a convex combination of the existing sets as presented in Algorithms 1 and 2 will do. Thus, the convergence properties of the algorithm do not break by changing this step.

Algorithm 3 Classic Frank-Wolfe Algorithm (minimization step, stopping criterion and update step adapted for $P_\alpha^{L^{2,0}}(f)$)

1: **for** $k = 0, \dots, n$ **do**
2: $\quad \eta^{[k]} = -\frac{1}{\alpha}(K^0)^\#(K^0 u^{[k]} - f)$;
3: $\quad$ Minimize: $E_* \in \underset{E \text{ rectangular}}{\operatorname{argmax}} \frac{1}{\operatorname{Per}(E;\Omega)} |\int_E \eta^{[k]}|$
4: $\quad$ **if** $|\int_{E_*} \eta^{[k]}| \leq \operatorname{Per}(E_*;\Omega)$ **then**
5: $\qquad u^{[k]}$ solution of $(P_\alpha^{L^{2,0}}(f))$. Stop.
6: $\quad$ **else**:
7: $\qquad E^{[k+1]} \leftarrow (E_1^{[k]}, \dots, E_{N^{[k]}}^{[k]}, E_*)$
8: $\qquad a^{[k+1]} \leftarrow \underset{a \in \mathbb{R}^{N^{[k]}+1}}{\operatorname{argmin}} J_\alpha^{f,0}\left(\sum_{i=1}^{N^{[k]}+1} a_i \mathring{\mathbb{1}}_{E_i^{[k+1]}}\right)$
9: $\qquad$ remove atoms with zero amplitude
10: $\qquad N^{[k+1]} \leftarrow$ number of atoms in $E^{[k+1]}$
11: $\qquad u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \mathring{\mathbb{1}}_{E_i^{[k+1]}}$
12: $\quad$ **end if**
13: **end for**

Remark 3.36. Following the argumentation of Remark 3 in [23], we may assume that for all $a \in \mathbb{R}^N$ the total variation may be expressed via

$$\mathrm{TV}(\sum_{i=1}^N a_i \mathring{\mathbb{1}}_{E_i};\Omega) = \mathrm{TV}(\sum_{i=1}^N a_i \mathbb{1}_{E_i};\Omega) = \sum_{i=1}^N |a_i| \operatorname{Per}(E_i;\Omega).$$

This yields that $J_\alpha^{f,0}\left(\sum_{i=1}^{N^{[k]}+1} a_i \mathring{\mathbb{1}}_{E_i^{[k+1]}}\right)$ can be interpreted as a functional of $a \in \mathbb{R}^{N^{[k]}+1}$ via

$$\mathcal{J}_\alpha^{[k+1]}(a) = J_\alpha^{f,0}\left(\sum_{i=1}^{N^{[k]}+1} a_i \mathring{\mathbb{1}}_{E_i^{[k+1]}}\right) = \frac{1}{2}|K^0_{E^{[k+1]}}a - f|_2^2 + \alpha \sum_{i=1}^{N^{[k]}+1} \mathrm{Per}(E_i^{[k+1]};\Omega)\cdot |a_i|. \tag{37}$$

Here, $K^0_{E^{[k+1]}} : \mathbb{R}^{N^{[k]}+1} \to Y$ and its preadjoint $(K^0_{E^{[k+1]}})^\# : Y \to \mathbb{R}^{N^{[k]}+1}$ are given by

$$K^0_{E^{[k+1]}}a = \sum_{i=1}^{N^{[k]}+1} a_i K^0\left(\mathring{\mathbb{1}}_{E_i^{[k+1]}}\right)$$
$$(K^0_{E^{[k+1]}})^\# y = \left(\langle \mathring{\mathbb{1}}_{E_i^{[k+1]}}, (K^0)^\# y\rangle\right)_{i=1,\dots,N^{[k]}+1}.$$

Proof of Proposition 3.35. The only aspect that needs to be verified in order to prove that the stated stopping criterion is valid is that the optimality in line 8 ensures $\int_\Omega \eta^{[k]}u^{[k]} = \mathrm{TV}(u^{[k]};\Omega)$.
Let $u^{[k]} = \sum_{i=1}^{N^{[k]}} a_i \mathring{\mathbb{1}}_{E_i^{[k]}}$. As the $(a_i)_{i=1,\dots,N^{[k]}}$ are optimal for $\mathcal{J}_\alpha^{[k]}$, one gets

$$0 \in \partial \mathcal{J}_\alpha^{[k]}(a) = \left(K^0_{E^{[k]}}\right)^\# \left(K^0_{E^{[k]}}a - f\right) + \alpha\partial \underbrace{\left(\sum_{i=1}^{N^{[k]}} \mathrm{Per}(E_i^{[k]};\Omega)\cdot |a_i|\right)}_{=:T(a)}$$
$$\Leftrightarrow -\frac{1}{\alpha}\left(K^0_{E^{[k]}}\right)^\# \left(K^0_{E^{[k]}}a - f\right) \in \partial T(a).$$

Note that T is 1-homogeneous. Together with Lemma A.2 from the appendix,

$$T(a) = \left\langle -\frac{1}{\alpha}\left(K^0_{E^{[k]}}\right)^\# \left(K^0_{E^{[k]}}a - f\right), a\right\rangle.$$

Putting this together one can conclude

$$
\begin{aligned}
\mathrm{TV}(u^{[k]}) &= \sum_{i=1}^{N^{[k]}} \mathrm{Per}(E_i^{[k]};\Omega)\cdot|a_i| = T(a) = \left\langle a, -\frac{1}{\alpha}(K^0_{E^{[k]}})^{\#}(K^0_{E^{[k]}}a - f)\right\rangle \\
&= \left\langle (a_i)_{i=1,\dots,N^{[k]}}, \Big(\langle \mathring{\mathbb{1}}_{E_i^{[k]}}, \underbrace{-\frac{1}{\alpha}(K^0)^{\#}(K^0u^{[k]}-f)}_{=\eta^{[k]}}\rangle\Big)_{i=1,\dots,N^{[k]}}\right\rangle \\
&= \sum_{i=1}^{N^{[k]}} a_i\cdot\int_\Omega \mathring{\mathbb{1}}_{E_i^{[k]}}\eta^{[k]} = \int_\Omega \sum_{i=1}^{N^{[k]}} a_i \mathring{\mathbb{1}}_{E_i^{[k]}}\eta^{[k]} \\
&= \int_\Omega u^{[k]}\eta^{[k]},
\end{aligned}
$$

which proves that $\eta^{[k]} \in \partial\mathrm{TV}(u^{[k]};\Omega)$ and thus ensures optimality of the solution $u^{[k]}$ once the stopping criterion is satisfied. □

3.4. Sliding step

As proposed in [3] in the context of recreating point masses, it is useful to introduce a sliding step to the classic Frank-Wolfe algorithm. The generated image function $u^{[k]}$, that is produced by the classic Frank-Wolfe algorithm is not going to be sparse but will consist of a multitude of simple sets approximating the indicator sets of the ground truth image $u^\dagger$. This can prevented by the introduction of a *sliding step*. In each iteration of the Frank-Wolfe algorithm there will be an additional minimization step over sets and weights at the same time initialized with the sets and weights computed previously. This means, in this additional step the sets are allowed to move and errors in location from former steps may be corrected. Thus, there is not such a strong need to compensate for small errors with new indicator functions that are created in the next iterations but the possibility of making corrections to existing indicator functions. The final algorithm fully adapted to solve problem $P_\alpha^{L^{2,0}}(f)$ including a sliding step therefore reads as follows:

Algorithm 4 Sliding Frank-Wolfe Algorithm (minimization step, stopping criterion and update step adapted for $P_\alpha^{L^{2,0}}(f)$)

for $k = 0,\dots,n$ **do**
  $\eta^{[k]} = -\frac{1}{\alpha}(K^0)^{\#}(K^0u^{[k]} - f)$;
  Minimize: $E_* \in \underset{E\text{ rectangular}}{\operatorname{argmax}} \frac{1}{\mathrm{Per}(E;\Omega)}|\int_E \eta^{[k]}|$
  if $|\int_{E_*}\eta^{[k]}| \le \mathrm{Per}(E_*;\Omega)$ **then**
    $u^{[k]}$ solution of $(P_\alpha^{L^{2,0}}(f))$. Stop.
  else:
    $E^{[k+1]} \leftarrow (E_1^{[k]},\dots,E_{N^{[k]}}^{[k]}, E_*)$;

8: $a^{[k+1]} \leftarrow \underset{a \in \mathbb{R}^{N^{[k]}+1}}{\operatorname{argmin}} \; J_\alpha^{f,0}\left(\sum_{i=1}^{N^{[k]}+1} a_i \mathring{\mathbb{1}}_{E_i^{[k+1]}}\right)$;
9: remove atoms with zero amplitude;
10: $N^{[k+1]} \leftarrow$ number of atoms in $E^{[k+1]}$;
11: perform a local descent on $(a, E) \mapsto J_\alpha^{f,0}\left(\sum_{i=1}^{N^{[k+1]}} a_i \mathring{\mathbb{1}}_{E_i^{[k+1]}}\right)$ initialized with $(a^{[k+1]}, E^{[k+1]})$;
12: repeat the operations of lines 8-10;
13: $u^{[k+1]} \leftarrow \sum_{i=1}^{N^{[k+1]}} a_i^{[k+1]} \mathring{\mathbb{1}}_{E_i^{[k+1]}}$;
14: **end if**
15: **end for**

Note that Algorithm 4 is still a valid application of the standard Frank-Wolfe algorithm. The additional reduction of the objective in line 11 does not break convergence properties and line 12 ensures that also the stopping criterion from Proposition 3.35 remains valid. The improvement resulting from the additional sliding step will be numerically tested and compared to the classic Frank-Wolfe algorithm (Algorithm 3) in the next section.

