Smoothing for signals with discontinuities using higher order Mumford-Shah models

Martin Storath∗ Lukas Kiefer‡ Andreas Weinmann‡§

Abstract

Minimizing the Mumford-Shah functional is frequently used for smoothing signals or time series with discontinuities. A significant limitation of the standard Mumford-Shah model is that linear trends – and in general polynomial trends – in the data are not well preserved. This can be improved by building on splines of higher order which leads to higher order Mumford-Shah models. In this work, we study these models in the univariate situation: we discuss important differences to the first order Mumford-Shah model, and we obtain uniqueness results for their solutions. As a main contribution, we derive fast minimization algorithms for Mumford-Shah models of arbitrary orders. We show that the worst case complexity of all proposed schemes is quadratic in the length of the signal. Remarkably, they thus achieve the worst case complexity of the fastest solver for the piecewise constant Mumford-Shah model (which is the simplest model of the class). Further, we obtain stability results for the proposed algorithms. We complement these results with a numerical study. Our reference implementation processes signals with more than 10,000 elements in less than one second.

Keywords: piecewise smooth approximation, discontinuous signals, complexity penalized estimation, changepoint estimation, segmented least squares, spline smoothing, Mumford-Shah model, Potts model, Blake-Zisserman model.

AMS subject classification (MSC2010): 65D10, 65K05, 62G08, 65K10, 65D07.

1 Introduction

Smoothing is an important processing step when working with measured signals or time series. For signals without discontinuities, it is standard to use smoothing splines for this task. In various applications however, the signals possess discontinuities. Such applications are, for example, the cross-hybridization of DNA [19, 30, 51], the reconstruction of brain stimuli [65], single-molecule fluorescence resonance energy transfer [33], cellular ion channel functionalities [29], photo-emission spectroscopy [23] and the rotations of the bacterial flagellar motor [52]. (Further examples can be found in [23, 38, 39] and the references therein.) Frequently, it is important to preserve the discontinuities since they typically indicate a significant change. Unfortunately the locations of the discontinuities are in general unknown; they have to be estimated along with the signal. A slightly different point of view of the task is approximating the data by a piecewise smooth function whose segment boundaries are unknown.

In the literature, various approaches for this task have been proposed. There are schemes which estimate the discontinuities locally and adapt the corresponding fitting operator to the local situation in an explicit way; for instance [1, 3, 27]. Other approaches use variational methods: an energy functional is considered and a corresponding minimizer yields a smoothed approximation to the
data. Examples for discontinuity preserving methods are total variation (TV) or Rudin-Osher-Fatemi (ROF) models [50]. They may be thought of as a kind of piecewise constant approximation. In recent years, there has been a lot of research on higher order TV type methods to take linear and higher order trends in the data into account [16]; in particular, total generalized variation (TGV) approaches [11] have turned out to be effective for this task. ROF models and their higher order extensions, however, do not explicitly incorporate the notion of segments and boundaries. A model taking these notions explicitly into account is the Potts model [25, 44, 63] which is also known as piecewise constant Mumford-Shah model [42]; its two-phase variant is also known as Chan-Vese model [17]. It estimates a discontinuity set and a corresponding piecewise constant function which optimally approximates the data; the discontinuity set is not fixed a priori but chosen by the variational model. However, in various situations there are natural small variations in the data which are not caused by noise, and therefore require a signal model which is more general than piecewise constant functions. The celebrated Mumford-Shah model [8, 41, 42] addresses this need by allowing for piecewise smooth signal estimates. Mumford-Shah and Potts models have been used in signal and image processing [8, 25, 41, 63] for a long time. Their major fields of application are discontinuity preserving smoothing as well as segmentation. They also have been successfully used for stabilizing the reconstructions of inverse problems, for instance for electrical impedance tomography [49]. The regularizing properties (in the sense of inverse problems) of these functionals have been studied in [32, 45, 46, 48]. The discretization of these models has been studied in [14, 15]. As related work, we also mention [4] and [21] where existence results on minimizers are established, and [60] where versions for manifold-valued data spaces been investigated. We further note that besides $\ell^2$-based energies also $\ell^p$-based variants with $p \geq 1$ have been considered [20, 24, 28, 36, 53, 59]. Algorithmic approaches to the multivariate Mumford-Shah and Potts problems include, for instance, approaches via graduated non-convexity [8, 9], approximation by elliptic functionals [2, 6, 49], graph cuts [10], active contours [56], convex relaxations [55], iterative thresholding algorithms [20], and ADMM splitting schemes [28, 54]. Due to the NP hardness of the Mumford-Shah problem and the Potts problem in the multivariate case, all the above mentioned algorithms are approximative strategies in the multivariate case; in the univariate case dynamic programming strategies yield exact solutions as discussed in more detail later on.

A significant limitation of the classical Mumford-Shah model is that it does not well preserve locally linear or polynomial trends in the data. Instead, it tends to produce spurious discontinuities when the slope of the signal is too high, which has been termed the “gradient limit effect” by Blake and Zisserman [8]. The reason for this is that it penalizes deviations from a piecewise constant spline. The preservation of linear or polynomial trends can be accomplished by passing to splines of higher order. This leads to considering higher order Mumford-Shah models which – instead of penalizing the deviation from a piecewise constant function – penalize the deviation from a piecewise polynomial. The multivariate discrete situations are particularly interesting in image processing for edge preserving smoothing of images with locally linear or polynomial trends. The univariate discrete situations are particularly interesting when smoothing time series data which, as a matter of fact, are discrete and univariate, and which frequently arise in connection with biological applications such as [43]. We note that such univariate problems also arise as basic building blocks when applying an ADMM splitting to multivariate higher order Mumford-Shah functionals in analogy to [28, 54]. Further, they appear in majorization-minimization approaches for edge-preserving regularization of inverse problems [58].

We intend to systematically study higher order Mumford-Shah models, where, in this paper, we consider the univariate and discrete situation. It is given by the minimization problem

$$\left( u^*, I^* \right) = \arg\min_{u \in \mathbb{R}^N, I} \|u - f\|_2^2 + \sum_{I \in \mathcal{I}} \beta^k \|\nabla^k u_I\|_2^2 + \gamma |I|, \quad (P_{k,\beta,\gamma})$$

Here, $f \in \mathbb{R}^N$ denotes the given data, and the minimum is computed with respect to the target variables $u$ and $I$, where $u$ is a discrete univariate signal of length $N$ and $I$ is a partition of the domain $\Omega = \{1, \ldots, N\}$. (The connection between $u$ and $I$ is discussed in detail later in Section 2.2.)
The symbol \( \nabla^ku_I \) denotes the \( k \)-th order finite difference operator applied to the vector \( u \) restricted to the “interval” \( I \) of a partition \( I \) of the domain. The functional value comprises a cost term for the data deviation, a cost term for the inner energy of a spline on the single segments, and a cost term for the complexity of the partition. The minimizing signal \( u^* \) is a piecewise \( k \)-th order discrete spline approximation to \( f \) with elasticity parameter \( \beta \) which has discontinuities or breakpoints at the boundaries given by the partition \( I^* \). Choosing a large parameter value \( \beta \) leads to stronger smoothing on the segments, and choosing a large parameter value \( \gamma \) leads to less segments. It is interesting to look at the cases for very large parameters of \( \beta \) and \( \gamma \). As the kernel of \( \nabla^k \) consists of polynomials of maximum degree \( k - 1 \), the limit situation \( \beta \to \infty \) can be written as

\[
\arg\min_{u,I} \| u - f \|_I^2 + \gamma|I|, \quad \text{s.t. } u_I \text{ is a polynomial of maximum degree } k - 1 \text{ for all } I \in I. \quad (P_{k,\infty,\gamma})
\]

As the case \( k = 1 \) is widely known as the Potts model we refer to this special case as higher order Potts model. On the other hand, for sufficiently large \( \gamma \), it reduces to the (discrete) \( k \)-th order spline approximation

\[
\arg\min_{u} \| u - f \|_2^2 + \beta|\nabla^ku|_2^2, \quad (P_{k,\beta,\infty})
\]

which is a classical method for smoothing data; see [57, 61]. Depending on the application, there are different points of view for Mumford-Shah-type models: On the one hand, the optimal partition \( I^* \) can serve as a basis for identifying segment neighborhoods [5] and as an indicator for changepoints of the signal [34]. On the other hand, the corresponding optimal signal \( u^* \) can serve as a smoother for a signal with discontinuities [8, 64].

In the literature, the members of the higher order Mumford-Shah family (\( P_{k,\beta,\gamma} \)) have been considered for the special cases \( k = 1, 2 \) and for \( \beta < \infty \) or \( \beta = \infty \) (strict piecewise polynomial model) by various, separate individual studies. The simplest instance, the piecewise constant variant (\( P_{1,\infty,\gamma} \)) denotes the univariate Potts model named after Renfrey Potts [44]. A problem closely related to the Potts problem (\( P_{1,\infty,\gamma} \)) was studied by Bruce [12] in the context of scalar minimum error quantization. Bellman and Roth [7] studied curve fitting by segmented straight lines which corresponds to the strict piecewise linear model (\( P_{2,\infty,\gamma} \)) which we refer to as the affine Potts model or the piecewise affine Mumford-Shah model. Kleinberg and Tardos [35] discuss this approach under the name segmented least squares problem. The first order problems (\( P_{1,\beta,\gamma} \)) for arbitrary parameters \( \beta \) have been studied in the seminal works of Mumford and Shah [41, 42]. (This motivates the denomination higher order Mumford-Shah model for the family (\( P_{k,\beta,\gamma} \)).) At approximately the same time, Blake and Zisserman [8] have studied (\( P_{1,\beta,\gamma} \)) which they called the weak string model. In the same work, they pointed out a shortcoming of the first order model, namely the already mentioned gradient limit effect; this means that the first order model penalizes large slopes and therefore produces spurious extra intervals when steep slopes are present in the signal. Blake and Zisserman [8] also introduced a second order extension, called the weak rod model. This model is more general than the model (\( P_{2,\beta,\gamma} \)) considered here because it has an extra penalty for discontinuities in the first derivative. We refer to [13] for a recent survey on Blake-Zisserman models. To our knowledge, the models (\( P_{k,\beta,\gamma} \)) have not been systematically studied for arbitrary orders \( k \).

The computation of Mumford-Shah models involves minimization of a non-convex cost function. As pointed out above, all these problems are NP hard in the multivariate case. Remarkably, in the univariate case, the situation is different: it is well known that (\( P_{k,\beta,\gamma} \)) can be cast into a partitioning problem which can be solved by dynamic programming; see [5, 7, 9, 24, 31, 64]. The state-of-the-art solver has worst case complexity \( O(N^2\phi(N)) \) where \( \phi(N) \) comprises the costs of computing a spline approximation error on an interval of maximum length \( N \); see [24, 35, 64]. We note that Killick et al. [34] have established a pruning strategy to further accelerate the algorithm in a special yet practically relevant case: if the expected number of segments \( |I^*| \) grows linearly in \( N \) and if the expected log-likelihood fulfills certain estimates, detailed in [34], the expected complexity is \( O_p(N\phi(N)) \). Another pruning scheme has been established in [54]. Further, an algorithm for solving the first order Mumford-Shah problem for all parameters \( \gamma \) simultaneously was proposed
A straightforward approach for computing the approximation errors for the higher order Mumford-Shah problem \((P_{k,\beta,\gamma})\) exploits the band structure of the accompanying linear systems and thereby achieves the cost \(\phi(N) = N\); this leads to an overall worst case complexity of \(O(N^3)\). For the first order problem \((P_{1,\beta,\gamma})\), we proposed an algorithm of \(O(N^2)\) worst case complexity in [28] which utilizes a fast computation scheme for the approximation errors first proposed by Blake [9]. Unfortunately, as that scheme is based on algebraic recurrences, a generalization to arbitrary orders of \(k\) is difficult. The approximation errors associated with the higher order Potts model \((P_{k,\infty,\gamma})\) can be obtained in constant time by precomputations of moments so that \(\phi(N) = 1\); see [24, 37]. Although that approach gives reasonable results for the low orders \(k = 1, 2\), it gets numerically unstable for higher orders. A completely different dynamic programming approach of Viterbi-type was discussed in [8]. However, this approach requires to restrict the range of \(u\) to a finite set of \(R\) different real values a priori. Thus, it does not compute an exact minimizer, as the set is not known in advance; further, as the complexity is \(O(NR^{k+1})\) it has been deemed too expensive in practice for higher orders [8].

To summarize, an important advantage of higher order models is that they rather penalize the deviation from piecewise polynomial signals than the deviation from piecewise constant ones. As such, they improve the estimation of data with linear, quadratic or other polynomial trends. At the same time, higher order Mumford-Shah models are theoretically and practically more involved than the first order ones. In particular, no efficient and stable solvers for the involved non-convex optimization problems are available yet. Remarkably, this statement even holds for the particular instance of the second-order Mumford-Shah problems \((k = 2\), affine situation\)).

