Numerische
Mathematik

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

Martin Storath[1,2] · Lukas Kiefer[3,4] · Andreas Weinmann[4,5]

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 1 s.

✉ Martin Storath
martin.storath@fhws.de

1   Interdisciplinary Center for Scientific Computing (IWR), Universität Heidelberg, Heidelberg, Germany

2   Faculty of Applied Natural Sciences and Humanities, University of Applied Sciences Würzburg-Schweinfurt, Schweinfurt, Germany

3   Mathematical Imaging Group, Universität Heidelberg, Heidelberg, Germany

4   Department of Mathematics and Natural Sciences, Hochschule Darmstadt, Darmstadt, Germany

5   Institute of Computational Biology, Helmholtz Zentrum München, Munich, Germany

🪟 Springer

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 [20,32,56], the reconstruction of brain stimuli [74], single-molecule fluorescence resonance energy transfer [36], cellular ion channel functionalities [31], photo-emission spectroscopy [25] and the rotations of the bacterial flagellar motor [57]; see also [25,43,44] for further examples. 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.

One approach to this problem is to estimate the discontinuities locally and adapt the corresponding fitting operator to the local situation in an explicit way; for instance [1,3,29]. 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)/Rudin-Osher-Fatemi models [55] which yield a kind of piecewise constant approximation. To account for linear and higher order trends in the data, higher order TV methods have been proposed [11,16]. TV models and their higher order extensions, however, do not explicitly incorporate the notion of discontinuities and segment boundaries. A model taking these notions explicitly into account is the Mumford–Shah model. It simultaneously estimates a discontinuity set and a corresponding piecewise smooth approximation [47]. Its piecewise constant variant is also known as the Potts model [27,49,72], or as the Chan–Vese model for the case of two phases [17]. Mumford–Shah and Potts models are classical models for discontinuity preserving smoothing and for segmentation [8,27,46,72]. More recent applications are smoothing of video sequences [61] and segmentation with shape priors [33]. They are also used for stabilizing the reconstructions of inverse problems [35,50,51,53,54,67]. Corresponding existence results on minimizers were established in [4,23]. Besides the classical ℓ^2-based models, ℓ^p-based variants with $p \geq 1$ [22,26,30,40,58,68] and manifold-valued data spaces [58,66,69] have been considered. Discretizations have been studied in [14,15]. Mumford–Shah and Potts problems are known to be NP hard in the multivariate case and in the univariate inverse problem setup [10,63,67]. Approximate solution strategies are for example based on graduated non-convexity [8,9], approximation by elliptic functionals [2,6,54], graph cuts [10], active contours [62], convex relaxations [60], iterative thresholding algorithms [22], ADMM splitting schemes [30,59], and iterative Potts minimization [37,67]. 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 considering *higher order Mumford–Shah models*. They penalize the deviation from a piecewise polynomial instead of the

deviation from a piecewise constant function. The multivariate discrete higher order Mumford–Shah and Potts models are particularly interesting in image processing for edge preserving smoothing of images with locally linear or polynomial trends; for example, second order methods are applied for piecewise approximation and segmentation of images [18,65,77,78] or regularization of flow fields [24,76]. The univariate discrete models are particularly interesting for smoothing time series with discontinuities; examples with biological applications are for instance [48,57]. Further they arise as subproblems in univariate and multivariate splitting schemes [30,37,59,67].

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

$$(u^*, \mathcal{I}^*) = \operatorname*{argmin}_{u \in \mathbb{R}^N, \ \mathcal{I} \text{ partition}} \|u - f\|_2^2 + \sum_{I \in \mathcal{I}} \beta^{2k} \|\nabla^k u_I\|_2^2 + \gamma \, |\mathcal{I}|. \qquad (\mathcal{P}_{k,\beta,\gamma})$$

Here, $f \in \mathbb{R}^N$ denotes the given data, and the minimum is computed with respect to the target variables u and $\mathcal{I}$, where u is a discrete univariate signal of length N and $\mathcal{I}$ is a partition of the domain $\Omega = \{1, \dots, N\}$. (The connection between u and $\mathcal{I}$ is discussed in detail later in Sect. 2.2.) The symbol $\nabla^k u_I$ denotes the kth order finite difference operator applied to the vector u restricted to the "interval" I of a partition $\mathcal{I}$ 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 (measured in terms of the number of segments $|\mathcal{I}|$). The minimizing signal u^* is a piecewise kth order discrete spline approximation to f with elasticity parameter β which has discontinuities or breakpoints at the boundaries given by the partition $\mathcal{I}^*$. Choosing a large parameter value β leads to stronger smoothing on the segments, and choosing a large parameter value γ leads to less segments. It is interesting to look at the cases for very large parameters of β and γ. As the kernel of ∇^k consists of polynomials of maximum degree $k - 1$, the limit situation $\beta \to \infty$ can be written as

$$\operatorname*{argmin}_{u,\mathcal{I}} \|u - f\|_2^2 + \gamma \, |\mathcal{I}|,$$

$$\text{s.t. } u_I \text{ is a polynomial of maximum degree } k - 1 \text{ for all } I \in \mathcal{I}. \qquad (\mathcal{P}_{k,\infty,\gamma})$$