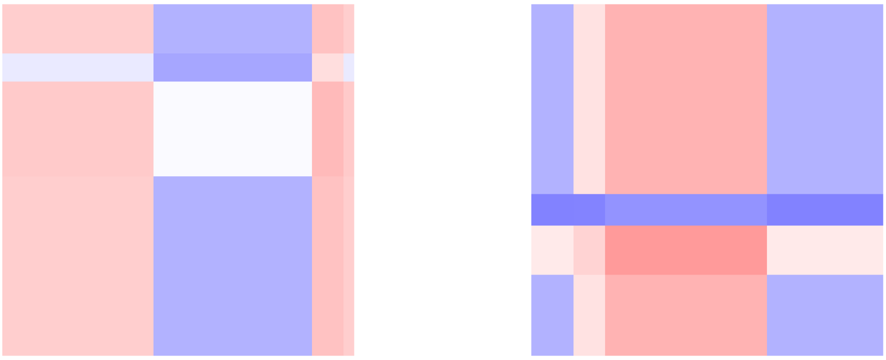
4. Numerical implementation and results

4.1. Implementation details

The complete implementation and source code used in this thesis can be accessed via `https://github.com/lenasme/SlidingFrankWolfe_Thesis_Schmedt`.
To test Algorithm 3 and its sliding version, Algorithm 4, we need to implement the ground truth as described in equation (7) with the additional property that the integral vanishes over $\Omega = [0,1]^2$, i.e. $u^\dagger \in \mathrm{BV}^0(\Omega)$. To do so, we make use of the implementation that was proposed in [2]. With this construction of $u^\dagger$ the consistent gradient direction is maintained and thus the requirements of Theorem 2.13 are satisfied, so convergence in the described sense can be expected. To reduce the computational effort, we restrict to at most three jumps in vertical as well as in horizontal direction. This means $u^\dagger$ is of the form

$$u^\dagger = \sum_{m=1}^{M} \sum_{n=1}^{N} u^\dagger_{mn} \mathring{\mathbb{1}}_{[x_m, x_{m+1}[\times[y_n, y_{n+1}[}$$

with $M, N \leq 4$. Moreover the minimal distance between the jump points is restricted to the interval $I = [0.075, 0.095]$.



ground truth 1 ground truth 2

Figure 5: Two examples of a valid ground truths with periodic boundary conditions.

In order to apply Algorithm 4 it is necessary to develop an implementation of

- the minimization in line 2 to find the optimal set E_*;
- the minimization in line 8 to find the optimal weights $a^{[k+1]}$ for existing sets $E^{[k+1]}$;
- the local descent in line 11 on

$$(a, E) \mapsto \left(\frac{1}{2} \Big| \sum_{i=1}^{N^{[k]}+1} a_i K^0(\mathring{\mathbb{1}}_{E_i^{[k+1]}}) - f \Big|_2^2 + \alpha \sum_{i=1}^{N^{[k]}+1} \mathrm{Per}(E_i^{[k+1]}; \Omega) \cdot |a_i| \right).$$

Line 2:

We split the minimization step of the algorithm that is carried out in line 2 into two separate minimizations. Firstly, we perform a coarse convex minimization which ensures that we end up in a global minimum. Afterwards, the optimal set $E_* \in \underset{E \text{ simple}}{\operatorname{argmin}} \frac{\operatorname{Per}(E;\Omega)}{|\int_E \eta^{[k]}|}$ is determined directly, initialized with the results from the first minimization step.

Concerning the coarse convex optimization, the initial minimization problem that needs to be solved in line 2 is to find a minimizer

$$(s,v) \text{ s.t. } \operatorname{TV}(v;\Omega) \leq s \leq M \quad \text{ of } \quad D\tilde{J}_\alpha^{f,0}(t,u)(s,v) = \alpha s + \int_\Omega (K^0)^\#(K^0 u - f) \cdot v \, \mathrm{d}x. \tag{I}$$

This was derived in equation (11) but is not a convex problem. However, given a minimizer (s^*, v^*) of (I), the pair $(\frac{s^*}{M}, -\frac{v^*}{M})$ is a solution of

$$\underset{\operatorname{TV}(v)\leq s\leq 1}{\operatorname{argmin}} \quad s \underbrace{-\frac{1}{\alpha}\int_\Omega K^\#(Ku-f)\cdot v \, \mathrm{d}x}_{=\int_\Omega \eta\cdot v \, \mathrm{d}x}. \tag{II}$$

Note that clearly $s = \operatorname{TV}(v)$ as a larger choice of s would be suboptimal. Moreover, for any minimizer (s,v) of (II), $\operatorname{TV}(v) = 1$. Suppose, this was not the case, i.e. $\operatorname{TV}(v) = r < 1$. Then, $\tilde{v} = \frac{v}{r}$ still satisfies $\operatorname{TV}(\tilde{v}) \leq 1$ and

$$\operatorname{TV}(\tilde{v}) + \int_\Omega \eta \cdot \tilde{v} \, \mathrm{d}x = \frac{\operatorname{TV}(v) + \int_\Omega \eta \cdot v \, \mathrm{d}x}{r} < \operatorname{TV}(v) + \int_\Omega \eta \cdot v \, \mathrm{d}x.$$

This contradicts the assumed optimality of v, so indeed for any minimizing pair (s,v) of (II) we have $\operatorname{TV}(v) = s = 1$. Thus, the first summand in (II) has no influence on the minimizer and the optimization reduces to

$$\underset{\operatorname{TV}(v)\leq 1}{\operatorname{argmin}} \int_\Omega \eta \cdot v \, \mathrm{d}x. \tag{III}$$

This is a convex problem and thus the minimization terminates in a global minimum. Moreover, a solution of (III) just differs from a solution of (I) by a scalar, which has no influence on the shape but solely on the intensity of v. Therefore, (III) can be solved instead of (I) in order to determine an optimal set E_*.

The algorithm that solves (III) on a coarse grid and extracts the boundaries of a level set of the solution v was done in [23] by using a primal-dual algorithm. Denote the created set by $\widetilde{E}$.

To compute an optimal simple set E_* minimizing the quotient $\frac{\operatorname{Per}(E;\Omega)}{|\int_E \eta^{[k]}|}$, we restrict ourselves to rectangular sets instead of general simple sets as justified in Section 3.2.2. Starting from an initial rectangle induced by the boundary of $\widetilde{E}$, which ensures global optimality, we refine the set through a gradient-based optimization procedure, a Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm with bounds (L-BFGS-B).