### 1.1 Contribution

This work deals with the analysis and with solvers for the higher order Mumford-Shah and Potts problems \((P_{k,\beta,\gamma})\). First, we discuss basic properties of higher order Mumford-Shah models, we prove that the solutions are unique for almost all input data, and we discuss connections with related models. A main contribution is a fast non-iterative algorithm for minimizing the higher-order Mumford-Shah and Potts models of arbitrary order. The proposed schemes are based on dynamic programming and recurrence relations. We prove that the proposed algorithms have the same worst case complexity as the state-of-the-art solver for minimizing the (simpler) piecewise constant Mumford-Shah model, i.e., their runtime grows quadratically with respect to the length of the signal. Our reference implementation processes signals of length over 10,000 in less than one second on a standard desktop computer. Further, we derive stability results for the proposed algorithms. Eventually, we provide a numerical study where we in particular compare with the first order model with respect to runtime and reconstruction quality. An illustration on the smoothing effect of higher order Mumford-Shah models in comparison to classical splines and to first order models is given in Figure 1.

### 1.2 Organization of the paper

In Section 2 we describe and discuss higher order Mumford-Shah and higher order Potts models. In Section 3, we develop a fast solver for higher order Mumford-Shah problems and for higher order Potts problems, and we analyze the stability. A numerical study is given in Section 4.

### 2 Higher order Mumford-Shah and Potts models

We start with some basic notations and definitions. Our goal is to recover an unknown signal \(g \in \mathbb{R}^N\) from its noisy samples

\[
f_n = g_n + \eta_n, \quad n = 1, ..., N,\]

where the \(\eta_n\) are independently distributed Gaussian random variables of zero mean and variance \(\sigma^2\). We write \(l : r\) for a discrete “interval” from \(l\) to \(r\), i.e. \(l : r = [l, l + 1, \ldots r]\). It is convenient to
use the Matlab-type notation \( x_I = x_{i,r} = (x_i, x_{i+1}, \ldots, x_r) \) for indexing. We say that \( I \) is a partition of \( \Omega = 1 : N \) into intervals, if \( I \cap J = \emptyset \) for all \( I, J \in I \), if \( \bigcup_{I \in I} I = \Omega \), and if all \( I \in I \) are discrete intervals of the form \( I = I : r \), with \( 1 \leq I \leq r \leq N \). As we will only work with partitions into intervals here, we briefly call \( I \) a partition. Further, we use the notation \( \|u\|_2^2 = \sum_{n=1}^{N} u_n^2 \) for \( u \in \mathbb{R}^N \) to denote the squared Euclidean length of \( u \).

2.1 First order Mumford-Shah models and the gradient limit effect

Before considering higher order models, we first review some important properties of first order models. The first order Mumford-Shah problem \((\mathcal{P}_{1,\beta,\gamma})\) on the discrete domain \( \Omega \) can be written as

\[
(u^*, I^*) = \arg\min_{u \in \mathbb{R}^N, I \text{ partition of } \Omega} \sum_{n=1}^{N} (u_n - f_n)^2 + \beta^2 \sum_{I \in I} (\|u_I\|_2 - (u_I))^2 + \gamma |I|.
\]  

The two-fold minimization with respect to the signal \( u \) and the partition \( I \) is instructive but cumbersome in practice. We can remove the explicit dependence on the signal \( u \). The formulation in terms of partitions reads

\[
I^* = \arg\min_{I \text{ partition of } 1:N} \sum_{I \in I} (\mathcal{E}^I + \gamma), \quad \text{with } \mathcal{E}^I = \min_{v \in \mathbb{R}^I} \sum_{i=1}^{I} (v_i - f_i)^2 + \beta^2 \sum_{I \in I} (\|v_{I+1} - v_i\|_2^2).
\]

Here \( \mathcal{E}^I \) describes the error of the best first order discrete spline approximation on the interval \( I \). This formulation is typically used for derivation of algorithms based on dynamic programming [9, 24].
Figure 2: Limitations of the classical Mumford-Shah model $\mathcal{P}_{1,\beta,\gamma}$. (a) The reconstruction from Figure 1c with the optimized parameters $\gamma = 0.04$, $\beta = 1.3625$ is displayed. It provides unsatisfactory smoothing and detection of discontinuities. (b) Decreasing the complexity penalty (here $\gamma = 0.02$) leads to more but dislocated discontinuities. (c) Increasing the elasticity parameter (here $\beta = 3$) leads to stronger smoothing, but to spurious segments as well. In either case, the first order model shows the tendency to create spurious discontinuities at steep slopes which is known as the gradient limit effect.

We remark that the first order Mumford-Shah model (2) is typically formulated in terms of the number of “jumps” instead of the number of segments. As the number of jumps is equal to the number of partitions minus 1, the two formulations are related by the additive constant $\gamma$ so that the minimizers are equal.

For the first order model (2), it is possible to rewrite the functional without partitions. The corresponding formulation reads

$$ u^* = \arg\min_{u \in \mathbb{R}^N} \sum_{n=1}^{N} (u_n - f_n)^2 + \sum_{n=1}^{N-1} \min(\beta^2 (u_{n+1} - u_n)^2, \gamma). \quad (4) $$

This formulation is useful for derivation of algorithms based on iterative thresholding techniques [20]. The key property that makes the formulation in terms of $u$ possible is that, for the first order model, the signal $u^*$ and the partition $I^*$ are equivalent in the sense that $I^*$ can be recovered from $u^*$ and vice-versa.

As mentioned in the introduction, a major limitation of the classical Mumford-Shah model is that data with locally linear or polynomial trends are not well approximated. This undesirable effect, known as gradient limit effect, is illustrated in Figure 2. We observe that the solution of the first order model (using model parameters optimized to the $\ell^2$ error) do not catch all discontinuities. We also see that tuning the model parameters towards allowing for more discontinuities leads to spurious discontinuities at steep slopes. This shows that the first order models are not rich enough for dealing with data having locally linear or polynomial trends.

### 2.2 Basic properties of higher order Mumford-Shah models

We denote by $\nabla^k \in \mathbb{R}^{(q-k) \times q}$ the matrix that acts as $k$-th order finite difference on the vector $u_I$, where $q = |I|$. To fix ideas, the matrices $\nabla^k$ are given for $k = 1$ and $k = 2$ by

$$ \nabla = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ & \ddots & \ddots \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(q-1) \times q} \quad \text{and} \quad \nabla^2 = \begin{pmatrix} 1 & -2 & 1 \\ & 1 & -2 & 1 \\ & & \ddots & \ddots \\ & & & 1 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{(q-2) \times q}. $$

For higher orders $k \geq 3$, the row pattern is equal to $t * t * \ldots * t$ which denotes the $k$-fold convolution of the finite difference vector $t = (-1, 1)$ with itself.
Using this notation the higher order Mumford-Shah problem \( (\mathcal{P}_{k,\beta,\gamma}) \) on the discrete domain \( \Omega \) can be written as

\[
(u^*, \mathcal{I}^*) = \operatorname*{argmin}_{u \in \mathbb{R}^\Omega, \mathcal{I} \text{ partition of } \Omega} \sum_{n=1}^{N} (u_n - f_n)^2 + \beta^{2k} \sum_{I \in \mathcal{I}} \sum_{i=1}^{n} (\nabla^k u_i)^2 + \gamma |\mathcal{I}|.
\]

(5)

As mentioned in the introduction, the functional comprises a data penalty term, a smoothness penalty term, and a complexity penalty term. A minimizer \( u^* \) is a \( k \)-th order discrete spline approximation to data \( f \) on each segment of the partition \( \mathcal{I}^* \). The parameter \( \gamma \) determines the penalty for opening a new segment. The parameter \( \beta \) controls influence of the smoothness penalty. The order \( k \) is the derivative order of the (discrete) spline. An important effect is that polynomials of order \( k - 1 \) on a segment do not have any smoothness penalty.

**Formulation as partitioning problem.** As in the first order case, it is convenient to formulate the higher order Mumford-Shah problem (5) in terms of the partition only; it reads

\[
\mathcal{I}^* = \operatorname*{argmin}_{\mathcal{I} \text{ partition of } \Omega} \sum_{I \in \mathcal{I}} (\mathcal{E}^I + \gamma),
\]

(6)

where \( \mathcal{E}^I \) denotes the approximation error of the \( k \)-th order (discrete) smoothing spline on \( I \) given by

\[
\mathcal{E}^I = \min_{v \in \mathbb{R}^\Omega} \|v - f_I\|^2_2 + \beta^{2k} \|\nabla^k v\|^2_2 = \min_{v \in \mathbb{R}^\Omega} \sum_{i=1}^{n} (v_i - (f_I)_i)^2 + \sum_{i=1}^{n} \beta^{2k} (\nabla^k v_i)^2.
\]

(7)

Note that the \( k \)-th order finite difference \( \nabla^k \) is only well defined for vectors of length greater than \( k \), so \( \mathcal{E}^I = 0 \) if \( |I| \leq k \). The minimizing estimate \( u^* \) can be recovered from an optimal partition \( \mathcal{I}^* \) by solving

\[
u^*_I = \operatorname*{argmin}_{v \in \mathbb{R}^\Omega} \|v - f_I\|^2_2 + \beta^{2k} \|\nabla^k v\|^2_2, \quad \text{for all } I \in \mathcal{I}^*.
\]

(8)

Hence, if we have computed an optimal partition \( \mathcal{I}^* \) of the domain \( \Omega \), its accompanying signal estimate \( u^* \) is uniquely determined.

To express the relation between \( u^* \) and \( \mathcal{I}^* \), it is convenient to introduce a formulation in terms of block matrices. A partition \( \mathcal{I} \) defines a block diagonal matrix \( L_I \) by

\[
L_I = \begin{pmatrix}
L_{|\mathcal{I}|_1} & & \\
& L_{|\mathcal{I}|_2} & \\
& & \ddots \\
& & & L_{|\mathcal{I}|_{|\mathcal{I}|}}
\end{pmatrix}, \quad \text{with} \quad L_m = \nabla^k \in \mathbb{R}^{(n-k)\times n}.
\]

(9)

and with \( k \)-th order finite difference matrices \( \nabla^k \) of the appropriate size \( (n-k) \times n \) defined as above. Here, \( |\mathcal{I}|_m \) denotes the cardinality of the \( m \)-th element of the partition \( \mathcal{I} \). If \( n \leq k \), we use the convention that \( L_m \) is an “empty” block of length \( n \). The number of columns of \( L_I \) is equal to \( N \), and the number of rows depends on the size of the intervals with minimum length \( k + 1 \) of the partition, i.e., \( \sum_{m=1}^M \max(|I_m| - k, 0) \). For example, for \( k = 2 \) the partition \( \mathcal{I} = \{(1,\ldots,4), (5,6), (7,\ldots,9)\} \) defines the matrix

\[
L_I = \begin{pmatrix}
1 & -2 & 1 \\
1 & -2 & 1 \\
1 & -2 & 1
\end{pmatrix} \in \mathbb{R}^{3\times 9}.
\]

The block matrix notation (9) allows to formulate the minimization problem (5) in the compact form

\[
\operatorname*{argmin}_{u \in \mathbb{R}^\Omega, \mathcal{I} \text{ partition}} \beta^{2k} \|L_I u\|^2_2 + \|u - f\|^2_2 + \gamma |\mathcal{I}|.
\]

(10)
For a fixed partition $I$, taking derivatives with respect to $u$ reveals that a minimizer $u_I$ satisfies the linear system

$$2\beta^2 L_I^T L_I u_I + 2(u_I - f) = 0. \tag{11}$$

As the system has full rank for all $\beta \geq 0$ we get the unique solution

$$u_I = S_{I,\beta} f, \quad \text{where } S_{I,\beta} = (\beta^2 L_I^T L_I + \text{id})^{-1}. \tag{12}$$

We omit the subscript if the dependence on $I$ or $\beta$ is clear. Plugging (12) into (5) gives the explicit expression for the functional value of the higher order Mumford-Shah functional restricted to the partition $I$ as

$$G'_I(f) = \beta^2 \|L_I S_I f\|_2^2 + \|S_I f - f\|_2^2 + \gamma |I|. \tag{13}$$

Hence, the minimizing partition $I^*$ is given as the minimizing argument of $I \mapsto G'_I(f)$.

In contrast to the first order model, expressing the problem only in terms of $u$ just like in (4) is not feasible. A reason for this is that one solution $u$ may be the result of different partitions with different numbers of segments. A simple example is the data $f = (0, 1, 0)$. For $\gamma < 2/3$ and for $\beta$ sufficiently large, a minimizer is given by $u = f$. The partitions $\{1, 2\}, \{(1), (2, 3)\}$ and $\{(1), (2), (3)\}$ lead to this $u = f$.

Minimum functional values and minimum segment lengths. We record the following basic property about minimizers. Its proof follows an argument similar to the one used in [8] for a continuous domain second order problem.

