A fast primal-dual-active-jump method for minimization in $\text{BV}((0,T)\mathbb{R}^d)$

Philip Trautmann$^a$ and Daniel Walter$^b$

$^a$Institute for Mathematics and Scientific Computing, Karl-Franzens-Universität, Graz, Austria; $^b$Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria

ABSTRACT

We analyse a solution method for minimization problems over a space of $\mathbb{R}^d$-valued functions of bounded variation on an interval $I$. The presented method relies on piecewise constant iterates. In each iteration, the algorithm alternates between proposing a new point at which the iterate is allowed to be discontinuous and optimizing the magnitude of its jumps as well as the offset. A sublinear $O(1/k)$ convergence rate for the objective function values is obtained in general settings. Under additional structural assumptions on the dual variable, this can be improved to a locally linear rate of convergence $O(\zeta^k)$ for some $\zeta < 1$. Moreover, in this case, the same rate can be expected for the iterates in $L^1(I;\mathbb{R}^d)$.

1. Introduction

We consider minimization problems of the form

$$\inf_{u \in \text{BV}(I;\mathbb{R}^d)} j(u) = [F(Ku) + \beta \|u'\|_M]. \quad (P)$$

where the minimizer is sought for in the space of $\mathbb{R}^d$-valued functions of bounded variation on an interval $I = (0, T)$ where $T > 0$. Here $K$ denotes a linear and continuous operator mapping to a Hilbert space of observations $Y$ and $F$ is assumed to be a convex smooth loss function. Given $\beta > 0$, the second term in the objective functional penalizes the total variation norm $\|\cdot\|_M$ of the distributional derivative $u'$ which is an element of the space of $\mathbb{R}^d$-valued regular vector measures $M(I;\mathbb{R}^d)$. It is well known that such a penalization favours piecewise constant minimizers $\bar{u}$ or, equivalently, functions whose distributional derivative is given by a finite number $\bar{N}$ of vector-valued Dirac Delta functionals

$$\bar{u}' = \sum_{i=1}^{\bar{N}} \bar{\mu}_i \bar{v}_i \delta_{\bar{t}_i}. \quad (1)$$
Here, the \( \tilde{t}_i \) denote the discontinuities of \( \tilde{u} \), i.e. there holds \( \tilde{u} \equiv \tilde{u}(\tilde{t}_i) \in \mathbb{R}^d \) on \([\tilde{t}_i, \tilde{t}_{i+1}]\), and \( \bar{v} \in \mathbb{R}^d, |\bar{v}_i|_{\mathbb{R}^d} = 1 \), and \( \bar{\mu}_i > 0 \) describe the direction and the length of the jump of \( \tilde{u} \) occurring at \( \tilde{t}_i \). This structural property of \((P)\) makes it appealing for a variety of practical applications. For example, we point out PDE-constrained optimal control problems, \([1–3]\) and the denoising of scalar signals, \([4,5]\). For the precise functional analytic setting, we refer to Section 3.

1.1. Contribution

The aim of this paper is to analyse a simple yet efficient iterative solution algorithm for problem \((P)\), the ‘Primal-Dual-Active-Jump’ method. It relies on the observation that a piecewise constant function \( \tilde{u} \) with derivative (1) is a minimizer of \((P)\) if the associated dual variable \( \tilde{p}(\cdot) = \int_0^s K^* \nabla F(Ku)(s) \, ds \) satisfies

\[
\sup_{t \in I} |\tilde{p}(t)|_{\mathbb{R}^d} \leq \beta, \quad |\tilde{p}(\tilde{t}_i)|_{\mathbb{R}^n} = \beta, \quad \bar{v}_i = \tilde{p}(\tilde{t}_i)/\beta
\]

i.e. the discontinuities of \( \tilde{u} \) align with maximizers of \( |\tilde{p}(\cdot)|_{\mathbb{R}^d} \) and the direction of each jump is given by the normalized dual variable evaluated at \( \tilde{t}_i \), see Corollary 3.6. Now, the proposed method generates sequences of piecewise constant iterates \( u_k \) with

\[
u'_k = \sum_{i=1}^{N_k} \mu_i^k v_i^k \delta_{t_i}^k
\]

as well as associated active sets \( \mathcal{A}_k = \{(\mu_i^k, v_i^k)\}_{i=1}^{N_k} \) which store the jumps \( v_i^k := v_i \delta_{t_i}^k \) of \( u_k \) and the associated magnitudes \( \mu_i^k > 0 \). Each iteration then proceeds in three phases: First, we check whether \( u_k \) and the associated dual variable \( p_k(\cdot) = \int_0^s K^* \nabla F(Ku_k)(s) \, ds \) satisfy (2). If this is not the case we ‘allow’ for an additional jump \( \hat{\nu}_k = v_k \delta_{\hat{t}_k} \) in the iterate \( u_k \). Motivated by (2), the position and direction are chosen such that

\[
|p_k(\hat{t}_k)|_{\mathbb{R}^d} = \sup_{t \in I} |p_k(t)|_{\mathbb{R}^d}, \quad \bar{v}_k = p_k(\hat{t}_k)/|p_k(\hat{t}_k)|_{\mathbb{R}^d},
\]

assuming that \( p_k \neq 0 \). Subsequently, we determine the next iterate \( u_{k+1} \) by solving the finite-dimensional convex minimization problem

\[
\min_{\mu_i \geq 0, a_u \in \mathbb{R}^d} \left[ F(Ku) + \beta \sum_{i=1}^{N_k+1} \mu_i \right]
\]

s.t. \( u' = \mu_{N_k+1} \hat{\nu}^k + \sum_{i=1}^{N_k} \mu_i v_i^k \), \( a_u = \frac{1}{T} \int_0^T u(t) \, dt \).
Loosely speaking, this can be connected to a discretization of \((P)\) in which we replace the whole space \(BV(I;\mathbb{R}^d)\) by the set of all \(u \in BV(I;\mathbb{R}^d)\) with

\[ u' = \mu_{N_k+1} + \sum_{i=1}^{N_k} \mu_i v_i^k \quad \text{for some} \quad \mu \in \mathbb{R}^{N_k+1}, \quad \mu \geq 0, \]

i.e. we fix the jumps and only try to find optimal magnitudes. Note that the jump magnitudes characterize such a function up to a constant shift \(a_u \in \mathbb{R}^d\) which is optimized alongside \(\mu_i \geq 0\). Finally, the active set is updated by removing all jumps whose associated magnitude was set to zero. The theoretical contribution of the present manuscript is twofold. First, we prove that, without any structural assumption on the solutions of \((P)\), the generated sequence \(u_k\) indeed converges, on subsequences, to minimizers of \((P)\) and the functional values \(j(u_k)\) converge sublinearly to the minimum value. Second, under appropriate structural assumptions on the optimal dual variable, similar to \([2,3]\), we deduce the local linear convergence of \(j(u_k)\) and of the iterates \(u_k\) with respect to the strict topology on \(BV(I;\mathbb{R}^d)\). More specifically, we assume the existence of a unique minimizer \(\bar{u}\) of \((P)\) with derivative \((1)\) and whose finitely many discontinuities coincide with the maxima of \(|\tilde{p}(\cdot)|_{\mathbb{R}^d}\). Moreover, \(|\tilde{p}(\cdot)|_{\mathbb{R}^d}\) is assumed to grow quadratically in the vicinity of \(\bar{t}_{i}\).

1.2. Related work

The efficient algorithmic solution of \((P)\) is a delicate issue for a variety of reasons. On the one hand, this is attributed to the appearance of the BV seminorm which makes the objective functional nonsmooth. Moreover, \(j\) lacks coercivity with respect to \(u\) which is often a vital tool in the derivation of fast convergence result for minimization schemes. On the other hand, we point out that \(BV(I;\mathbb{R}^d)\) is nonreflexive. Many well-studied algorithms for nonsmooth optimization rely on the reflexive structure of the underlying space and thus do not yield direct extensions to the problem at hand.

A first straightforward approach to circumventing the aforementioned difficulties consists of discretizing the space \(BV(I;\mathbb{R}^d)\) in \((P)\). More in detail, instead of minimizing over all \(u \in BV(I;\mathbb{R}^d)\), one could only consider piecewise constant \(u_h\) that solely jump in the nodes \(0 = t_0 < t_1 < \cdots < t_{N_h} = T\) of a partition of \(I\). This reduces \((P)\) to a finite-dimensional convex minimization problem with a nonsmooth group sparsity regularization term, \([6]\). The solution of discretized 1D BV problems has been addressed e.g. in \([7–10]\). Nonetheless, such reasoning often leads to algorithms that exhibit mesh-dependency meaning that their convergence behaviour critically depends on the partition of \(I\) and might degenerate as \(N_h \to \infty\). To mitigate these effects, a second line of works, see e.g. \([11,12]\), proposes the regularization of \((P)\) by adding \((\varepsilon/2)\|u'\|_{L^2_I}^2\) for \(0 < \varepsilon << 1\) and minimizing for \(u\) in the Sobolev space \(H^1(I;\mathbb{R}^d)\). Since the total
variation norm of $u'$ remains present in objective functional, the derivative of minimizers can still be expected to exhibit sparsity, i.e. its support is small. However, due to the Sobolev seminorm penalty, minimizers cannot be piecewise constant if $\varepsilon > 0$. For this reason, algorithmic approaches based on regularization are usually accompanied by a path-following strategy for $\varepsilon \to 0$ which requires additional analysis. If $(P)$ is restricted to mean-value free BV functions, it can be equivalently reformulated as a minimization problem over the space of $\mathbb{R}^d$-valued vector measures. Over the past years, there has been an increasing body of work on the efficient solution of such problems using exchange type algorithms, [13–17], which rely on iterates comprised of finitely many Dirac Delta functionals. These alternate between proposing a new Dirac Delta (i.e. a "jump" in our terminology) and approximately solving finite-dimensional convex and/or nonconvex subproblems to achieve sufficient descent. Most recently, linear convergence of such methods relying on convex subproblems has been addressed in [16], for $d = 1$, and [17], for the general vector-valued case. Our approach is closest related to the earlier work [17] but differs in the treatment of the convex subproblems. More in detail while the present method relies solely on optimizing the magnitudes $\mu_i$ in each iteration ($\#A_k$ DOF), the linear convergence result of [17] also requires the optimization of the jump directions ($d\#A_k$ DOF). Hence we obtain the same theoretical convergence guarantees while solving smaller subproblems. Let us also mention the finite step convergence results of [15,16]. However, these require 'point-moving', i.e. an additional step in which the jump positions are optimized. This constitutes a nonconvex problem and is therefore not considered in the present work. The idea of using exchange type methods for 1D BV penalties has previously been proposed in [18] together with a sublinear convergence result. Moreover, we also point out [19,20], where the authors propose a predu al approach for the solution of BV-propblems in higher dimensions. The considered ansatz circumvents the complicated structure of the BV-space by solving a suitable Hilbert space predu al problem with infinitely many bilateral constraints. For these, the derivative of the BV-minimizer plays the role of a Lagrangian multiplier. In contrast to the present manuscript; however, the algorithmic approaches in these earlier works require the strong convexity of $F \circ K$ which rules out, e.g. applications with finite-dimensional observations. In Section 4.2, we briefly shine light on the predu al approach in the one dimensional, vector-valued setting considered in the present work. In particular, we interpret our algorithm as a primal-(pre)dual method which iteratively solves the predu al problem by replacing its infinitely many constraints by a finite number of successively updated ones. This result also motivates the name ‘Primal-Dual-Active-Jumps’ for the presented method. Finally, we point out the denoising problem for a scalar signal $y_d \in L^2(I)$. In our setting, this corresponds to the case of $d = 1$, $F(\cdot) = (1/2)\|\cdot - y_d\|_2^2$ and $K = \text{Id}$. For this particular instance of $(P)$ the unique minimizer can be determined directly using a taut-string-method, see e.g. [21,22]. To the best of our knowledge, this method does, however, not yield
extensions to the case of a general observation operator $K$ and the vector-valued case $d > 1$.

### 1.3. Outline of the paper

The relevant notation used throughout the paper is introduced in Section 2. In Section 3, we equivalently reformulate $(P)$ as a minimization problem over the distributional derivative and the mean value of $u$. Subsequently, this equivalence is used to derive first-order necessary and sufficient optimality conditions. A detailed description of the proposed solution algorithm for $(P)$ can be found in Section 4. The convergence of the method is addressed in Section 5. Finally, Section 6 finishes the paper with numerical experiments illustrating our theoretical findings.

### 2. Notation & definitions

In the following set $I = (0, T)$ for some $T > 0$ and fix $d \in \mathbb{N}$. Denote by $(\cdot, \cdot)_{\mathbb{R}^d}$ the euclidean inner product on $\mathbb{R}^d$ and let $|\cdot|_{\mathbb{R}^d}$ denote the corresponding norm. By $C_0(I; \mathbb{R}^d)$ together with the usual supremum norm

$$
\|\varphi\|_C = \sup_{t \in I} |\varphi(t)|_{\mathbb{R}^d} \quad \forall \varphi \in C_0(I; \mathbb{R}^d)
$$

we refer to the Banach space of $\mathbb{R}^d$-valued continuous functions on $I$ that vanish at its boundary. Its topological dual space is identified with the space of regular vector measures $\mathcal{M}(I; \mathbb{R}^d)$. The corresponding duality pairing is denoted by $\langle \cdot, \cdot \rangle$. For example, if $q$ is a discrete measure, i.e.

$$
q = \sum_{i=1}^{N} q_i \delta_{t_i}
$$

where $q_i \in \mathbb{R}^d$ and $\delta_{t_i}$ denotes the Dirac Delta functional supported on $t_i \in I$, then

$$
\langle \varphi, q \rangle = \sum_{i=1}^{N} (\varphi(t_i), q_i)_{\mathbb{R}^d}.
$$

The space $\mathcal{M}(I; \mathbb{R}^d)$ is equipped with the canonical dual norm

$$
\|q\|_{\mathcal{M}} = \sup_{\|\varphi\|_C=1} \langle \varphi, q \rangle.
$$

We call $u \in L^1(I; \mathbb{R}^d)$ a function of bounded variation if its distributional derivative $u'$ is representable by an element of $\mathcal{M}(I; \mathbb{R}^d)$, i.e.

$$
(u, \varphi')_{L^2} = -\langle \varphi, u' \rangle \quad \forall \varphi \in C^1_c(I; \mathbb{R}^d),
$$

where $C^1_c(I; \mathbb{R}^d)$ denotes the set of $\mathbb{R}^d$-valued, continuously differentiable functions with compact support on $I$. The set of $\mathbb{R}^d$-valued functions of bounded
variation on $I$ is now defined as

$$\text{BV}(I; \mathbb{R}^d) = \left\{ u \in L^1(I; \mathbb{R}^d) \mid \|u'\|_M < \infty \right\}.$$  

Equipping $\text{BV}(I; \mathbb{R}^d)$ with the norm

$$\|u\|_{\text{BV}} = \|u'\|_M + \|u\|_{L^1} \quad \forall u \in \text{BV}(I; \mathbb{R}^d),$$

where

$$\|u\|_{L^1} = \int_I |u(s)|_{\mathbb{R}^d} \, ds,$$

makes it a Banach space which continuously embeds into $L^p(I)$, $p \in [1, \infty]$, the embedding being compact for $p < \infty$. Compact embeddings are denoted by ‘$\hookrightarrow$’. Given a function $u \in L^1(I; \mathbb{R}^d)$, the vector of the mean values of its components is defined as

$$a_u = \frac{1}{T} \int_I u(t) \, dt,$$

where integration has to be understood in the sense of Bochner. From e.g. [23, Theorem 3.44], we conclude the existence of constants $C_1, C_2 > 0$ with

$$C_1(|a_u|_{\mathbb{R}^d} + \|u'\|_M) \leq \|u\|_{\text{BV}} \leq C_2(|a_u|_{\mathbb{R}^d} + \|u'\|_M) \quad \forall u \in \text{BV}(I; \mathbb{R}^d).$$

Following, e.g. [23, Remark 3.12] $\text{BV}(I; \mathbb{R}^d)$ can be identified as the topological dual space of a separable Banach space. A sequence $\{u_k\}_{k \in \mathbb{N}} \subset \text{BV}(I; \mathbb{R}^d)$ is called weak* convergent in $\text{BV}(I; \mathbb{R}^d)$ with limit $\bar{u}$ if

$$\|u_k - \bar{u}\|_{L^1} \to 0, \quad u_k' \rightharpoonup^* \bar{u},$$

where $\rightharpoonup^*$ denotes weak* convergence in $\mathcal{M}(I; \mathbb{R}^d)$. Due to the sequential Banach-Alaoglu theorem, every bounded sequence in $\text{BV}(I; \mathbb{R}^d)$ admits a weak* convergent subsequence. Furthermore a weak* convergent sequence $\{u_k\}_{k \in \mathbb{N}} \subset \text{BV}(I; \mathbb{R}^d)$ is called convergent with respect to the strict topology on $\text{BV}(I; \mathbb{R}^d)$, or shortly strictly convergent, if additionally $\|u_k'\|_M \to \|\bar{u}\|_M$ holds. This is indicated by ‘$\rightharpoonup^s$’. The strict topology on $\text{BV}(I; \mathbb{R}^d)$ is induced by the metric

$$d(u_1, u_2) = \|u_1 - u_2\|_{L^1} + \|u_1\|_M - \|u_2\|_M.$$

Last, given an open interval $(t, 1)$ for some $t \in [0, 1)$ its characteristic function is defined by

$$\chi_t = \begin{cases} 0 & \text{on } I \setminus (t, 1) \\ 1 & \text{else} \end{cases}.$$  

There holds $\chi_t \in \text{BV}(I)$ with $\chi'_t = \delta_t$, $t > 0$, and $\chi'_0 = 0$, respectively.
3. Optimization problem

The following assumptions concerning \((P)\) are made throughout this paper.