The terms that are passed to the optimization must depend on the outer coordinates of the axis-aligned rectangle $E = [x_{\min}, x_{\max}] \times [y_{\min}, y_{\max}]$, denoted by the tuple $(x_{\min}, x_{\max}, y_{\min}, y_{\max}) \in [0, N]^4$, where N is the size of the underlying grid. Therefore, the perimeter is given by

$$\mathrm{Per}(E;\Omega) = 2 \cdot ((x_{\max} - x_{\min}) + (y_{\max} - y_{\min}))$$

and its partial derivatives by

$$\tfrac{\partial \mathrm{Per}(E;\Omega)}{\partial x_{\min}} = -2 \qquad \tfrac{\partial \mathrm{Per}(E;\Omega)}{\partial x_{\max}} = 2 \qquad \tfrac{\partial \mathrm{Per}(E;\Omega)}{\partial y_{\min}} = -2 \qquad \tfrac{\partial \mathrm{Per}(E;\Omega)}{\partial y_{\max}} = 2.$$

For the computation of $|\int_E \eta|$ recall that $\eta = -\frac{1}{\alpha}(K^0)^{\#}(K^0 u - f)$ with K^0 being the truncated Fourier transform with cut-off frequency Φ and $(K^0)^{\#}$ its inverse mapping into the space with vanishing integral. By not taking the factor $-\frac{1}{\alpha}$ into account as it does not influence the minimizing set and denoting $w = K^0 u - f$ for simplicity, $\int_E \tilde{\eta}\, \mathrm{d}x := \int_E (K^0)^{\#} w\, \mathrm{d}x$ is of the following form:

$$\int_E \tilde{\eta}\, \mathrm{d}x = \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \tilde{\eta}\, \mathrm{d}y\, \mathrm{d}x = \sum_{k_1=-\Phi}^{\Phi} \sum_{k_2=-\Phi}^{\Phi} w_{k_1,k_2} \cdot I(k_1, k_2)$$

with

$$I(k_1,k_2) = \begin{cases} \frac{1}{-(2\pi)^2 k_1 k_2}\Big(\exp(2\pi i \frac{k_1 x_{\max} + k_2 y_{\max}}{N}) - \exp(2\pi i \frac{k_1 x_{\max} + k_2 y_{\min}}{N}) \\ \qquad - \exp(2\pi i \frac{k_1 x_{\min} + k_2 y_{\max}}{N}) + \exp(2\pi i \frac{k_1 x_{\min} + k_2 y_{\min}}{N})\Big) & \text{if } k_1, k_2 \neq 0 \\ \frac{1}{N 2\pi i k_2}\Big(\Big(\exp(2\pi i \frac{k_2 y_{\max}}{N}) - \exp(2\pi i \frac{k_2 y_{\min}}{N})\Big) \cdot x_{\max} \\ \qquad + \Big(\exp(2\pi i \frac{k_2 y_{\min}}{N}) - \exp(2\pi i \frac{k_2 y_{\max}}{N})\Big) \cdot x_{\min}\Big) & \text{if } k_1 = 0, k_2 \neq 0 \\ \frac{1}{N 2\pi i k_1}\Big(\Big(\exp(2\pi i \frac{k_1 x_{\max}}{N}) - \exp(2\pi i \frac{k_1 x_{\min}}{N})\Big) \cdot y_{\max} \\ \qquad + \Big(\exp(2\pi i \frac{k_1 x_{\min}}{N}) - \exp(2\pi i \frac{k_1 x_{\max}}{N})\Big) \cdot y_{\min}\Big) & \text{if } k_2 = 0, k_1 \neq 0 \\ 0 & \text{if } k_1 = k_2 = 0. \end{cases}$$

Here, the vanishing integral of $\tilde{\eta}$ is ensured by setting $I(0,0) = 0$. This excludes the entry $w_{0,0}$ from the frequency space denoting the mean value of $\tilde{\eta}$ as explained in Remark 3.22.

The partial derivatives of the integral with respect to the coordinates can be computated analytically in a straightforward manner, as exemplified for $x_{\min}$ below.

$$\frac{\partial}{\partial x_{\min}} \int_E \tilde{\eta} \, \mathrm{d}x = \sum_{k_1=-\Phi}^{\Phi} \sum_{k_2=-\Phi}^{\Phi} w_{k_1,k_2} \cdot \frac{\partial}{\partial x_{\min}} I(k_1, k_2)$$

with

$$\frac{\partial}{\partial x_{\min}} I(k_1,k_2) = \begin{cases} \frac{i}{-2\pi k_2 N}\Big(- \exp(2\pi i \frac{k_1 x_{\min} + k_2 y_{\max}}{N}) \\ \qquad + \exp(2\pi i \frac{k_1 x_{\min} + k_2 y_{\min}}{N})\Big) & \text{if } k_1, k_2 \neq 0 \\ \frac{1}{N 2\pi i k_2}\Big(\exp(2\pi i \frac{k_2 y_{\min}}{N}) - \exp(2\pi i \frac{k_2 y_{\max}}{N})\Big) & \text{if } k_1 = 0, k_2 \neq 0 \\ \frac{1}{N^2}\Big(\Big(- \exp(2\pi i \frac{k_1 x_{\min}}{N})\Big) \cdot y_{\max} \\ \qquad + \Big(\exp(2\pi i \frac{k_1 x_{\min}}{N})\Big) \cdot y_{\min}\Big) & \text{if } k_2 = 0, k_1 \neq 0 \\ 0 & \text{if } k_1 = k_2 = 0. \end{cases}$$

Finally, the gradient that is transferred to the L-BFGS-B is obtained via quotient rule:

$$\nabla \frac{\operatorname{Per}(E;\Omega)}{|\int_E \tilde{\eta}|} = \operatorname{sign}\left(\int_E \tilde{\eta}\right) \cdot \frac{\nabla(\operatorname{Per}(E;\Omega)) \cdot \int_E \tilde{\eta} - \nabla(\int_E \tilde{\eta}) \cdot \operatorname{Per}(E;\Omega)}{(\int_E \tilde{\eta})^2}$$

Line 8:

Let the current function at iteration k be of the form $\sum_{i=1}^{N^{[k]}+1} \tilde{a}_i^{[k+1]} \mathring{\mathbb{1}}_{E_i^{[k+1]}}$ with $E^{[k+1]} = (E_1^{[k]}, \ldots, E_{N^{[k]}}^{[k]}, E_*)$, $\tilde{a}^{[k+1]} = (a_1^{[k]}, \ldots, a_{N^{[k]}}^{[k]}, 1)$ and $N^{[k]} + 1$ the number of rectangular sets. The aim of line 8 of Algorithm 4 is to improve the weights $\tilde{a}_i^{[k+1]}$ after the new rectangle E_* is added. Note, that we optimize over all weights and not just the new one corresponding to E_*. Again, a L-BFGS-B is used with optimization parameters $a \in [-1,1]^{N^{[k]}+1}$. The objective that is passed to the algorithm is of the form

$$a \mapsto \left(\frac{1}{2} | \sum_{i=1}^{N^{[k]}+1} a_i K^0(\mathring{\mathbb{1}}_{E_i^{[k+1]}}) - f|_2^2 + \alpha \sum_{i=1}^{N^{[k]}+1} \operatorname{Per}(E_i^{[k+1]}; \Omega) \cdot |a_i| \right)$$

as stated in equation (37) and can be formulated equivalently by denoting $E_j^{[k+1]} = [x^j_{\min}, x^j_{\max}] \times [y^j_{\min}, y^j_{\max}]$ and plugging in the truncated Fourier transform by

$$a \mapsto \Bigg(\frac{1}{2} \sum_{k_1,k_2=-\Phi}^{\Phi} \Big| \sum_{j=1}^{N^{[k]}+1} a_j \cdot \underbrace{\int_{x^j_{\min}}^{x^j_{\max}} \int_{y^j_{\min}}^{y^j_{\max}} \exp(-2\pi i (\frac{x k_1 + y k_2}{N}))\, \mathrm{d}y\, \mathrm{d}x}_{=:H^j(k_1,k_2)} - f_{k_1,k_2} \Big|_2^2 + \alpha \sum_{j=1}^{N^{[k]}+1} 2|a_j| \cdot ((x^j_{\max} - x^j_{\min}) + (y^j_{\max} - y^j_{\min})) \Bigg). \tag{38}$$

Note that $H^j(k_1,k_2)$ and f_{k_1,k_2} are complex as they stem from the Fourier frequency space. Thus, the partial derivatives of $\frac{1}{2}\sum_k |\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2}|_2^2$ in the directions $r \in (a, x_{\min}, x_{\max}, y_{\min}, y_{\max})$ are of the form

$$\begin{aligned} &\frac{\partial}{\partial r} \frac{1}{2} \sum_k |\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2}|_2^2 \\ &= \frac{\partial}{\partial r} \frac{1}{2} \sum_k \Big((\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2}) \cdot \overline{(\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2})} \Big) \\ &= \frac{1}{2} \sum_k \Big(\frac{\partial}{\partial r} (\sum_j a_j H^j(k_1,k_2)) \cdot \overline{(\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2})} \\ &\quad + (\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2}) \cdot \overline{\frac{\partial}{\partial r} (\sum_j a_j H^j(k_1,k_2))} \Big) \\ &= \sum_k \Re \Big(\frac{\partial}{\partial r} (\sum_j a_j H^j(k_1,k_2)) \cdot \overline{(\sum_j a_j H^j(k_1,k_2) - f_{k_1,k_2})} \Big) \end{aligned} \tag{39}$$