As the case $k = 1$ is known as the Potts model (as a tribute to the work of Potts [49]) we refer to $(\mathcal{P}_{k,\infty,\gamma})$ as *higher order Potts model*. On the other hand, for sufficiently large γ, it reduces to the (discrete) kth order spline approximation

$$\operatorname*{argmin}_{u} \|u - f\|_2^2 + \beta^{2k} \|\nabla^k u\|_2^2, \qquad (\mathcal{P}_{k,\beta,\infty})$$

which is a classical method for smoothing data; see [64,70]. Depending on the application, there are different points of view for Mumford–Shah-type models: On the one hand, the optimal partition $\mathcal{I}^*$ can serve as a basis for identifying segment neighborhoods [5] and as an indicator for changepoints of the signal [38]. On the other

hand, the corresponding optimal signal u^* can serve as a smoother for a signal with discontinuities [8,73].

In the literature, the members of the higher order Mumford–Shah family $(\mathcal{P}_{k,\beta,\gamma})$ have been considered for the cases $k = 1, 2$ and for $\beta < \infty$ or $\beta = \infty$ (strict piecewise polynomial model) by various individual studies. The classical (first order) Potts model $(\mathcal{P}_{1,\infty,\gamma})$ and closely related models were studied in various works; for example [12, 38,42,73,75]. The strict piecewise linear model $(\mathcal{P}_{2,\infty,\gamma})$ was studied by Bellman and Roth [7]; see also [39]. We refer to it as affine Potts model or piecewise affine Mumford–Shah model. The first order problems $(\mathcal{P}_{1,\beta,\gamma})$ for arbitrary parameters β have been studied in the seminal works of Mumford and Shah [46,47]. (This motivates the denomination higher order Mumford–Shah model for the family $(\mathcal{P}_{k,\beta,\gamma})$.) The same problem was studied at around the same time by Blake and Zisserman [8] under the name weak string model. They also introduced a second order extension, called the weak rod model [8]. This model is more general than the model $(\mathcal{P}_{2,\beta,\gamma})$ considered here because it has an extra penalty for discontinuities in the first derivative. We refer to [13,78] for a recent investigation of Blake-Zisserman models in 2D. To our knowledge, the models $(\mathcal{P}_{k,\beta,\gamma})$ have not been systematically studied for arbitrary orders k.

Although $(\mathcal{P}_{k,\beta,\gamma})$ is a non-convex problem it can be solved exactly by dynamic programming; see [5,7,9,26,34,73]. 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 [26,39,73]. Killick et al. [38] proposed a pruning strategy to accelerate the algorithm in a special yet practically relevant case: if the expected number of segments $|\mathcal{I}^*|$ grows linearly in N and if the expected log-likelihood fulfills certain estimates, detailed in [38], the expected complexity is $\mathcal{O}_P(N\phi(N))$. Another pruning scheme has been established in [59]. An algorithm for solving the first order Mumford–Shah problem for all parameters γ simultaneously was proposed in [26]. It is straightforward to devise an algorithm for $(\mathcal{P}_{k,\beta,\gamma})$ of complexity $\mathcal{O}(N^3)$. For the first order problem $(\mathcal{P}_{1,\beta,\gamma})$, we proposed an $\mathcal{O}(N^2)$ algorithm [30] which utilizes a fast computation scheme for the approximation errors proposed by Blake [9]. Unfortunately, as that scheme is based on algebraic recurrences, a generalization to arbitrary orders of k seems difficult. By precomputations of moments one can achieve $\phi(N) = \mathcal{O}(1)$; see [26,41]. Although that approach gives reasonable results for the low orders $k = 1, 2$, it gets numerically unstable for higher orders. A different dynamic programming approach was discussed in [8] which however only computes an approximate minimizer.

Contribution This work deals with the analysis and with solvers for the higher order Mumford–Shah and Potts problems $(\mathcal{P}_{k,\beta,\gamma})$. 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 1 s 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.