**Assumption 3.1:** In the following let \(Y\) be a Hilbert space with inner product \((\cdot,\cdot)_Y\) and induced norm \(\|\cdot\|_Y\). There holds:

- The operator \(K: L^2(I; \mathbb{R}^d) \rightarrow Y\) is linear and continuous. Moreover, denoting by \(\{e_i\}_{i=1}^d\) the canonical basis of \(\mathbb{R}^d\), the set \(\{K(e_i\chi_I)\}_{i=1}^d\) is linearly independent.
- The mapping \(F: Y \rightarrow [0, +\infty)\) is strictly convex and radially unbounded, i.e. there holds
  \[\|y_k\|_Y \rightarrow +\infty \Rightarrow F(y_k) \rightarrow +\infty.\]
- \(F\) is continuously Fréchet differentiable. The Fréchet derivative \(F'(y) \in \mathcal{L}(Y, \mathbb{R})\) of \(F\) at \(y \in Y\) is identified with its Riesz representative \(\nabla F(y) \in Y\), i.e.
  \[F'(y)\delta y = (\nabla F(y), \delta y)_Y \quad \forall \delta y \in Y.\]

The assumptions on \(F\) are, e.g. fulfilled by quadratic misfits of the form \(F(y) = (1/2)\|y - y_d\|_Y^2\) where \(y_d \in Y\) is a given reference. Moreover, while the linear independence assumption on \(\{K(e_i\chi_I)\}_{i=1}^d\) might seem nonstandard at a first glance, it is indeed a necessary condition for the boundedness of the sublevel sets of the objective functional \(j\). In fact, if this is not the case, i.e. if \(\{K(e_i\chi_I)\}_{i=1}^d\) is linearly dependent, then every potential minimizer to \((P)\) is only defined up to a shift of the form \(\tilde{c}\kappa_I\) where \(\tilde{c} \neq 0\) satisfies \(K(\tilde{c}\kappa_I) = 0\). This further complicates the analysis in the following sections and is therefore omitted at the moment. Note that this particular assumption can be verified a priori, i.e. before \((P)\) is actually solved.

3.1. Optimality conditions

The derivation of most subsequent results relies on an equivalent reformulation of \((P)\) which will be introduced next. Define the linear and continuous operator

\[
B: \mathcal{M}(I; \mathbb{R}^d) \times \mathbb{R}^d \rightarrow L^2(I; \mathbb{R}^d), \quad (q, c) \mapsto \int_0^T dq - \frac{1}{T} \int_0^T \int_0^s dq \, ds + c,
\]

where integration has to be understood componentwise. We arrive at the following identification.

**Proposition 3.1:** For all \((q, c) \in \mathcal{M}(I; \mathbb{R}^d) \times \mathbb{R}^d\) we have \(B(q, c) \in BV(I; \mathbb{R}^d)\). The linear and continuous operator \(B: \mathcal{M}(I; \mathbb{R}^d) \times \mathbb{R}^d \rightarrow BV(I; \mathbb{R}^d)\) from (3) is an isomorphism.
**Proof:** The bounded invertibility of $B$ is imminent noting that its inverse is given by the operator

$$B^{-1}: \text{BV}(I; \mathbb{R}^d) \to \mathcal{M}(I; \mathbb{R}^d) \times \mathbb{R}^d, \quad u \mapsto (u', a_u).$$

Loosely speaking, the previous result states that any function of bounded variation on $I$ is uniquely characterized by its distributional derivative and mean values of its components. Thus $(P)$ is equivalent to the sparse minimization problem

$$\inf_{q \in \mathcal{M}(I; \mathbb{R}^d), c \in \mathbb{R}^d} [F(K(q, c)) + \beta \|q\|_\mathcal{M}]$$

where we abbreviate $K = K \circ B$. This equivalence is now used to show existence of minimizers to $(P)$.

**Proposition 3.2:** Let Assumption 3.1 hold. Then there exists at least one solution $\bar{u} \in \text{BV}(I; \mathbb{R}^d)$ to $(P)$.

**Proof:** First, we point out that there is $\kappa > 0$ with

$$|c|_{\mathbb{R}^d} \leq \kappa \|K(c \chi_I)\|_Y \quad \forall \ c \in \mathbb{R}^d$$

since the operator

$$K|_{\mathbb{R}^d}: \mathbb{R}^d \to \text{span}\{K(e_i \chi_I)\}_{i=1}^d,$$

$$K|_{\mathbb{R}^d}(c) = K(c \chi_I) = \sum_{i=1}^d c_i K(e_i \chi_I) \quad \forall \ c \in \mathbb{R}^d,$$

is boundedly invertible due to the linear independence of the set $\{K(e_i \chi_I)\}_{i=1}^d$. According to Proposition 3.1, a function $\bar{u} \in \text{BV}(I; \mathbb{R}^d)(I; \mathbb{R}^d)$ is a solution to $(P)$ if and only if $(\bar{u}', a_{\bar{u}})$ solves (4). Note that $0 \leq \inf(4)$. Let $\{(q_k, c_k)\}_k$ denote an infimizing sequence for (4). Since we have

$$\beta \|q_k\|_\mathcal{M} \leq F(K(q_k, c_k)) + \beta \|q_k\|_\mathcal{M} \to \inf(4)$$

and $\beta > 0$, the sequence $\{q_k\}_k$ is bounded in $\mathcal{M}(I; \mathbb{R}^d)$ and thus admits a weak* convergent subsequence, denoted by the same index, with limit $\bar{q} \in \mathcal{M}(I; \mathbb{R}^d)$. Recalling the definition of the operators $B$ and $K$ as well as the compact embedding of $\text{BV}(I; \mathbb{R}^d)$ into $L^2(I; \mathbb{R}^d)$, this implies $K(q_k, 0) \to K(\bar{q}, 0)$ in $Y$. Now assume that the corresponding sequence of constants $\{c_k\}_k$ satisfies $|c_k|_{\mathbb{R}^d} \to +\infty$. Due to (5), we then also have $\|K(c_k \chi_I)\|_Y \to +\infty$ and thus

$$+\infty = \liminf_{k \to \infty} \left(\|K(q_k, 0)\|^2_Y + \|K(c_k \chi_I)\|_Y(\|K(c_k \chi_I)\|_Y - 2\|K(q_k, 0)\|_Y)\right)$$

$$\leq \liminf_{k \to \infty} \|K(q_k, c_k)\|^2_Y.$$
Since, by assumption, \( F \) is radially unbounded, we then get \( F(K(q_k, c_k)) \to +\infty \). Hence \( \{c_k\}_k \) is bounded and, again by extracting a subsequence denoted by the same index, we have \( q_k \rightharpoonup^* \bar{q} \) in \( \mathcal{M}(I; \mathbb{R}^d) \) and \( c_k \to \bar{c} \) for some \( \bar{c} \in \mathbb{R}^d \). Finally, since the total variation norm is weak* lower semicontinuous, we arrive at
\[
\inf(4) \leq F(K(\bar{q}, \bar{c})) + \beta \|\bar{q}\|_M \leq \liminf_{k \to \infty} [F(K(q_k, c_k)) + \beta \|q_k\|_M] = \inf(4),
\]
i.e. \((\bar{q}, \bar{c})\) is a minimizer of \((4)\) and \( \tilde{u} := B(\bar{q}, \bar{c}) \) is a solution of \((P)\).

Next, we derive first-order necessary and sufficient optimality conditions for problem \((P)\). For this purpose, we characterize the ‘pre-adjoint’ operator \( B_\ast \). Consider the system of auxiliary ordinary differential equations
\[
-\omega'' = \phi \quad \text{in} \ (0, T), \quad \omega'(0) = \omega'(T) = 0, \quad \int_0^T \omega(t) \, dt = 0, \quad (6)
\]
where \( \phi \in L^2(I; \mathbb{R}^d) \) with \( a\phi = 0 \). Clearly, this problem admits a unique solution \( \omega \in H^2(I; \mathbb{R}^d) \hookrightarrow C^1(I; \mathbb{R}^d) \) and \( \omega' \in C_0(I; \mathbb{R}^d) \).

**Lemma 3.3:** The linear and continuous operator \( B \) from \((3)\) is the Banach space adjoint of
\[
B_\ast : L^2(I; \mathbb{R}^d) \to C_0(I; \mathbb{R}^d) \times \mathbb{R}^d, \quad \varphi \mapsto \left( \omega', \int_0^T \varphi(s) \, ds \right), \quad (7)
\]
where \( \omega \in C^1(I; \mathbb{R}^d) \) fulfils (6) for \( \phi = \varphi - a\varphi \).

**Proof:** Obviously, the operator \( B_\ast \) is linear and continuous. Let \( \varphi \in L^2(I; \mathbb{R}^d) \) and a pair \((q, c) \in \mathcal{M}(I; \mathbb{R}^d) \times \mathbb{R}^d\) be given. We obtain
\[
\langle \omega', q \rangle + \left( c, \int_0^T \varphi(t) \, dt \right) = \int_0^T \left( \varphi(s), \int_0^s dq \right) \, ds - \left( a\varphi, \int_0^T \int_0^s dq \, ds \right) + \left( c, \int_0^T \varphi(t) \, dt \right) = \left( \varphi, B(q, 0) \right)_{L^2} + \left( c, \int_0^T \varphi(t) \, dt \right) = \left( \varphi, B(q, c) \right)_{L^2}.
\]
Here we used \( B(q, 0) \in BV(I; \mathbb{R}^d) \) with \( B(q, 0)' = q \) as well as the integration by parts formula in the second equality. This establishes the result.

Combining the equivalence of \((P)\) and \((4)\) as well as the characterization of \( B_\ast \), we arrive at the following necessary and sufficient first-order optimality conditions for \((P)\).
**Theorem 3.4:** Let \( \bar{u} \in \text{BV}(I; \mathbb{R}^d) \) be given. Further, define

\[
\tilde{p}(t) = \int_0^t K^* \nabla F(K\bar{u})(s) \, ds \in C(\bar{I}; \mathbb{R}^d).
\]

Then \( \bar{u} \) is an optimal solution to \((P)\) if and only if

\[
\|\tilde{p}\|_C \in \begin{cases} 
\{\beta\} & \bar{u}' \neq 0, \\
[0, \beta] & \text{else},
\end{cases}, \quad \tilde{p}(T) = 0
\]

as well as

\[
\langle \tilde{p}, \bar{u}' \rangle = \beta \|\bar{u}'\|_M.
\]

**Proof:** A function \( \bar{u} \in \text{BV}(I; \mathbb{R}^d) \) is an optimal solution to \((P)\) if and only if the pair

\[
(\bar{q}, \bar{c}) = \left( \bar{u}', \frac{1}{T} \int_0^T \bar{u}(s) \, ds \right)
\]

is a minimizer of \((4)\). Since \( j \) is convex and \( F \) is Fréchet differentiable, optimality of \((\bar{q}, \bar{c})\) is equivalent to

\[
(-\nabla F(K(\bar{q}, \bar{c})), K(q - \bar{q}, 0))_{L^2(I; \mathbb{R}^d)} + \beta \|\bar{q}\|_M \leq \beta \|q\|_M, \quad \forall q \in \mathcal{M}(I; \mathbb{R}^d)
\]

as well as

\[
(-\nabla F(K(\bar{q}, \bar{c})), K(0, \delta \bar{c}))_{L^2(I; \mathbb{R}^d)} = 0 \quad \forall \delta \bar{c} \in \mathbb{R}.
\]

Let \( \bar{\omega} \in C^1(\bar{I}; \mathbb{R}^d), \bar{I} := [0, T] \), denote the solution of \((6)\) for \( \varphi = -K^* \nabla F(K\bar{u}) \in L^2(I) \). Note that \( K^* = B_\ast K^* \). Utilizing the characterization of \( B_\ast \), see Lemma 3.3, as well as the definition of the convex subdifferential, conditions \((10)\) and \((11)\) can be rewritten as

\[
\bar{\omega}' \in \beta \partial \|\bar{q}\|_M, \quad \tilde{p}(T) = \int_0^T -K^* \nabla F(K\bar{u})(t) \, dt = 0,
\]

respectively. It is well known that the subdifferential inclusion is equivalent to

\[
\|\bar{\omega}'\|_C \in \begin{cases} 
\{\beta\} & \bar{q} \neq 0, \\
[0, \beta] & \text{else},
\end{cases}, \quad \langle \bar{\omega}', \bar{q} \rangle = \beta \|\bar{q}\|_M.
\]

Due to the fundamental theorem of analysis, there exists a vector \( c \in \mathbb{R}^d \) with

\[
\bar{\omega}'(t) = \int_0^t K^* \nabla F(K\bar{u})(s) \, ds - t \int_0^T K^* \nabla F(K\bar{u})(s) \, ds + c
\]

\[
= \int_0^t K^* \nabla F(K\bar{u})(s) \, ds + c
\]
for all \( t \in \bar{I} \). From \( \tilde{\omega}'(0) = 0 \) we deduce \( c = 0 \). Thus, we conclude \( \tilde{p} = \tilde{\omega}' \) on \( \bar{I} \).

Combining all the previous observations now finishes the proof. \[\square\]

Throughout this paper, and with a slight abuse of notation, we will refer to the function \( \tilde{p} \in H^1_0(I; \mathbb{R}^d) \cap C_0(I; \mathbb{R}^d) \) as the \textit{optimal dual variable} or the \textit{optimal dual state}. Indeed, it corresponds to a solution of a suitable predual problem

\[
\inf_{p \in H^1_0(I; \mathbb{R}^d)} \inf_{y \in Y} F^*(y) \quad \text{s.t.} \quad \|p'\|_C \leq \beta
\]

where \( F^* \) denotes the convex conjugate of \( F \) and for which \( \bar{u}' \) plays the role of a Lagrangian multiplier. The proof follows standard arguments and is thus skipped for the sake of brevity.

**Proposition 3.5:** Let \( \bar{u} \) denote a solution of (\( \mathcal{P} \)) and set \( \bar{p}(\cdot) = \int_0^\cdot K^*\nabla F(\bar{u}) \). Then \( \bar{p} \) is a minimizer of (\( \mathcal{D} \)). Moreover, (\( \mathcal{P} \)) is the Fenchel dual problem to (\( \mathcal{D} \)) and we have strong duality, i.e.

\[
\min \mathcal{D} = -\min(\mathcal{P}).
\]

Following e.g. [24, Proposition 6.23], conditions (8) and (9) imply the following \textit{support condition} on the derivative of a minimizer of (\( \mathcal{P} \)):

\[
\text{supp } \bar{u}' \subset \{ t \in I \mid |\bar{p}(t)|_{\mathbb{R}^d} = \beta \}.
\] (13)

Loosely speaking, this implies that \( \bar{u} \) is constant outside of the level set defined above. Using this observation, we can draw conclusions on the structure of \( \bar{u} \) by looking at the dual variable \( \bar{p}. \) For example, if the level set in (13) is nonempty but finite, then \( \bar{u} \) is piecewise constant and its ‘jumps’, i.e. its points of discontinuity, align themselves with global extrema of \( \bar{p} \). As a consequence, \( \bar{p} \) can be regarded as a \textit{jump detector} for the minimizer \( \bar{u} \). This can be seen as a one-dimensional analogue to the multidimensional case, [19,20], where the (pre-)dual variable serves as an \textit{edge detector} to identify the jumping parts of minimizers. We formalize this property in the following corollary.

**Corollary 3.6:** Let \( \bar{u} \in BV(I; \mathbb{R}^d) \) be a minimizer of (\( \mathcal{P} \)) and let \( \bar{p} \) be defined as in Theorem 3.4. Assume that

\[
\{\bar{t}_i\}_{i=1}^N = \{ t \in I \mid |\bar{p}(t)|_{\mathbb{R}^d} = \beta \}
\] (14)

for some \( N \in \mathbb{N} \) and \( \{\bar{t}_i\}_{i=1}^N \subset I \). Then \( \bar{u}' \in M(I; \mathbb{R}^d) \) is of the form

\[
\bar{u}' = \sum_{i=1}^N \bar{\mu}_i \bar{v}_i \quad \text{where} \quad \bar{v}_i = (\bar{p}(\bar{t}_i)/\beta)\delta_{\bar{t}_i},
\]

i.e. \( \bar{u} \) is piecewise constant on \( I \).
Proof: This can be proven analogously to [24, Corollary 6.25].