**Lemma 1.** Let $I^*$ be a minimizing partition of (5). Then the minimal functional value is given by

$$G'_I(f) = \|f\|_2^2 - f^T S_{I,\beta} f + \gamma |I^*|.$$  

**Proof.**

Let $\bar{u} = S_{I,\beta} f$. Expanding the functional yields

$$G'_I(f) = \beta^2 \|L_I \bar{u}\|_2^2 + \|\bar{u} - f\|_2^2 + \gamma |I^*|$$

$$= \beta^2 \bar{u}^T L_I^T L_I \bar{u} + (\bar{u} - f)^T (\bar{u} - f) + \gamma |I^*|$$

$$= \beta^2 \bar{u}^T L_I^T L_I \bar{u} + \bar{u}^T (\bar{u} - f) + f^T (\bar{u} - f) + \gamma |I^*|$$

$$= \bar{u}^T (\beta^2 L_I^T L_I \bar{u} + (\bar{u} - f)) - f^T (\bar{u} - f) + \gamma |I^*|$$

$$= -f^T (\bar{u} - f) + \gamma |I^*| = \|f\|_2^2 - f^T \bar{u} + \gamma |I^*|,$$

where we used the minimality property (11) in the penultimate line. $\square$

Next we show that there is always an optimal partition $I^*$ which has at most one segment with less than $k$ elements:

**Lemma 2.** For each partition $I$ there is a partition $I'$ such that all segments $I' \in I'$ (except possibly the leftmost one) have length greater or equal than $k$ and that

$$\sum_{I \in I} \mathcal{E}^I \leq \sum_{I \in I} \mathcal{E}^l \quad \text{and} \quad |I'| \leq |I|.$$  

In particular $G'_I(f) \leq G'_I(f)$.

**Proof.** Let $I$ be a partition and let $I$ be its rightmost segment such that $|l| < k$. Denote by $i$ the left boundary index of $I$. If $i = 1$ we are done. Otherwise, we transfer the element $i - 1$ from the left neighboring segment to the segment $I$ and denote the partition modified in this way by $I'$. (If the neighboring segment gets empty, we remove it from the partition.) On the one hand, $|I'| \leq |I|$. On the other hand, since $|I| \leq l - 1| l - 1| \leq k$ we have that $\mathcal{E}^I[l-1] = 0$. Repeating the above procedure a finite number of times, we end up with a partition $I''$ whose segments have length greater or equal to $k$, except possibly the leftmost segment. $\square$
2.3 Higher order Potts models

As mentioned in the introduction, the higher order Potts model \((P_{k,\infty,\gamma})\) can be seen as the limit case of the higher order Mumford-Shah model for \(\beta \to \infty\). The main difference to the higher order Mumford-Shah model is that the approximation on a segment is performed by a polynomial of maximum degree \(k - 1\) instead of a \(k\)-th order spline. In consequence, the approximation error on a segment \(I = I : r\) is given by

\[
E^I_\gamma = \min_{v \text{ polynomial of degree } \leq k - 1 \text{ on } I} \|v - f_I^*\|_2^2. \tag{14}
\]

Complementing (7) for the Mumford-Shah problem, (14) is a least squares problem in the coefficients of the polynomial.

On the one hand, higher order Potts models are more restrictive than genuine higher order Mumford-Shah models since they enforce piecewise polynomial solutions. On the other hand, due to the stronger prior, they are more robust to noise. From the computational side, one parameter less has to be determined for the Potts model.

It was observed in [22], that the second order Potts model can be appealingly formulated in terms of the \(L_0\)-"norm" of an affine parameter field. For the higher order Potts model this can be accomplished as follows. Let \(C\) be a \(\mathbb{R}^k\)-valued function on \(\Omega\) such that \(C(n) = (a_0, \ldots, a_{k-1})^T\) describes a (column-)vector of polynomial coefficients for each \(n \in \Omega\). Further, let \(|\nabla C|_0 = |\{n : C(n) \neq C(n + 1)\}|\) count the number of changes of the polynomial parameter field \(C\). Then the higher order Potts model can be formulated in terms of \(C\) only:

\[
C^* = \arg\min_C \gamma \|\nabla C\|_0 + \sum_{n=1}^N (1, n, \ldots, n^{k-1}) C(n) - f_n^* \|^2. \tag{15}
\]

A minimal partition \(I^*\) can be recovered by extracting the intervals of \(C^*\) with a constant functional value. A corresponding signal is obtained by \(u_n^* = (1, n, \ldots, n^{k-1}) C(n)\).

Being the limit case \(\beta \to \infty\), the higher order Potts models has similar properties as the Mumford-Shah model. In particular, if the data can be described by a polynomial of order \(k - 1\) on a segment, then that segment does not get any approximation penalty. In consequence, the assertion of Lemma 2 holds true for the higher order Potts models as well.

2.4 Existence and uniqueness of minimizers

It is straightforward to show the existence of minimizers.

**Theorem 3.** The higher-order Mumford-Shah/Potts model \((P_{k,\beta,\gamma})\) has a minimizer for each \(k \in \mathbb{N}, \gamma > 0, \beta \in (0, \infty]\).

**Proof.** For a fixed partition \(I\), the problem \((P_{k,\beta,\gamma})\) reduces to least squares problems on the intervals of \(I\) which all possess minimizers. As there are only finitely many partitions on \(\Omega\), there is at least one solution with a minimal functional value. \(\square\)

Uniqueness of the solution is more intricate. The next example shows that the solutions of the higher order Mumford-Shah models \((P_{k,\beta,\gamma})\) need not be unique. For simplicity, we consider only the case \(k = 2\), but analogous examples can be given for any order \(k \geq 3\).

**Example 4.** Consider data \(f = (0, 1, 0)\) and \(k = 2\). The optimal signal corresponding to the partition \(I^3 = \{(1), (2), (3)\}\) is given by \(u^3 = f\) and it has the functional value \(2\gamma\). The optimal solution of a partition with two segments is given by \(u^2 = (0, 1, 0)\) as well, and \(u^2\) has the lower functional value \(\gamma\). One can show that the the partition \(I^1 = \{(1, 2, 3)\}\) yields the signal \(u^1 = \frac{1}{1 + \epsilon^3} (2\beta^4, 1 + 2\beta^4, 2\beta^4)^T\) and that the functional value is given by \(\frac{4\beta^4}{1 + \epsilon^3}\). Setting this equal to the energy of the two-segment
solution, γ, gives us the critical value \( \frac{4\pi^2}{\gamma + \beta} \) which is equivalent to \( \beta^2 = \gamma / (4 - 6\gamma) \). Thus, for each \( \gamma < 2/3 \) there is \( \beta > 0 \) such that both the two-segment and the one-segment solutions are minimizers, and that \( u^1 \neq u^2 \).

Fortunately, configurations as described above are very improbable, as we will see next. As preparation we introduce a notion of equivalent partitions. We say that two partitions \( I, J \) are equivalent, i.e.,

\[
I \sim J :\iff (I \in \mathcal{I} \text{ and } |I| > k \Rightarrow I \in \mathcal{J} \text{ and } J \in \mathcal{J} \text{ and } |J| > k \Rightarrow J \in \mathcal{J}),
\]

if these partitions have the same intervals of minimum length \( k + 1 \). (The smaller intervals are irrelevant.) Equivalent partitions \( I, J \) define the same block matrices \( L_I, L_J \), i.e., \( L_I = L_J \). Therefore, using (12),

\[
u_{f,t} = S_{f,\beta} f = S_{J,\beta} f = u_{f,t}
\]

which tells that the minimizers w.r.t. the equivalent partitions \( I, J \) are given by the same function. Together, each equivalence class of partitions \( \{I\} \) defines a unique restricted minimizer \( u_{f,t} \). Further, to each \( \{I\} \) there is a unique matrix \( L_I \). The latter correspondence is even one-to-one. Summing up,

both assignments \( \{I\} \rightarrow L_I, \{I\} \rightarrow u_{f,t} \) are well-defined, and \( \{I\} \rightarrow L_I \) is one-to-one.

(18)

In particular, the minimization problem (10) may be recast in the form

\[
\arg\min_{u \in \mathbb{R}^n \setminus \{I\}} F_{\{I\}}(u), \text{ where } F_{\{I\}}(u) = \beta^2 \|L_I u\|_2^2 + \|u - f\|_2^2 + \gamma \|I\|.
\]

Here, we let \( \|I\| = \min_{J \in \mathcal{J}} |J| \).

Using this notation, the functional (13) is well-defined w.r.t. the equivalence classes so that we can write

\[
G_{\{I\}}(f) = \beta^2 \|L_I S_I f\|_2^2 + \|S_I f - f\|_2^2 + \gamma \|I\|.
\]

(20)

With these preparations we get the following result on the uniqueness of minimizers:

**Theorem 5.** Let \( \gamma > 0, \beta \in (0, \infty), \text{ and } k \in \mathbb{N} \). The minimizer \( u^* \) of \( (P_{k,\beta,\gamma}) \) is unique for almost all

input data \( f \).

**Proof.** Using the notation introduced right above, we may conclude that the solution of \( (P_{k,\beta,\gamma}) \) is unique for any \( f \in \mathcal{F} \) where the set \( \mathcal{F} \) is given by

\[
f \in \mathcal{F} :\iff \text{there is a partition } I^* \text{ such that } F_{\{I^*\}}(u_{f,t}^*) < F_{\{I\}}(u_{f,t}) \text{ for all } I \neq \{I^*\}.
\]

(21)

We are going to show that the complement \( \mathcal{F}^C \) of \( \mathcal{F} \) in euclidean space is a negligible set in the sense that it has Lebesgue measure zero. Depending on the equivalence class of the partition \( I \), we get that the minimal function value constraint to \( \{I\} \) for data \( f \) is as given by \( G_{\{I\}}(f) \) defined in (20) as \( G_{\{I\}}(f) = \beta^2 \|L_I S_I f\|_2^2 + \|S_I f - f\|_2^2 + \gamma \|I\| \). Hence, \( \mathcal{F}^C \subset \mathcal{H} \), where

\[
\mathcal{H} = \{f : \text{there are } I, I' \text{ with } [I] \neq [I'] \text{ such that } G_{\{I\}}(f) - G_{\{I'\}}(f) = 0\}.
\]

(22)

For fixed \( I, I' \), both \( G_I, G_{I'} \) are quadratic forms w.r.t. the input \( f \). Since \( [I] \neq [I'] \) we have by (18) that the quadratic form \( f \mapsto G_{\{I\}}(f) - G_{\{I'\}}(f) \) is nonzero. Therefore, by the Morse-Sard theorem, the set \( \{f : G_{\{I\}}(f) - G_{\{I'\}}(f) = 0\} \) has Lebesgue measure zero. Forming the finite union w.r.t. \( I, I' \), we see that \( \mathcal{H} \) has Lebesgue measure zero. In turn, the subset \( \mathcal{F}^C \) is a negligible set in the sense that it has Lebesgue measure zero which completes the proof. \( \square \)
2.5 Related models

Relation to complexity-constrained models. Along with the complexity penalized models \((\mathcal{P}_{k,\beta,\gamma})\)

it is natural to study the constrained variant

\[
(u^*, I^*) = \arg \min_{u \in \mathbb{R}^N, I \leq J} \|u - f\|_2^2 + \beta 2^k \sum_{\ell \in I} \|\nabla^k u_\ell\|_2^2. \tag{C_{k,\beta,\gamma}}
\]

In fact, both variants are closely related: Let us denote by \((u^J, I^J)\) a solution of \((C_{k,\beta,\gamma})\) for parameter \(J\). From the solutions for \(J = 1, \ldots, N\), one can recover a solution of \((\mathcal{P}_{k,\beta,\gamma})\) by simply choosing the solution \((u^*, I^*)\) with the optimal functional value in \((\mathcal{P}_{k,\beta,\gamma})\); that is,

\[
J^* \in \arg \min_{J = 1, \ldots, N} \gamma J + \|u^J - f\|_2^2 + \beta 2^k \sum_{\ell \in I^J} \|\nabla^k u_\ell^J\|_2^2.
\]

In [9], this relation was used for deriving a solver for the first order problem \((\mathcal{P}_{1,\beta,\gamma})\).

It is a particularly useful consequence of this relation that the set of solutions of \((C_{k,\beta,\gamma})\) for all \(J = 1, \ldots, N\), can be used to compute minimizers of \((\mathcal{P}_{k,\beta,\gamma})\) for all \(\gamma > 0\) simultaneously, in the sense that we can determine a finite number of intervals for \(\gamma\) where the corresponding solution does not change.

Relation to \(\ell_0\)-penalized problems. The classical first order Potts model \((\mathcal{P}_{1,\infty,\gamma})\) can also be written in terms of \(\ell_0\)-"norm" of the target variable \(u\) as

\[
u^* = \arg \min_{u \in \mathbb{R}^N} \gamma \|\nabla u\|_0 + \|u - f\|_2^2, \tag{23}
\]

where \(\|v\|_0\) denotes the number of non-zero elements of a vector; that is, \(\|v\|_0 = |\{n : v_n \neq 0\}|\).