Organization of the paper   In Sect. 2 we describe and discuss higher order Mumford–Shah and higher order Potts models. In Sect. 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 Sect. 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

$$f_n = g_n + \eta_n, \quad n = 1, ..., N, \tag{1}$$

where the η_n are independently distributed Gaussian random variables of zero mean and variance σ^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 $x_I = x_{l:r} = (x_l, x_{l+1}, \ldots, x_r)$ for indexing. We say that $\mathcal{I}$ is a *partition of* $\Omega = 1 : N$ *into intervals*, if $I \cap J = \emptyset$ for all $I, J \in \mathcal{I}$, if $\bigcup_{I \in \mathcal{I}} I = \Omega$, and if all $I \in \mathcal{I}$ are discrete intervals of the form $I = l : r$, with $1 \leq l \leq r \leq N$. As we will only work with partitions into intervals here, we briefly call $\mathcal{I}$ 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 Ω can be written as

$$(u^*, \mathcal{I}^*) = \underset{u \in \mathbb{R}^N, \, \mathcal{I} \text{ partition of } \Omega}{\text{argmin}} \sum_{n=1}^{N} (u_n - f_n)^2$$

$$+ \beta^2 \sum_{I \in \mathcal{I}} \sum_{i=1}^{|I|-1} ((u_I)_{i+1} - (u_I)_i)^2 + \gamma \, |\mathcal{I}|. \tag{2}$$

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

$$\mathcal{I}^* = \operatorname*{argmin}_{\mathcal{I} \text{ partition of } 1:N} \sum_{I \in \mathcal{I}} \left(\mathcal{E}^I + \gamma \right),$$

$$\text{with } \mathcal{E}^I = \min_{v \in \mathbb{R}^{|I|}} \sum_{i=1}^{|I|} (v_i - f_i)^2 + \beta^2 \sum_{I \in \mathcal{I}} \sum_{i=1}^{|I|-1} (v_{i+1} - v_i)^2. \qquad (3)$$

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 [9,26].

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

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

This formulation is useful for derivation of algorithms based on iterative thresholding techniques [22]. 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 $\mathcal{I}^*$ are equivalent in the sense that $\mathcal{I}^*$ 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 Fig. 1. We observe that the solution of the first order model (using model parameters optimized to the ℓ^2 error) does 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 and Potts models

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

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

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

(a) Noisy signal with discontinuities

(b) Result of $(\mathcal{P}_{1,\beta,\gamma})$ with optimal β, γ

(c) Smaller complexity penalty γ

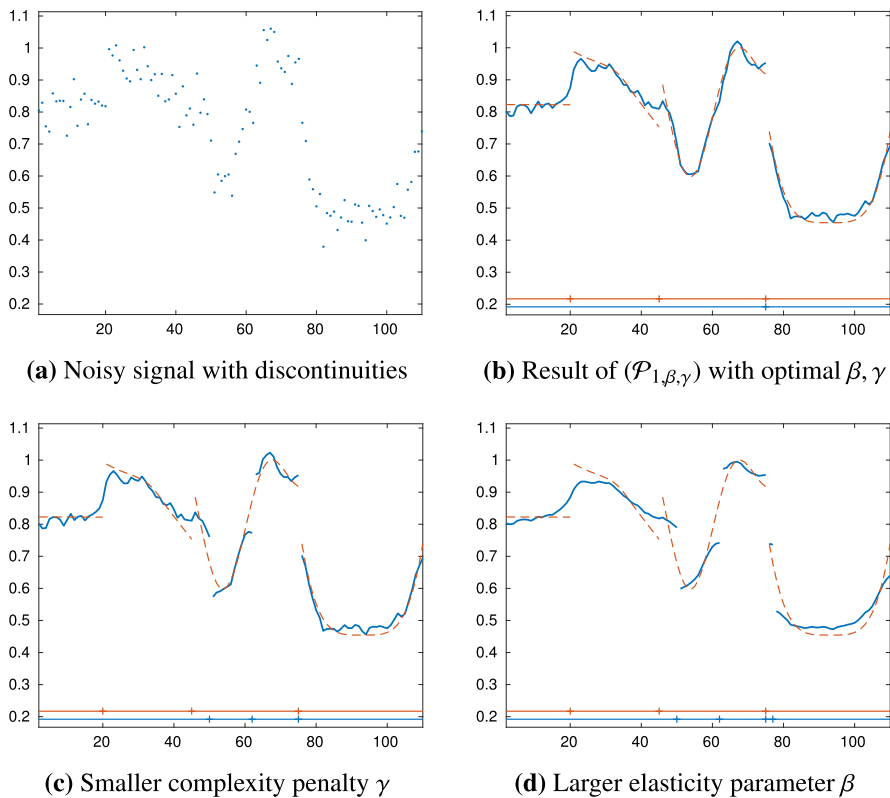**(d)** Larger elasticity parameter β

Fig. 1 Limitations of the classical (first order) Mumford–Shah model: **b–d** Reconstructions of the signal in **a** by the first order Mumford–Shah model $(\mathcal{P}_{1,\beta,\gamma})$. The red dashed lines depict the ground truth; the streaks at the bottom indicate the discontinuities or segments of the ground truth (top, red) and the computed segmentations (bottom, blue). **b** The model parameters were optimized with respect to the ℓ^2 error to the ground truth ($\gamma = 0.04$, $\beta = 1.3625$). The result provides unsatisfactory smoothing and detection of discontinuities. **c** Decreasing the complexity penalty (here $\gamma = 0.02$) leads to more but dislocated discontinuities. **d** 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 (colour figure online)