Finally, we point out that the optimal observation \( \tilde{y} \in Y \) in (P) and thus also the dual variable \( \tilde{p} \in C_0(I; \mathbb{R}^d) \), see Theorem 3.4, are unique.

Corollary 3.7: Let \( \bar{u}_1, \bar{u}_2 \in BV(I; \mathbb{R}^d) \) denote two minimizers of (P). Moreover, denote by \( \bar{y}_1 = K\bar{u}_1, \bar{y}_2 = K\bar{u}_2 \) and \( \bar{p}_1, \bar{p}_2 \in C_0(I; \mathbb{R}^d) \) the associated observations and dual variables, see Theorem 3.4, respectively. Then \( \bar{y}_1 = \bar{y}_2 \) and \( \bar{p}_1 = \bar{p}_2 \).

Proof: The uniqueness of the optimal observation, and thus also that of the dual variable, directly follows from the strict convexity of \( F \).

4. Algorithmic solution

This section is devoted to the description of an efficient solution algorithm for (P). The method we propose relies on the iterative update of a finite active set \( A_k = \{ \mu^k_i, v^k_i \}_{i=1}^{N_k} \) comprised of ‘jumps’ \( v^k_i \in \mathcal{M}(I; \mathbb{R}^d) \) and the associated ‘magnitudes’ \( \mu^k_i > 0 \). Each jump is of the form \( v^k_i = v^k_i \delta^k_i \) for a position \( t^k_i \in I \) and a direction \( v^k_i \in \mathbb{R}^d, |v^k_i|_{\mathbb{R}^d} = 1 \).

Given an offset \( c^k \in \mathbb{R}^d \) the kth iterate is defined as

\[
 u_k = B \left( \sum_{i=1}^{N_k} \mu^k_i v^k_i, c^k \right). \tag{15}
\]

If \( A_k = \emptyset \), i.e. \( u^k = c^k \chi_0 \), we adopt the convention \( N_k = 0 \).

4.1. The algorithm

We shortly describe the individual steps of the algorithm in the following. Given the current active set \( A_k \) and iterate \( u_k \) we first compute the current dual variable \( p_k(\cdot) = \int_0^\cdot K^* \nabla F(Ku_k) \, ds \) as well as one of its global extrema \( \hat{t}_k \in I \). Next, assuming that \( \|p_k\|_C > 0 \), see Proposition 4.1, we define the new candidate jump \( \hat{v}^k := (p_k(\hat{t}_k)/\|p_k\|_C)\delta^k_{\hat{t}_k} \) and find improved jump heights \( \mu^{k+1/2} \in \mathbb{R}^{N_k+1}_+ \) and a new offset \( c^{k+1} \in \mathbb{R}^d \) from solving

\[
 \min_{(\mu, c) \in \mathbb{R}^{N_k+1}_+ \times \mathbb{R}^d} \left[ F \left( K \left( \mu^{N_k+1} \hat{v}^k + \sum_{i=1}^{N_k} \mu_i v^k_i, c \right) \right) + \beta \sum_{i=1}^{N_k+1} \mu_i \right] \quad (P_{A_k})
\]

which represents a finite-dimensional problem for the minimization of a convex and smooth function subject to one-sided box constraints. This type of problems can be tackled by a variety of efficient solution algorithms such as interior points or generalized Newton methods provided that the objective functional is...
sufficiently smooth. Now the new jump is added to the active set and the jump heights are updated setting
\[ A_{k+1/2} := (\mu_{N_k+1}^{k+1/2}, \nu_k^k) \cup \left\{(\mu_i^{k+1/2}, v_i^k)\right\}_{i=1}^{N_k}. \]

Finally, we prune the active set by removing all jumps whose associated jump magnitude was set to zero, i.e.
\[ A_{k+1} := \left\{(\mu_i^{k+1}, v_i^{k+1})\right\}_{i=1}^{N_k+1} = \left\{(\mu, v) \in A_{k+1/2} \mid \mu > 0 \right\} \]
and increment the iteration counter \( k \) by one. A summary can be found in Algorithm 1. Note that this also incorporates a ‘warm-up’ iteration without a point insertion step. This ensures that \( p_k \in C_0(I; \mathbb{R}^d) \) and that \( u_k \) is a minimizer of (P) if and only if \( \|p_k\|_C \leq 1 \) for all \( k \geq 1 \). The latter is used as a termination criterion in Algorithm 1. We formalize these observations in the following proposition.

**Proposition 4.1:** Denote by
\[ A_k = \left\{(\mu_i^k, v_i^k)\right\}_{i=1}^{N_k} = \left\{(\mu_i^k, v_i^k, \delta_{t_i})\right\}_{i=1}^{N_k}, \quad u_k = B \left( \sum_{i=1}^{N_k} \mu_i^k v_i^k, c^k \right) \]
the sequences of active sets and iterates generated by Algorithm 1 for \( k \geq 1 \). Moreover, set \( p_k(\cdot) = \int_0^T K^* \nabla F(Ku_k) \, ds \). Then there holds \( p_k \in C_0(I; \mathbb{R}^d) \) as well as
\[ \langle p_k, v_i^k \rangle = \langle p_k(t_i^k), v_i^k \rangle_{\mathbb{R}^d} = \beta. \]
In particular, \( \langle p_k, u_i^k \rangle = \beta \sum_{i=1}^{N_k} \mu_i^k \) and \( \|p_k\|_C \geq \beta \) if \( A_k \neq \emptyset \). Moreover, if \( \|p_k\|_C \leq \beta \) then \( u_k \) is a minimizer of (P). In particular, this holds if \( (\tilde{\mu}, \tilde{\nu}) \in A_k \) for some \( \tilde{\mu} > 0 \).

**Proof:** By step 2. and 7., respectively, of Algorithm 1 we have \( \mu_i^k > 0 \). Moreover, \( (\mu^k, c^k) \) is a minimizer of
\[ \min_{(\mu, c) \in \mathbb{R}_+^{N_k} \times \mathbb{R}^d} \left[ F \left( K \left( \sum_{i=1}^{N_k} \mu_i v_i^k, c \right) \right) + \beta \sum_{i=1}^{N_k} \mu_i \right]. \]
It is verified that the first-order necessary and sufficient optimality conditions for this problem imply
\[ p_k(T) = 0, \quad \langle p_k, v_i^k \rangle = \langle p_k(t_i^k), v_i^k \rangle_{\mathbb{R}^d} = \beta, \quad i = 1, \ldots, N_k. \]
Consequently, we get
\[ \langle p_k, u_i^k \rangle = \sum_{i=1}^{N_k} \mu_i^k \langle p_k, v_i^k \rangle = \beta \sum_{i=1}^{N_k} \mu_i^k. \]
Algorithm 1 Primal-dual-active-jump method (PDAJ) for (P)

**Input:** Active set $A_0 = \{(\mu_i^0, v_i^0)\}_{i=1}^{N_0}$, iterate $u_0 = B(\sum_{i=1}^{N_0} \mu_i^0 v_i^0, c^0)$

**Output:** Minimizer $\bar{u}$ to (P).

1. Find $(\mu^{1/2}, c^1) \in \mathbb{R}_{+}^{N_0} \times \mathbb{R}^d$ by solving
   \[
   \min_{(\mu, c) \in \mathbb{R}_{+}^{N_0} \times \mathbb{R}^d} \left[ F \left( \mathcal{K} \left( \sum_{i=1}^{N_0} \mu_i v_i^0, c \right) \right) + \beta \sum_{i=1}^{N_0} \mu_i \right].
   \]

2. Prune active set and update iterate:
   \[A_1 = \left\{ (\mu_i^1, v_i^1) \right\}_{i=1}^{N_1} = \left\{ (\mu_i^{1/2}, v_i^0) \mid \mu_i^{1/2} > 0 \right\}, \quad u_1 = B \left( \sum_{i=1}^{N_0} \mu_i^1 v_i^1, c^1 \right).\]

   for $k = 1, 2, \ldots$ do

3. Compute $p_k \in C_0(I; \mathbb{R}^d)$ and $\hat{t}_k \in I$ with
   \[p_k(\cdot) = \int_0^\cdot K^* \nabla F(Ku_k(s)) \, ds, \quad |p_k(\hat{t}_k)|_{\mathbb{R}^d} = \|p_k\|c = \max_{t \in I} |p_k(t)|_{\mathbb{R}^d}.\]
   if $\|p_k\|c \leq \beta$ then
   4. Terminate with $\bar{u} = u_k$ a minimizer of (P).
   end if

5. Find $(\mu^{k+1/2}, c^{k+1}) \in \mathbb{R}_{+}^{N_k+1} \times \mathbb{R}^d$ from $(P_{A_k})$ for $\hat{v}_k = (p_k(\hat{t}_k))/\|p_k\|c)\delta_{\hat{t}_k}$.

6. Update active set and iterate:
   \[A_{k+1/2} := (\mu_{N_k+1}^{k+1/2}, v_{N_k+1}^{k+1/2}) \cup \left\{ (\mu_i^{k+1/2}, v_i^{k+1/2}) \right\}_{i=1}^{N_k}, \]
   \[u_{k+1} := B \left( \mu_{N_k+1}^{k+1/2} \hat{v}_k + \sum_{i=1}^{N_k} \mu_i^{k+1/2} v_i^{k+1/2}, c^{k+1} \right).\]

7. Prune active set:
   \[A_{k+1} := \left\{ (\mu_i^{k+1}, v_i^{k+1}) \right\}_{i=1}^{N_k+1} = \left\{ (\mu, v) \in A_{k+1/2} \mid \mu > 0 \right\}.\]
   and set $k = k + 1$. 
end for
as well as
\[ \beta = \langle p_k, v^k \rangle \leq \|p_k\|_C \]
for every \((\mu^k_{\cdot}, v^k_{\cdot}) \in A_k\). Finally, assume that \(\|p_k\|_C \leq \beta\) so that \(A_k = \emptyset\), i.e. \(u'_k = 0\), and \(u_k\) satisfies the first-order optimality conditions for \((P)\), see Theorem 3.4. Hence, in this case, \(u_k\) is a minimizer of \((P)\). The same holds true if \(\|p_k\|_C = \beta\) and \(u'_k \neq 0\). Last, let \(\|p_k\|_C < \beta\). If \(\|p_k\|_C < \beta\) we note that \(A_k = \emptyset\), i.e. \(u'_k = 0\), and \(u_k\) fulfils the sufficient first-order optimality conditions from Theorem 3.4. Hence, in this case, \(u_k\) is a minimizer of \((P)\). Consequently, \(u_k = B(\sum_{i=1}^{N_k} \mu_i^k v^k_{\cdot}, c^k)\) satisfies
\[ \|u'_k\|_M = \sum_{i=1}^{N_k} \mu_i^k |v^k_{\cdot}|_R^d = \sum_{i=1}^{N_k} \mu_i^k. \]
Together with \(\langle p_k, u'_k \rangle = \sum_{i=1}^{N_k} \mu_i^k \) we finish noting that \(u_k\) fulfils the sufficient first-order optimality conditions from Theorem 3.4. Finally, if \((\hat{\mu}, \hat{v}^k) \in A_k\) for some \(\hat{\mu} > 0\), then we have
\[ \beta = \langle p_k, \hat{v}^k \rangle = \|p_k\|_C \]
and thus \(u_k\) is again a minimizer of \((P)\) following the previous observations. ■

4.2. A motivation by convex duality

To end this section, we briefly motivate the individual steps of Algorithm 1 from the perspective of the predual problem \((D)\). For this purpose, we first note that for every \(p \in C_0(I; \mathbb{R}^d)\) there holds
\[ \|p\|_C = \max_{t \in I} |p(t)|_R^d = \max_{t \in I} \max_{|v|_R^d = 1} \langle p(t), v \rangle_\mathbb{R}^d. \]
Consequently, \((D)\) can be equivalently rewritten as
\[ \inf_{p \in C_0(I; \mathbb{R}^d)} \inf_{y \in Y_{ad}(p)} F^*(y) \quad \text{s.t.} \quad \max_{t \in I} \max_{|v|_R^d = 1} \langle p(t), v \rangle_\mathbb{R}^d \leq \beta \]
where for \(p \in C_0(I; \mathbb{R}^d)\) we abbreviate
\[ Y_{ad}(p) := \left\{ y \in Y \mid p(\cdot) = \int_0^t K^* y(s) \, ds \right\}. \]
Now, given finitely many positions \( \{ t_i \}_{i=1}^N \subset I \) and jump directions \( \{ v_i \}_{i=1}^N \subset \mathbb{R}^d = 1 \), consider the set \( A = \{ v_i \delta t_i \}_{i=1}^N \) as well as the problem

\[
\inf_{p \in C_0(I;\mathbb{R}^d)} \inf_{y \in Y_{ad}(p)} F^*(y) \quad \text{s.t.} \quad (p(t_i), v_i)_{\mathbb{R}^d} \leq \beta, \quad i = 1, \ldots, N. \tag{\text{DA}}
\]

Here, loosely speaking, the infinitely many constraints of \( \text{(D)} \) are replaced by a finite subset described by the elements of \( A \). Consequently, there always holds

\[
\inf(\text{DA}) \leq \min(\mathcal{D}). \tag{17}
\]

Furthermore, define

\[
\min_{(\mu, c) \in \mathbb{R}_+^N \times \mathbb{R}^d} \left[ F \left( K \left( \sum_{i=1}^{N} \mu_i v_i, c \right) \right) + \beta \sum_{i=1}^{N} \mu_i \right]. \tag{\text{PA}}
\]

In the following, we show that, reminiscent to Proposition 3.5, \( \text{(PA)} \) corresponds to the predual problem to \( \text{(DA)} \) and strong duality holds. We omit the proof for the sake of brevity.

**Proposition 4.2:** Let \( (\hat{\mu}, \hat{c}) \) denote a minimizing pair of \( \text{(PA)} \) and define \( \hat{p} \) by

\[
\hat{p}(\cdot) = \int_0^\cdot K^* \nabla F \left( K \left( \sum_{i=1}^{N} \hat{\mu}_i v_i, \hat{c} \right) \right) (s) \, ds.
\]

Problem \( \text{(PA)} \) is equivalent to the Fenchel dual problem of \( \text{(DA)} \) in the sense that both, \( \text{(PA)} \) and the Fenchel dual, admit the same solution set. Moreover, strong duality holds, i.e. we have

\[
\inf(\text{DA}) = -\inf(\text{PA}).
\]

Finally, there holds

\[
\hat{p}(T) = 0, \quad (\hat{p}(t_i), v_i)_{\mathbb{R}^d} - \beta \leq 0, \quad \hat{\mu}_i \geq 0 \quad i = 1, \ldots, N \tag{18}
\]

as well as

\[
\sum_{i=1}^{N} \hat{\mu}_i \left( (\hat{p}(t_i), v_i)_{\mathbb{R}^d} - \beta \right) = 0, \tag{19}
\]

and \( \hat{p} \) is a solution to \( \text{(DA)} \).

In particular, due to (18) and (19), \( \hat{\mu} \) can be interpreted as a Lagrangian multiplier for the inequality constraints in \( \text{(DA)} \). These observations offer a primal-dual interpretation of the method proposed in Algorithm 1. Let

\[
\mathcal{A}_k = \left\{ (\mu_i^k, v_i^k) \right\}_{i=1}^{N_k}, \quad u_k = B \left( \sum_{i=1}^{N_k} \mu_i^k v_i^k, c^k \right)
\]

denote the active set and iterate of Algorithm 1 in iteration \( k \geq 1 \). Moreover, set \( A_k := \{ v_i^k \}_{i=1}^{N_k} \). Given the current predual variable \( p_k(\cdot) = \int_0^\cdot K^* \nabla F(Ku_k)(s) \, ds \),
we first check whether it is a solution to \((D)\). According to (17), this holds if and only if \(p_k\) is admissible for \((D)\), i.e. if

\[
\|p_k\|_C = \max_{t \in I} \max_{|v|_{\mathbb{R}^d} = 1} (p_k(t), v)_{\mathbb{R}^d} \leq \beta.
\]

If this is not the case, we determine one of the constraints in \((D)\) that are violated the most by \(p_k\). That is, we solve

\[
\max_{t \in I} \max_{|v|_{\mathbb{R}^d} = 1} [(p(t), v)_{\mathbb{R}^d} - \beta].
\]

Since \(\|p_k\|_C > 0\), a pair \((\hat{t}_k, v_k)\) is a maximizer to this problem if and only if

\[
\hat{t}_k \in \arg\max_{t \in I} |p_k(t)|_{\mathbb{R}^d}, \quad v_k = \frac{p_k(\hat{t}_k)}{\|p_k\|_C}.
\]

This corresponds to the computation of the new candidate jump \(\hat{\nu}_k\) in step 3. and step 5. of Algorithm 1. Subsequently, the new constraint is added to the subproblem and we solve \((\mathcal{D}_{A_k \cup \{\hat{\nu}_k\}})\) for the next dual variable \(p_{k+1}\) as well as an associated Lagrange multiplier \(\mu_{k+1/2}\). Due to the duality result in Proposition 4.2, this is equivalent to solving \((\mathcal{P}_{A_k})\) in step 5. of Algorithm 1. Finally, we remove all constraints for which the associated Lagrangian multiplier satisfies \(\hat{\mu}_{k+1/2} = 0\), i.e. the corresponding constraint is either not active at \(p_k\) or we do not have strong complementarity. This corresponds to the pruning step, step 7., in Algorithm 1.