In line 8 of the algorithm, $r = a$ and for any $i \in \{1, \dots, N^{[k]} + 1\}$:

$$\frac{\partial}{\partial a_i} \sum_{j=1}^{N^{[k]}+1} a_j H^j(k_1,k_2) = a_i \cdot H^i(k_1,k_2),$$

so the overall gradient that is passed to the L-BFGS-B consists of $N^{[k]} + 1$ partial derivatives of the form

$$\begin{aligned} \frac{\partial}{\partial a_i} = \sum_{k_1,k_2=-\Phi}^{\Phi} \Re \Bigg(a_i \cdot H^i(k_1,k_2) \cdot \overline{\Big((\sum_{j=1}^{N^{[k]}+1} a_j H^j(k_1,k_2) - f_{k_1,k_2}) \Big)} \Bigg) \\ + 2\alpha \cdot ((x^i_{\max} - x^i_{\min}) + (y^i_{\max} - y^i_{\min})) \cdot \operatorname{sign}(a_i). \end{aligned} \tag{40}$$

The double integral $H^j(k_1,k_2) = \int_{x^j_{\min}}^{x^j_{\max}} \int_{y^j_{\min}}^{y^j_{\max}} \exp(-2\pi i (\frac{x k_1 + y k_2}{N}))\, \mathrm{d}y\, \mathrm{d}x$ that appears in the objective as well as in the gradient can be computed analytically and is implemented similar to $I(k_1, k_2)$ from line 2 of the algorithm in the following way:

$$H^j(k_1,k_2) = \begin{cases} \frac{N^2}{-(2\pi)^2 k_1 k_2}\Big(\exp(-2\pi i \frac{k_1 x^j_{\text{max}} + k_2 y^j_{\text{max}}}{N}) \\ \qquad - \exp(-2\pi i \frac{k_1 x^j_{\text{max}} + k_2 y^j_{\text{min}}}{N}) \\ \qquad - \exp(-2\pi i \frac{k_1 x^j_{\text{min}} + k_2 y^j_{\text{max}}}{N}) \\ \qquad + \exp(-2\pi i \frac{k_1 x^j_{\text{min}} + k_2 y^j_{\text{min}}}{N})\Big) & \text{if } k_1, k_2 \neq 0 \\ \frac{N}{2\pi i k_2}\Big(\Big(- \exp(-2\pi i \frac{k_2 y^j_{\text{max}}}{N}) \exp(-2\pi i \frac{k_2 y^j_{\text{min}}}{N})\Big) \cdot x^j_{\text{max}} \\ \qquad + \Big(\exp(-2\pi i \frac{k_2 y^j_{\text{max}}}{N}) - \exp(-2\pi i \frac{k_2 y^j_{\text{min}}}{N})\Big) \cdot x^j_{\text{min}}\Big) & \text{if } k_1 = 0, k_2 \neq 0 \\ \frac{N}{2\pi i k_1}\Big(\Big(- \exp(-2\pi i \frac{k_1 x^j_{\text{max}}}{N}) + \exp(-2\pi i \frac{k_1 x^j_{\text{min}}}{N})\Big) \cdot y^j_{\text{max}} \\ \qquad + \Big(\exp(-2\pi i \frac{k_1 x^j_{\text{max}}}{N}) - \exp(-2\pi i \frac{k_1 x^j_{\text{min}}}{N})\Big) \cdot y^j_{\text{min}}\Big) & \text{if } k_2 = 0, k_1 \neq 0 \\ 0 & \text{if } k_1 = k_2 = 0. \end{cases}$$

Line 11:

The objective that is supposed to be reduced in line 11 of Algorithm 4 is similar to equation (38). What differs here is that we do not solely optimize over the weights a_i but at the same time over all boundaries of existing rectangles. Again, a L-BFGS-B algorithm is used to find parameters $(a_i, x^i_{\text{min}}, x^i_{\text{max}}, y^i_{\text{min}}, y^i_{\text{miax}}) \in [-1, 1] \times [0, N]^4$ with $i \in \{1, \ldots, N^{[k+1]}\}$ that reduce the objective. With the notation from above the objective can be written via

$$(a, x_{\text{min}}, x_{\text{max}}, y_{\text{min}}, y_{\text{max}}) \mapsto \Bigg(\frac{1}{2} \sum_{k_1,k_2=-\Phi}^{\Phi} \Big| \sum_{j=1}^{N^{[k]}+1} a_j \cdot H^j(k_1,k_2) - f_{k_1,k_2} \Big|_2^2 \\ + \alpha \sum_{j=1}^{N^{[k]}+1} 2|a_j| \cdot ((x^j_{\text{max}} - x^j_{\text{min}}) + (y^j_{\text{max}} - y^j_{\text{min}})) \Bigg).$$

What remains is the computation of the partial derivatives concerning the coordinates, as $\frac{\partial}{\partial a_i}$ has already been determined in equation (40). As $H^i(k_1, k_2)$ depends on the coordinates of the i-th rectangle, we need to compute partial derivatives of $H^i(k_2, k_2)$ for $x^i_{\text{min}}, x^i_{\text{max}}, y^i_{\text{min}}, y^i_{\text{max}}$. Again, we will just formulate this explicitly for x^i_{min} as the remaining coordinates follow similarly by calculation.

$$\frac{\partial}{\partial x^i_{\min}} H^i(k_1,k_2) = \begin{cases} \frac{iN}{-2\pi k_2}\Big(\exp(-2\pi i \frac{k_1 x^i_{\min} + k_2 y^i_{\max}}{N}) \\ \qquad - \exp(-2\pi i \frac{k_1 x^i_{\min} + k_2 y^i_{\min}}{N})\Big) & \text{if } k_1, k_2 \neq 0 \\ \frac{N}{2\pi i k_2}\Big(\exp(-2\pi i \frac{k_2 y^i_{\max}}{N}) - \exp(-2\pi i \frac{k_2 y_{\min}}{N})\Big) & \text{if } k_1 = 0, k_2 \neq 0 \\ \Big(- \exp(-2\pi i \frac{k_1 x^i_{\min}}{N})\Big) \cdot y^i_{\max} \\ \qquad + \Big(\exp(-2\pi i \frac{k_1 x^i_{\min}}{N})\Big) \cdot y^i_{\min} & \text{if } k_2 = 0, k_1 \neq 0 \\ 0 & \text{if } k_1 = k_2 = 0. \end{cases}$$

4.2. Numerical results

In this section the created algorithms are tested on the image sources created in Figure 5. We are going to display how the classic Frank-Wolfe algorithm (Algorithm 3) and its sliding variation (Algorithm 4) work and how their reconstructed results differ.

The noise strength $\frac{1}{2}|f - f^{\dagger}|_2^2 \leq \delta$ that is considered in the experiments is $\delta = 10^{-2}$ and according to [2] the regularization parameter is chosen as $\alpha = C\sqrt{\delta}$ with $C = \frac{1}{0.028}$. In both presented examples the classic Frank-Wolfe algorithm as well as the sliding version are applied on input data generated by a truncated Fourier transform with cut-off frequency $\Phi = 5$ and $\Phi = 18$. The reconstructed images and their corresponding deviations from the ground truth are illustrated in Figures 6 to 9.

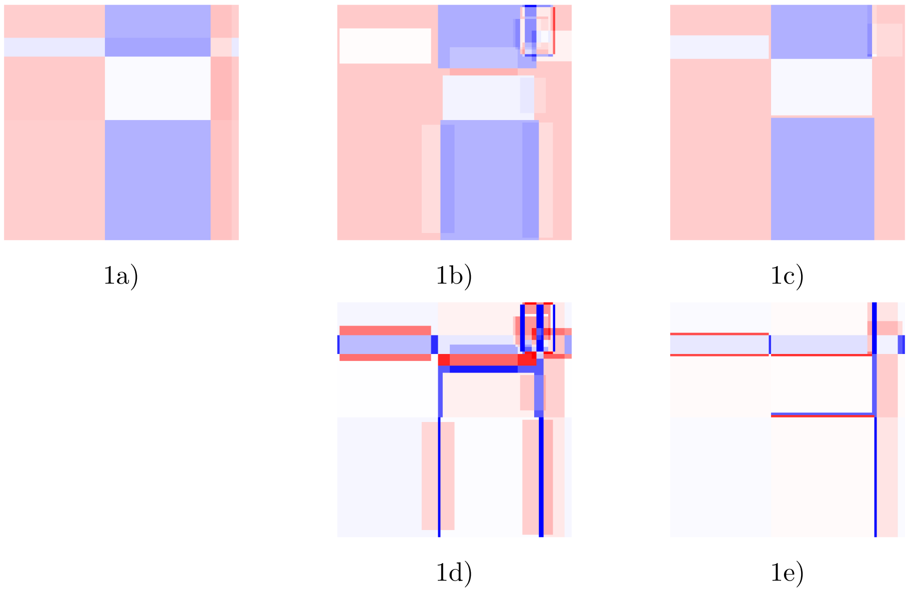


Figure 6: Output of the reconstruction with classic Frank-Wolfe algorithm for different cut-off frequencies Φ. **1a)** ground truth 1, **1b)** classic FW, $\Phi = 5$, **1c)** classic FW, $\Phi = 18$, **1d)** difference of 1a) and 1b), **1e)** difference of 1a) and 1c).