As mentioned before in (15), the higher order Potts model can be written in terms of the \(\ell_0\) norm of a polynomial parameter field. We point out that plugging \(\nabla^k\) in (23) does not lead to an equivalent of the higher order Potts model \((\mathcal{P}_{k,\infty,\gamma})\); that is, in general for \(k \geq 2\)

\[
\arg \min_{w \in \mathbb{R}^N} \gamma \|\nabla^k w\|_0 + \|w - f\|_2^2 \neq u^*, \quad \text{where} \ u^* \ \text{is the minimizer of} \ (\mathcal{P}_{k,\infty,\gamma}). \tag{24}
\]

For \(k = 2\), the difference can be seen in the following example: Let \(f = (-1, -1, 1, 1)\). The optimal signal when restricting to the segmentation \(I^2 = \{(1, 2), (3, 4)\}\) is given by \(u_{I^2} = f\) and thus the approximation error is equal to 0. Optimal signals with respect to other partitions with two elements yield a higher approximation error. A simple calculation gives that the best linear approximation on the one-segment partition \(I^1 = \{(1, 2, 3, 4)\}\) is given by \(u_{I^1} = (-\frac{6}{5}, -\frac{2}{5}, \frac{2}{5}, \frac{5}{5})\) so that \(E^{(1,0)} = \frac{5}{5}\). The functional values are given by \(\frac{5}{5} + \gamma\) for \((I^1, u_{I^1})\) and by \(2\gamma\) for \((I^2, u_{I^2})\). Hence, \((\mathcal{P}_{2,\infty,\gamma})\) has the solution \((I^1, u_{I^1})\) for \(\gamma > \frac{4}{5}\) and \((I^2, u_{I^2})\) for \(\gamma < \frac{4}{5}\). (They are both optimal for \(\gamma = \frac{4}{5}\).) In contrast, as \(\|\nabla^2 u_{I^1}\|_0 = 0\) and \(\|\nabla^2 u_{I^2}\|_0 = \|(2, -2)\|_0 = 2\), the critical value for the model in (24) is \(\gamma = \frac{5}{5}\). Thus, the solutions of (24) for \(k = 2\) and \((\mathcal{P}_{2,\infty,\gamma})\) are different for \(\gamma \in (\frac{5}{5}, \frac{4}{5})\). The intuition behind that difference is that in (24) for \(k = 2\) the number of kinks of \(u\) are penalized, whereas in \((\mathcal{P}_{2,\infty,\gamma})\) the number of changes in the affine parameters are penalized.

3 Fast and stable solver for higher order Mumford-Shah problems

We develop efficient and stable solvers for higher order Mumford-Shah and Potts problems \((\mathcal{P}_{k,\beta,\gamma})\) for all \(\gamma > 0, \beta \in (0, \infty]\), and \(k \geq 1\). First we recall a dynamic programming scheme commonly used for partitioning problems. Then we develop a recurrence scheme for computing the required approximation errors which is key for the efficiency of the algorithm. Eventually, we analyze the stability of the algorithm.
3.1 Dynamic programming scheme for partitioning problems

Let us denote the functional in (6) by \( P \), i.e.,
\[
P(I) = \sum_{i \in I} (\mathcal{E}_i^r + \gamma).
\]

Note that the functional is well-defined also for a partition \( I \) on the reduced domain \( 1 : r \), which we will utilize in the following. Let the minimal functional value for the domain \( 1 : r \) be denoted by
\[
P^*_r = \min_{I: \text{partition on } 1 : r} P(I).
\]

The value \( P^*_r \) for the domain \( 1 : r \) satisfies the Bellman equation
\[
P^*_r = \min_{l=1,\ldots,r} \left\{ \mathcal{E}_l^r + \gamma + P^*_{l-1} \right\},
\]
where we let \( P^*_0 = 0 \). Recall that \( \mathcal{E}_l^r = 0 \) if \( r - l + 1 \leq k \), so the minimum on the right hand side actually only has to be taken over the values \( l = 1, \ldots, r - k \). By the dynamic programming principle, we successively compute \( P^*_1, P^*_2 \), until we reach \( P^*_N \). As our primary interest is the optimal partition \( J^* \), rather than the minimal functional value \( P^*_N \), we keep track of a corresponding partition. An economic way to do so is to store at step \( r \) the minimizing argument \( l^* \) of (25) as the value \( J_r \), so that \( J \) encodes the boundaries of an optimal partition; see [24].

The above procedure has the complexity \( O(N^2 \phi(N)) \) where \( \phi \) is an upper bound for the effort of computing the approximation errors \( \mathcal{E}_l^r \). The straightforward way to compute \( \mathcal{E}_l^r \) is solving the least squares system (7) which leads to \( \phi(N) = N \). We develop a strategy that achieves \( \phi(N) = 1 \) in the next section.

In the following, we recall two strategies from [54] and [34], respectively, to prune the search space. In [54], a pruning strategy was introduced by exploiting the relation \( \mathcal{E}_l^r \leq \mathcal{E}_{l'}^r \) if \( l' \leq l \). From (25) follows immediately that if the current value \( P_r \) for \( P^*_r \) satisfies
\[
P_r < \mathcal{E}_l^r + \gamma
\]
for some \( l \), one can omit checking all \( l' < l \) for this \( r \), thus, \( P^*_r = P_r \). That is, we do not have to compute \( \mathcal{E}_{l'}^r \). Another way to prune the dynamic program follows from the observation that the approximation errors satisfy the inequality \( \mathcal{E}_{l+1}^r + \mathcal{E}_{l+1}^{r+1} \leq \mathcal{E}_l^r \), for all \( l \leq s < r \). Killick et al. [34] deduced that if
\[
P^*_r \leq P^*_l + \mathcal{E}_{l+1}^{r+1},
\]
then \( l \) cannot be an optimal last changepoint at a future timepoint \( r \). That means, the intervals \( l + 1 : r \) for all \( r = l + 1, \ldots, N \) cannot be reached and consequently \( l \) does not need to be considered again for any future timepoint \( r \).

3.2 Fast computation of the approximation errors for higher order Mumford-Shah problems

Here, we develop a recurrence formula for computing the \( \mathcal{E}_l^r \) needed in (25). For notational simplicity, we describe the basic scheme for the left bound \( l = 1 \), i.e., computing \( \mathcal{E}_1^r \) for \( r = 1, \ldots, N \). The procedure works analogously for any \( l > 1 \).

Recall that \( \mathcal{E}_1^r = 0 \) if \( r \leq k \); so we may assume that \( r > k \) in the following. Our starting point is to rewrite the minimization problem (7) for \( I = 1 : N \) in matrix form as
\[
\mathcal{E}_1^N = \min_{v \in \mathbb{R}^N} \|Av - y\|_2^2.
\]
Here, \[ A = \begin{pmatrix} E_N \\ \mathbf{e}^{\mathcal{F}} \end{pmatrix} \in \mathbb{R}^{(2N-k)\times N} \quad \text{and} \quad y = (f_1, \ldots, f_N, 0)^T \in \mathbb{R}^{2N-k}, \] (29)
and \( E_N \) represents the identity matrix of dimension \( N \). Further, determining \( E^{1:r} \) for \( r < N \) amounts to solving the least squares problem of smaller size
\[ E^{1:r} = \min_{v \in \mathbb{R}^r} \| A^{(r)} v - y^{(r)} \|_2^2, \] (30)
where \( A^{(r)} \) is the submatrix of \( A \) given by
\[ A^{(r)} = \begin{pmatrix} A_{1:r,1:r} \\ A_{(N+1:N+r-k),(1:r)} \end{pmatrix}, \quad \text{and} \quad y^{(r)} = \begin{pmatrix} f_{1:r} \\ 0 \end{pmatrix}. \]
Note that we do not have to compute a minimizer \( v \) of (30) to evaluate \( E^{1:r} \). Instead, we develop a recurrence formula computing \( E^{1:r} \) directly based on Givens rotations. As preparation, we use the symbols \( Q^{(r)} \) and \( R^{(r)} \) to denote the QR decomposition of \( A^{(r)} \), i.e.
\[ A^{(r)} = Q^{(r)} \begin{pmatrix} R^{(r)} \\ 0 \end{pmatrix}, \]
with an orthogonal matrix \( Q^{(r)} \) and an upper triangular matrix \( R^{(r)} \). As the \( \ell^2 \)-norm is invariant to orthogonal transformations we may represent \( E^{1:r} \) as
\[
E^{1:r} = \min_{v \in \mathbb{R}^r} \left\| \begin{pmatrix} R^{(r)} \\ 0 \end{pmatrix} v - (Q^{(r)})^T y^{(r)} \right\|_2^2 = \min_{v \in \mathbb{R}^r} \left\| R^{(r)} v - ((Q^{(r)})^T y^{(r)})_{1:1} \right\|_2^2 + \left\| ((Q^{(r)})^T y^{(r)})_{1:2r-k} \right\|_2^2 = \left\| ((Q^{(r)})^T y^{(r)})_{1:2r-k} \right\|_2^2. \]
(31)
The first term in the second line vanishes since the corresponding linear system can be solved exactly. All terms in the last line of (31) are explicitly given and do not involve minimization. Our goal is to recursively compute \( E^{1:r+1} \) without explicitly computing QR decompositions and without carrying out the summation involved in the last line of (31).
The first step is the determination of recurrence coefficients. To this end, assume that we have computed the QR decomposition of \( A^{(r)} \). We consider the auxiliary matrix \( W^{(r)} \) containing the upper triangular matrix \( R^{(r)} \) and the beginning of the \((N + r - k + 1)\)-th row of \( A \):
\[ W^{(r)} = \begin{pmatrix} R^{(r)} \\ 0 & 1 \\ 0 & 0 \\ A_{N+r-k+1,(1:r+1)} \end{pmatrix}. \]
By the band structure of \( A \), only the last \( k + 1 \) entries of \( A_{N+r-k+1,(1:r+1)} \) are non-zero. We aim at bringing \( W^{(r)} \) to upper tridiagonal form using orthogonal transformations without modifying the already present zeros. To this end, we employ Givens rotations. (Note that Householder reflections would destroy the existing zero entries.) Recall that a Givens rotation \( G = G(j, m, \theta) \) is equal to the identity matrix with the \( 2 \times 2 \) submatrix \((G_{11}; G_{12}; G_{21}; G_{22})\) replaced by a planar rotation matrix; that is,
\[ G(j, m, \theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(\theta) & \cdots & \sin(\theta) & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\sin(\theta) & \cdots & \cos(\theta) & \cdots & 0 \\ \vdots & & \cdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}, \] (32)
where \( \theta \) denotes the rotation angle. In order to eliminate the matrix entry \( a = A_{mj} \) by the pivot element \( b = A_{jj} \) we use left multiplication by the Givens rotation \( G(j, m, \Theta_{mj}) \) with the parameters

\[
\cos(\Theta_{mj}) = \frac{b}{\rho}, \quad \sin(\Theta_{mj}) = \frac{a}{\rho}
\]

(33)

where \( \rho = \text{sign}(b) \sqrt{a^2 + b^2} \). Here, we have used the notation \( \Theta_{mj} \) to denote the rotation angle of the corresponding Givens rotation. Since \( G(j, m, \theta) \) only operates on the \( j \)-th and \( m \)-th row of a matrix it does not destroy the zeros already present in other lines. Hence, we eliminate the last row of \( W^{(l)} \) by using \( k + 1 \) Givens rotations with parameters chosen according to (33) to obtain \( R^{(l+1)} \). This shows how to recursively compute \( R^{(l+1)} \) given \( R^{(l)} \). The relevant quantities we need in the following are the rotation angles \( \Theta_{mj} \) which serve as the recurrence coefficients.

Having computed \( \Theta_{mj} \), we now are able to carry out the error update step from \( \mathcal{E}^{1:r} \) to \( \mathcal{E}^{1:r+1} \) in \( O(1) \): Assume that we have computed \( \mathcal{E}^{1:r} \) and the vector \( q^{(r)} \) defined by

\[
q^{(r)} = (Q^{(r)})^T \tilde{y}^{(r)}.
\]

The \( q^{(r)} \) satisfy the recurrence relation

\[
q^{(r+1)} = G^{(r+1)} \begin{pmatrix} q_{1:r}^{(r)} \\ f_{r+1}^{(r)} \\ q_{r+1:2r-k}^{(r)} \\ 0 \end{pmatrix},
\]

(34)

where \( G^{(r+1)} \) denotes the elimination matrix composed of the above \( k + 1 \) Givens rotations; that is,

\[
G^{(r+1)} = \prod_{j=1}^{k+1} G(r - k + j, 2(r + 1) - k, \Theta_{N+r+1-k,r-k+j}),
\]

(35)

where we use the convention \( \prod_{j=1}^{k} Z_j = Z_k Z_{k-1} \cdots Z_1 \). Further, as \( G^{(r+1)} \) only operates on the first \( r + 1 \) lines and the last line of the vector on the right hand side of (34) it follows that

\[
\|q_{r+1:2r-k}^{(r+1)}\|_2^2 = \|q_{r+1:2r-k}^{(r)}\|_2^2 + (q_{2r+1-k}^{(r+1)})^2.
\]

(36)

Therefore, the error update is given by

\[
\mathcal{E}^{1:r+1} = \mathcal{E}^{1:r} + (q_{2r+1-k}^{(r+1)})^2.
\]

(37)

To summarize the update scheme consists of computing \( q^{(k+1)} \) by (34) and updating \( \mathcal{E}^{1:r+1} \) by (37).

The errors \( \mathcal{E}^{1:r} \) can be updated in the same fashion by applying the above procedure to the data \( \tilde{f} = (f_1, \ldots, f_r) \). An important practical aspect is that the recurrence coefficients \( \Theta_{mj} \) do not depend on the data. Thus, we only need to compute the \( (N - k)(k + 1) \) recurrence coefficients once and can reuse them for computing all \( \mathcal{E}^{1:r} \).

We briefly discuss the accuracy of the error update scheme (37). As Givens rotations are orthogonal they have the optimal condition number one. Hence, there is no inherent error amplification in the elimination steps. The practical accuracy of the error update is illustrated by the following numerical experiment. We compute the approximation errors of a polynomial of degree \( k - 1 \). As these are in the null space of \( \mathbb{V}^k \) the approximation errors \( \mathcal{E}^{1:r} \) are exact to 0 for all \( r = 1, \ldots, N \). Figure 3 shows that the proposed procedure reproduces the exact results up to machine precision.