### 5. Convergence analysis

This section addresses the convergence of Algorithm 1. The presentation is again split into two parts. In Section 5.1, we provide the subsequential strict convergence of \(u_k\) towards minimizers of \((\mathcal{P})\) as well as a first convergence result for the residuals

\[
r_j(u_k) := j(u_k) - \min_{u \in BV(I;\mathbb{R}^d)} j(u).
\]

In the second part, Section 5.2, we prove that under additional structural assumptions on the optimal dual variable \(\tilde{p}(\cdot) = \int_0^\cdot K^* \nabla F(\tilde{y})(s)\,ds\), \((\mathcal{P})\) admits a unique minimizer \(\tilde{u}\) and the iterates \(u_k\) generated by Algorithm 1 satisfy

\[
r_j(u_k) + \|u_k - \tilde{u}\|_{L^1} + \|u_k'\|_{\mathcal{M}} - \|\tilde{u}'\|_{\mathcal{M}} \leq c \xi^k
\]

for some \(\xi \in (0, 1)\) and all \(k \in \mathbb{N}\) large enough. Since some of the proofs in the following section are quite technical, we focus on the main results and postpone the proof of auxiliary results to Appendix.
### 5.1. Global sublinear convergence

In the following let

\[ A_k = \left\{ \left( \mu^k_i, v^k_i \right) \right\}_{i=1}^{N_k}, \quad u_k = B \left( \sum_{i=1}^{N_k} \mu^k_i v^k_i, c^k \right), \quad y_k = K u_k, \]

\[ p_k(\cdot) = \int_0^\cdot K^*(K y_k)(s) \, ds \]

denote the active set, iterate, observation and dual variable in iteration \( k \) of Algorithm 1, respectively. Since \( F \geq 0 \), see Assumption 3.1, the BV semi-norm of all elements in the sublevel set

\[ E_{u_k} = \left\{ u \in BV(I; \mathbb{R}^d) \mid j(u) \leq F(K u_k) + \beta \sum_{i=1}^{N_k} \mu^k_i \right\} \]

is bounded by

\[ \| u' \|_{\mathcal{M}} \leq \left( F(K u_k) + \beta \| u'_k \| \right) / \beta \leq M_k := \left( F(K u_k) + \beta \sum_{i=1}^{N_k} \mu^k_i \right) / \beta. \]

By construction, there holds

\[ F(K u_{k+1}) + \beta \sum_{i=1}^{N_{k+1}} \mu^{k+1}_i \leq F(K u_k) + \beta \sum_{i=1}^{N_k} \mu^k_i. \]

i.e. \( M_k \) is monotonically decreasing. We require additional regularity assumptions on the loss functional \( F \).

**Assumption 5.1:** The following two conditions hold:

(A1) The gradient \( \nabla F \) is Lipschitz, i.e. there is \( L > 0 \) such that

\[ \| \nabla F(y_1) - \nabla F(y_2) \|_Y \leq L \| y_1 - y_2 \|_Y \quad \forall y_1, y_2 \in Y. \]

(A2) The functional \( F : Y \to \mathbb{R} \) is strongly convex around the optimal observation, i.e. there exist a neighbourhood \( \mathcal{N}(\bar{y}) \) of \( \bar{y} \) in \( Y \) and \( \gamma_0 > 0 \) with

\[ F(y) \geq F(\bar{y}) + (\nabla F(\bar{y}), y - \bar{y})_Y + \gamma_0 \| y - \bar{y} \|_Y^2 \quad \forall y \in \mathcal{N}(\bar{y}). \]
This is, e.g. again fulfilled for the quadratic loss function $F(\cdot) = (1/2) \| \cdot - y_d \|_Y^2$ with a reference observation $y_d \in Y$. Now define the auxiliary residual

$$\hat{r}_j(u_k) := F(Ku_k) + \beta \sum_{i=1}^{N_k} \mu_i^k - \min_{u \in BV(I; \mathbb{R}^d)} j(u).$$

Note that $r_j(u_k) \leq \hat{r}_j(u_k)$ holds due to

$$\| u'_k \|_\mathcal{M} = \left\| \sum_{i=1}^{N_k} \mu_i^k v_i^k \right\|_\mathcal{M} \leq \sum_{i=1}^{N_k} \mu_i^k$$

using that $\| v_i^k \|_\mathcal{M} = 1$. The following version of the classical descent lemma holds.

**Lemma 5.1:** Let $u_k \in BV(I; \mathbb{R}^d)$, $p_k \in C_0(I; \mathbb{R}^d)$ and $\hat{r}^k \in \mathcal{M}(I; \mathbb{R}^d)$ be generated by Algorithm 1. Then we have

$$\hat{r}_j(u_{k+1}) - \hat{r}_j(u_k) \leq \min_{s \in [0,1]} \left[ -sM_k (\| p_k \|_C - \beta) + \frac{LC(K, M_1)s^2}{2} \right]$$

for all $k \geq 1$ and some $C(K, M_1) > 0$ which only depends on the operator norm of $K$ and $M_1$.

Using Lemma 5.1, we prove the subsequential convergence of $u_k$ towards minimizers of $(P)$ as well as the sublinear convergence of $r_j(u_k)$.

**Theorem 5.2:** Let $u_k \in BV(I; \mathbb{R}^d)$ and $p_k \in C_0(I; \mathbb{R}^d)$ be generated by Algorithm 1. Then we have

$$r_j(u_k) \leq \hat{r}_j(u_k) \leq M_k(\| p_k \|_C - \beta)$$

(22)

Moreover, Algorithm 1 either terminates after finitely many steps with $u_k$ a solution to $(P)$ or we have

$$r_j(u_k) \leq \hat{r}_j(u_k) \leq \frac{\hat{r}_j(u_1)}{1 + qk}$$

where $q = \frac{1}{2} \min \left\{ 1, \frac{\hat{r}_j(u_1)}{LC(K, M_1)} \right\}$

(23)

for all $k \geq 1$. In this case, $u_k$ admits at least one strict accumulation point and each such point is a solution to $(P)$. Moreover, we have $Ku_k \to \tilde{y}$ in $Y$ as well as $p_k \to \tilde{p}$ in $C_0(I; \mathbb{R}^d)$. If the minimizer $\tilde{u}$ to $(P)$ is unique then $u_k \rightharpoonup \tilde{u}$ on the whole sequence.

**Proof:** Let $\tilde{u}$ denote an arbitrary minimizer of $(P)$. Since $F$ is convex we estimate

$$\hat{r}_j(u_k) \leq (-K^* \nabla F(Ku_k), \tilde{u} - u_k)_Y + \beta \left( \sum_{i=1}^{N_k} \mu_i^k - \| \tilde{u} \|_\mathcal{M} \right) = \langle p_k, \tilde{u}' \rangle - \beta \| \tilde{u} \|_\mathcal{M}$$
where the last equality follows by integrating by parts. Finally, note that \( \bar{u} \in E_{u_k} \) and thus
\[
\langle p_k, \bar{u}' \rangle - \beta \| \bar{u} \|_{\mathcal{M}} \leq \| \bar{u} \|_{\mathcal{M}} (\| p_k \|_{C} - \beta) \leq M_k (\| p_k \|_{C} - \beta)
\]
yielding (22).

Now assume that Algorithm 1 does not converge after finitely many steps. Then \( \| p_k \|_{C} \geq \beta \), see Proposition 4.1, and \( r_j(u_k) > 0 \) for all \( k \). Combining (21) with (22) yields
\[
\tilde{r}_j(u_{k+1}) \leq \tilde{r}_j(u_k) + \min_{s \in [0,1]} \left[ -\tilde{s}_j(u_k) + \frac{LC(\mathcal{K}, M_1)s^2}{2} \right].
\]
The minimizer of the one-dimensional problem on the right-hand side is given by
\[
s^*_k = \min\{1, \tilde{r}_j(u_k)/(LC(\mathcal{K}, M_1))\}.
\]
Substituting this into (21) yields
\[
\tilde{r}_j(u_{k+1}) \leq \tilde{r}_j(u_k) + \begin{cases} 
\frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} & s^*_k = 1 \\
\frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} & \text{else}
\end{cases}
\]
Distinguishing both cases and dividing both sides by \( \tilde{r}_j(u_1) \) finally yields
\[
\frac{\tilde{r}_j(u_{k+1})}{\tilde{r}_j(u_1)} \leq \frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} - \frac{1}{2} \min \left\{ \frac{1}{\tilde{r}_j(u_k)}, \frac{1}{LC(\mathcal{K}, M_1)} \right\} \tilde{r}_j(u_k)^2
\]
\[
= \frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} - \frac{1}{2} \min \left\{ \frac{\tilde{r}_j(u_1)}{\tilde{r}_j(u_k)}, \frac{\tilde{r}_j(u_1)}{LC(\mathcal{K}, M_1)} \right\} \left( \frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} \right)^2
\]
\[
\leq \frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} - \frac{1}{2} \min \left\{ 1, \frac{\tilde{r}_j(u_1)}{LC(\mathcal{K}, M_1)} \right\} \left( \frac{\tilde{r}_j(u_k)}{\tilde{r}_j(u_1)} \right)^2
\]
where the last inequality follows due to the monotonicity of \( \tilde{r}_j(u_k) \).

Invoking [25, Lemma 3.1] yields (23). Consequently, we have \( r_j(u_k) \to 0 \) and \( \{u_k\}_k \) is a minimizing sequence for (P). As in the proof of Proposition 3.2, we conclude that \( u_k \) is bounded in BV(I; \( \mathbb{R}^d \)). Thus, it admits at least one weak* convergent subsequence, denoted by the same index, with limit \( \bar{u} \in \text{BV}(I; \mathbb{R}^d) \), i.e. \( u_k \to \bar{u} \) in \( L^1(I; \mathbb{R}^d) \) and \( u_k' \to^* \bar{u}' \) in \( \mathcal{M}(I; \mathbb{R}^d) \). Since \( \text{BV}(I; \mathbb{R}^d) \hookrightarrow c \ L^2(I; \mathbb{R}^d) \) we also conclude \( y_k \to K\bar{u} \) in \( Y \) as well as
\[
p_k \to \int_0^1 K^*\nabla F(K\bar{u})(s) \, ds \text{ in } C_0(I; \mathbb{R}^d).
\]
Finally, we note that \( j \) is weak* lower semicontinuous on \( \text{BV}(I; \mathbb{R}^d) \). Consequently \( r_j(\bar{u}) = 0 \) and \( \bar{u} \) is a minimizer of (P). Finally, since \( F(Ku_k) \to F(K\bar{u}) \), we also get \( \| u'_k \|_{\mathcal{M}} \to \| \bar{u}' \|_{\mathcal{M}} \) yielding the strict convergence of \( u_k \) towards \( \bar{u} \).
Thus, we have shown that any weak* accumulation point of $u_k$ is indeed a strict accumulation point and a minimizer of $(\mathcal{P})$. Recalling that the optimal observation $\tilde{y}$ as well as the optimal dual variable $\tilde{p}$ are unique we conclude $y_k \rightarrow \tilde{y}$ in $Y$ and $p_k \rightarrow \tilde{p}$ in $C_0(I; \mathbb{R}^d)$ for the whole sequence. If $\tilde{u}$ is the unique minimizer of $(\mathcal{P})$ then it is also the unique strict accumulation point of $u_k$ and thus $u^k \rightharpoonup \tilde{u}$ on the whole sequence.

If $F$ is strongly convex around $\tilde{y}$, see Assumption 5.1 A2, then the convergence guarantee for the residual from Theorem 5.2 also carries over to the observations and dual variables.

**Proposition 5.3:** Let Assumption 5.1 hold. Then we have

$$\|y_k - \tilde{y}\|_Y + \|p_k - \tilde{p}\|_C + \|p_k\|_C - \|\tilde{p}\|_C \leq c r_j(u_k)^{1/2}$$

for all $k \in \mathbb{N}$ large enough.

**Proof:** Let $\mathcal{N}(\tilde{y})$ denote the neighbourhood from Assumption 5.1 A2. Since $y_k \rightarrow \tilde{y}$ in $Y$, see Theorem 5.2, there holds $y_k \in \mathcal{N}(\tilde{y})$ for all $k \in \mathbb{N}$ large enough. Consequently Assumption 5.1 A2 yields

$$r_j(u_k) \geq (\nabla F(\tilde{y}), y_k - \tilde{y})_Y + \beta(\|u'_k\|_M - \|\tilde{u}'\|_M) + \gamma_0 \|y_k - \tilde{y}\|_Y^2.$$

Finally noting that

$$(\nabla F(K\tilde{u}), y_k - \tilde{y})_Y + \beta(\|u'_k\|_M - \|\tilde{u}'\|_M)$$

$$= \langle \tilde{p}, \tilde{u}' - u'_k \rangle + \beta(\|u'_k\|_M - \|\tilde{u}'\|_M) \geq 0,$$

see Theorem 3.4, we get

$$\|y_k - \tilde{y}\|_Y \leq (1/\gamma_0)^{1/2} r_j(u_k)^{1/2}.$$

The remaining estimates follow from

$$\|p_k\|_C - \|\tilde{p}\|_C \leq \|p_k - \tilde{p}\|_C \leq c \|K^*(\nabla F(y_k) - \nabla F(\tilde{y}))\|_{L^2}$$

$$\leq c \|K^*\|_{Y,L^2} \|\nabla F(y_k) - \nabla F(\tilde{y})\|_Y$$

$$\leq cL \|K^*\|_{Y,L^2} \|y_k - \tilde{y}\|_Y.$$

**5.2. Local linear convergence**

Next, we prove that Algorithm 1 converges linearly provided that additional structural requirements on the optimal dual variable $\tilde{p}$ hold. First we assume that $\tilde{p}$ only admits a finite number $N$ of global extrema $\{\tilde{t}_i\}_{i=1}^N$. Together with a linear independence assumption on $\{\tilde{p}(\tilde{t}_i)\}_{i=1}^N$ this ensures the existence of a unique, piecewise constant minimizer of $(\mathcal{P})$. 
\textbf{Assumption 5.2:} Recall the definition of the optimal dual variable \( \bar{p} = \int_0^T K^* \nabla F(\bar{y}) \, ds \). Assume that there is \( N \in \mathbb{N} \) and \( \{ \bar{t}_i \}_{i=1}^N \subset I \) with
\[
\{ \bar{t}_i \}_{i=1}^N = \left\{ t \in I \mid |\bar{p}(t)|_{\mathbb{R}^d} = \| \bar{p} \|_{\mathcal{C}} = \beta \right\}.
\] (24)
Moreover, let \( \{ e_i \}_{i=1}^N \subset \mathbb{R}^d \) denote the canonical basis of \( \mathbb{R}^d \). The set
\[
\{ K(\bar{p}(\bar{t}_i) \chi_{\bar{t}_i}) \}_{i=1}^N \cup \{ K(e_i \chi_i) \}_{i=1}^d \subset Y
\] (25)
is linearly independent.

Here, the linear independence assumption can be interpreted as a tightening of Assumption 3.1 which now ensures the uniqueness of the minimizer to \((P)\).

\textbf{Corollary 5.4:} Let Assumption 5.2 hold. Then the minimizer \( \bar{u} = B(\bar{u}', a_{\bar{u}}) \) to \((P)\) is unique and \( \bar{u}' \) is given by
\[
\bar{u}' = \sum_{i=1}^N \tilde{\mu}_i \tilde{v}_i = \sum_{i=1}^N \tilde{\mu}_i \tilde{v}_i \delta_{\tilde{t}_i} \quad \text{where} \quad \tilde{\mu}_i \geq 0, \quad \tilde{v}_i = \frac{\bar{p}(\bar{t}_i)}{\beta}
\]
for all \( i = 1, \ldots, N \).

\textbf{Proof:} Introduce the linear and continuous operator \( \hat{K} : \mathbb{R}^N \times \mathbb{R}^d \to Y \) by
\[
\hat{K}(\mu, C) = K(C \chi_0) + \sum_{i=1}^N K\left( (\bar{p}(\bar{t}_i)/\beta) \chi_{\bar{t}_i} \right) \quad \forall \mu \in \mathbb{R}^N, \quad C \in \mathbb{R}^d.
\]
Then \( \hat{K} \) is injective according to (25). According to Corollary 3.6 and (24) every minimizer \( \hat{u} \) of \((P)\) is of the form
\[
\hat{u} = B\left( \sum_{i=1}^N (\bar{p}(\bar{t}_i)/\beta) \delta_{\bar{t}_i}, a_{\bar{u}} \right) = \tilde{C} \chi_0 + \sum_{i=1}^N \tilde{\mu}_i \chi_{\bar{t}_i}
\]
where \( \tilde{\mu} \in \mathbb{R}_+^N \) and \( \tilde{C} \) is implicitly given by
\[
\tilde{C} = a_{\bar{u}} - \frac{1}{T} \sum_{i=1}^N \tilde{\mu}_i (\bar{p}(\bar{t}_i)/\beta)) (T - \bar{t}_i),
\]
see the definition of the operator \( B, (3) \), and its inverse \( B^{-1} \), respectively. Due to the optimality of \( \hat{u} \) for \((P)\) we verify that \( (\tilde{\mu}, \tilde{C}) \) is a minimizing pair for
\[
\min_{\mu \in \mathbb{R}_+^N, C \in \mathbb{R}^d} \left[ F(\hat{K}(\mu, C)) + \beta \sum_{i=1}^N \mu_i \right].
\] (26)
The proof is finished noting that (26) admits a unique minimizer since \( F \circ \hat{K} \) is strictly convex. \( \blacksquare \)
Note that the linear independence assumption in Assumption 5.2 trivially holds if $K$ is injective, see the example in Section 6.2, or can often be verified a posteriori, i.e. after a solution to $(P)$ is computed, see Section 6.1, by checking the rank of the operator $\hat{K}$.