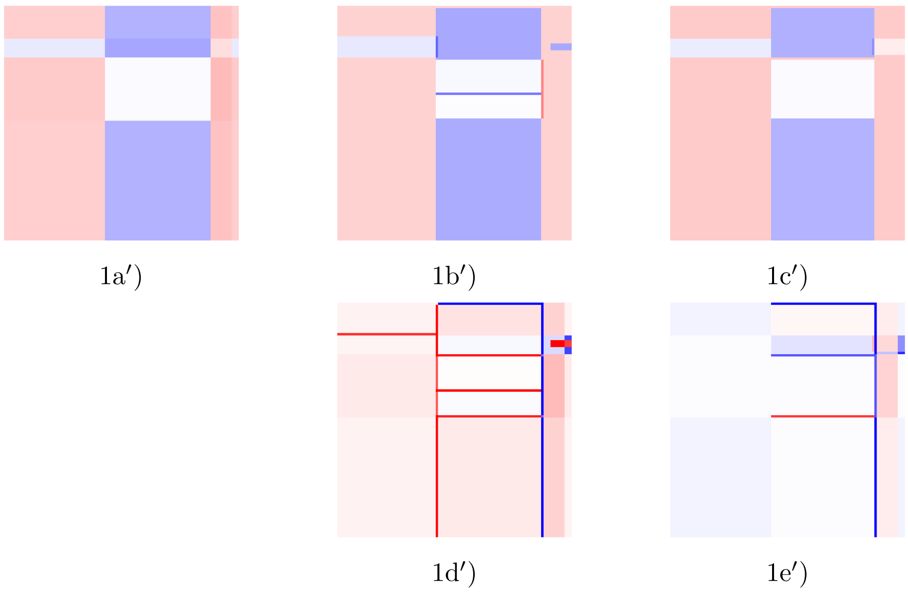


Figure 7: Output of the reconstruction with sliding Frank-Wolfe algorithm for different cut-off frequencies Φ. **1a′)** ground truth 1, **1b′)** classic FW, $\Phi = 5$, **1c′)** classic FW, $\Phi = 18$, **1d′)** difference of 1a′) and 1b′), **1e′)** difference of 1a′) and 1c′).

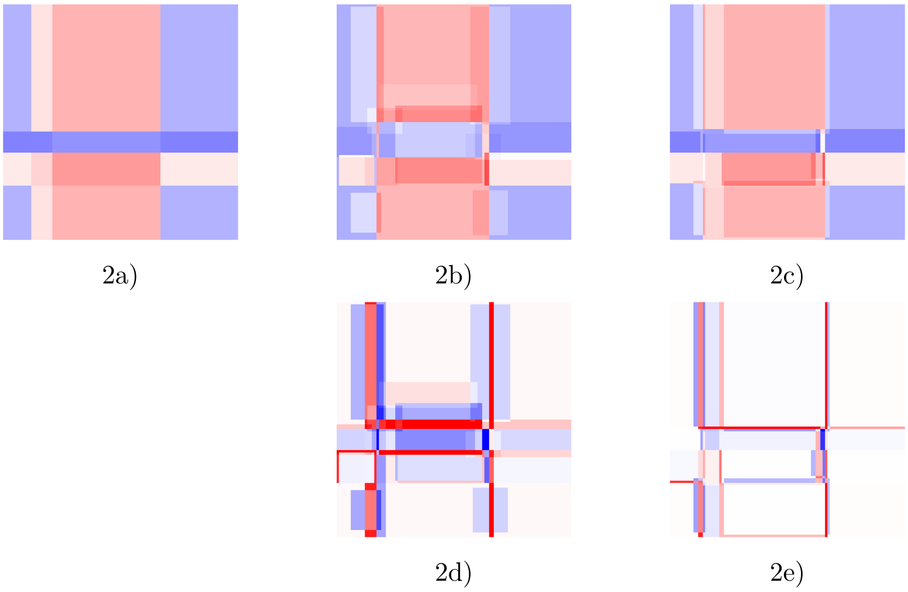


Figure 8: Output of the reconstruction with classic Frank-Wolfe algorithm for different cut-off frequencies Φ. **2a)** ground truth 2, **2b)** classic FW, $\Phi = 5$, **2c)** classic FW, $\Phi = 18$, **2d)** difference of 2a) and 2b), **2e)** difference of 2a) and 2c).

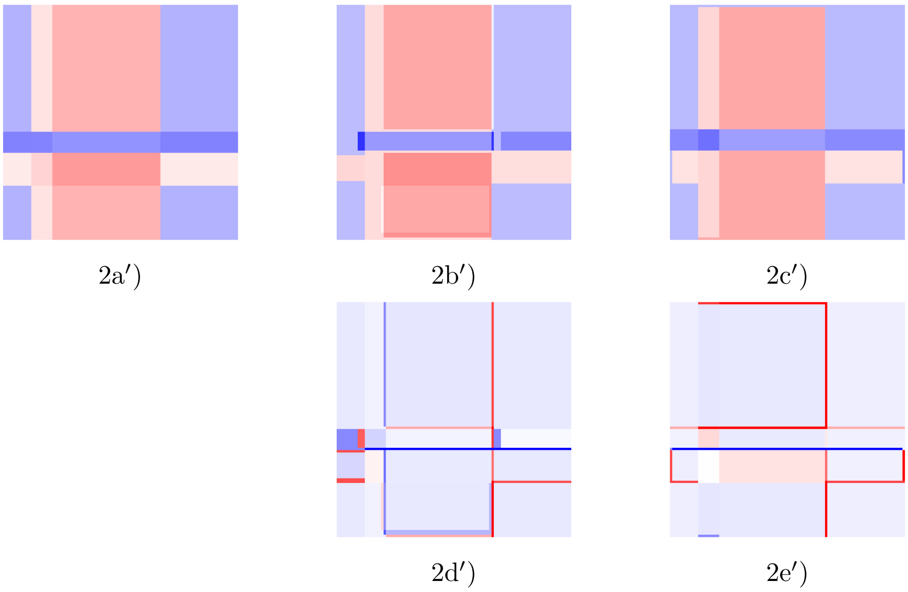


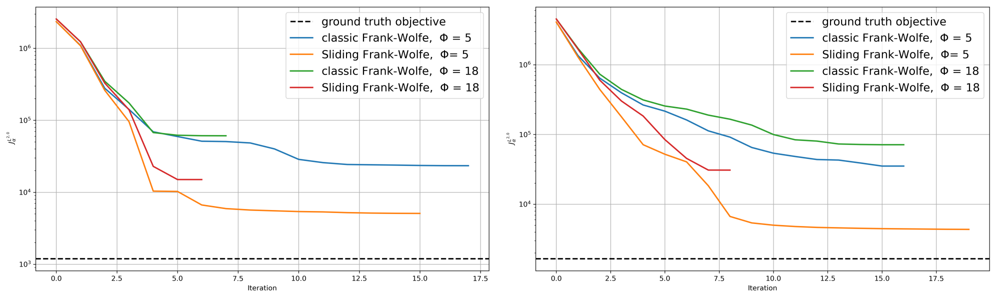
Figure 9: Output of the reconstruction with sliding Frank-Wolfe algorithm for different cut-off frequencies Φ. **2a′)** ground truth 2, **2b′)** classic FW, $\Phi = 5$, **2c′)** classic FW, $\Phi = 18$, **2d′)** difference of 2a′) and 2b′), **2e′)** difference of 2a′) and 2c′).

The reduction of the objective

$$J_\alpha^{L^{2,0}} = \frac{1}{2} | \sum_{i=1}^{N^{[k]}+1} a_i K^0(\mathring{\mathbb{1}}_{E_i^{[k+1]}}) - f|_2^2 + \alpha \sum_{i=1}^{N^{[k]}+1} \mathrm{Per}(E_i^{[k+1]}; \Omega) \cdot |a_i|$$

as well as the number of iterations that are required to obtain the results from Figures 6 to 9 are displayed in Figure 10. Here the maximum number of iterations is limited by N=20 due to the computation time. Note that the scale of the y-axis is algorithmic and the bottom black dashed line represents the objective value of the ground truth. In its computation the fidelity term is ignored as it is of magnitude $\delta = 10^{-2}$ and thus negligible.