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 $(\mathcal{P}_{k,\beta,\gamma})$ can be written as

$$
(u^*, \mathcal{I}^*) = \underset{u \in \mathbb{R}^N, \, \mathcal{I} \text{ partition of } \Omega}{\mathrm{argmin}} \underbrace{\sum_{n=1}^{N} (u_n - f_n)^2}_{\text{data penalty}}
$$

$$
+ \beta^{2k} \underbrace{\sum_{I \in \mathcal{I}} \sum_{i=1}^{|I|-k} (\nabla^k u_I)_i^2}_{\text{smoothness penalty}} + \underbrace{\gamma \, |\mathcal{I}|}_{\text{complexity penalty}}. \tag{5}
$$

(a) Noisy signal with discontinuities

(b) Smoothing spline

(c) Classical Mumford–Shah ($\mathcal{P}_{1,\beta,\gamma}$)

(d) Higher order Mumford–Shah ($\mathcal{P}_{2,\beta,\gamma}$)

(e) Higher order Mumford–Shah ($\mathcal{P}_{3,\beta,\gamma}$)

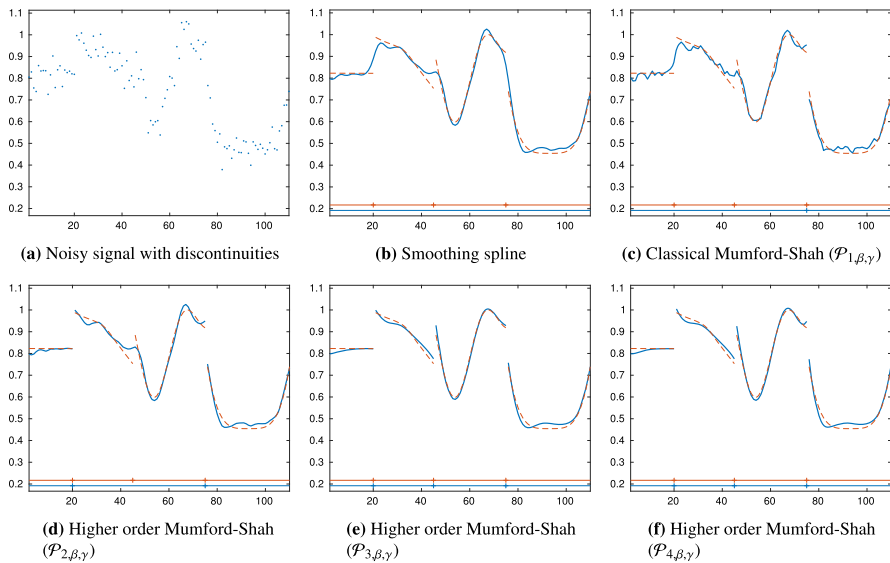**(f)** Higher order Mumford–Shah ($\mathcal{P}_{4,\beta,\gamma}$)

Fig. 2 Smoothing a noisy signal with discontinuities using various models. All model parameters are chosen with respect to optimal ℓ^2-error. As in Fig. 1, the ground truth is depicted as red dashed line and the estimate as blue solid line. **b** Classical spline approximation smoothes out the discontinuities. **c** The classical Mumford–Shah model allows for discontinuities, but the estimate misses most of them and the result remains noisy. **d–f** The higher order Mumford–Shah results provide improved smoothing and segmentation. In particular, the third and the fourth order models get the discontinuities of the groundtruth (colour figure online)

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 kth order discrete spline approximation to data f on each segment of the partition $\mathcal{I}^*$. The parameter γ determines the penalty for opening a new segment. The parameter β controls influence of the smoothness penalty. The order k is the derivative order of the (discrete) spline. Note that polynomials of order $k-1$ on a segment do not have any smoothness penalty.

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 Fig. 2.

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

$$\mathcal{I}^* = \operatorname*{argmin}_{\mathcal{I} \text{ partition of } 1:N} \sum_{I \in \mathcal{I}} (\mathcal{E}^I + \gamma), \tag{6}$$

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

$$
\begin{aligned}
\mathcal{E}^I &= \min_{v \in \mathbb{R}^{|I|}} \|v - f_I\|_2^2 + \beta^{2k} \|\nabla^k v\|_2^2 \\
&= \min_{v \in \mathbb{R}^{|I|}} \sum_{i=1}^{|I|} (v_i - (f_I)_i)^2 + \sum_{i=1}^{|I|-k} \beta^{2k} (\nabla^k v)_i^2.
\end{aligned}
\tag{7}
$$