**Remark 5.1:** Recalling the strong convexity of $F$ from Assumption 5.1, it is worth pointing out that, the injectivity of $\hat{K}$, and thus the linear independence assumption in Assumption 5.2, is equivalent to the quadratic growth behaviour

$$\beta \sum_{i=1}^{N} (\mu_i - \tilde{\mu}_i) + F(K(\mu, C)) - F(K(\tilde{\mu}, \tilde{C})) \geq \theta_1 \left( |C - \tilde{C}|_{\mathbb{R}^d}^2 + |\mu - \tilde{\mu}|^2_{\mathbb{R}^N} \right)$$

in the vicinity of $(\tilde{\mu}, \tilde{C})$ for some $\theta_1 > 0$. Moreover, while Corollary 5.4 shows that the linear independence assumption is sufficient for the uniqueness of the solution to $(P)$, it is also a necessary condition under slightly stronger assumptions. More in detail, if there is any solution to $(P)$ with $\bar{\mu}_i > 0$, $i = 1, \ldots, N$, but

$${\{K(\tilde{p}(\bar{t}_i)x_{\bar{t}_i})\}}_{i=1}^{N} \cup {\{K(e_i\chi I)\}}_{i=1}^{d}$$

is linearly dependent, then the solution is not unique. This can be argued along the lines of [24, Theorem 6.31], hence we omit a detailed proof at this point.

According to Assumption 5.2 and the continuity of $|\tilde{p}|_{\mathbb{R}^d}$ there is $\sigma > 0$ as well as a radius $R > 0$ such that the intervals $(\bar{t}_i - R, \bar{t}_i + R) \subset I$, $i = 1, \ldots, N$, are pairwise disjoint and

$$|\tilde{p}(t)|_{\mathbb{R}^d} \leq \beta - \sigma \quad \forall t \in \bar{I} \setminus \bigcup_{i=1}^{N} (\bar{t}_i - R, \bar{t}_i + R). \quad (27)$$

Now we impose a final set of assumptions which requires the positivity of $\tilde{\mu}_i$ as well as the quadratic growth of $|\tilde{p}|_{\mathbb{R}^d}$ around its global maximizers. From the perspective of the dual problem, this first condition corresponds to a *strict complementarity condition* and the second one is equivalent to a *second-order-sufficient-condition (SSC)* for the maxima $\bar{t}_i$ of $|\tilde{p}(\cdot)|_{\mathbb{R}^d}$.

**Assumption 5.3:** For all $i = 1, \ldots, N$, there holds $\bar{\mu}_i > 0$ as well as

$$\theta_0 |t - \bar{t}_i|^2 \leq \beta - |\tilde{p}(t)|_{\mathbb{R}^d} \quad \forall t \in (\bar{t}_i - R, \bar{t}_i + R)$$

where $R > 0$ denotes the radius from (27). Moreover, $K^* \in L(Y; L^\infty(I; \mathbb{R}^d))$.

**Remark 5.2:** Define the scalar-valued function $\bar{P}(t) = |\tilde{p}(t)|_{\mathbb{R}^d}$ and assume that $\tilde{p} \in C^2(I; \mathbb{R}^d)$. Then it is verified that $\bar{P}$ is also at least two times continuously differentiable on $(\bar{t}_i - R, \bar{t}_i + R)$ if $R > 0$ is chosen small enough. In particular,
this implies $\bar{P}(\bar{t}_i) = \beta$, $\bar{P}'(\bar{t}_i) = 0$ and $\bar{P}''(\bar{t}_i) \leq 0$. Thus, by potentially choosing $R > 0$ even smaller as well as Taylor approximation of $\bar{P}$ we arrive at

$$\bar{P}(t) \leq \beta - \frac{\bar{P}''(\bar{t}_i)}{4} |t - \bar{t}_i|^2$$

for all $t \in (\bar{t}_i - R, \bar{t}_i + R)$. Hence the quadratic growth condition of Assumption 5.3 is fulfilled if $\bar{P}''(\bar{t}_i) > 0$, $i = 1, \ldots, N$.

The following quadratic growth behaviour of the linear functional induced by $\bar{p}$ is a direct consequence.

**Lemma 5.5:** Let Assumption 5.3 hold. Then there is $\gamma_1 > 0$ such that

$$\gamma_1 \left( |t - \bar{t}_i|^2 + |v - \bar{v}_i|^2 |_{\mathbb{R}^d} \right) \leq \beta - \langle \bar{p}, v \delta_t \rangle \quad \forall t \in (\bar{t}_i - R, \bar{t}_i + R), \quad |v|_{\mathbb{R}^d} = 1$$

and all $i = 1, \ldots, N$.

Moreover, we deduce the following Lipschitz property of $K$.

**Lemma 5.6:** There holds

$$\| K(v_1 \delta_{t_1} - v_2 \delta_{t_2}, 0) \|_Y \leq c \left( |t_1 - t_2| + |v_1 - v_2|_{\mathbb{R}^d} \right)$$

for all $t_1, t_2 \in I$, $v_1, v_2 \in \mathbb{R}^d$, $|v_1|_{\mathbb{R}^d} = |v_2|_{\mathbb{R}^d} = 1$.

**Main result**

The following theorem summarizes the main results of the following sections.

**Theorem 5.7:** Let $u_k$ be generated by Algorithm 1 and let Assumption 3.1–5.3 hold. Then Algorithm 1 either terminates after finitely many steps with $u_k = \bar{u}$ or there is $\zeta \in (0, 1)$ such that

$$r_j(u_k) + \| u_k - \bar{u} \|_{L^1} + \| u_k' \|_{\mathcal{M}} - \| \bar{u}' \|_{\mathcal{M}} \leq c \zeta^k$$

for all $k \in \mathbb{N}$ large enough.

Since the proof of this improved convergence behaviour is rather technical we give a short outline before going into detail. Utilizing the strict convergence of $u_k$ towards $\bar{u}$ as well as the isolation of the global extrema of $\bar{p}$ we conclude that the iterate $u_k$ only jumps in the vicinity of $\{\bar{t}_i\}_{i=1}^N$. More in detail, for sufficiently large $k$, these observations yield a partition of $\{1, \ldots, N_k\}$ into nonempty, pairwise disjoint sets $A^i_k$, $i = 1, \ldots, N$, such that

$$(\mu^k_i, v^k_j \delta_{t^k_j}) \in \mathcal{A}_k, \quad j \in A^i_k \Rightarrow t^k_j \in (\bar{t}_i - R, \bar{t}_i + R).$$

Moreover, the ‘closeness’ of the jumps $v^k_j, j \in A^i_k$, and the optimal one $\bar{v}_i$, i.e. the distance between the positions $t^k_j$ and $\bar{t}_i$ as well as the misfit between the associated directions $v^k_i - \bar{v}_i$, can be quantified in terms of the auxiliary residual $\hat{r}_j(u_k)$,
see Lemma 5.12. Similarly, in Proposition 5.13, we show that the new candidate jump \( \hat{\nu}^k \), see step 5. in Algorithm 1, lies in the vicinity of some \( \hat{\tau}_i \in \{ \tau_i \}_{i=1}^N \). Finally, as in the proof of Lemma 5.1, we then rely on an auxiliary iterate \( \hat{u}_{k,s} = B(\hat{u}'_{k,s}, \nu^k) \), \( s \in (0, 1) \), where

\[
\hat{u}'_{k,s} = (1 - s) \sum_{j \in A^i_k} \mu^k_j v^k_j + s \left( \sum_{j \in A^i_k} \mu^k_j \right) \frac{p_k(\hat{\tau}_i)}{\|p_k\|_C} \delta_{\hat{\tau}_i} + \sum_{i=1}^N \sum_{j \in A^i_k} \mu^k_j v^k_j
\]

The descent properties of this auxiliary iterate are then exploited in Lemma 5.15 to prove an improved version of the descent lemma, Lemma 5.1, which finally yields the linear convergence of \( r_j(u_k) \). The linear convergence of \( u_k \) w.r.t. to the strict topology is then concluded as a by-product, see Lemmas 5.17 and 5.18.

Remark 5.3: To finish this section, let us briefly compare \( \hat{u}_{k,s} \) with the auxiliary iterate \( u_{k,s} = B(u'_{k,s}, \nu^k) \), where

\[
u'_{k,s} = sM_0 \nu_k + (1 - s) \sum_{i=1}^N \sum_{j \in A^i_k} \mu^k_j v^k_j = (1 - s)u'_{k} + sM_0 \nu_k
\]

which is used in the proof of Lemma 5.1. Loosely speaking, to obtain \( u'_{k,s} \), we take ‘mass’ from all Dirac Delta functionals in \( u'_{k} \), i.e. the height of all jumps in the iterate is decreased, and move it to the new candidate jump \( \hat{\nu}^k \). In contrast, the construction of \( \hat{u}_{k,s} \) can be viewed as a local update of \( u_k \) since mass is only taken away from those jumps \( u^k_j \) supported in \( (\hat{\tau}_i - R, \hat{\tau}_i + R) \). On the complement, \( I \setminus (\hat{\tau}_i - R, \hat{\tau}_i + R) \), we have \( \hat{u}_{k,s} = u_k \). This allows for a refined analysis of the descent achieved by Algorithm 1 in each iteration.

**Linear convergence of the residual**

For the sake of readability, we tacitly assume that Algorithm 1 does not converge after finitely many steps. The following proposition summarizes some immediate consequences of this assumption.

**Proposition 5.8:** Assume that Algorithm 1 does not terminate after finitely many steps. Then there holds \( u_k \rightharpoonup_s \tilde{u} \), \( \|p_k\|_C \geq \beta \) and \( A_k \neq \emptyset \) for all \( k \in \mathbb{N} \) large enough.

**Proof:** Since the minimizer of \( (P) \) is unique, see Corollary 5.4, we get \( u_k \rightharpoonup_s \tilde{u} \) from Theorem 5.2. In particular, this implies \( u'_k \rightharpoonup^* \tilde{u}' \) in \( \mathcal{M}(I; \mathbb{R}^d) \) and thus \( u'_k \neq 0 \) for \( k \) large enough. This also yields \( A_k \neq \emptyset \) and \( \|p_k\|_C \geq \beta \), see Proposition 4.1.

Now we use the isolation of the global extrema of \( \hat{p} \), see (27), as well as the uniform convergence of \( p_k \) from Proposition 5.3 to conclude that \( |p_k(\cdot)|_{\mathbb{R}^d} \) is bounded away from \( \beta \) outside of the intervals \( (\hat{\tau}_i - R, \hat{\tau}_i + R) \).
Corollary 5.9: Let $\sigma > 0$ and $R > 0$ as in (27) be given. Moreover, let $p_k$ be generated by Algorithm 1. For all $k \in \mathbb{N}$ large enough we have

$$|p_k(t)|_{\mathbb{R}^d} \leq \beta - \frac{\sigma}{2} \quad \forall \, t \in \bar{I} \setminus \bigcup_{i=1}^{N}(\bar{t}_i - R, \bar{t}_i + R).$$

Proof: Choose an arbitrary but fixed $t \in \bar{I} \setminus \bigcup_{i=1}^{N}(\bar{t}_i - R, \bar{t}_i + R)$. We estimate

$$|p_k(t)|_{\mathbb{R}^d} \leq |\bar{p}(t)|_{\mathbb{R}^d} + |p_k(t)|_{\mathbb{R}^d} - |\bar{p}(t)|_{\mathbb{R}^d} \leq \beta - \sigma + \|p_k - \bar{p}\|_{C} \leq \beta - \frac{\sigma}{2}$$

for all $k \in \mathbb{N}$ large enough. Here we use (27) in the second inequality and the uniform convergence of $p_k$, see Proposition 5.3, in the last one. ■

Using this estimate, we prove that the iterate $u_k$ solely jumps in the vicinity of the optimal jump positions $\bar{t}_i$.

Proposition 5.10: Denote by

$$A_k = \left\{ (\mu^k_i, v^k_i) \right\}_{i=1}^{N_k} = \left\{ (\mu^k_i, v^k_i \delta_{\bar{t}_i}) \right\}_{i=1}^{N_k}$$

the sequence of active sets generated by Algorithm 1. For all $k \in \mathbb{N}$ large enough there exist pairwise disjoint index sets $A^i_k$ with $\bigcup_{i=1}^{N} A^i_k = \{1, \ldots, N_k\}$ and $t^j_k \in (\bar{t}_i - R, \bar{t}_i + R), j \in A^i_k$.

Next, we prove that the sets $A^i_k$ are nonempty for large $k \in \mathbb{N}$. This means that each optimal jump $\bar{v}_i$ is approximated by at least one jump in the iterate $u_k$. Moreover, the “lumped” height $\sum_{j \in A^i_k} \mu^k_j$ of all jumps $v^k_j, j \in A^i_k$, converges to the optimal jump height $\bar{\mu}_i$. For this purpose define the restricted measures

$$U'_{k,i} := \sum_{j \in A^i_k} \mu^k_j v^k_j.$$  (28)

Lemma 5.11: Let $U'_{k,i}$ be defined as in (28). Then there holds

$$U'_{k,i} \rightharpoonup * \bar{\mu}_i \delta_{\bar{t}_i}, \quad \sum_{j \in A^i_k} \mu^k_j \rightarrow \bar{\mu}_i.$$

In particular this implies $A^i_k \neq \emptyset$ for all $k \in \mathbb{N}$ and $\sum_{i=1}^{N_k} \mu^k_i \rightarrow \|\bar{u}'\|_M$. Up to now we have only given qualitative statements on the approximation of $\bar{v}_i$ by jumps $v^k_j$ of the iterate $u_k$. In order to improve on the convergence result of Theorem 5.2 we also need a quantitative estimate for this observation. For this purpose, we recall that both, $\bar{v}_i$ and $v^k_j$, are vector-valued Dirac Delta functionals. Thus, a suitable way to compare these jumps is given in terms of the differences
$t^k_j - \bar{t}_i$ and $v^k_j - \bar{v}_i$ of jump positions and directions, respectively. This can be quantified using the quadratic growth behaviour of $\bar{p}$ from Lemma 5.5.

Lemma 5.12: There holds

$$\sum_{i=1}^{N} \sum_{j \in A^i_k} \mu^k_j \left( |t^k_j - \bar{t}_i| + |v^k_j - \bar{v}_i|_{\mathbb{R}^d} \right) \leq c \sqrt{\hat{r}_j(u_k)}$$

(29)

for all $k \in \mathbb{N}$ large enough.

Proof: Let $\gamma_1$ denote the constant from Lemma 5.5. Recalling that

$$\sum_{i=1}^{N_k} \mu^k_i = \sum_{i=1}^{N} \sum_{j \in A^i_k} \mu^k_j$$

we estimate

$$\frac{\gamma_1}{2} \sum_{i=1}^{N_k} \mu^k_i \left( \sum_{i=1}^{N} \sum_{j \in A^i_k} \mu^k_j \left( |t^k_j - \bar{t}_i| + |v^k_j - \bar{v}_i|_{\mathbb{R}^d} \right) \right)^2 \leq \frac{\gamma_1}{2} \sum_{i=1}^{N_k} \mu^k_i \left( |t^k_j - \bar{t}_i| + |v^k_j - \bar{v}_i|_{\mathbb{R}^d} \right)^2 \leq \gamma_1 \sum_{i=1}^{N_k} \mu^k_i \left( |t^k_j - \bar{t}_i|^2 + |v^k_j - \bar{v}_i|^2_{\mathbb{R}^d} \right) \leq \sum_{i=1}^{N_k} \mu^k_i \left( \beta - \langle \bar{p}, v^k_j \rangle \right) = \beta \sum_{i=1}^{N_k} \mu^k_i - \langle \bar{p}, u'_k \rangle.$$

Here, the first inequality follows from Jensen’s inequality, the second is due to Young’s inequality and the final one follows from Lemma 5.5. Moreover, due to the convexity of $F$ we estimate

$$\hat{r}_j(u_k) = F(Ku_k) + \beta \sum_{i=1}^{N_k} \mu^k_i - F(K\bar{u}) - \beta \|\bar{u}\|_M$$

$$\geq \beta \sum_{i=1}^{N_k} \mu^k_i - \beta \|\bar{u}\|_M + (\nabla F(K\bar{u}), K\bar{u} - K\bar{u})_Y.$$
Now we rewrite
\[
(\nabla F(K\tilde{u}), K\tilde{u}_k - K\bar{u})_Y - \beta \|\tilde{u}\|_\mathcal{M} = \langle \bar{p}, \tilde{u}' - u'_k \rangle - \beta \|\tilde{u}\|_\mathcal{M} = -\langle \bar{p}, u'_k \rangle.
\]
using the first-order optimality conditions for \(\tilde{u}\), see Theorem \ref{thm:optimality_conditions}, as well as integration by parts. Summarizing all previous observations, we arrive at
\[
\left( \sum_{i=1}^N \sum_{j \in A_k^i} \mu_j^k (|t_j^k - \bar{t}_i| + |\psi_j^k - \bar{u}_i|) \right)^2 \leq \frac{2 \sum_{i=1}^N \mu_i^k \bar{\gamma}_j(u_k)}{\gamma_1} \leq \frac{4 \|\tilde{u}\|_\mathcal{M} \bar{\gamma}_j(u_k)}{\gamma_1}.
\]
Taking the square root on both sides of the inequality yields the claimed statement. \hfill \blacksquare

A similar estimate holds for the new candidate jump \(\hat{\gamma}^k\) computed in step 3. of Algorithm \ref{alg:algorithm}.