Looking at the differences between the SFW and classic FW we notice that the SFW reduces the objective more than the classic FW does for either ground truths. With both tested cut-off frequencies the SFW produces a lower objective than the classic FW for any of the tested $\Phi \in \{5, 18\}$. Comparing the results for different cut-off frequencies but the same algorithm, it is remarkable that with a lower Φ the objective is reduced more. This can be explained with the fidelity term of $J_\alpha^{f,0}$. K^0 produces an image with at most $(2\Phi + 1)^2$ nonzero entries, where a deviation from f may occur. This means for inaccurate results the error of the fidelity term rises with increasing cut-off frequency Φ, which can be noticed in Figure 10 for both tested ground truths.

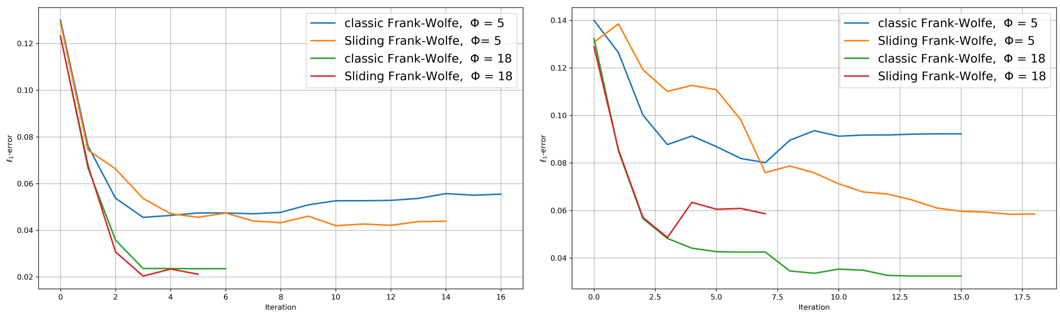


a) b)

Figure 10: Development of the objective $J_\alpha^{f,0}$ in the course of iterations. **a)** considered ground truth 1 (reconstructions from Figures 6 and 7); **b)** considered ground truth 2 (reconstructions from Figures 8 and 9).

Figure 11 displays the development of the ℓ_1-error of the reconstructions. One can observe that the two exemplary ground truths provide different results regarding the ℓ_1-error estimation.

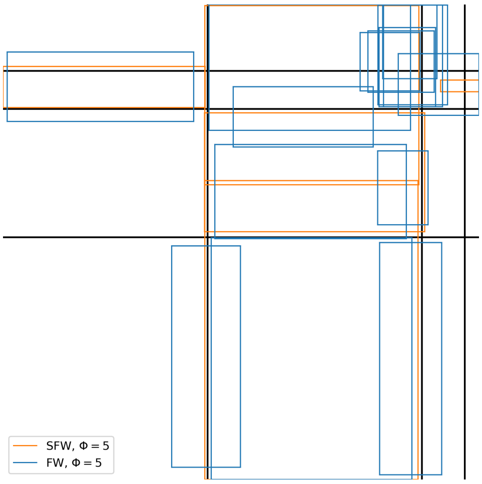
Figure 11a) shows that regardless of the accuracy of the input image, i.e. for $\Phi = 5$ and $\Phi = 18$, the SFW provides a result with lower ℓ_1-error than the standard FW. Moreover, the SFW in case of $\Phi = 18$ yields the reconstruction with smallest ℓ_1-error in least iterations at the same time. This differs in the second example. In case of $\Phi = 5$, the SFW still produces the reconstruction with lower ℓ_1-error in comparison to the classic FW, but for $\Phi = 18$, the SFW result is considerably worse than the classic FW result in terms of ℓ_1-error.



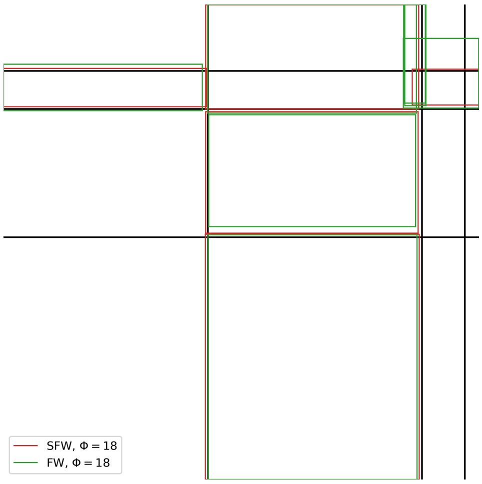
a) b)

Figure 11: ℓ_1-error of the reconstructed image in the course of iterations. **a)** considered ground truth 1 (reconstructions from Figures 6 and 7); **b)** considered ground truth 2 (reconstructions from Figures 8 and 9).

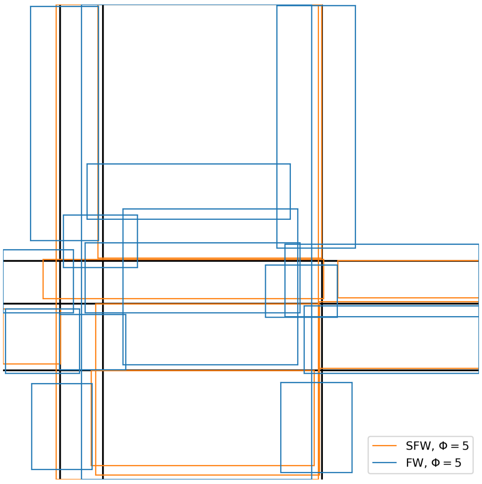
The main advantage of the sliding Frank-Wolfe is its accuracy when it comes to the recovery of the jump points of the ground truth images and its sparseness. Figure 12 illustrates the gradient supports of the reconstructed images for both examples as well as the gradient support of the ground truth image.



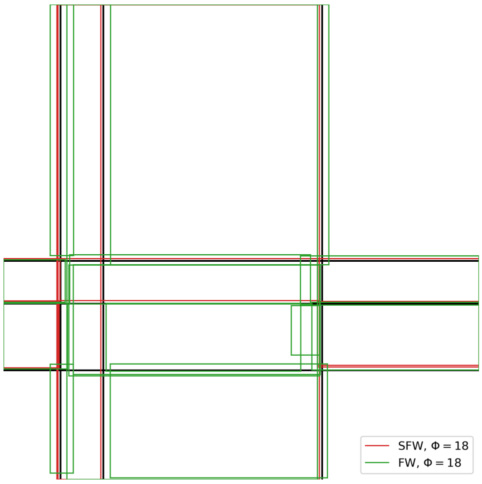
a) ground truth 1, $\Phi = 5$



b) ground truth 1, $\Phi = 18$



c) ground truth 2, $\Phi = 5$



d) ground truth 2, $\Phi = 18$

Figure 12: Visualization of the gradient support of the reconstructions and in black the gradient support of the ground truths. The number of rectangles the reconstructed images are made of: **a)** SFW: 5, FW: 15; **b)** SFW: 5, FW: 7; **c)** SFW: 8, FW: 15; **d)** SFW: 7, FW: 14.

5. Conclusion and outlook

In this thesis, we analyzed the classic Frank-Wolfe algorithm as well as the sliding Frank-Wolfe algorithm applied on TV regularized inverse problems, in particular in order to reconstruct piecewise constant functions from a truncated Fourier image. As seen in the last chapter, a key benefit of the approach including a sliding step is its ability to accurately recover sharp edges from the underlying signal and establish a sparse reconstruction.

There are numerous instances where this is highly desirable. As indicated in the introduction, the accurate localization of tissue boundaries can be crucial for diagnosis in medical imaging. Similar to this, the interpretation of subsurface structures in geophysical or seismic imaging can be greatly impacted by the ability to accurately detect layer transitions. Furthermore, the numerical findings reveal that the SFW outperforms the traditional FW in difficult setups with little data information, in addition to producing visually sharper reconstructions. This indicates that the SFW is quite suitable in practical scenarios as often only incomplete and inexact measurement data, such as a truncated Fourier transform with small cut-off frequency, are given.

Despite these promising outcomes, several open directions remain. One natural next step is to improve the computational performance of the algorithm especially when extensive data needs to be processed. This particularly includes the computation of gradients, which is required in every minimization step. The computation becomes very costly for a high cut-off frequency Φ of the measurement operator or for rising number of sets that need to be reconstructed.
Additionally, the implemented SFW algorithm could be improved in order to reduce the ℓ_1-error of the reconstruction. As previously stated, this is not of primary interest; however, it would contribute to improving the reconstructions overall accuracy and visual quality.

All things considered, the strategy that has been presented establishes a basis for additional investigation into edge-preserving techniques for inverse problems, which could serve a wide range of purposes.