Note that the kth order finite difference ∇^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

$$
u_I^* = \operatorname*{argmin}_{v \in \mathbb{R}^{|I|}} \|v - f_I\|_2^2 + \beta^2 \|\nabla^k v\|_2^2, \quad \text{for all } I \in \mathcal{I}^*.
\tag{8}
$$

Hence, if we have computed an optimal partition $\mathcal{I}^*$ of the domain Ω, 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 $L_{\mathcal{I}}$ by

$$
L_{\mathcal{I}} = \begin{pmatrix} L_{|\mathcal{I}_1|} & & & \\ & L_{|\mathcal{I}_2|} & & \\ & & \ddots & \\ & & & L_{|\mathcal{I}_M|} \end{pmatrix}, \quad \text{with} \quad L_n = \nabla^k \in \mathbb{R}^{(n-k)\times n},
\tag{9}
$$

and with kth order finite difference matrices ∇^k of the appropriate size $(n-k) \times n$ defined as above. Here, $|\mathcal{I}_m|$ denotes the cardinality of the mth element of the partition $\mathcal{I}$. If $n \leq k$, we use the convention that L_n is an "empty" block of length n. The number of columns of $L_{\mathcal{I}}$ 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

$$
L_{\mathcal{I}} = \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ & 1 & -2 & 1 & & & & & \\ & & & & & & 1 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 9}.
$$

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

$$
\operatorname*{argmin}_{u \in \mathbb{R}^N, \, \mathcal{I} \text{ partition}} \beta^{2k} \|L_{\mathcal{I}} u\|_2^2 + \|u - f\|_2^2 + \gamma \, |\mathcal{I}|.
\tag{10}
$$

For a fixed partition $\mathcal{I}$, taking derivatives with respect to u reveals that a minimizer $u_{\mathcal{I}}$ satisfies the linear system

$$2\beta^{2k}L_{\mathcal{I}}^{T}L_{\mathcal{I}}u_{\mathcal{I}} + 2(u_{\mathcal{I}} - f) = 0. \tag{11}$$

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

$$u_{f,\mathcal{I}} = S_{\mathcal{I},\beta}f, \qquad \text{where } S_{\mathcal{I},\beta} = (\beta^{2k}L_{\mathcal{I}}^{T}L_{\mathcal{I}} + \mathrm{id})^{-1}. \tag{12}$$

We omit the subscript if the dependence on $\mathcal{I}$ or β 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 $\mathcal{I}$ as

$$\mathcal{G}'_{\mathcal{I}}(f) = \beta^{2k}\|L_{\mathcal{I}}S_{\mathcal{I}}f\|_{2}^{2} + \|S_{\mathcal{I}}f - f\|_{2}^{2} + \gamma\,|\mathcal{I}|. \tag{13}$$

Hence, the minimizing partition $\mathcal{I}^{*}$ is given as the minimizing argmuent of $\mathcal{I} \mapsto \mathcal{G}'_{\mathcal{I}}(f)$.

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 partitions with different numbers of segments. A simple example is the data $f = (0, 1, 0)$. For $\gamma < 2/3$ and for β sufficiently large, a minimizer is given by $u = f$. The partitions $\{(1, 2), (3)\}$, $\{(1), (2, 3)\}$ 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 $\mathcal{I}^{*}$ be a minimizing partition of (5). Then the minimal functional value is given by*

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

Proof Let $\tilde{u} = S_{\mathcal{I}^{*},\beta}f$. Expanding the functional yields

$$\begin{aligned}
\mathcal{G}'_{\mathcal{I}^{*}}(f) &= \beta^{2k}\|L_{\mathcal{I}^{*}}\tilde{u}\|_{2}^{2} + \|\tilde{u} - f\|_{2}^{2} + \gamma\,|\mathcal{I}^{*}| \\
&= \beta^{2k}\tilde{u}^{T}L_{\mathcal{I}^{*}}^{T}L_{\mathcal{I}^{*}}\tilde{u} + (\tilde{u} - f)^{T}(\tilde{u} - f) + \gamma\,|\mathcal{I}^{*}| \\
&= \beta^{2k}\tilde{u}^{T}L_{\mathcal{I}^{*}}^{T}L_{\mathcal{I}^{*}}\tilde{u} + \tilde{u}^{T}(\tilde{u} - f) + f^{T}(\tilde{u} - f) + \gamma\,|\mathcal{I}^{*}| \\
&= \tilde{u}^{T}(\beta^{2k}L_{\mathcal{I}^{*}}^{T}L_{\mathcal{I}^{*}}\tilde{u} + (\tilde{u} - f)) - f^{T}(\tilde{u} - f) + \gamma\,|\mathcal{I}^{*}| \\
&= -f^{T}(\tilde{u} - f) + \gamma\,|\mathcal{I}^{*}| = \|f\|_{2}^{2} - f^{T}\tilde{u} + \gamma\,|\mathcal{I}^{*}|,
\end{aligned}$$

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

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

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

$$\sum_{I' \in \mathcal{I}'} \mathcal{E}^{I'} \le \sum_{I \in \mathcal{I}} \mathcal{E}^I \quad and \quad |\mathcal{I}'| \le |\mathcal{I}|.$$

In particular $\mathcal{G}'_{\mathcal{I}'}(f) \le \mathcal{G}'_{\mathcal{I}}(f)$.