**Proposition 5.13:** Let \(\hat{\gamma}^k = (p_k(\hat{t}_k)/\|p_k\|_C)\delta_{\hat{t}_k}\) with \(|p_k(\hat{t}_k)|_{\mathbb{R}^d} = \|p_k\|_C\) be given. For all \(k \in \mathbb{N}\) large enough there is a \(k\)-dependent index \(\hat{\gamma} \in \{1, \ldots, N\}\) such that \(\hat{t}_k \in A_k^1\) and
\[
|\hat{t}_k - \hat{\gamma}| + |p_k(\hat{t}_k)/\|p_k\|_C - \bar{v}_\gamma|_{\mathbb{R}^d} \leq c \sqrt{r_j(u_k)} \tag{30}
\]

**Proof:** According to Proposition \ref{prop:optimality_conditions} there holds \(\|p_k\|_C \to \beta\). Thus we conclude \(\hat{t}_k \in (\hat{\gamma} - R, \hat{\gamma} + R)\) for some \(\hat{\gamma} \in \{1, \ldots, N\}\) from Corollary \ref{cor:corollary}. Applying Lemma \ref{lem:lemma5} we get
\[
\frac{\gamma_1}{2} \left( |\hat{t}_k - \hat{\gamma}| + |p_k(\hat{t}_k)/\|p_k\|_C - \bar{v}_\gamma|_{\mathbb{R}^d} \right)^2 \leq \gamma_1 \left( |\hat{t}_k - \hat{\gamma}|^2 + |p_k(\hat{t}_k)/\|p_k\|_C - \bar{v}_\gamma|_{\mathbb{R}^d}^2 \right) \leq \beta - \langle \bar{p}, \hat{\gamma}^k \rangle.
\]
For abbreviation, set \(\tilde{u}'_k := \bar{v}_\gamma \delta_{\hat{t}_k}\). Next note that
\[
\beta - \langle \bar{p}, \hat{\gamma}^k \rangle = \langle \bar{p}, \tilde{u}'_k - \hat{\gamma}^k \rangle \leq \langle \bar{p} - p_k, \tilde{u}'_k - \hat{\gamma}^k \rangle = (\nabla F(y_k) - \nabla F(\gamma), K(\tilde{u}'_k - \hat{\gamma}^k, 0)_Y
\]
since
\[
\|p_k\|_C = \langle p_k, \hat{\gamma}^k \rangle \geq \langle p_k, \tilde{u}'_k \rangle.
\]
Utilizing Proposition \ref{prop:optimality_conditions} and Lemma \ref{lem:algorithm}, we finally arrive at
\[
\frac{\gamma_1}{2} \left( |\hat{t}_k - \hat{\gamma}| + |p_k(\hat{t}_k)/\|p_k\|_C - \bar{v}_\gamma|_{\mathbb{R}^d} \right)^2 \leq \|\nabla F(\gamma) - \nabla F(y_k)\|_Y \|K(\tilde{u}'_k - \hat{\gamma}^k, 0)|_Y \leq c \sqrt{r_j(u_k)} \left( |\hat{t}_k - \hat{\gamma}| + |p_k(\hat{t}_k)/\|p_k\|_C - \bar{v}_\gamma|_{\mathbb{R}^d} \right). \hfill \blacksquare
\]
Now fix \( k \in \mathbb{N} \) large enough and let \( \hat{\tau} \in \{1, \ldots, N\} \) be the index from Proposition 5.13. Further recall the index sets \( A^i_k, i = 1, \ldots, N \), from Proposition 5.10. For every \( s \in [0, 1] \) define the locally lumped measure

\[
\hat{\mu}^k_{s, N_k + 1} = \sum_{j=1}^{N_k} \hat{\mu}^k_{sj} v^k_j
\]

where \( \hat{\mu}^k_s \in \mathbb{R}^{N_k+1} \) is defined as

\[
\hat{\mu}^k_s = \mu^k, \quad \forall j \in A^\hat{\tau}_k, \quad i \neq \hat{\tau}, \quad \hat{\mu}^k_s = (1 - s) \mu^k, \quad \forall j \in \hat{A}^\hat{\tau}_k
\]

as well as

\[
\hat{\mu}^k_{s, N_k + 1} = s \left( \sum_{j \in \hat{A}^\hat{\tau}_k} \mu^k_j \right).
\]

Set \( \hat{u}_{k,s} = B(\hat{u}'_{k,s}, c_k) \). By construction, there holds

\[
\hat{\tau}_j(u_{k+1}) - \hat{\tau}_j(u_k) \leq F(K \hat{u}_{k,s}) - F(K \hat{u}_k) + \beta \left( \sum_{j=1}^{N_k+1} \hat{\mu}^k_{sj} - \sum_{j=1}^{N_k} \mu^k_j \right).
\]

The following properties of \( \hat{u}_{k,s} \) follow directly.

**Lemma 5.14:** Let \( u_k \) and \( p_k \) be generated by Algorithm 1. Moreover, let \( \hat{u}_{k,s} \) be defined as above. Then there holds

\[
\langle p_k, u'_k - \hat{u}'_{k,s} \rangle = -s \left( \sum_{j \in \hat{A}^\hat{\tau}_k} \mu^k_j \right) (\|p_k\|c - \beta), \quad \sum_{j=1}^{N_k+1} \hat{\mu}^k_{sj} = \sum_{j=1}^{N_k} \mu^k_j.
\]

As a final step, we now use \( \hat{u}_{k,s} \) to prove a refined descent estimate for Algorithm 1.

**Lemma 5.15:** For all \( k \in \mathbb{N} \) large enough there holds

\[
\hat{\tau}_j(u_{k+1}) - \hat{\tau}_j(u_k) \leq \min_{s \in [0, 1]} \left[ \left( s^2 c_1 - s \left( \min_{i=1, \ldots, N} \hat{\mu}_i / 2M_0 \right) \right) \hat{\tau}_j(u_k) \right]
\]

for some \( c_1 > 0 \) independent of \( k \) and \( s \).
**Proof:** Let \( s \in [0, 1] \) be arbitrary but fixed. We estimate

\[
\hat{r}_j(u_{k+1}) - \hat{r}_j(u_k) \leq F(K\hat{u}_{k,s}) - F(K\hat{u}_k) + \beta \left( \sum_{j=1}^{N_{k+1}} \hat{\mu}_{s,j} - \sum_{j=1}^{N_{k}} \mu_j^k \right)
\]

\[
= F(K\hat{u}_{k,s}) - F(K\hat{u}_k)
\]

where the last equality holds due to Lemma 5.14. As in the proof of Lemma 5.1 we now find

\[
F(K\hat{u}_{k,s}) - F(K\hat{u}_k) \leq \langle p_k, u'_k - \hat{u}'_{k,s} \rangle + \frac{L}{2} \|K(u_k - \hat{u}_{k,s})\|^2_Y
\]  \hspace{1cm} (33)

where \( L > 0 \) denotes the Lipschitz constant of \( \nabla F \). Summarizing the previous observations and again utilizing Lemma 5.14 we thus get

\[
\hat{r}_j(u_{k+1}) - \hat{r}_j(u_k) \leq -s \left( \sum_{j \in \hat{A}_k^i} \mu_j^k \right) (\|p_k\|c - \beta) + \frac{L}{2} \|K(u_k - \hat{u}_{k,s})\|^2_Y.
\]

Next, we use (22) as well as \( \sum_{j \in \hat{A}_k^i} \mu_j^k \to \bar{\mu}_i, \ i = 1, \ldots, N \), see Lemma 5.11, to establish the upper bound

\[
-s \left( \sum_{j \in \hat{A}_k^i} \mu_j^k \right) (\|p_k\|c - \beta) \leq -s \left( \min_{i=1,\ldots,N} \bar{\mu}_i/2M_0 \right) \hat{r}_j(u_k).
\]

Finally, it remains to estimate the difference of the observations associated to \( \hat{u}_{k,s} \) and \( u_k \), respectively. Again, set \( \tilde{u}_i^l := \tilde{v}_i^l \hat{r}_i \). For this purpose we note

\[
\|K(u_k - \hat{u}_{k,s})\|_Y \leq \sum_{j \in \hat{A}_k^i} \mu_j^k \|K(v_j^k - \hat{v}_j^k, 0)\|_Y
\]

\[
\leq \sum_{j \in \hat{A}_k^i} \mu_j^k \left( \|K(v_j^k - \hat{u}_i^l, 0)\|_Y + \|K(\hat{u}_i^l - \hat{v}_j^k, 0)\|_Y \right)
\]

\[
\leq \sum_{j \in \hat{A}_k^i} \mu_j^k \left( |t_j^k - \hat{t}_l| + |v_j^k - \hat{v}_l|_{\mathbb{R}^d} + |\hat{v}_j^k - \hat{t}_l| + |p_k(\hat{t}_k)/\|p_k\|c - \hat{v}_l|_{\mathbb{R}^d} \right)
\]

\[
\leq sc \left( 1 + \left( \sum_{j \in \hat{A}_k^i} \mu_j^k \right) \right) \sqrt{\hat{r}_j(u_k)}
\]

where Lemma 5.6 is used in the third inequality and Lemma 5.12 as well as Proposition 5.13 in the final one. Again pointing out that \( \sum_{j \in \hat{A}_k^i} \mu_j^k \) is uniformly
bounded independently of $\hat{\gamma}$ and $k \in \mathbb{N}$, see Lemma 5.11, we finally arrive at
$$\|K(u_k - \hat{u}_{k,s})\|_Y^2 \leq s^2 c \hat{r}_j(u_k)$$
and thus
$$\hat{r}_j(u_{k+1}) - \hat{r}_j(u_k) \leq \left(s^2 c - s \left( \min_{i=1,\ldots,N} \frac{\mu_i}{2M_0} \right) \right) \hat{r}_j(u_k).$$
Minimizing both sides w.r.t $s \in [0, 1]$ yields the desired result.

Using this improved descent estimate, we prove the linear convergence of the auxiliary residual $\hat{r}_j(u_k)$.

**Theorem 5.16**: Let Assumptions 3.1–5.3 hold. Then there is $\zeta \in (0, 1)$ such that
$$r_j(u_k) \leq \hat{r}_j(u_k) \leq c \zeta^k$$
for all $k \in \mathbb{N}$ large enough.

**Proof**: According to Lemma 5.15 there is $K \in \mathbb{N}$ such that
$$\hat{r}_j(u_{k+1}) \leq \min_{s \in [0,1]} \left[ (1 + s^2 c_1 - sc_2) \hat{r}_j(u_k) \right] \quad \forall \ k \geq K$$
where we set
$$c_2 := \left( \min_{i=1,\ldots,N} \frac{\mu_i}{2M_0} \right)$$
for abbreviation. Explicitly calculating the minimum reveals
$$\min_{s \in [0,1]} \left( 1 + s^2 c_1 - sc_2 \right) \leq \zeta := 1 - \frac{c_2}{2} \min \left\{ 1, \frac{c_2}{2c_1} \right\}$$
and thus
$$r_j(u_k) \leq \hat{r}_j(u_{k+1}) \leq \zeta^{k-K} \hat{r}_j(u_K)$$
for all $k \geq K$.

**Linear convergence of the iterates**

In this last subsection, we aim to quantify the strict convergence of $u_k$ towards $\bar{u}$. More in detail we utilize Theorem 5.16 to prove
$$\|u_k - \bar{u}\|_{L^1} + \|u'_k\|_{\mathcal{M}} - \|\bar{u}'\|_{\mathcal{M}} \leq c \zeta_2^k$$
for some $\zeta_2 \in (0, 1)$ and all $k \in \mathbb{N}$ large enough. For this purpose, we rely on the following auxiliary estimates.
Lemma 5.17: For all $k \in \mathbb{N}$ large enough there holds
\[
\| u_k' \|_M - \| \bar{u}' \|_M \leq c \sqrt{r_j(u_k)} + \sum_{i=1}^{N} \left| \sum_{j \in A_k^i} \mu_k^j - \bar{\mu}_i \right|
\]

Proof: Recall the definition of the restricted measures $U'_{k,i}$ from (28). Then there holds
\[
\| u_k' \|_M - \| \bar{u}' \|_M \leq \sum_{i=1}^{N} \left[ \| U'_{k,i} \|_M - \sum_{j \in A_k^i} \mu_k^j \right] + \sum_{j \in A_k^i} \left| \mu_k^j - \bar{\mu}_i \right|
\]

Now, fix an arbitrary $i \in \{1, \ldots, N\}$. Given two indices $j_1, j_2 \in A_k^i$ we note that
\[
\| \mu_k^{j_1} u_k^{j_1} + \mu_k^{j_2} u_k^{j_2} \|_M = \| \mu_k^{j_1} u_k^{j_1} + \mu_k^{j_2} u_k^{j_2} \|_{\mathbb{R}^d}
\]

if $t_{j_1}^k \neq t_{j_2}^k$ and
\[
\| \mu_k^{j_1} u_k^{j_1} + \mu_k^{j_2} u_k^{j_2} \|_M = \| \mu_k^{j_1} u_k^{j_1} + \mu_k^{j_2} u_k^{j_2} \|_{\mathbb{R}^d}
\]

if $t_{j_1}^k = t_{j_2}^k$. Similarly, we conclude the existence of a partition of $A_k^i$ into pairwise disjoint, nonempty sets $I_h^k$, $h = 1, \ldots, n_k$, with
\[
\| U'_{k,i} \|_M - \sum_{j \in A_k^i} \mu_k^j \leq \sum_{h=1}^{n_k} \left[ \sum_{j \in I_h^k} \mu_k^j \right] - \sum_{j \in I_h^k} \bar{\mu}_i \leq \sum_{h=1}^{n_k} \left[ \sum_{j \in I_h^k} \mu_k^j \right] v_k^j - \bar{v}_i \leq c \sqrt{r_j(u_k)}
\]

where we use the inverse triangle inequality in the second inequality and
\[
\sum_{h=1}^{n_k} \sum_{j \in I_h^k} \mu_k^j | v_k^j - \bar{v}_i | \leq \sum_{j \in A_k^i} \mu_k^j | v_j^k - \bar{v}_i | \leq c \sqrt{r_j(u_k)}
\]
as well as Lemma 5.12 in the final inequality. Summarizing all previous observations and noting that the index $i$ was chosen arbitrarily finishes the proof. ■
A similar estimate holds for the $L^1$ distance of the iterates to the minimizer $\tilde{u}$.