A. Subsidiary statements

Total variation computation

Lemma A.1. *The anisotropic perimeter of a connected convex and bounded set E equals the length of the edges of the surrounding axis-aligned rectangle.*

Proof. Without loss of generality consider the two-dimensional setting. Let E be a bounded connected set and $\Omega \subset \mathbb{R}^2$ be an open set. Recall that the perimeter of E in Ω is defined by

$$\begin{aligned}\mathrm{Per}(E;\Omega) &= \sup\left\{\int_E \mathrm{div}\phi(x)\ \mathrm{d}x \mid \phi \in C_c^1(\Omega;\mathbb{R}^2);\ \|\phi\|_\infty \le 1\right\}\\ &= \sup\left\{\int_{\partial E} \phi\cdot\nu\ \mathrm{d}A \mid \phi \in C_c^1(\Omega;\mathbb{R}^2); \|\phi\|_\infty \le 1\right\},\end{aligned}$$

where ν denotes the outer unit normal and A the surface measure.
Ignoring the constraints one can observe that the supremum of $\int_{\partial E}\phi\cdot\nu\,\mathrm{d}A$ is attained for $\phi_i = \mathrm{sign}(\nu_i)$ with $i \in \{1,2\}$ on ∂E. Then,

$$\int_{\partial E}\phi\cdot\nu\,\mathrm{d}A = \int_{\partial E}|\nu|\,\mathrm{d}A = \int_{\partial E}|\nu_1| + |\nu_2|\,\mathrm{d}A.$$

With the definitions as introduced in Figure 13 such a ϕ can be defined on the whole domain Ω by

$$\phi_1(x) = \begin{cases}-1 & \text{on } \partial E_3,\\ 1 & \text{on } \partial E_4,\\ \in[-1,1] & \text{otherwise}\end{cases} \qquad \phi_2(x) = \begin{cases}1 & \text{on } \partial E_1,\\ -1 & \text{on } \partial E_2,\\ \in[-1,1] & \text{otherwise.}\end{cases}$$

However, this ϕ cannot be a $C_c^1(\Omega;\mathbb{R}^2)$-function, as ϕ_1 is discontinuous on ∂E, where ∂E_3 and ∂E_4 meet, i.e. where $x_2 = x_2^b$ and $x_2 = x_2^t$. In these points, there is a jump from -1 to 1 or vice versa. Exactly the same is true for ϕ_2, in the points where ∂E_1 and ∂E_2 meet. Nevertheless, e.g. by [15, Cor. 4.23], for $\Omega \subset \mathbb{R}^N$ an open set, $C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$ for any $1 \le p < \infty$. As Ω is open and bounded and ϕ_1 and ϕ_2 clearly are $L^1(\Omega)$-functions, the corollary is applicable. Thus, there are sequences of $C_c^\infty(\Omega)$-functions $(\phi_1)_n$ and $(\phi_2)_n$ that approximate ϕ_1 and ϕ_2 with arbitrary accuracy and one can conclude that ϕ is indeed the supremum of the set above.
To prove the desired statement it remains to show $\mathrm{Per}(E;\Omega) = \int_{\partial E}|\nu_1| + |\nu_2|\ \mathrm{d}A \overset{!}{=} 2|x_1^r - x_1^l| + 2|x_2^t - x_2^b|$. Consider $\int_{\partial E}|\nu_1|\,\mathrm{d}A$ and $\int_{\partial E}|\nu_2|\ \mathrm{d}A$ separately.

- $\int_{\partial E}|\nu_2|\ \mathrm{d}A$:
  Let ∂E_1 and ∂E_2 be defined as before in Figure 13. Then $\int_{\partial E}|\nu_2|\ \mathrm{d}A = \int_{\partial E_1}|\nu_2|\ \mathrm{d}A + \int_{\partial E_2}|\nu_2|\ \mathrm{d}A$.

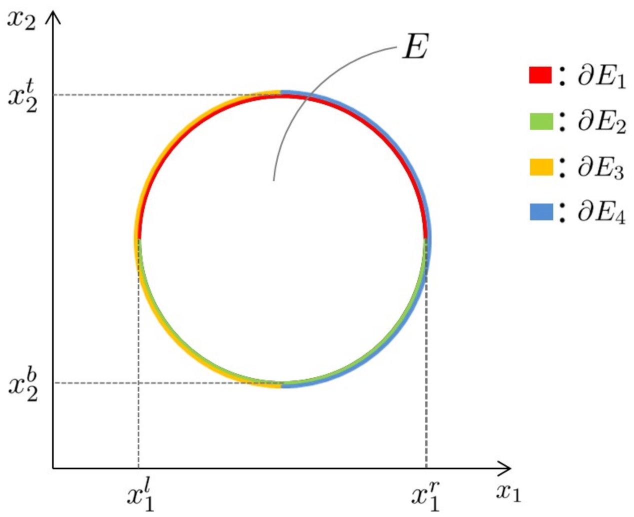


Figure 13: Two partitions of the boundary of E and identification of the outermost points x_1^l, x_1^r, x_2^b and x_2^t of E. Note, that $\partial E = \partial E_1 \cup \partial E_2 = \partial E_3 \cup \partial E_4$ with $\partial E_1 \cap \partial E_2 = \emptyset$ and $\partial E_3 \cap \partial E_4 = \emptyset$.

Concerning the first summand, let ∂E_1 be parametrized by $x_2 = p_1(x_1)$. Then:

$$\int_{\partial E_1} |\nu_2| \, \mathrm{d}A = \int_{x_1^l}^{x_1^r} |\nu_2(p_1(x_1))| \cdot \sqrt{1 + |p_1'(x_1)|^2} \, \mathrm{d}x_1$$

with $\nu(p_1(x_1)) = \frac{1}{\sqrt{1+|p_1'(x_1)|^2}} \cdot \begin{pmatrix} -p_1'(x_1) \\ 1 \end{pmatrix}$. This gives

$$\begin{aligned} \int_{\partial E_1} |\nu_2| \, \mathrm{d}A &= \int_{x_1^l}^{x_1^r} \frac{1}{\sqrt{1 + |p_1'(x_1)|^2}} \cdot \sqrt{1 + |p_1'(x_1)|^2} \, \mathrm{d}x_1 \\ &= \int_{x_1^l}^{x_1^r} 1 \, \mathrm{d}x_1 = |x_1^r - x_1^l|. \end{aligned}$$

Similarly, for ∂E_2 parametrized by $x_2 = p_2(x_1)$, the outer unit normal is of the form $\nu(p_2(x_1)) = \frac{1}{\sqrt{1+|p_2'(x_1)|^2}} \cdot \begin{pmatrix} p_2(x_1)' \\ -1 \end{pmatrix}$. With the former computations this yields

$$\int_{\partial E_2} |\nu_2| \, \mathrm{d}A = \int_{x_1^l}^{x_1^r} \frac{|-1|}{\sqrt{1 + |p_2(x_1)'|^2}} \cdot \sqrt{1 + |p_2(x_1)'|^2} \, \mathrm{d}x_1 = |x_1^r - x_1^l|.$$

Putting this together, one gets:

$$\int_{\partial E} |\nu_2| \, \mathrm{d}A = 2 \cdot |x_1^r - x_1^l|.$$

- $\int_{\partial E} |\nu_1| \, \mathrm{d}A$:
  With the same argumentation as above and the partition of ∂E into ∂E_3 and ∂E_4 with associated parametrisations $x_1 = p_3(x_2)$ and $x_1 = p_4(x_2)$ one gets a quite similar result as before. Note that the outer unit normal vectors of the parametrisations are given by

$$\nu(p_3(x_2)) = \frac{1}{\sqrt{1 + |p_3(x_2)'|^2}} \cdot \begin{pmatrix} -1 \\ -p_3(x_2)' \end{pmatrix}$$
$$\text{and} \quad \nu(p_4(x_2)) = \frac{1}{\sqrt{1 + |p_4(x_2)'|^2}} \cdot \begin{pmatrix} 1 \\ p_4(x_2)' \end{pmatrix}.$$

  Plugging this in, one gets

$$\begin{aligned} \int_{\partial E} |\nu_1| \, \mathrm{d}A &= \int_{\partial E_3} |\nu_1| \, \mathrm{d}A + \int_{\partial E_4} |\nu_1| \, \mathrm{d}A \\ &= \int_{x_2^b}^{x_2^t} \frac{|-1|}{\sqrt{1 + |p_3(x_2)'|^2}} \cdot \sqrt{1 + |p_3(x_2)'|^2} \, \mathrm{d}x_2 \\ &\qquad + \int_{x_2^b}^{x_2^t} \frac{1}{\sqrt{1 + |p_4(x_2)'|^2}} \cdot \sqrt{1 + |p_4(x_2)'|^2} \, \mathrm{d}x_2 \\ &= 2 \cdot |x_2^t - x_2^b| \end{aligned}$$

Putting everything together, one can conclude

$$\mathrm{Per}(E;\Omega) = \int_{\partial E} |\nu_1| + |\nu_2| \, \mathrm{d}A = 2 \cdot |x_1^r - x_1^l| + 2 \cdot |x_2^t - x_2^b|$$

which equals the length of the edges of the surrounding rectangle and shows the claim.