Next we explain how to include the two pruning strategies from Section 3.1. The first strategy with condition (26) requires to run over the \( l \)-index in a descending way, i.e. in the order \( l = r, r - 1, \ldots, 1 \). On the other hand, the second strategy demands checking (27) for all \( 1 < l < r \) after \( P_r \) was determined. Consequently, if (26) holds for some \( l \) at \( (1 : r) \), condition (27) cannot be checked for \( l' < l \) since \( \mathcal{E}^{1:l'} \) has not been computed yet. In order to overcome this issue without

---

14
obliterating the first pruning, we proceed as follows. At domain \( (1 : r) \), run through all \( l \) in the descending list \( L \) and update successively the corresponding approximation errors to \( E^{l,r} \). After each update step, check whether (27) is satisfied for the current upper interval bound of the error and if so, delete \( l \) from \( L \) and start again with the next entry in \( L \). By this, it is not necessary to adapt the second pruning strategy essentially: check condition (26) after testing if a not pruned \( l \) is the current optimal last changepoint. By combining the pruning strategies we effectively decrease the total number of error updates (37) that have to be performed; see Section 4.2 for a numerical study.

We provide a pseudocode for the solver in Algorithm 1. Let us summarize the above derivation:

**Theorem 6.** Let \( f \in \mathbb{R}^N \), \( k \in \mathbb{N} \), and \( \beta, \gamma > 0 \). Algorithm 1 computes a global minimizer of \( (P_{k,\beta,\gamma}) \). The worst case time complexity is \( O(N^2) \).

**Proof.** It follows from the Bellman equation (25) that the algorithm computes indeed a global minimizer. The double loop over the the \( l \) and \( r \) indices has quadratic worst-case complexity. It remains to show that for each \( r \in (1 : N) \) we can compute \( E^{r-1,r}, E^{r-2,r}, \ldots, E^{1,r} \) in \( O(1) \) per element. As each line of \( A \) has at most \( k + 1 \) entries, the elimination of one line requires \( k + 1 \) elimination steps. By the band structure of \( A \) each elimination step by Givens rotations creates only new non-zeros in a band of \( k + 1 \) entries above the diagonal \( A_{ii} \), \( i = 1, \ldots, N \). Thus, computing the recurrence coefficients needs only \( O(k^2 N) \) operations. As applying a Givens rotation to a vector only needs a constant amount of operations, the multiplication in (34) is in \( O(k) \). (Note that the matrix \( G^{r+1} \) is not explicitly created.) Hence, executing the recurrence (37) is \( O(k) \). It follows that computing the errors for all \( O(N^2) \) intervals sums up to \( O(kN^2) \). The reconstruction step from a partition is in \( O(kN) \) as it reduces to solving a least squares system of band matrices whose number of rows sum up at most \( 2N - k \). As \( k \) is fixed the overall worst case time complexity is \( O(N^2) \). \( \square \)

### 3.3 Fast computation of the approximation errors for higher order Potts problems

We describe a stable yet fast procedure to compute the approximation errors for the higher order Potts problems (14). To this end, we first rewrite (14) in terms of the polynomial coefficients \( p \in \mathbb{R}^k \) as

\[
E^{l,r} = \min_{p \in \mathbb{R}^k} ||B_{1,r,1,r} p - f_{1,r}||_2^2, \tag{38}
\]

**Figure 3:** Approximation errors \( E^{1,r}, r = 1, \ldots, N \), for (a) a linear polynomial and order \( k = 2 \), for (b) a quadratic polynomial and order \( k = 3 \), and for (c) a cubic polynomial and order \( k = 4 \). The theoretical approximation errors are equal to zero; the approximation errors computed using the recurrence formula (37) are accurate up to machine precision.
Algorithm 1: Solver for the higher order Mumford-Shah problem and higher order Potts problem

Input: Data \( f \in \mathbb{R}^N \); model parameters \( k \in \mathbb{N}, \beta \in (0, \infty], \gamma > 0 \);

Output: Global minimizer \( (u^*, I^*) \) of \((\mathcal{P}_{k,\beta,\gamma})\) or \((\mathcal{P}_{k,\alpha,\gamma})\)

/* Precomputations */

1. \((\mathcal{P}_{k,\beta,\gamma})\): \(A\) from (29)
2. \((\mathcal{P}_{k,\alpha,\gamma})\): \(B\) from (39)

/* Row-wise transform the matrix to upper triangular form using successive Givens rotations and store the rotation angles in \(\Theta\). */

3. Compute \(E^{1,r}\) for all \(r = 1, \ldots, N\) with \(\Theta\)

4. Initialize lists \(L = [2], R = [2], \mathcal{E} = [0]\)

5. for \(r \leftarrow 2, \ldots, N\) do
   /* Initialization */
   \(L \leftarrow 0, P^*_r \leftarrow E^{1,r}\)

6. /* Find optimal \(P^*_r\) using (25) */
   for \(i = 1, \ldots, \text{length of } L\) do
     /* Update approximation error */
     \(E_i \leftarrow E^{i,r}\)
   end

7. /* Pruning (26) */
   if \(P^*_{i-1} + E_i \geq P_r^*\) then
     Delete: \(l_i\) from \(L\), \(r_i\) from \(R\) and \(E_i\) from \(E\)
     go to 7
   end

8. /* Update lists */
   Prepend: \(r + 1\) to \(L\), \(r + 1\) to \(R\), \(0\) to \(E\)

9. /* Recover partition \(I^*\) from segment boundary locations \(J\) */

10. \(r \leftarrow N, I^* \leftarrow 0\)

11. while \(r > 0\) do
       \(l \leftarrow J(r) + 1, I^* \leftarrow I^* \cup \{l : r\}, r \leftarrow l - 1\)

12. /* Reconstruction of \(u^*\) by solving linear systems on segments (using QR decomposition and reusing \(\Theta\) for speedup) */

13. for \(I \in I^*\) do
       \(\mathcal{P}_{k,\beta,\gamma}\): Solve \(u^*_I = \arg\min_{v \in \mathbb{R}^N} \|v - f_I\|^2_2 + \beta_2 \|\nabla v\|^2_2\)
       \(\mathcal{P}_{k,\alpha,\gamma}\): Solve \(u^*_I = \arg\min_{v \in \mathbb{R}^N} \|v - f_I\|^2_2\) such that \(v\) is polynomial of degree \(k - 1\) on \(I\)

14. end
where $B$ is the $R^{N \times k}$ matrix defined by

$$B = \begin{pmatrix}
1 & 1 & \cdots & 1^{k-1} \\
1 & 2 & \cdots & 2^{k-1} \\
\vdots & \vdots & \ddots & \vdots \\
1 & N-1 & \cdots & (N-1)^{k-1} \\
1 & N & \cdots & N^{k-1}
\end{pmatrix} \in R^{N \times k}. \quad (39)$$

As in Section 3.2, we describe the method for the prototypical case $l = 1$. Furthermore, we assume that $r > k$ since otherwise $E^{1,r} = 0$. Denoting the submatrix $B_{1,r+1:k}$ by $B^{(r)}$ and its QR decomposition by $Q^{(r)}, R^{(r)}$, we obtain in analogy to (31) that

$$E^{1,r} = \min_{p \in \mathbb{R}^k} \left\| \begin{pmatrix} R^{(r)} \\ 0 \end{pmatrix} p - (Q^{(r)})^T f_{1,r} \right\|^2_2 = \|q^{(r)}_{k+1,r}\|^2_2. \quad (40)$$

where $q^{(r)}$ is given by

$$q^{(r)} = (Q^{(r)})^T f_{1,r}.$$ 

The recurrence coefficients for the error update $\Theta_{r+1,j}$ for $j = 1, \ldots, k$ are the Givens rotation angles for eliminating the entry $R_{r+1,j}$ with the pivot element $R_{j,r}$. Now assume that we have computed $q^{(r)}$ and $E^{1,r}$. Then, $q^{(r+1)}$ can be expressed by the recurrence relation

$$q^{(r+1)} = G^{(r+1)} \begin{pmatrix} q^{(r)} \\ f_{1,r} \end{pmatrix}, \quad (41)$$

where $G^{(r+1)}$ comprises the Givens rotations $G(j, r + 1, \Theta_{r+1,j})$ for $j = 1, \ldots, k$; that is,

$$G^{(r+1)} = \prod_{j=1}^{k} G(j, r + 1, \Theta_{r+1,j}). \quad (42)$$

where we again use the convention $\prod_{j=1}^{k} Z_j = Z_k Z_{k-1} \cdots Z_1$. As $G^{(r+1)}$ operates only on the first $k$ entries and the last entry of $q^{(r)}$, we obtain by (40)

$$E^{1,r+1} = \|q^{(r+1)}_{k+1,r}\|^2_2 = \|q^{(r)}_{k+1,r}\|^2_2 + (q^{(r+1)}_{r+1})^2 + E^{1,r} + (q^{(r+1)}_{r+1})^2. \quad (43)$$

**Remark 7.** For the higher order Potts problems, there are also closed formulae for the evaluation of the errors $E_{i,r}$ which one might consider to use directly. Such formulae are derived in [24] and in [37], for the first and the second order Potts problem, respectively. Using computer algebra, we have derived such formulae for $k = 3$ and $k = 4$. By precomputing moments, the errors $E_{i,r}$ can then be computed in $O(1)$ per element. The results are typically acceptable for the piecewise constant and piecewise affine linear problems ($k = 1, 2$) and moderate signal lengths. Unfortunately, for higher orders or longer signals, the approach based on the precomputations of moments is prone to numerical instability. This is illustrated by the following experiment (cf. Figure 4). We consider the parabolic signal $f_n = n^2/100$, $n = 0, \ldots, N$ where $N = 100$. The true approximation errors for the higher order Potts model of order $k = 3$ are given by $E_{i,r} = 0$ for all $l, r$ with $1 \leq l \leq r \leq N$. Figure 4 shows that the results for $E_{i,r}$ are distorted when using the approach based on the precomputations of moments, in particular if $r$ is close to $N$. The errors $E_{i,N}$ are even more severely affected because of loss of significance. We observe that – in contrast to the moment precomputation approach – the proposed method gives accurate results up to machine precision.

### 3.4 Stability results

In this section, we investigate the stability of the proposed algorithm. We start out with some basic lemmas we will need later on. In the following, we consider the functional $G(f)$ defined in (20) and omit the brackets and simply write $G(f)$.
Figure 4: Approximation errors for a parabolic signal for the higher order Potts model of order $k = 3$. The graphs show the results based on precomputed moments (blue) and based on the proposed scheme (red). The true approximation errors are all equal to zero. Left: Errors $E^{1,r}$ for $r = 1,...,N$. The computation based on precomputed moments is distorted beyond machine precision and gives even negative values. Right: Errors $E^{3,N}$ for $l = 1,...,N-1$. The values derived from precomputed moments are strongly distorted when $l$ approaches $N$. The proposed computation based on Givens rotations is accurate up to machine precision in either case.

Lemma 8. We consider data $f \in \mathbb{R}^N$. If there is a partition $I'$ and $\varepsilon > 0$ such that

$$\mathcal{G}_I(f) < \mathcal{G}_I(f) - \varepsilon \quad \text{for all} \quad I \not\subseteq [I']$$

then there is an euclidean $\delta$-ball $B(f, \delta)$ around $f$ such that for any $g \in B(f, \delta)$ holds: for data $g$ there is a unique optimal solution $u = u_{g,I'}$ of the problem $(P_{k,\delta,\gamma})$, and the corresponding partition $I'$ fulfills $[I'] = [I']$. We may choose $\delta$ by

$$\delta := \min \left( \frac{\varepsilon}{2(2\beta)^{2k} (\|f\| + 1/2)}, \frac{1}{2} \right).$$

Proof. The essential argument here relies on the continuity of the quadratic forms $x \mapsto \mathcal{G}_I(x)$ given by

$$\mathcal{G}_I(f) := \beta^{2k} \|A_I S f\|^2_2 + \|S f - f\|^2_2$$

which are the main parts of the $\mathcal{G}_I$ given by (20). Each $\mathcal{G}_I$ may be represented w.r.t. the euclidean standard scalar product $(\cdot, \cdot)$ via a symmetric matrix $A_I$ as $\mathcal{G}_I(x) = \langle A_I x, x \rangle$. The operator norm of $A_I$ equals the norm of the corresponding bilinear form which in turn, since the $\mathcal{G}_I$ are positive (semi-definite), corresponds to

$$\|A_I\| = \sup_{\|x\| = 1} \mathcal{G}_I(x).$$

We first let $\delta'$ be defined by

$$\delta' := \min \left( \frac{1}{2}, \frac{\varepsilon}{2 \max_{I'} \|A_I\| (\|f\| + 1/2)} \right).$$