**Lemma 5.18:** Define constants

$$\tilde{C} = -\frac{1}{T} \int_0^T \int_0^s d\tilde{u'} ds + a_{\tilde{u}}, \quad C^k = -\frac{1}{T} \int_0^T \int_0^s d\tilde{u'} ds + a_{u_k}. \quad (35)$$

For all $k \in \mathbb{N}$ large enough there holds

$$\|u_k - \tilde{u}\|_{L^1} \leq c \sqrt{\tau_j(u_k)} + T|C^k - \tilde{C}| + \sum_{i=1}^N \left| \sum_{j \in A_k^i} \mu^k_j \chi^{k}_{j} - \bar{\mu}_i \chi^i \right|_{L^1}.$$ 

**Proof:** According to the definition of the $B$ operator, (3), we have

$$\tilde{u} = \bar{C} + \sum_{i=1}^N \bar{\mu}_i \bar{v}_i \chi^i, \quad u_k = C^k + \sum_{i=1}^N \sum_{j \in A_k^i} \mu^k_j \chi^{k}_{j},$$

and thus

$$\|u_k - \tilde{u}\|_{L^1} \leq T|C^k - \tilde{C}| + \sum_{i=1}^N \left| \sum_{j \in A_k^i} \mu^k_j \chi^{k}_{j} - \bar{\mu}_i \chi^i \right|_{L^1}.$$ 

Now fix an arbitrary index $i \in \{1, \ldots, N\}$. We estimate

$$\left\| \sum_{j \in A_k^i} \mu^k_j \chi^{k}_{j} - \bar{\mu}_i \bar{v}_i \chi^i \right\|_{L^1} \leq T \left\| \sum_{j \in A_k^i} \mu^k_j - \bar{\mu}_i \right\|_{L^1} + \left\| \sum_{j \in A_k^i} \mu^k_j \chi^{k}_{j} - \bar{\mu}_i \chi^i \right\|_{L^1}$$

$$\leq T \left\| \sum_{j \in A_k^i} \mu^k_j - \bar{\mu}_i \right\|_{L^1} + \sum_{j \in A_k^i} \| \mu^k_j \chi^{k}_{j} - \bar{\nu}_i \chi^i \|_{L^1}$$

using that

$$\| \bar{v}_i \chi^i \|_{L^1} = \int_{\tilde{t}_i}^T |\bar{v}_i|_{\mathbb{R}^d} ds = (T - \tilde{t}_i) \leq T.$$

Moreover, from (A2) and (29), we conclude

$$\sum_{j \in A_k^i} \mu^k_j \| \chi^{k}_{j} - \bar{\nu}_i \chi^i \|_{L^1} \leq \sum_{j \in A_k^i} \mu^k_j \left( |t^k_j - \tilde{t}_i| + T|\chi^{k}_{j} - \bar{\nu}_i \chi^i|_{\mathbb{R}^d} \right) \leq c \sqrt{\tau_j(u_k)}.$$ 

Summarizing all previous observations yields the desired estimate. ■
Thus, to prove (34), it suffices to quantify the error $C_k - \tilde{C}$ as well as the difference between $\sum_{j \in A^i_k} \mu^k_j$ and $\tilde{\mu}_i$. This is done in the following proposition.

**Proposition 5.19:** For all $k \in \mathbb{N}$ large enough there holds

$$|\tilde{C} - C^k| + \left| \sum_{j \in A^i_k} \mu^k_i - \tilde{\mu}_i \right| \leq c \sqrt{r_j(u_k)}.$$

**Proof:** Define $\tilde{u}_k = C_k + \sum_{i=1}^N (\sum_{j \in A^i_k} \mu^k_j) \bar{\nu}_i \bar{\chi}_i$, as well as the vector of lumped coefficients $\tilde{\mu}^k \in \mathbb{R}^N$, $\tilde{\mu}^k_i = \sum_{j \in A^i_k} \mu^k_j$. Recall the definition of the injective operator $\hat{K}$ from the proof of Corollary 5.4. Then $\hat{K}(\tilde{\mu}^k - \tilde{\mu}, C^k - \tilde{C}) = K(u_k - \tilde{u})$ and thus

$$|\tilde{C} - C^k| + \left| \sum_{j \in A^i_k} \mu^k_i - \tilde{\mu}_i \right| \leq c \|K(\tilde{u}_k - \tilde{u})\|_Y.$$

Applying Proposition 5.3 yields

$$\|K(\tilde{u}_k - \tilde{u})\|_Y \leq \|K(u_k - \tilde{u})\|_Y + \|K(\tilde{u}_k - u_k)\|_Y \leq \sqrt{r_j(u_k)}/\gamma_0 + \|K(\tilde{u}_k - u_k)\|_Y.$$

Finally, we estimate

$$\|K(\tilde{u}_k - u_k)\|_Y \leq \sum_{i=1}^N \sum_{j \in A^i_k} \mu^k_j \|K(\nu^k_j - \tilde{u}_i, 0)\|_Y \leq c \sum_{i=1}^N \sum_{j \in A^i_k} \mu^k_j \left( |t^k_j - \bar{t}_i| + |\nu^k_j - \bar{\nu}_i|_{\mathbb{R}^d} \right) \leq c \sqrt{r_j(u_k)}$$

using Lemma 5.6 in the second inequality and Lemma 5.12 in the final one. □

Combining the previous results, we are in the position to prove linear convergence of $u_k$ with respect to the strict topology on $\text{BV}(I; \mathbb{R}^d)$.

**Theorem 5.20:** Let Assumptions 5.1, 5.2, 5.3 hold. Then we have

$$\|u_k - \tilde{u}\|_{L^1} + \|u'_k\|_{M} - \|\tilde{\nu}'\|_{\mathcal{M}} \leq c \zeta^k_2$$

for some $\zeta_2 \in (0, 1)$ and all $k \in \mathbb{N}$ large enough.

**Proof:** The statement directly follows from Lemmas 5.17 and 5.18 taking Proposition 5.19 into account. □
6. Numerical examples

The last section is devoted to the numerical illustration of our theoretical results. For this purpose two examples are discussed. First we address the inverse problem of identifying a piecewise constant signal from finitely many data samples. The forward operator \( K \) is modelled by convolution with a Gaussian kernel. Second we consider an optimal control problem for the linear wave equation. Here the control enters as the time-dependent signals of two spatially fixed actuators. In this case, the fidelity term is given by the \( L^2 \)-misfit between the solution to the wave equation and a desired state \( y_d \) over the whole space-time cylinder.

6.1. Deconvolution from finitely many measurements

As a first example consider

\[
\min_{u \in BV(I)} j(u) := \left[ \frac{1}{2} \sum_{i=1}^{9} (k(\rho_i) * u - y_d^i)^2 + \beta \|u'\|_{\mathcal{M}} \right]
\] (36)

where \( I = (0, 1) \), \( y_d \in \mathbb{R}^9 \) is a given finite-dimensional data vector and

\[
k(\rho_i) * u = \frac{1}{\sqrt{2\pi\sigma}} \int_0^1 u(t) e^{-\frac{(t-\rho_i)^2}{2\sigma^2}} \, dt, \quad \rho_i = i \cdot 0.1, \quad i = 1, \ldots, 9.
\]

The deconvolution problem (36) can be embedded in the general setting \((P)\) by choosing

\[
Y = \mathbb{R}^9, \quad F(\cdot) = \frac{1}{2} \sum_{i=1}^{9} ((\cdot)_i - y_i)^2, \quad (Ku)_i = k(\rho_i) * u.
\]

In this case, the \( K = K \circ B \) operator is given by

\[
K(q, c)_i = (k(\rho_i), B(q, c))_{L^2} = -\langle \psi_i, q \rangle + \psi_i(1)(c + \langle t, q \rangle), \quad i = 1, \ldots, 9,
\]

for \( \psi_i \) defined as

\[
\psi_i(t) = \frac{1}{2} \left( \text{erf} \left( \frac{t - \rho_i}{\sqrt{2\sigma}} \right) + \text{erf} \left( \frac{\rho_i}{\sqrt{2\sigma}} \right) \right)
\]

where \( \text{erf}(\cdot) \) denotes the error function. Moreover, we verify \( K^*: \mathbb{R}^m \to C^2(I) \) and

\[
p_k(\cdot) = \int_0^1 K^* \nabla F(Ku_k) = \sum_{i=1}^{9} \psi_i(\cdot)(k(\rho_i) * u_k - y_d^i).
\]

Now, note that the practical realization of Algorithm 1 requires the computation of a global maximum of \(|p_k(\cdot)|\). Since the norm of the dual variable is, in general, nonconcave, this might be computationally expensive or even intractable. For
Figure 1. Ground truth, reconstruction and dual variable. (a) Ground truth $u^\dagger$ and $\bar{u}$. (b) Second derivative $\bar{p}''$ and (c) Dual variable $\bar{p}$.

this reason, we resort to the following heuristic: More in detail, we find solutions of $p_k'(t) = 0$ using a Newton method starting at equally spaced points $t_i^0 = i \cdot 0.1$, $i = 1, \ldots, 9$ . Then $\hat{t}_k$ is chosen from the set of computed solutions by comparing the corresponding function values. Similar approaches have been used for minimization problems in spaces of measures, see, e.g. [16]. While this does obviously not ensure the global optimality of $\hat{t}_k$, the approach seems to work well in practice. In particular, at least for the present example, comparing the computed jump position $\hat{t}_k$ with a plot of $p_k$ reveals its global optimality. The solution of the finite-dimensional subproblems relies on a semismooth Newton method for the ‘normal map’ reformulation of its first-order sufficient optimality conditions, see e.g. [26]. In each iteration, the method is warmstarted using the current magnitudes $\mu_i^k$ and the mean value $c^k$ to construct a good starting point. Moreover, we further enhance its practical performance by incorporating a heuristic globalization strategy based on damped Newton steps.

6.1.1. Structural assumptions on $\bar{p}$

We considered (36) for $\beta \approx 10^{-5}$ and observations $y_d = Ku^\dagger + \zeta$ where $u^\dagger \in \text{BV}(I)$ and $\zeta \in \mathbb{R}^9$ is a random perturbation. Since no analytic solution for Problem (36) is available, we choose $\bar{u} = u_k$ as a reference, where $\bar{k} \geq 1$ is the smallest iteration index in Algorithm 1 with $\Phi_{\bar{k}} \leq 10^{-13}$ and

$$\Phi_{\bar{k}} := M_{\bar{k}}(\|p_{\bar{k}}\|c - \beta), \quad M_{\bar{k}} = F(Ku_{\bar{k}}) + \beta \sum_{i=1}^{#A_{\bar{k}}} \mu_i^{\bar{k}}.$$

According to Theorem 5.2, this implies $0 \leq r_j(u_{\bar{k}}) \leq 10^{-13}$. The ground truth $u^\dagger$ and $\bar{u}$ are depicted in Figure 1(a). Before addressing the performance of Algorithm 1 we numerically verify Assumptions 5.2 and 5.3. For this purpose, we plot the dual variable $\bar{p}$ as well as its second derivative $\bar{p}''$ in Figure 1(b,c). The functional values corresponding to the jumps of $\bar{u}$ are marked by red crosses.

First we point out that $\|\bar{p}\|c = \beta$ and $\bar{p}$ achieves its global maximum/minimum in three distinct points $\{\hat{t}_i\}_{i=1}^3$ which coincide with the jumps of
In particular, the optimal magnitudes satisfy $\bar{\mu}_i > 0$. Moreover, the operator $\hat{K}$ from the proof of Corollary 5.4 has full rank which is equivalent to the linear independence of (25), i.e. Assumption 5.2 holds. Second, there holds $\bar{p}''(\bar{t}_i) \neq 0$. Hence, see Remark 5.2, the quadratic growth condition of Assumption 5.3 holds.

### 6.1.2. Practical performance of algorithm 1

In order to assess the performance of Algorithm 1 we plot the residuals $r_j(u_k)$ alongside the sublinear convergence rate from Theorem 5.2 as well as a linear rate with $\zeta = 0.33$ in Figure 2(a). Next to it, in Figure 2(b), we report on the convergence of the iterates $u_k$ in $L_1(I)$ and the norms $\|u'_k\|_{M}$. As predicted by Theorems 5.16 and 5.20 all considered quantities converge at least linearly. Moreover, we plot the evolution of the size of the active set in Figure 2(c). Note that $\#A_k$ is not strictly increasing. This is testament to the efficiency of the pruning step 7. of Algorithm 1 in combination with the full resolution of the subproblem $(P_{A_k})$ in step 5. Finally, we compare Algorithm 1 to the Fast iterative shrinkage-thresholding algorithm (FISTA) from [27,28]. However, in contrast to our proposed method, its practical application to (36) requires a discretization of the interval $(0, 1)$. For this purpose we consider a uniform partition.
of $[0, 1]$ into subintervals $[t_i, t_{i+1}]$, $i = 1, \ldots, N_h - 1$, where $t_0 = 0$ and $t_i = t_{i-1} + h$, else, with $h = 1/(N_h - 1)$. Subsequently, we replace $\text{BV}(I)$ in (36) by the finite-dimensional subspace

$$
\text{BV}_h(I) = \left\{ u \in \text{BV}(I) \mid u = B \left( \sum_{i=1}^{N_h-1} \mu_i^h \delta_{t_i^h}, c \right), \ \mu \in \mathbb{R}^{N_h-1}, \ c \in \mathbb{R} \right\}
$$

and apply FISTA with constant stepsize as described in [27]. Additionally, we also use this comparison to study the behaviour of Algorithm 1 under perturbations and apply it to the discretized problem. In this context, we restrict the search for the new candidate jump position $\hat{t}_k$ in step 3. of Algorithm 1 to the set of nodes of the partition. More in detail we choose $\hat{t}_k \in \{t_i^h\}_{i=1}^{N_h-1}$ such that

$$
|p_k(\hat{t}_k)| = \max_{i=1,\ldots,N_h-1} |p_k(t_i^h)|.
$$

The other steps of the method remain the same. In Figure 3(a) we plot the behaviour of the residual $r_j(u_k) = j(u_k) - \min_{u \in \text{BV}_h(I)} j(u)$ for FISTA and our method with different grid widths $h = 10^{-3}, 10^{-2}, 10^{-1}$. Additionally, we also include Algorithm 1 without discretization in the plot. This is formally denoted by "$h = 0$". In both methods, the same starting point $u_0$ is used. We observe that Algorithm 1 solves the problem on each refinement level in a few iterations while the convergence of FISTA significantly slows down after the first iterations. Moreover, Algorithm 1 exhibits strong mesh-independence, i.e. its convergence is stable w.r.t. to $h$ and is essentially governed by its behaviour on the continuous problem. In contrast, the convergence behaviour of FISTA degenerates as $h$ gets smaller. Let us, however, point out that the per iteration cost of both algorithms is wildly different. In fact, the practical realization of FISTA only requires the computation of one proximal operator per iteration, which can be done analytically, while Algorithm 1 relies on a Newton-based heuristic to determine a global extremum of $p_k$ as well as the full resolution of $(\mathcal{P}_{A_k})$. To respect the different cost per iteration of both methods we also give a comparison in terms of the computational time in Figure 3(b). For this purpose, we plot the convergence history of Algorithm 1 (up to optimality) and of FISTA (first 200 iterations) as a function of time. We observe that the more complicated subproblems in Algorithm 1 do not lead to highly increased computational times. This is, on the one hand, a consequence of the use of an efficient second order optimization scheme for $(\mathcal{P}_{A_k})$ in combination with a warmstart. On the other hand this is also attributed to the observation that the active set size $\#A_k$, and thus the dimension of the subproblems $(\mathcal{P}_{A_k})$, is essentially independent of the underlying discretization. We omit an additional plot showcasing the convergence of $\#A_k$ on the different discretization levels since the resulting curves align themselves with the plot in Figure 2(c).
Figure 4. Optimal control and convergence of relevant quantities. (a) Reference $u^*$ and $\bar{u}$. (b) Norm of dual variable $\bar{p}$. (c) Residual error over $k$. (d) Norm/L1-error over $k$.

6.2. Optimal control of the wave equation

In this section, we apply the proposed method for the solution of a PDE-constraint optimization problem of the form

$$\min_{u \in BV(I;\mathbb{R}^2), y \in L^2(I \times \Omega)} J(u) = \frac{1}{2} \| y - y_d \|_{L^2(I \times \Omega)}^2 + \beta \| u' \|_{M(I;\mathbb{R}^2)}$$

(37)

where the vector-valued control $u$ is connected to the state variable $y$ by a linear wave equation of the form

$$\begin{aligned}
&\partial_t y - \Delta y = u_1(t)\delta_{x_1}(x) + u_2(t)\delta_{x_2}(x) \quad \text{in } I \times \Omega, \\
&\partial_n y = 0 \quad \text{on } (0, T) \times \partial \Omega, \\
&y(0) = 0, \quad \partial_t y(0) = 0, \quad \text{on } \Omega
\end{aligned}$$

(38)

with $x_1 = (0.5, 0.5), x_2 = (-0.5, -0.5), \beta = 10^{-5}, \Omega = (-1,1)^2$ and $T = 1$. The desired state $y_d \in L^2(I \times \Omega)$ is given by $y_d = y_d^* + \zeta$ where $y_d^*$ is the unique
solution of (38) for the reference source \( u^* = (u_1^*, u_2^*) \) given by

\[
\begin{align*}
  u_1^*(t) &= \begin{cases} 
    0.05 & 0 < t \leq 0.25, \\ 
    0.65 & 0.25 < t \leq 0.5, \\ 
    0.15 & 0.5 < t \leq 0.75, \\ 
    0.35 & 0.75 < t \leq 1 
  \end{cases} \\
  u_2^*(t) &= \begin{cases} 
    0.775 & 0 < t \leq 0.25, \\ 
    -0.025 & 0.25 < t \leq 0.5, \\ 
    0.975 & 0.5 < t \leq 0.75, \\ 
    0.275 & 0.75 < t \leq 1 
  \end{cases}
\end{align*}
\]

and \( \zeta \in L^2(I \times \Omega) \), \( \| \zeta \|_{L^2(I \times \Omega)} / \| \nu^* \|_{L^2(I \times \Omega)} = 0.05 \), is a noise term.