□

Property of subgradients of 1-homogeneous functional

Lemma A.2 (Euler's Homogeneous Function Theorem, [25, Thm. 3.1.21])**.** *Let a function T be convex and subdifferentiable on its domain. If it is homogeneous of degree $p \geq 1$, i.e. if*

$$T(\tau x) = \tau^p \, T(x) \qquad \forall x \in domT, \tau \geq 0, \tag{41}$$

then

$$\langle g, x \rangle = pT(x) \qquad \forall x \in domT, \forall g \in \partial T(x). \tag{42}$$

Proof. The proof stems from [25] and is stated here for completeness. Let $x \in \text{dom}T$ and $g \in \partial T(x)$. Then for any $\tau \geq 0$ one gets with the provided p-homogeneity (Equation (41)) and the definition of the subdifferential (Definition 3.23)

$$\tau^p \, T(x) = T(\tau x) \geq \langle g, \tau x - x \rangle + T(x) = (\tau - 1)\langle g, x \rangle + T(x). \tag{43}$$

As this inequality holds for all $\tau \geq 0$, consider $\tau < 1$ and $\tau > 1$ seperately to estimate Equation (42) in both directions.

- $\tau > 1$: This choice of τ yields with Equation (43)

  $$\frac{\tau^p - 1}{\tau - 1} T(x) \geq \langle g, x \rangle.$$

  Now take the limit $\tau \downarrow 1$. This gives with the l'Hospital rule (e.g. [26, Thm. 10], originally stated as Satz 10):

  $$\lim_{\tau \downarrow 1} \frac{\tau^p - 1}{\tau - 1} T(x) = \lim_{\tau \downarrow 1} \frac{p\tau^{p-1}}{1} T(x) = pT(x) \geq \langle g, x \rangle.$$

- $\tau < 1$: Again, Equation (43) gives

  $$\frac{1 - \tau^p}{1 - \tau} T(x) \leq \langle g, x \rangle.$$

  Taking the limit $\tau \uparrow 1$ the l'Hospital rule yiels

  $$\lim_{\tau \uparrow 1} \frac{1 - \tau^p}{1 - \tau} T(x) = \lim_{\tau \uparrow 1} \frac{-p\tau^{p-1}}{-1} T(x) = pT(x) \leq \langle g, x \rangle.$$

From the above considerations we can now conclude that $pT(x) = \langle g, x \rangle$ holds for all $x \in \text{dom}T$ and every $g \in \partial T(x)$. □

References

[1] Benedikt Wirth. Inverse problems. lecture notes, 2021.

[2] Martin Holler and Benedikt Wirth. Exact reconstruction and reconstruction from noisy data with anisotropic total variation, 2023. URL https://arxiv.org/abs/2207.04757.

[3] Quentin Denoyelle, Vincent Duval, Gabriel Peyré, and Emmanuel Soubies. The sliding frank-wolfe algorithm and its application to super-resolution microscopy, 2018.

[4] Antigoni Panagiotopoulou and Vassilis Anastassopoulos. Super-resolution image reconstruction techniques: Trade-offs between the data-fidelity and regularization terms. *Inf. Fusion*, 13:185–195, 2012. URL https://api.semanticscholar.org/CorpusID:11410357.

[5] Luigi Ambrosio, Diego Pallara, and Nicola Fusco. Functions of bounded variation and free discontinuity problems. 2000. URL https://api.semanticscholar.org/CorpusID:118499431.

[6] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Monographs in mathematics. Birkhäuser, 1984. ISBN 9783764331535. URL https://books.google.de/books?id=wgPEQgAACAAJ.

[7] Paul Rodriguez. Total variation regularization algorithms for images corrupted with different noise models: A review. *Journal of Electrical and Computer Engineering*, 2013, 01 2013. doi: 10.1155/2013/217021.

[8] Sebastian Pokutta. The frank-wolfe algorithm: a short introduction, 2023.

[9] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956. URL https://EconPapers.repec.org/RePEc:wly:navlog:v:3:y:1956:i:1-2:p:95-110.

[10] Fabian Pedregosa, Geoffrey Negiar, Armin Askari, and Martin Jaggi. Linearly convergent frank-wolfe with backtracking line-search, 2022.

[11] C.D. Aliprantis and K.C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Studies in Economic Theory. Springer Berlin Heidelberg, 2013. ISBN 9783662030042. URL https://books.google.de/books?id=TxjpCAAAQBAJ.

[12] Wenqi Lu, Jinming Duan, David Orive-Miguel, Lionel Herve, and Iain B Styles. Graph- and finite element-based total variation models for the inverse problem in diffuse optical tomography, 2019. URL https://arxiv.org/abs/1901.01969.

[13] M.J. Fabian. *Functional Analysis and Infinite-Dimensional Geometry*. CMS Books in Mathematics. Springer, 2001. ISBN 9780387952192. URL https://books.google.de/books?id=Hcqm4_lW4EkC.

[14] Giuseppe Buttazzo. Problems with free interfaces and free discontinuities: Variational problems. *Encyclopedia of Mathematical Physics (ISBN 978-0-1251-2666-3), Elsevier, Amsterdam (2006).*, 2:411–416, 2006. URL http://cvgmt.sns.it/paper/929/. cvgmt preprint.

[15] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext. Springer New York, 2010. ISBN 9780387709130. URL https://books.google.de/books?id=GAA2XqOIIGoC.

[16] Kristian Bredies and Marcello Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *Calculus of Variations and Partial Differential Equations*, 59(1), December 2019. ISSN 1432-0835. doi: 10.1007/s00526-019-1658-1. URL http://dx.doi.org/10.1007/s00526-019-1658-1.

[17] Giacomo Cristinelli, José A. Iglesias, and Daniel Walter. Conditional gradients for total variation regularization with pde constraints: a graph cuts approach, 2023. URL https://arxiv.org/abs/2310.19777.

[18] Paolo Bonicatto and Nikolay A. Gusev. On the structure of divergence-free measures on $\mathbb{R}^2$, 2020. URL https://arxiv.org/abs/1912.10936.

[19] Titu Andreescu, Cristinel Mortici, and Marian Tetiva. *The Intermediate Value Theorem*, pages 189–200. Springer New York, New York, NY, 2017. ISBN 978-0-8176-4629-5. doi: 10.1007/978-0-8176-4629-5_11. URL https://doi.org/10.1007/978-0-8176-4629-5_11.

[20] L. Narici and E. Beckenstein. *Topological Vector Spaces.* Chapman & Hall/CRC Pure and Applied Mathematics. CRC Press, 2010. ISBN 9781584888673. URL https://books.google.de/books?id=BYLOBQAAQBAJ.

[21] J. Borwein and Q. Zhu. *Techniques of Variational Analysis.* CMS Books in Mathematics. Springer New York, 2006. ISBN 9780387282718. URL https://books.google.de/books?id=Fquue78kJeEC.

[22] José A Iglesias, Gwenael Mercier, and Otmar Scherzer. A note on convergence of solutions of total variation regularized linear inverse problems. *Inverse Problems*, 34(5):055011, April 2018. ISSN 1361-6420. doi: 10.1088/1361-6420/aab92a. URL http://dx.doi.org/10.1088/1361-6420/aab92a.

[23] Yohann de Castro, Vincent Duval, and Romain Petit. Towards off-the-grid algorithms for total variation regularized inverse problems, 2022. URL https://arxiv.org/abs/2104.06706.

[24] E.H. Lieb and M. Loss. *Analysis.* Crm Proceedings & Lecture Notes. American Mathematical Society, 2001. ISBN 9780821827833. URL https://books.google.de/books?id=Eb_7oRorXJgC.

[25] Yurii Nesterov. *Lectures on Convex Optimization.* Springer Publishing Company, Incorporated, 2nd edition, 2018. ISBN 3319915770.

[26] Otto Forster. *Analysis 1*. Springer Spektrum, Wiesbaden, 2016. URL https://doi.org/10.1007/978-3-658-40130-6.

Decleration of Academic Integrity

I hereby confirm that this thesis, entitled *The Frank-Wolfe algorithm for inverse problems with anisotropic TV regularization* is solely my own work and that I have used no sources or aids other than the ones stated. All passages in my thesis for which other sources, including electronic media, have been used, be it direct quotes or content references, have been acknowledged as such and the sources cited. I am aware that plagiarism is considered an act of deception which can result in sanction in accordance with the examination regulations.

(date, signature of student)

I consent to having my thesis cross-checked with other texts to identify possible similarities and to having it stored in a database for this purpose.

I confirm that I have not submitted to following thesis in part or whole as an eximination paper before.

(date, signature of student)