We want to estimate $\mathcal{G}_I(f)$ from above for $g$ in a $\delta'$-ball around $f$. For brevity, we write $\mathcal{G}_I(g) = \mathcal{G}_I(g) + \gamma N_{I'}$, where we let $N_{I'} := \|I'\|$. Then we may estimate

$$\mathcal{G}_I(g) = \mathcal{G}_I(g) + \gamma [I'] = \mathcal{G}_I(f) + \mathcal{G}_I(f - g) + 2 \langle A_I f, f - g \rangle + \gamma [I']$$

$$\leq \mathcal{G}_I(f) - \varepsilon + \delta'^2 \|A_I\| + \delta' \|A_I\| (\|f\| + 1) + \gamma [I']$$

$$\leq \mathcal{G}_I(f) - \varepsilon + 2 \max_{I'} \|A_I\| (\|f\| + \delta') + \gamma [I']$$

$$\leq \mathcal{G}_I(f) - \varepsilon + \varepsilon + \gamma [I'] = \mathcal{G}_I(g) + \gamma [I'] = \mathcal{G}_I(g).$$
For the first inequality, we applied (44) for $\mathcal{G}_f(f)$ and used the assumption that $g \in B(f, \delta')$. For the second inequality, we employed (48). In order to relate (48) with (45), we now estimate $\max_I \|A_I\|$ using basic spectral theory for self-adjoint bounded operators. Since $A_I$ is the matrix representing the bilinear form corresponding to $\mathcal{G}_f$, we may estimate using (46) that

$$\|A_I\| \leq \beta^{2k} \|L_f^T L_f\| \|S_f\|^2 + \|I - S_f\|^2,$$

for any partition $I$, (50)

with the definitions of $L_f$ given in (9) and of $S_f$ given in (12); here we only employed the triangle inequality and the submultiplicativity of operator norms. By (12), $S_f = (\beta^{2k} L_f^T L_f + \text{id})^{-1}$. Since $L_f^T L_f$ is self-adjoint and positive, the spectrum of $\beta^{2k} L_f^T L_f + \text{id}$ is contained in $[1, \infty)$. Hence, its inverse $S_f$ has its spectrum contained in $[0, 1]$. Being again self-adjoint, and positive, $\|S_f\| \leq 1$. Further, since $S_f$ has its spectrum contained in $[0, 1]$, $I - S_f$ has its spectrum contained in $[0, 1]$ as well. Then, with the same argument, $\|I - S_f\| \leq 1$. In order to estimate $L_f^T L_f$, we consider $L_I$ in (9), and notice that $L_f^T L_f$ is block diagonal with entries consisting of convolutions of $k$th differences with themselves. Thus the row-sums as well as the column sums of $L_f^T L_f$ are bounded by $2^k$. The using the Schur criterion, the operator norm of $L_f^T L_f$ w.r.t. euclidean norm in the base space can be estimated by $2^k$. Summing up, we conclude invoking these estimates in (50) that

$$\|A_I\| \leq \beta^{2k} 2^{2k} + 1,$$

for any partition $I$. (51)

We can now show the assertion of the lemma. If $g \in B(f, \delta)$, then $g \in B(f, \delta')$, by the estimate (51) relating (48) with (45). In consequence, the estimate (49) applies to $g$. Hence the solution for $g$ is unique and given by (12); in particular, the corresponding equivalence class of partitions equals $I'$ which shows the assertion. □

Next, we need a backward stability result for the QR algorithm [26]. We present it adapted to our setup as needed later on. In analogy to (12), we denote the linear mapping from $f_I$ (restricted to the interval $I$) to the solution $u_I$ by $S_I$.

**Theorem 9.** The QR algorithm $S_{I,r}$ needed for computing the $E^{I,r}$ is backward stable, i.e., given data $f_{I,r}$ living on the subinterval $I : r$, there is a perturbation $\tilde{f}_{I,r}$ of $f_{I,r}$ such that

$$S_{I,r} (f_{I,r}) = S_{I,r} (\tilde{f}_{I,r}) \quad \text{with} \quad \|\tilde{f}_{I,r} - f_{I,r}\| \leq \delta_{I,r},$$

(52)

where $\delta_{I,r}$ depends on the machine precision $\tau$ and on the norm $\|f_{I,r}\|$ via

$$\delta_{I,r} \leq 6\tau \sqrt{r-l+1} \left( \frac{9(r-l+1) - 5}{4} - k \right) (1 + 6\tau)^{3(r-l)-k},$$

(53)

Here, the QR algorithm is understood as in the analysis setup of [26, 62].

**Proof.** If $r - l < k$, then $S_{I,r} (f_{I,r}) = S_{I,r} (\tilde{f}_{I,r}) = 0$. So we may assume $r - l \geq k$. Recall that calculating $E^{I,r}$ corresponds to computing the residual vector of the least squares problem with system matrix $A \in \mathbb{R}^{(2r-l+1) \times (r-l+1)}$ from (29) and data $f_{I,r}$. In [26] it is shown that

$$\tilde{R} = \tilde{Q}^T (A + \Delta A), \quad \|\Delta A\| \leq \mu_{I,r}(\tau) \|A\|_{\text{r}},$$

$$\mu_{I,r}(\tau) = 6\tau \sqrt{r-l+1} \left( \frac{9(r-l+1) - 5}{4} - k \right) (1 + 6\tau)^{3(r-l)-k},$$

where $\tilde{R}$ is the computed upper triangular matrix by means of Givens rotations and note that $\tilde{Q}^T$ is the orthogonal matrix that is the product of exact Givens rotations we apply. Analogously, for the data vector it is shown in [26] that

$$\tilde{Q}^T \tilde{f}_{I,r} = \tilde{Q}^T (f_{I,r} + \Delta f_{I,r}), \quad \|\Delta f_{I,r}\| \leq \mu_{I,r}(\tau) \|f_{I,r}\|,$$

hence $S_{I,r} (f_{I,r}) = S_{I,r} (f_{I,r} + \Delta f_{I,r})$ which implies (53). □
Corollary 10. We consider bounded data \( f \in \mathbb{R}^N, \|f\| < C \). For any partition \( I \) considered in the proposed algorithm for the higher order Potts and Mumford-Shah problem, there is a perturbation \( \tilde{f} \) of \( f \) such that

\[
\tilde{S}_I(f) = S_I(\tilde{f}) \quad \text{where} \quad \|\tilde{f} - f\| < \delta_I(\tau),
\]

where \( \delta_I(\tau)^2 = \sum_i \delta_{I,x}(\tau)^2 \) depends on the machine precision \( \tau \) via the dependence of the \( \delta_{I,x} \) on \( \tau \) given in Theorem 9 and on \( C \), but not on \( f \). More precisely, \( \delta_I(\tau) \) can be estimated from above by

\[
\delta_I(\tau)^2 \leq 36C^2\tau^2 \sum_i (r_i - l_i + 1) \left( \frac{9(r_i - l_i + 1) - 5}{4} - k \right)^2 (1 + 6\tau)^6(\tau - 2k).
\]

Proof. The statement is a consequence of Theorem 9 since, for fixed partition, the proposed algorithm computes optimal solutions \( u_{I,\tilde{f}} = S_I(\tilde{f}) \) (cf. (12)) using the QR algorithm on intervals. In particular, (55) is a consequence of (53). \( \square \)

For the formulation of the next statement, we use the notation \( \tilde{G}_I \) to denote the algorithm to compute the energy \( \tilde{G}_I \) given by (20). Further, we use the notation \( g(\tau) \) to bound the approximation error between \( \tilde{G}_I \) and \( G_I \) for all \( I \) in dependence of the precision \( \tau \).

Proposition 11. We consider bounded data \( f \in \mathbb{R}^N, \|f\| < C \), and assume that (44) is fulfilled for \( f \). Let

\[
\delta(\tau) = \max \delta_I(\tau) \leq 6C\tau N^2 \left( \frac{9N - 5}{4} - k \right) (1 + 6\tau)^3(N - 1)^{-k}
\]

given in (10) and assume that \( \tau \) is small enough such that \( \delta(\tau) < \delta/2 \) with \( \delta \) given by (45) and such that \( g(\tau) \leq \epsilon/4 \) with \( \epsilon \) given in (44). Then, the higher order Potts and Mumford-Shah problem \( \mathcal{P}_{k,\beta,\gamma} \) has a unique minimizer \( u_I \), and the proposed algorithm for computing this minimizer of the higher order Potts and Mumford-Shah problem \( \mathcal{P}_{k,\beta,\gamma} \) is backward stable in the sense that

\[
\tilde{u}_I = u_I \quad \text{where} \quad \|\tilde{f} - f\| < \delta(\tau).
\]

Here, \( \tilde{u}_I \) is the result produced by the proposed algorithm for data \( f \) and \( u_I \) is the (unique) solution of the higher order Potts and Mumford-Shah problem \( \mathcal{P}_{k,\beta,\gamma} \) for perturbed data \( \tilde{f} \).

Remark 12. A more explicit relation of \( \epsilon \) and the precision \( \tau \) without using the \( \delta \)'s sufficient for the assumptions of Proposition 11 to hold is given by

\[
(1 + 6\tau)^3N^{-k}\tau < \frac{\epsilon}{12CN(2\beta)^{2k}(C + \frac{1}{2}) (\frac{9N - 5}{4} - k)} \quad \text{if} \quad \epsilon \leq (2\beta)^{2k}(C + \frac{1}{2}),
\]

\[
(1 + 6\tau)^3N^{-k}\tau < \frac{1}{12CN (\frac{9N - 5}{4} - k)} \quad \text{if} \quad \epsilon > (2\beta)^{2k}(C + \frac{1}{2}),
\]

\[
\mu_{[1,\infty]}(\tau) < \frac{1}{2} \left( \frac{4C^2N + \epsilon}{C^2N} \right)^{\frac{1}{2}} - 1
\]

w.r.t. \( \mu_{[1,\infty]}(\tau) \) from the proof of Theorem 9. Conditions (58) and (59) are sufficient for \( \delta(\tau) < \delta/2 \) which is an immediate implication of combining (44) and (56). From (60) follows \( g(\tau) \leq \epsilon/4 \) since: for any admissible \( I : r \) we have

\[
\left\| \left| \left| \tilde{Q}^T f_{I,r} \right| \right|^2 - \left| \left| Q^T f_{I,r} \right| \right|^2 \right\| \leq \left( \left| \left| Q^T f_{I,r} \right| + \left| \left| \tilde{Q}^T f_{I,r} \right| \right| \right) \left| \left| Q^T f_{I,r} \right| - \left| \left| \tilde{Q}^T f_{I,r} \right| \right| \right|
\]

\[
\leq \left( C + \left| \left| \tilde{Q}^T f_{I,r} \right| \right| \right) \left| \left| Q^T f_{I,r} - \tilde{Q}^T f_{I,r} \right| \right|
\]

and

\[
\left| \left| \tilde{Q}^T f_{I,r} \right| \right| \leq \left| \left| Q^T f_{I,r} \right| \right| + \left| \left| Q^T f_{I,r} - \tilde{Q}^T f_{I,r} \right| \right| \leq C + \left| \left| Q^T f_{I,r} - \tilde{Q}^T f_{I,r} \right| \right| \leq C + \mu_{[1,\infty]}(C).
Combining both yields
\[
g(\tau) = |\tilde{G}_I(f) - G_I(f)| \leq \sum_{l \in I} |\tilde{E}_l - E_l| \\
\leq \sum_{l \in I} \|Q^T f_{l|z} - Q^T f_{l|z}\|^2 \leq NC^2 \mu_{[1,N]}(\tau) (2 + \mu_{[1,N]}(\tau)).
\]

An easy computation shows that (60) is equivalent to requiring the latter to be smaller than \(\varepsilon/4\).

**Proof of Proposition 11.** By the proof of Theorem 5, the solution \(u_I\) of \((\mathcal{P}_{k,\beta,\gamma})\) is unique for data \(f\). We denote the equivalence class of partitions corresponding to this optimal solution \(u_I\) by its representer \(I'\). As a first step, we show that the solution \(\tilde{u}_I\) computed by the proposed algorithm for data \(f\) has partition \(I'\) as well. To that end, we first notice that by Corollary 10, there is \(\tilde{f}_I\) with \(\|\tilde{f}_I - f\| < \delta_I(\tau)\) such that \(\tilde{S}_I(f) = S_I(\tilde{f}_I)\), for any partition \(I\) which is considered by the algorithm. In particular, using the notation of Theorem 9,
\[
\|\tilde{S}_{I,\tau}(f_{I,\tau}) - S_{I,\tau}(f_{I,\tau})\| = \|S_{I,\tau}(\tilde{f}_{I,\tau}) - S_{I,\tau}(f_{I,\tau})\|
\leq \|S_{I,\tau}\| \|\tilde{f}_{I,\tau} - f_{I,\tau}\| \leq \|\tilde{f}_{I,\tau} - f_{I,\tau}\| < \delta_{I,\tau}.
\]

For the second before last inequality, we used that \(\|S_{I,\tau}\| \leq 1\) which we have shown in the proof of Lemma 8. In consequence, summing over all intervals of \(I\) of length at least \(k + 1\), we obtain from (61) that
\[
\|\tilde{S}_I(f) - S_I(f)\| = \|S_I(\tilde{f}_I) - S_I(f)\| \leq \|\tilde{f}_I - f\| < \delta_I(\tau).
\]

for any partition \(I\) which is considered by the algorithm. Then, using the notation \(\tilde{G}_I\) for the algorithmic variant of \(G_I\), we have (with the notation as in Lemma 8) that
\[
\tilde{G}_I(\tilde{f}_I) \leq G_I(\tilde{f}_I) + |\tilde{G}_I(\tilde{f}_I) - G_I(\tilde{f}_I)|
\leq G_I(\tilde{f}_I) + g(\tau)
\leq G_I(\tilde{f}_I) - \varepsilon + \delta_I(\tau)^2 \|A_I\| + \delta_I(\tau) \|A_I\| \|f\| + g(\tau)
\leq G_I(\tilde{f}_I) - \varepsilon + 2\delta(\tau) \max f \|A_I\| \|f\| + \delta(\tau) + g(\tau)
\leq G_I(\tilde{f}_I) - \varepsilon + \varepsilon/2 + g(\tau)
\leq G_I(\tilde{f}_I) - \varepsilon + \varepsilon/2 + 2g(\tau) \leq \tilde{G}_I(\tilde{f}_I).
\]