Using the \( L^2(I \times \Omega) \) regularity of \( y \) from [29,30], we can eliminate the PDE-constraint by introducing the linear continuous solution operator \( K \in L(L^2(I; \mathbb{R}^2), L^2(I \times \Omega)) \) which maps \( u \) to \( y \). The adjoint operator \( K^* \in L(L^2(I \times \Omega), L^2(I; \mathbb{R}^2)) \) of \( K \) is defined by the mapping \( \phi \mapsto (\Phi(\cdot, x_1), \Phi(\cdot, x_2)) \) where \( \Phi \) is the solution of the corresponding adjoint state equation

\[
\begin{align*}
  \frac{\partial}{\partial t} \Phi - \Delta \Phi &= \phi & \text{in } I \times \Omega, \\
  \frac{\partial}{\partial t} \Phi &= 0 & \text{on } (0,T) \times \partial \Omega, \\
  \Phi(T) &= 0, \quad \frac{\partial}{\partial t} \Phi(T) = 0 & \text{on } \Omega
\end{align*}
\]

for \( \phi \in Y \). The operator \( K^* \) is well defined according to [29,30].

In order to apply Algorithm 1 to (37) we need to discretize the wave equation, and thus the operator \( K \), using a finite element method. For this purpose, consider ansatz and test spaces spanned by products of piecewise linear and continuous functions on a uniform time grid \( 0 = t_0 < t_1 < \cdots < t_{N_t} = 1 \) with \( t_{i+1} = t_i + \tau \), \( \tau > 0 \), in \( I \) and a uniform spatial triangulation of \( \Omega \) with gridsize \( h \). The associated discretization of the forward operator is denoted by \( K_h \). The adjoint equation is discretized consistently. Finally, as in the FISTA comparison in Section 6.1.2, we replace the control space by piecewise constant functions on the time grid

\[
BV_{\tau}(I; \mathbb{R}^2) = \left\{ u \in BV(I; \mathbb{R}^2) \mid u = B \left( \sum_{i=1}^{N_t-1} \mu_i^T \delta_{t_i}, c \right), \mu^T \in \mathbb{R}^{N_t-1}, c \in \mathbb{R}^2 \right\}
\]

and then apply a discretized version of Algorithm 1 to the problem. More in detail, given the current iterate \( u_k \), we first compute the dual variable \( p_k(\cdot) = \int_0^1 K_h^*(K_h u_k - y_d)(s) \, ds \). This requires one solution of the (discretized) state and adjoint equations, respectively. Subsequently, we choose \( \hat{t}_k \in \{ t_i^\tau \}_{i=1}^{N_t-1} \) such that

\[
|p_k(\hat{t}_k)| = \max_{i=1, \ldots, N_t-1} |p_k(t_i^\tau)|.
\]

Note that \( K_h^*(K_h u_k - y_d) \) is piecewise linear and continuous on \( I \). Hence, the required integration to evaluate \( p_k \) in the temporal grid points can be done analytically. The finite dimensional subproblems in step 5. are again solved by a semismooth Newton method. We plot the computed function \( \tilde{u} \) alongside the reference \( u^* \) as well as the norm of the optimal dual variable \( \tilde{p}(\cdot) = \int_0^1 K_h^*(K_h \tilde{u} - \)
$y_d(s) ds$ in Figure 4(a,b). Upon a closer inspection, in contrast to the first example, we now observe local clustering of the jumps of $\tilde{u}$. More in detail, in the vicinity of every jump of the reference function $u_\ast$, $\tilde{u}$ admits two jumps supported on neighbouring grid nodes. Similar discretization effects for sparse deconvolution problems have been observed in [31]. Alongside the optimal control we also report on the convergence history of the residual $r_j(u_k)$, the $L^1$-distance of $u_k$ and $\tilde{u}$ as well as the error of the norms in Figure 4(c,d). Again we observe a linear rate of convergence for all considered quantities.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This work was supported by Horizon 2020 Framework Programme[ERC advanced grant 668998 (OCLOC)].

**ORCID**

Daniel Walter http://orcid.org/0000-0003-3647-0728

**References**

[1] Casas E, Kruse F, Kunisch K. Optimal control of semilinear parabolic equations by BV-functions. SIAM J Control Optim. 2017;55:1752–1788.

[2] Engel S, Vexler B, Trautmann P. Optimal finite element error estimates for an optimal control problem governed by the wave equation with controls of bounded variation. IMA J Numer Anal. 2020;41(4): 2639–2667.

[3] Hafemeyer D, Mannel F, Neitzel I, et al. Finite element error estimates for one-dimensional elliptic optimal control by BV-functions. Math Control Relat Fields. 2020;10:333–363.

[4] Little MA, Jones NS. Generalized methods and solvers for noise removal from piecewise constant signals. I. background theory. Proc R Soc A Math Phys Eng Sci. 2011;467:3088–3114.

[5] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D Nonlin Phenom. 1992;60:259–268.

[6] Huang J, Zhang T. The benefit of group sparsity. Ann Stat. 2010;38:1978–2004.

[7] Barbero I, Sra S. Fast newton-type methods for total variation regularization. In: Proceedings of the 28th international conference on machine learning. ICML; 2011. p. 313–320.

[8] Condat L. A direct algorithm for 1-D total variation denoising. IEEE Signal Process Lett. 2013;20:1054–1057.

[9] Karahanoglu FI, Bayram L, Van De Ville D. A signal processing approach to generalized 1-D total variation. IEEE Trans Signal Process. 2011;59:5265–5274.

[10] Wahlberg B, Boyd S, Annergren M, et al. An ADMM algorithm for a class of total variation regularized estimation problems*. IFAC Proc Vol. 2012;45:83–88.

[11] Clason C, Kunisch K. A duality-based approach to elliptic control problems in non-reflexive Banach spaces. ESAIM Control Optim Calc Var. 2011;17:243–266.
Appendix 1. Proofs for section 5

This section contains the proofs for the auxiliary results in Section 5.
Proof of Lemma 5.1: For every $s \in (0, 1)$ define the auxiliary iterate $u_{k,s} = B(u_{k,s}, \hat{c})$ where

$$u_{k,s} = \mu_s^{k+1} v^k + \mu_s^{k+1} \nu^k,$$

where

$$\mu_s^k := (1 - s) \mu_k^k, \quad \text{and} \quad \mu_s^{k+1} \in \mathbb{R}^{N_k+1}.$$  

Since $u_{k+1}$ is constructed using a minimizing pair of $(\mathcal{P}_{\mathcal{A}})$ we have

$$\tilde{r}_j(u_{k+1}) - \tilde{r}_j(u_k) \leq F(Ku_{k,s}) - F(Ku_k) + \beta \left( \sum_{i=1}^{N_k+1} \mu_s^{k+1} - \sum_{i=1}^{N_k} \mu_i^k \right).$$

By construction, the second term on the right-hand side is equal to

$$\beta \left( \sum_{i=1}^{N_k+1} \mu_s^{k+1} - \sum_{i=1}^{N_k} \mu_i^k \right) = s \beta \left( M_k - \sum_{i=1}^{N_k} \mu_i^k \right).$$

Using a Taylor’s expansion of the first term $F(Ku_{k,s}) - F(Ku_k)$ and utilizing the Lipschitz continuity of $\nabla F$ yields

$$F(Ku_{k,s}) - F(Ku_k) \leq s \langle p_k, u' - M_k \hat{v}^k \rangle + \frac{Ls^2}{2} ||K(u' - \hat{v})||_Y^2.$$  

Finally note that due to $\hat{v}^k = (p_k(\hat{t}_k)/\|p_k\|_C)\xi_{\hat{t}_k}$ with $|p_k(\hat{t}_k)|_\mathbb{R}^d = \|p_k\|_C$ and Proposition 4.1 we have

$$s \langle p_k, u'_{k} - M_k \hat{v} \rangle = s \left( \beta \sum_{i=1}^{N_k} \mu_i^k - M_k \|p_k\|_C \right).$$

Summarizing all previous observations and minimizing w.r.t $s \in [0, 1]$, we arrive at

$$\tilde{r}_j(u_{k+1}) - \tilde{r}_j(u_k) \leq \min_{s \in [0, 1]} \left[ -s M_k \|p_k\|_C - \beta \right] + \frac{Ls^2}{2} \|K(u'_{k} - M_k \hat{v}^k, 0)\|_Y^2.$$  

Now, since $K$ is continuous, there holds

$$\|K(u'_{k} - M_k \hat{v}^k, 0)\|_Y^2 \leq \left( \|K\| \|u'_{k} - M_k \hat{v}^k\|_{\mathcal{A}} + \beta \right)^2 \leq \left( \|K\| (\|u'_{k}\|_{\mathcal{A}} + M_k) \right)^2 \leq (\|K\| 2M_2)^2 \leq 4 \|K\|^2 M_1^2,$$

where $\|K\|$ denotes the operator norm of $K: \mathcal{M}(I; \mathbb{R}^d) \times \mathbb{R}^d \rightarrow Y$ and $u'_{k} \|_{\mathcal{A}} \leq M_k \leq M_1$ is used in the final inequality. Setting $C(\bar{K}, M_1) := 4 \|K\|^2 M_1^2$ finishes the proof.

Proof of Lemma 5.5: Fix $i = 1, \ldots, N$ and $t \in (\bar{t}_i - R, \bar{t}_i + R)$ as well as $\nu \in \mathbb{R}^d$ with $|\nu|_{\mathbb{R}^d} = 1$. From Assumption 5.3 and $|\nu|_{\mathbb{R}^d} = 1$, we get

$$\beta \langle \bar{p}_t, \nu \delta_t \rangle \geq \beta \geq \bar{p}_t |t - \bar{t}_i|^2.$$  

We further estimate

$$\beta \left( 1 - \langle \bar{p}_t/\beta, \nu \delta_t \rangle \right) = \frac{\beta}{2} \left( 2 - 2(\bar{p}_t/\beta, \nu)_{\mathbb{R}^d} \right) \geq \frac{\beta}{2} \left( |\nu|_{\mathbb{R}^d}^2 - 2(\bar{p}_t/\beta, \nu)_{\mathbb{R}^d} + |\bar{p}_t/\beta|_{\mathbb{R}^d}^2 \right) = \frac{\beta}{2} \left( |\bar{p}_t/\beta|_{\mathbb{R}^d}^2 - |\nu|_{\mathbb{R}^d}^2 \right)$$

using $|\bar{p}_t/\beta|_{\mathbb{R}^d}^2 \leq 1$ in the first inequality. Finally, we have

$$|\bar{p}_t - \bar{p}_t(s)|_{\mathbb{R}^d} = \int_{t_i}^t K^* \nabla F(K\tilde{u})(s) \, ds \leq |t - \bar{t}_i| \|K^* \nabla F(K\tilde{u})\|_{L^\infty(I; \mathbb{R}^n)}. \quad (A1)$$

The claimed statement now follows from noting that

$$|t - \bar{t}_i|^2 + |\nu - \tilde{v}_i|_{\mathbb{R}^d}^2 \leq |t - \bar{t}_i|^2 + (|\nu - \tilde{v}_i|_{\mathbb{R}^d})^2.$$
\[
|t - \bar{t}_i|^2 + (|v - \bar{p}(t)/\beta|_{\mathbb{R}^d} + |(\bar{p}(t) - \bar{p}(\bar{t}_i))/\beta|_{\mathbb{R}^d})^2
\]
\[
\leq |t - \bar{t}_i|^2 + 2 (|v - \bar{p}(t)/\beta|_{\mathbb{R}^d} + |(\bar{p}(t) - \bar{p}(\bar{t}_i))/\beta|_{\mathbb{R}^d})^2
\]
where \(\bar{v}_i = \bar{p}(\bar{t}_i)/\beta\) is used in the first inequality and Young’s inequality is used in the second one.

**Proof of Lemma 5.6:** W.l.o.g assume that \(t_2 \geq t_1\). Using the additional regularity of \(K^*\) from Assumption 5.3, we get
\[
\|K(v_1\delta_{t_1} - v_2\delta_{t_2}, 0)\|_Y = \sup_{\|y\|_1} (\langle K(v_1\delta_{t_1} - v_2\delta_{t_2}, 0), y \rangle) \leq \sup_{\|y\|_1} \|K^*y\|_{L^\infty} \|B(v_1\delta_{t_1} - v_2\delta_{t_2}, 0)\|_{L^1},
\]
Now, recall that
\[
B(v_2\delta_{t_2}, 0) = v_i\chi_{t_i} - \frac{1}{T} \int_0^T v_i\chi_{t_i}(s) \, ds = v_i\chi_{t_i} - \frac{1}{T} v_i(T - t_i),
\]
i = 1, 2, and thus
\[
\|B(v_1\delta_{t_1} - v_2\delta_{t_2}, 0)\|_{L^1} \leq \|v_1\chi_{t_1} - v_2\chi_{t_2}\|_{L^1} + |v_1(T - t_1) - v_2(T - t_2)|_{\mathbb{R}^d}.
\]
The proof is finished noting that
\[
\|v_1\chi_{t_1} - v_2\chi_{t_2}\|_{L^1} \leq \|(v_1 - v_2)\chi_{t_2}\|_{L^1} + \|v_1\chi_{t_1} - \chi_{t_2}\|_{L^1}
\]
\[
= \int_{t_2}^T |v_1 - v_2|_{\mathbb{R}^d} \, ds + \int_{t_1}^{t_2} |v_1|_{\mathbb{R}^d} \, ds = (T - t_2)|v_1 - v_2|_{\mathbb{R}^d} + |t_1 - t_2|
\]
\[
\leq T|v_1 - v_2|_{\mathbb{R}^d} + |t_1 - t_2| \quad (A2)
\]
as well as
\[
|v_1(T - t_1) - v_2(T - t_2)|_{\mathbb{R}^d} \leq |v_1|_{\mathbb{R}^d}|t_1 - t_2| + |T - t_2||v_1 - v_2|_{\mathbb{R}^d} \leq |t_1 - t_2| + T|v_1 - v_2|_{\mathbb{R}^d}.
\]

**Proof of Proposition 5.10:** Let \((\mu_j^k, v_j^k) = (\mu_j^k, v_j^k\delta_{i_j}) \in A_k\) be arbitrary. Utilizing the first-order optimality condition for the subproblem \((\mathcal{P}_A_k)\), see Proposition 4.1, we have
\[
\beta = (p_k(t_j^k), v_j^k)_{\mathbb{R}^d} \leq |p_k(t_j^k)|_{\mathbb{R}^d}.
\]
Thus, together with Corollary 5.9, we conclude \(t_j^k \in (\bar{t}_i - R, \bar{t}_i + R)\) for exactly one \(i \in \{1, \ldots, N\}\). The existence of the index sets \(A_k^i\) is now imminent.

**Proof of Lemma 5.11:** Let \(i = 1, \ldots, N\) be arbitrary but fixed and let \(\chi \in C_0(I)\) be such that \(\chi(t) = 1, t \in (\bar{t}_i - R, \bar{t}_i + R)\), as well as \(\chi(t) = 0, t \in (\bar{t}_j - R, \bar{t}_j + R), j \neq i\). Moreover, denote by \(\nu \in C_0(I; \mathbb{R}^d)\) an arbitrary test function. Then we have \(\chi \nu \in C_0(I; \mathbb{R}^d)\) and thus
\[
\langle \nu, U'_{k,i} \rangle = \langle \chi\nu, U'_{k,i} \rangle \to \langle \chi\nu, \bar{u} \rangle = \langle \nu, \bar{u}_i \rangle
\]
due to \(u_k' \rightharpoonup^* \bar{u}_i\), see Theorem 5.2. Consequently \(U'_{k,i} \rightharpoonup^* \bar{u}_i\). Similarly, we conclude
\[
\beta \sum_{j \in A_k^i} \mu_j^k = \langle \chi p_k, u_k' \rangle = \langle \bar{p}, \bar{u}_i \rangle = \beta \bar{u}_i
\]
using the first-order optimality conditions for $u_k$ and $\bar{u}$, see Proposition 4.1 and Theorem 3.4, respectively, as well as $p_k \to \bar{p}$ in $C_0(I; \mathbb{R}^d)$, see Proposition 5.3. Thus, $A_k^i \neq \emptyset$ for all $k \in \mathbb{N}$ large enough. The last statement now follows due to

$$\sum_{i=1}^{N_k} \mu_i^k = \sum_{i=1}^{N} \sum_{j \in A_i^k} \mu_j^k.$$ 

**Proof of Lemma 5.14:** Note that

$$u_k' - \tilde{u}_{k,s} = s \sum_{j \in A_k^i} \mu_j^k v_j - s \left( \sum_{j \in A_k^i} \mu_j^k \right) \bar{v}^k$$

and thus

$$\langle p_k, u_k' - \tilde{u}_{k,s} \rangle = s \left( \sum_{j \in A_k^i} \mu_j^k \langle p_k, v_j \rangle - s \left( \sum_{j \in A_k^i} \mu_j^k \right) \langle p_k, \bar{v}^k \rangle \right) = -s \left( \sum_{j \in A_k^i} \mu_j^k \right) \left( \|p_k\|_C - \beta \right)$$

using that $\langle p_k, v_j^k \rangle = \beta$, see Proposition 4.1, and $\langle p_k, \bar{v}^k \rangle = \|p_k\|_C$. The statement on $\sum_{j=1}^{N_k+1} \gamma_j^k$ is imminent.