Proof Let $\mathcal{I}$ be a partition and let I be its right-most segment such that $|I| < k$. 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 $\mathcal{I}'$. (If the neighboring segment gets empty, we remove it from the partition.) On the one hand, $|\mathcal{I}'| \le |\mathcal{I}|$. On the other hand, since $|I \cup \{i - 1\}| \le k$ we have that $\mathcal{E}^{I \cup \{i-1\}} = 0$. Repeating the above procedure a finite number of times, we end up with a partition $\mathcal{I}''$ whose segments have length greater or equal to k, except possibly the leftmost segment. □

Higher order Potts models As mentioned in the introduction, the higher order Potts model $(\mathcal{P}_{k,\infty,\gamma})$ 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 kth order spline. In consequence, the approximation error on a segment $I = l : r$ is given by

$$\mathcal{E}^I = \min_{\substack{v \text{ polynomial of} \\ \text{degree } \le k-1 \text{ on } I}} \|v - f_I\|_2^2. \tag{14}$$

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.

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.

2.3 Existence and uniqueness of minimizers

It is straightforward to show the existence of minimizers.

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

Proof For a fixed partition $\mathcal{I}$, the problem ($\mathcal{P}_{k,\beta,\gamma}$) reduces to least squares problems on the intervals of $\mathcal{I}$ which all possess minimizers. As there are only finitely many partitions on Ω, 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 ($\mathcal{P}_{k,\beta,\gamma}$) 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 $\mathcal{I}^3 = \{(1), (2), (3)\}$ is given by $u^3 = f$ and it has the functional value 2γ. 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 γ. One can show that the the partition $\mathcal{I}^1 = \{(1, 2, 3)\}$ yields the signal $u^1 = \frac{1}{1+6\beta^4}(2\beta^4, 1 + 2\beta^4, 2\beta^4)^T$ and that the functional value is given by $\frac{4\beta^4}{1+6\beta^4}$. Setting this equal to the energy of the two-segment solution, γ, gives us the critical value $\frac{4\beta^4}{1+6\beta^4} = \gamma$ which is equivalent to $\beta^4 = \gamma/(4-6\gamma)$. 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 $\mathcal{I}, \mathcal{J}$ are equivalent, i.e.,

$$\mathcal{I} \sim \mathcal{J} \quad :\Leftrightarrow \quad (I \in \mathcal{I} \text{ and } |I| > k \Rightarrow I \in \mathcal{J} \quad \text{and} \quad J \in \mathcal{J} \text{ and } |J| > k \Rightarrow J \in \mathcal{J}), \tag{15}$$

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

$$u_{f,\mathcal{I}} = S_{\mathcal{I},\beta} f = S_{\mathcal{J},\beta} f = u_{f,\mathcal{J}} \tag{16}$$

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

$$\begin{aligned} \text{both assigments} \quad & [\mathcal{I}] \to L_{\mathcal{I}}, \ [\mathcal{I}] \to u_{f,\mathcal{I}} \\ \text{are well-defined, and} \quad & [\mathcal{I}] \to L_{\mathcal{I}} \quad \text{is one-to-one.} \end{aligned} \tag{17}$$

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

$$\operatorname*{argmin}_{u \in \mathbb{R}^N, [\mathcal{I}]} F_{[\mathcal{I}]}(u), \quad \text{where} \quad F_{[\mathcal{I}]}(u) = \beta^{2k} \|L_{\mathcal{I}} u\|_2^2 + \|u - f\|_2^2 + \gamma |[\mathcal{I}]|. \tag{18}$$

Here, we let

$$|[\mathcal{I}]| = \min_{\mathcal{J} \in [\mathcal{I}]} |\mathcal{J}|.$$

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

$$\mathcal{G}_{[\mathcal{I}]}(f) = \beta^{2k} \|L_{\mathcal{I}} S_{\mathcal{I}} f\|_2^2 + \|S_{\mathcal{I}} f - f\|_2^2 + \gamma |[\mathcal{I}]|. \qquad (19)$$

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

Theorem 5 *Let* $\gamma > 0$, $\beta \in (0, \infty]$, *and* $k \in \mathbb{N}$. *The minimizer* u^* *of* $(\mathcal{P}_{k,\beta,\gamma})$ *is unique for almost all input data* f.