Here, the third inequality is the central estimate which is obtained in analogy to the first part of the computation in (49) replacing the role of the vector \(g\) there (not to be confused with \(g(\tau)\)) by that of the perturbation \(\tilde{f}_I\) of \(f\) here. The fourth inequality is obtained in analogy to the second part of the computation in (49) with the role of the vector \(g\) there replaced by the perturbation \(f_I\) of \(f\). The second before last and last inequality follow by our assumptions made on \(\tau\). Together, (63) tells us that the solution \(\tilde{u}_I\) computed by the proposed algorithm has partition \(I'\) and
\[
\tilde{u}_I = \tilde{S}_I(f).
\]

Using again Corollary 10, we have
\[
S_I(f) = \tilde{S}_I(\tilde{f}) \text{ for } \|\tilde{f} - f\| < \delta_I(\tau),
\]

with the perturbation \(\tilde{f}\) of \(f\). We have that \(\|\tilde{f} - f\| < \delta(\tau) < \delta\) with \(\delta\) defined by (45). Therefore, we may now employ Lemma 8 to conclude that the solution of the higher order Potts and Mumford-Shah problem \((\mathcal{P}_{k,\beta,\gamma})\) denoted by \(u_I\) agrees with the optimal solution for the partition \(I'\) which we have denoted by \(u_{I',I} = S_I(\tilde{f})\), i.e.,
\[
u_I = u_{I',I} = S_I(\tilde{f}).
\]

Combined with (64) and (65), this shows (57) which completes the proof. \(\square\)
Lemma 13. We consider a nonzero quadratic form \( H \) in a ball of radius \( c \) in \( \mathbb{R}^N \). Then, the Lebesgue measure \( \lambda \) of the set \( H_{x,c} = \{ x : \|x\| \leq c, c - \varepsilon < H(x) < c + \varepsilon \} \) fulfills
\[
\lambda(H_{x,c}) \leq 2 \sqrt{\frac{\varepsilon}{\|A\|}} C^{N-1}
\] (67)
where \( \|A\| \) denotes the spectral norm of the representing matrix \( A \) of \( H \).

Proof. Without loss of generality, we may use an orthogonal transformation of the coordinate system to represent \( H \) by \( H(x) = \sum_{i} \alpha_i x_i^2 \) with the eigenvalues \( \alpha_i \) of the corresponding representing matrix of the quadratic form. We sort the \( \alpha_i \) by modulus, i.e., \( |\alpha_1| \geq |\alpha_2| \geq \ldots \). With respect to this coordinate system, we consider the \( C \)-ball with respect to the infinity norm \( D = \{ x : \|x\|_{\infty} \leq C \} \). We distinguish the eigenvalue \( \alpha_1 \) of highest modulus which agrees with the norm \( \|A\| \) of the representing matrix \( A \) of \( H \). We estimate the Lebesgue measure of \( \{ x : c - \varepsilon < H(x) < c + \varepsilon \} \) on the larger set \( D \) which provides an upper bound for that of \( H_{x,c} \). To this end, we notice that, for fixed \( x_2, \ldots, x_N \), we may estimate the univariate Lebesgue measure \( \lambda^1 \) of the section \( \{ x_1 : c - \varepsilon < \tilde{H}(x_1, x_2, \ldots, x_N) < c + \varepsilon \} \) by
\[
\lambda^1\left( c + \sum_{i=2}^{N} \frac{\alpha_i}{|\alpha_1|} x_i^2 - \varepsilon < \text{sign}(\alpha_1) x_i^2 < c + \sum_{i=2}^{N} \frac{\alpha_i}{|\alpha_1|} x_i^2 + \varepsilon \right) \leq 2 \sqrt{\frac{\varepsilon}{|\alpha_1|}}.
\] (68)
(Notice if \( \alpha_1 = 0 \) the quadratic form would be zero.) Hence, on \( D \), the Lebesgue measure of \( \{ x : c - \varepsilon < H(x) < c + \varepsilon \} \) is bounded by \( 2 \sqrt{\frac{\varepsilon}{|\alpha_1|}} C^{N-1} \) which implies the assertion of the lemma. \( \square \)

Theorem 14. Let \( \varepsilon > 0 \) be given and assume that the precision \( \tau \) fulfills the assumptions of Proposition 11. We consider the set of bounded data \( \{ f : \|f\| \leq C \} \) in \( \mathbb{R}^N \) for some \( C > 0 \). Then, up to a set of Lebesgue measure \( 2(\tau \varepsilon)^N \sqrt{\frac{1}{\sup_{\|x\|} \|x\|}} C^{N-1} \), \( A_f \) given by (47), \( \sigma_{N_k} \) the number of different means to choose intervals of length at least \( k+1 \) from a (discrete) set of length \( N \), the proposed algorithm for computing a minimizer of the higher order Potts and Mumford-Shah problem \( (P_{k,\beta,\gamma}) \) is backward stable in the sense that
\[
\bar{u}_f = u_f \quad \text{where} \quad \|\tilde{f} - f\| < \delta^*(\tau),
\] (69)
where \( \delta^*(\tau) \) is given by (56). Here, \( \bar{u}_f \) is the result produced by the proposed algorithm for data \( f \) and \( u_f \) is the (unique) solution of the higher order Potts and Mumford-Shah problem \( (P_{k,\beta,\gamma}) \) for perturbed data \( \tilde{f} \).

Proof. We proceed similar to the proof of Theorem 5 to show that the set of those data which do not fulfill (44) have Lebesgue measure smaller or equal to \( 2(\tau \varepsilon)^N \sqrt{\frac{1}{\sup_{\|x\|} \|x\|}} C^{N-1} \). We choose two different partitions \( I, I' \) with \( |I| \neq |I'| \), i.e., their equivalence classes do not agree, and consider the corresponding quadratic forms \( \mathcal{G}_I, \mathcal{G}_{I'} \). Their difference \( \mathcal{G}_I - \mathcal{G}_{I'} \) is again a quadratic form (plus a constant). By Lemma 13, the set where \( \mathcal{G}_I \) and \( \mathcal{G}_{I'} \) are closer than \( \varepsilon \) has Lebesgue measure \( 2 \sqrt{\frac{\varepsilon}{|\alpha|}} C^{N-1} \). Iterating this for all \( \binom{N}{2} \) different bilinear forms \( \mathcal{G}_I - \mathcal{G}_{I'} \) shows that the Lebesgue measure of those data where (44) is not fulfilled can be estimated from above by the quantity written in the formulation of the theorem. To the complementary set, we may now apply Proposition 11 to conclude the assertion of the theorem. \( \square \)

4 Numerical study

We conduct a numerical study on the reconstruction quality of the considered higher order Mumford-Shah and Potts models, and on the computation time of the proposed solvers. We implemented the solvers for the higher order Mumford-Shah and Potts models (Algorithm 1) in C++ with wrappers to Matlab using mex-files. All experiments were conducted on a desktop computer with 3.1 GHz Intel Core i5-2400 processor and 8 GB RAM.
Figure 5: Reconstructions of “Heavy Sine”-signal from noisy data. (a) Data corrupted by Gaussian noise of level 0.2. (b–f) Reconstructions for higher order Mumford-Shah and Potts models.

Figure 6: Reconstructions of “Blocks”-signal from noisy data. (a) Data corrupted by Gaussian noise of level 0.2. (b–f) Reconstructions using higher order Mumford-Shah and Potts models.
4.1 Reconstruction results

We first investigate the potential of the higher order Mumford-Shah and Potts models with respect to reconstruction quality. We employ commonly used test signals with discontinuities; see [18, 40]. We corrupt the signals by additive zero mean Gaussian noise with variance \( \sigma^2 \). We let the noise level \( \eta \) be given by \( \eta = \sigma N / \| g \|_2 \), where \( g \) denotes the clean signal. To obtain a meaningful comparison of the models’ potentials, we determined parameters \( \beta \) and \( \gamma \) such that the result \( u^* \) has the best relative \( \ell_2 \)-error \( e_{rel} \), given by

\[
e_{rel} = \| u^* - g \|_2 / \| g \|_2.
\]

(We use a full grid search over \( \gamma = (0, 1] \) with stepsize 0.001, and \( \beta \in (0, 25] \) with stepsize 0.025 and \( \beta = \infty \).) We are further interested in the quality of the computed partition \( I^* \). A commonly used measure for segmentation quality is the Rand index [47] which we briefly explain. The Rand index \( R_{ind} \) of two partitions \( I, I' \) is given by

\[
R_{ind}(I, I') = \left( \frac{N}{2} \right) \sum_{\{i,j: 1 \leq i < j \leq N\}} t_{ij}
\]

where \( t_{ij} \) is equal to one if there are \( I \in I \) and \( I' \in I' \) such that \( i \) and \( j \) are in both \( I \) and \( I' \), or if \( i \) is in both \( I \) and \( I' \) while \( j \) is in neither \( I \) and \( I' \). Otherwise, \( t_{ij} = 0 \). Further, \( N \) denotes the length of the signal. The Rand index is bounded from above by one and a higher value means a better match. A value of one means that \( I \) and \( I' \) agree. Here, we report the Rand index \( R_{ind} \) of the computed segmentation and the ground truth segmentation.\(^1\) It is worth recalling that a high parameter \( \beta \) leads to stronger smoothing on the segments, and that a high parameter \( \gamma \) leads to less segments.

The first signal is a sinusoidal with two steps (Figure 5). We observe that the first order model requires choosing a relatively small \( \beta \) parameter to avoid the gradient limit effect, i.e., creation of spurious segments at the larger slopes. As tradeoff, the resulting signal remains visibly affected by the noise and the second discontinuity is smoothed out. Increasing the order \( k \) to values greater than one leads to better results with respect to the segmentation quality. Furthermore, the relative error improves when increasing the order. It is worth mentioning that the reconstruction quality starts decaying from order \( k = 6 \) on which can be attributed to overfitting.

The second example is a piecewise constant signal (Figure 6). As the signal has no variation on the segments, the experiment confirms the intuition that large elasticity parameters – mostly \( \beta = \infty \) – are preferable. The best result is obtained by the first order Potts model as its search space is restricted to piecewise constant functions which perfectly matches the signal. Yet, using higher order models lead to very good segmentation results up to order \( k = 5 \) and good reconstructions up to order \( k = 3 \).

Figure 7 shows the reconstruction results for a piecewise smooth signal for different noise levels. We observe that the results of the first order model remain relatively noisy on the segments. A reason for this is that the elasticity parameter needs to be relatively small to prevent the gradient limit effect. Using the second order model improves the reconstruction results significantly but also tends to produce spurious segments, in particular at the parts of high curvature. Increasing the order to \( k \geq 3 \), leads to better segmentations and improved smoothing on the segments.

An example of the effect of higher order Mumford-Shah on real data times series is given in Figure 8. The data are time-averaged (hourly) wind speeds at the summit of highest German mountain Zugspitze from November to December 2016.\(^2\) We observe that strong changes of the windspeed result in breakpoints of the higher order Mumford-Shah estimate. Some breakpoints can be associated with a meaning: the break at 492 and the two breaks near 1154 and 1182 can be linked with the days of strongest squalls in November and December 2016, respectively.\(^3\)

\(^1\)For the numerical evaluation of the Rand index, we used the implementation of K. Wang and D. Corney available at the Matlab File Exchange.

\(^2\) The data were collected by German climate data center and are available via ftp at ftp://ftp-cdc.dwd.de/pub/
Figure 7: Smoothing of piecewise defined signal of increasing noise level $\eta$ (top row). We observe that the segmentation quality gets higher and that the noise is smoothed out better on the segments when using higher order Mumford-Shah models.
Figure 8: Top: Hourly averaged windspeeds [m/s] at the summit of the Zugspitze from November to December 2016. Bottom: Result of higher order Mumford-Shah model ($\mathcal{P}_{2,15}$).

4.2 Computation time

We investigate the computation time and the number of executed error updates depending on the signal length $N$. To this end, we generate two types of synthetic signals: signals with increasing number of discontinuities and signals with constant number of discontinuities.

The signals are generated as follows. For the first type we let for each $g_i$, $i = 1, \ldots, N$, the probability of a jump discontinuity be $p = 0.01$; that is, the length of each smooth segment $I$ of $g$ follows a geometric distribution with parameter $p$. Hence, the expected segment length is $1/p = 100$, and the expected number of segments grows linearly with respect to $N$. Within a segment $I$ the signal $g$ is polynomial of degree $k - 1$ with coefficients generated by the random variables $\frac{1}{(j+1)^2} X_j$, $j = 1, \ldots, k$, where $X_j$ are i.i.d. uniformly distributed on $[-1, 1]$. For the length $h$ of $I$, the domain of $g_I$ is $[0, hp]$ sampled with step size $p$. For spline order $k$, the degree of polynomials is set to $k - 1$.