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

$$
\begin{aligned}
f \in \mathcal{F} \quad :&\Leftrightarrow \quad \text{there is a partition } \mathcal{I}^* \text{ such that} \\
& F_{[\mathcal{I}^*]}(u_{f,\mathcal{I}^*}) < F_{[\mathcal{I}]}(u_{f,\mathcal{I}}) \text{ for all } \mathcal{I} \notin [\mathcal{I}^*].
\end{aligned} \qquad (20)
$$

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 $\mathcal{I}$, we get that the minimal function value constraint to $[\mathcal{I}]$ for data f as is given by $\mathcal{G}_{[\mathcal{I}]}(f)$ defined in (19) as $\mathcal{G}_{[\mathcal{I}]}(f) = \beta^{2k} \|L_{\mathcal{I}} S_{\mathcal{I}} f\|_2^2 + \|S_{\mathcal{I}} f - f\|_2^2 + \gamma |[\mathcal{I}]|$. Hence, $\mathcal{F}^C \subset \mathcal{H}$, where

$$\mathcal{H} = \{f : \text{ there are } \mathcal{I}, \mathcal{I}' \text{ with } [\mathcal{I}] \neq [\mathcal{I}'] \text{ such that } \mathcal{G}_{[\mathcal{I}]}(f) - \mathcal{G}_{[\mathcal{I}']}(f) = 0\}. \quad (21)$$

For fixed $\mathcal{I}, \mathcal{I}'$, both $\mathcal{G}_{\mathcal{I}}, \mathcal{G}_{\mathcal{I}'}$ are quadratic forms w.r.t. the input f. Since $[\mathcal{I}] \neq [\mathcal{I}']$ we have by (17) that the quadratic form $f \mapsto \mathcal{G}_{[\mathcal{I}]}(f) - \mathcal{G}_{[\mathcal{I}']}(f)$ is nonzero. Therefore, by the Morse–Sard theorem, the set $\{f : \mathcal{G}_{[\mathcal{I}]}(f) - \mathcal{G}_{[\mathcal{I}']}(f) = 0\}$ has Lebesgue measure zero. Forming the finite union w.r.t. $\mathcal{I}, \mathcal{I}'$, 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. $\qquad\square$

2.4 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

$$(u^*, \mathcal{I}^*) = \operatorname*{argmin}_{u \in \mathbb{R}^N, |\mathcal{I}| \leq J} \|u - f\|_2^2 + \beta^{2k} \sum_{I \in \mathcal{I}} \|\nabla^k u_I\|_2^2. \qquad (\mathcal{C}_{k,\beta,J})$$

In fact, both variants are closely related: Let us denote by $(u^J, \mathcal{I}^J)$ a solution of $(\mathcal{C}_{k,\beta,J})$ 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 $(u^{J^*}, \mathcal{I}^{J^*})$ with the optimal functional value in $(\mathcal{P}_{k,\beta,\gamma})$; that is,

$$J^* \in \operatorname*{argmin}_{J=1,\ldots,N} \gamma J + \|u^J - f\|_2^2 + \beta^{2k} \sum_{I \in \mathcal{I}^J} \|\nabla^k u_I^J\|_2^2.$$

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 $(\mathcal{C}_{k,\beta,J})$ 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 γ where the corresponding solution does not change.

Relations to ℓ_0-penalized problems The classical first order Potts model $(\mathcal{P}_{1,\infty,\gamma})$ can also be written in terms of ℓ_0-"norm" of the target variable u as

$$u^* = \operatorname*{argmin}_{u \in \mathbb{R}^N} \gamma \|\nabla u\|_0 + \|u - f\|_2^2, \tag{22}$$

where $\|v\|_0$ denotes the number of non-zero elements of a vector; that is $\|v\|_0 = |\{n : v_n \neq 0\}|$. We point out that plugging ∇^k in (22) does *not* lead to an equivalent of the higher order Potts model $(\mathcal{P}_{k,\infty,\gamma})$; that is, in general for $k \geq 2$

$$\operatorname*{argmin}_{w \in \mathbb{R}^N} \gamma \|\nabla^k w\|_0 + \|w - f\|_2^2, \neq u^*, \quad \text{where } u^* \text{ is the minimizer of } (\mathcal{P}_{k,\infty,\gamma}).$$
$$\tag{23}$$