The second type of signals is created by taking $N$ equidistant samples of the continuously defined signal shown in Figure 7. In all cases, the signals are corrupted by additive Gaussian noise with noise level $\eta = 0.1$. For every considered $N$, we computed 1000 realizations and report the mean computation time and the mean number of performed error updates, respectively.

The results for the first type of signals are shown in Figure 9. It is an important observation that the runtime and the errors updates exhibit linear growth in the signal length. Thus, Algorithm 1 does not show its worst case complexity. That means, that the utilized pruning strategies are highly effective. The results for the second type of signals are shown in Table 1. In contrast to the first type, the computation time grows approximately quadratic in the number of elements, which means that Algorithm 1 attains its worst case complexity. These results suggest that an increasing number of discontinuities is beneficial for the efficiency of Algorithm 1.

5 Conclusion

We have studied higher order Mumford-Shah and Potts models. Their central advantage compared with classical first order models is that they do not penalize polynomial trends of order $k - 1$ on the segments. This leads to improved estimation for data with piecewise linear or polynomial trends.

\[\text{CDC/observations\_germany/climate/hourly/wind/historical/ (station id: 02115).}\]

\textsuperscript{3}Monatsrückblick der Wetterwarte Garmisch-Partenkirchen/Zugspitze at http://www.schneefernerhaus.de/fileadmin/web_data/bilder/pdf/MontasrueckblickeZG/MORZG1116.pdf and at http://www.schneefernerhaus.de/fileadmin/web_data/bilder/pdf/MontasrueckblickeZG/MORZG1216.pdf}
We have shown that the defining functionals have unique minimizers for almost all input signals. We have proposed a fast solver for higher order Mumford-Shah and Potts models. We have obtained stability results. We have shown that the worst case complexity of the proposed algorithm is quadratic in the length of the signal for arbitrary orders \( k \geq 1 \). In the numerical experiments, we have further observed that the runtime grows only linear for signals with linearly increasing number of discontinuities. Further, the numerical experiments confirm the robustness and stability of the proposed method. Our reference implementation processes even long signals in reasonable time; for example signals of length 10,000 need less than one second. This way, the family of higher order Mumford-Shah and Potts models can serve as efficient smoothers for signals with discontinuities.
| $k$ | $2^9$ | $2^{10}$ | $2^{11}$ | $2^{12}$ | $N$ | $2^{13}$ | $2^{14}$ | $2^{15}$ | $2^{16}$ | $2^{17}$ |
|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|
| 1   | 0.0025 | 0.0091 | 0.0363 | 0.1132 | 1.6557 | 6.5899 | 28.8285 | 123.2784 |
| 2   | 0.0025 | 0.0064 | 0.0210 | 0.0714 | 0.4875 | 1.9040 | 7.5987 | 33.0966 | 139.5842 |
| 3   | 0.0030 | 0.0075 | 0.0196 | 0.0873 | 0.3435 | 2.3475 | 9.4270 | 40.4083 | 166.7910 |
| 4   | 0.0032 | 0.0078 | 0.0569 | 0.0935 | 0.3685 | 1.6173 | 10.1401 | 44.7423 | 179.0001 |

(a) Runtime [s] for (higher order) Mumford-Shah solver

| $k$ | $2^9$ | $2^{10}$ | $2^{11}$ | $2^{12}$ | $N$ | $2^{13}$ | $2^{14}$ | $2^{15}$ | $2^{16}$ | $2^{17}$ |
|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|
| 1   | 0.0010 | 0.0030 | 0.0055 | 0.0147 | 0.0502 | 0.1627 | 0.5713 | 2.0363 | 7.1102 |
| 2   | 0.0016 | 0.0037 | 0.0090 | 0.0292 | 0.1036 | 0.3693 | 1.3562 | 4.7746 | 16.8562 |
| 3   | 0.0019 | 0.0048 | 0.0116 | 0.0423 | 0.1613 | 0.6022 | 2.2298 | 8.2050 | 30.1515 |
| 4   | 0.0024 | 0.0057 | 0.0172 | 0.0517 | 0.1977 | 0.7536 | 2.8576 | 10.8020 | 41.7345 |

(b) Runtime [s] for (higher order) Potts solver

Table 1: Mean computation times of Algorithm 1 (in seconds) for the signal from Figure 7 sampled on $N$ points and corrupted by Gaussian noise with noise level 0.1. We observe that the runtime grows approximately quadratic in $N$; that is, the worst case complexity is attained. The relevant difference to the experiment in Figure 9 is that the number of discontinuities does not increase with $N$. Yet, the solver processes signals of size $2^{16}$ in less than one minute.

Acknowledgement

This work was supported by the German Research Foundation (DFG grants STO1126/2-1 and WE5886/4-1).

References

[1] S. Amat, Z. Li, and J. Ruiz. On an new algorithm for function approximation with full accuracy in the presence of discontinuities based on the immersed interface method. *Journal of Scientific Computing*, pages 1–35, 2017.

[2] L. Ambrosio and V. M. Tortorelli. Approximation of functional depending on jumps by elliptic functional via $\gamma$-convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990.

[3] F. Arandiga, A. Cohen, R. Donat, and N. Dyn. Interpolation and approximation of piecewise smooth functions. *SIAM Journal on Numerical Analysis*, 43(1):41–57, 2005.

[4] M. Artina, M. Fornasier, and F. Solombrino. Linearly constrained nonsmooth and nonconvex minimization. *SIAM Journal on Optimization*, 23(3):1904–1937, 2013.

[5] I. Auger and C. Lawrence. Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology*, 51(1):39–54, 1989.

[6] L. Bar, N. Sochen, and N. Kiryati. Variational pairing of image segmentation and blind restoration. In *ECCV*, pages 166–177. Springer, 2004.

[7] R. Bellman and R. Roth. Curve fitting by segmented straight lines. *Journal of the American Statistical Association*, 64 (327):1079–1084, 1969.

[8] A. Blake and A. Zisserman. *Visual reconstruction*. MIT Press Cambridge, 1987.

[9] A. Blake. Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):2–12, 1989.

[10] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[11] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[12] J. Bruce. Optimum quantization. Technical report, Massachusetts Institute of Technology, 1965.

[13] M. Carriero, A. Leaci, and F. Tomarelli. A survey on the Blake–Zisserman functional. *Milan Journal of Mathematics*, 83(2):397–420, 2015.
[14] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM Journal on Applied Mathematics*, 55(3):827–863, 1995.

[15] A. Chambolle. Finite-differences discretizations of the Mumford-Shah functional. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(02):261–288, 1999.

[16] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[17] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[18] D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, pages 425–455, 1994.

[19] A. Drobyshhev, C. Machka, M. Horsch, M. Seltmann, V. Liebscher, M. de Angelis, and J. Beckers. Specificity assessment from fractionation experiments (safe): a novel method to evaluate microarray probe specificity based on hybridisation stringencies. *Nucleic acids research*, 32(2):1–10, 2003.

[20] M. Fornasier and R. Ward. Iterative thresholding achieves free-discontinuity problems. *Foundations of Computational Mathematics*, 10(5):527–567, 2010.

[21] M. Fornasier, R. March, and F. Solombrino. Existence of minimizers of the mumford-shah functional with singular operators and unbounded data. *Annali di Matematica Pura ed Applicata*, 192(3):361–391, 2013.

[22] D. Fortun, M. Storath, D. Rickert, A. Weinmann, and M. Unser. Fast piecewise-affine motion estimation without segmentation. Preprint arXiv:1802.01872.

[23] K. Frick, A. Munk, and H. Sieling. Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):495–580, 2014.

[24] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler. Complexity penalized M-estimation: Fast computation. *Journal of Computational and Graphical Statistics*, 17(1):201–224, 2008.

[25] M. Frick, A. Munk, and H. Sieling. Multiscale change point inference. *SIAM Journal on Imaging Sciences*, 7(6):3565–3597, 2014.

[26] M. Fornasier, R. March, and F. Solombrino. Existence of minimizers of the mumford-shah functional with singular operators and unbounded data. *Annali di Matematica Pura ed Applicata*, 192(3):361–391, 2013.

[27] D. Fortun, M. Storath, D. Rickert, A. Weinmann, and M. Unser. Fast piecewise-affine motion estimation without segmentation. Preprint arXiv:1802.01872.

[28] K. Frick, A. Munk, and H. Sieling. Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):495–580, 2014.

[29] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler. Complexity penalized M-estimation: Fast computation. *Journal of Computational and Graphical Statistics*, 17(1):201–224, 2008.

[30] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[31] W. Gentleman. Least squares computations by Givens transformations without square roots. *J. Inst. Maths Applies*, 12:329–336, 1973.

[32] A. Harten. Multiresolution representation of data: A general framework. *SIAM Journal on Numerical Analysis*, 33(3):1205–1256, 1996.

[33] K. Hohm, M. Storath, and A. Weinmann. An algorithmic framework for Mumford-Shah regularization of inverse problems in imaging. *Inverse Problems*, 31(11):115011, 2015.

[34] T. Hotz, O. M. Schutte, H. Sieling, T. Polupanow, U. Diederichsen, C. Steinem, and A. Munk. Idealizing ion channel recordings by a jump segmentation multiresolution filter. *IEEE Transactions on NanoBioscience*, 12(4):376–386, 2013.

[35] P. Hupé, N. Stransky, J. Thiery, F. Radvanyi, and E. Barillot. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, 20(18):3413–3422, 2004.

[36] B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. Tsai. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, 12(2):105–108, 2005.

[37] M. Jiang, P. Maass, and T. Page. Regularizing properties of the Mumford-Shah functional for imaging applications. *Inverse Problems*, 30(3):035007, 2014.

[38] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annual Review of Biochemistry*, 77:51–76, 2008.

[39] R. Killick, P. Fearnhead, and I. Eckley. Optimal detection of changepoints with a linear computational cost. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(02):261–288, 1999.

[40] A. L. Nord, E. Gachon, R. Perez-Carrasco, J. A. Nirody, A. Barducci, R. M. Berry, and F. Pedaci. Catch bond drives
stator mechanosensitivity in the bacterial flagellar motor. *Proceedings of the National Academy of Sciences*, 114(49): 12952–12957, 2017.

[44] R. Potts. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1):106–109, 1952.

[45] R. Ramlau and W. Ring. A mumford–shah level-set approach for the inversion and segmentation of x-ray tomography data. *Journal of Computational Physics*, 221(2):539–557, 2007.

[46] R. Ramlau and W. Ring. Regularization of ill-posed mumford–shah models with perimeter penalization. *Inverse Problems*, 26(11):115001, 2010.

[47] W. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846, dec 1971.

[48] L. Rondi. On the regularization of the inverse conductivity problem with discontinuous conductivities. *Inverse Problems & Imaging*, 2(3):397–409, 2008.

[49] L. Rondi and F. Santosa. Enhanced electrical impedance tomography via the mumford–shah functional. *ESAIM: Control, Optimisation and Calculus of Variations*, 6:517–538, 2001.

[50] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[51] A. Snijders, N. Nowak, R. Segraves, et al. Assembly of microarrays for genome-wide measurement of DNA copy number by CGH. *Nature Genetics*, 29:263–264, 2001.

[52] Y. Sowa, A. Rowe, M. Leake, T. Yakushi, M. Homma, A. Ishijima, and R. Berry. Direct observation of steps in rotation of the bacterial flagellar motor. *Nature*, 437(7060):916–919, 2005.

[53] M. Storath, A. Weinmann, and M. Unser. Jump-penalized least absolute values estimation of scalar or circle-valued signals. *Information and Inference*, 6(3):225–245, 2017.

[54] M. Storath and A. Weinmann. Fast partitioning of vector-valued images. *SIAM Journal on Imaging Sciences*, 7(3): 1826–1852, 2014.

[55] E. Strekalovskiy, A. Chambolle, and D. Cremers. A convex representation for the vectorial Mumford-Shah functional. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1712–1719, 2012.

[56] A. Tsai, A. Yezzi, and A. S. Willsky. Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. *IEEE transactions on Image Processing*, 10(8):1169–1186, 2001.

[57] G. Wahba. *Spline models for observational data*. SIAM, 1990.

[58] A. Weinmann and M. Storath. Iterative Potts and Blake-Zisserman minimization for the recovery of functions with discontinuities from indirect measurements. *Proceedings of the Royal Society A*, 471(2176), 2015.

[59] A. Weinmann, M. Storath, and L. Demaret. The $\ell^1$-Potts functional for robust jump-sparse reconstruction. *SIAM Journal on Numerical Analysis*, 53(1):644–673, 2015.

[60] A. Weinmann, L. Demaret, and M. Storath. Mumford-Shah and Potts regularization for manifold-valued data. *Journal of Mathematical Imaging and Vision*, 55(3):428–445, 2016.

[61] E. Whittaker. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society*, 41:63–75, 1922.

[62] J. Wilkinson. *The algebraic eigenvalue problem*. Monographs on numerical analysis. Clarendon Press, Oxford, 1965.

[63] G. Winkler. *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*, volume 27. Springer Verlag, 2003.

[64] G. Winkler and V. Liebscher. Smoothers for discontinuous signals. *Journal of Nonparametric Statistics*, 14(1-2): 203–222, 2002.

[65] G. Winkler, O. Wittich, V. Liebscher, and A. Kempe. Don’t shed tears over breaks. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 107:57–87, 2005.