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 $\mathcal{I}^2 = \{(1, 2), (3, 4)\}$ is given by $u_{\mathcal{I}^2} = f$ 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 $\mathcal{I}^1 = \{(1, 2, 3, 4)\}$ is given by $u_{\mathcal{I}^1} = (-\frac{6}{5}, -\frac{2}{5}, \frac{2}{5}, \frac{6}{5})$ so that $\mathcal{E}^{(1:4)} = \frac{4}{5}$. The functional values are given by $\frac{4}{5} + \gamma$ for $(\mathcal{I}^1, u_{\mathcal{I}^1})$ and by 2γ for $(\mathcal{I}^2, u_{\mathcal{I}^2})$. Hence, $(\mathcal{P}_{2,\infty,\gamma})$ has the solution $(\mathcal{I}^1, u_{\mathcal{I}^1})$ for $\gamma > \frac{4}{5}$, and $(\mathcal{I}^2, u_{\mathcal{I}^2})$ for $\gamma < \frac{4}{5}$. (They are both optimal for $\gamma = \frac{4}{5}$.) In contrast, as $\|\nabla^2 u_{\mathcal{I}^1}\|_0 = 0$ and $\|\nabla^2 u_{\mathcal{I}^2}\|_0 = \|(2, -2)\|_0 = 2$, the critical value for the model in (23) is $\gamma = \frac{2}{5}$. Thus, the solutions of (23) for $k = 2$ and $(\mathcal{P}_{2,\infty,\gamma})$ are different for $\gamma \in (\frac{2}{5}, \frac{4}{5})$. The intuition behind that difference is that in (23) 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. The model in (23) was studied in [21] for the case $k = 2$.

It was observed in [24] that the second order Potts model can be 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 Ω 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:

$$C^* = \operatorname*{argmin}_{C} \ \gamma \, \|\nabla C\|_0 + \sum_{n=1}^{N} \left((1, n, \dots, n^{k-1}) \, C(n) - f_n \right)^2. \tag{24}$$

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

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.,

$$P(\mathcal{I}) = \sum_{I \in \mathcal{I}} (\mathcal{E}^I + \gamma).$$

Note that the functional is well-defined also for a partition $\mathcal{I}$ 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

$$P_r^* = \min_{\mathcal{I} \text{ partition on } 1:r} P(\mathcal{I}).$$

The value P_r^* for the domain $1 : r$ satisfies the Bellman equation

$$P_r^* = \min_{l=1,\dots,r} \left\{ \mathcal{E}^{l:r} + \gamma + P_{l-1}^* \right\}, \tag{25}$$

where we let $P_0^* = 0$. Recall that $\mathcal{E}^{1:r} = 0$ if $r - l + 1 \leq k$, so the minimum on the right hand side actually only has to be taken over the values $l = 1, \dots, r - k$. 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 $\mathcal{I}^*$, 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 [26].

The above procedure has the complexity $\mathcal{O}(N^2\phi(N))$ where ϕ is an upper bound for the effort of computing the approximation errors $\mathcal{E}^I$. The straightforward way to compute $\mathcal{E}^I$ is solving the least squares system (7). This leads to $\phi(N) = \mathcal{O}(N)$ as the involved matrices have a band structure. We develop a strategy that achieves $\phi(N) = \mathcal{O}(1)$ in the next section.

In the following, we recall two strategies from [38,59], respectively, to prune the search space. In [59], 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

$$P_r < \mathcal{E}^{l:r} + \gamma \tag{26}$$

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 $\mathcal{E}^{l:s} + \mathcal{E}^{s+1:r} \leq \mathcal{E}^{l:r}$, for all $l \leq s < r$. Killick et al. [38] deduced that if

$$P_s^* \leq P_l^* + \mathcal{E}^{l+1:s}, \tag{27}$$

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

$$\mathcal{E}^{1:N} = \min_{v \in \mathbb{R}^N} \|Av - y\|_2^2. \tag{28}$$

Here,

$$A = \begin{pmatrix} E_N \\ \beta^k \nabla^k \end{pmatrix} \in \mathbb{R}^{(2N-k) \times N} \quad \text{and} \quad y = (f_1, \ldots, f_N, 0)^T \in \mathbb{R}^{2N-k}, \tag{29}$$

and E_N represents the identity matrix of dimension N. Further, determining $\mathcal{E}^{1:r}$ for $r < N$ amounts to solving the least squares problem of smaller size

$$\mathcal{E}^{1:r} = \min_{v \in R^r} \left\| A^{(r)} v - y^{(r)} \right\|_2^2, \tag{30}$$

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 $\mathcal{E}^{1:r}$. Instead, we develop a recurrence formula computing $\mathcal{E}^{1:r}$ 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 ℓ^2-norm is invariant to orthogonal transformations we may represent $\mathcal{E}^{1:r}$ as

$$\begin{aligned} \mathcal{E}^{1:r} &= \min_{v \in \mathbb{R}^r} \left\| \begin{pmatrix} R^{(r)} \\ 0 \end{pmatrix} v - (Q^{(r)})^T y^{(r)} \right\|_2^2 \\ &= \min_{v \in \mathbb{R}^r} \| R^{(r)} v - ((Q^{(r)})^T y^{(r)})_{1:r} \|_2^2 + \| ((Q^{(r)})^T y^{(r)})_{r+1:2r-k} \|_2^2 \\ &= \| ((Q^{(r)})^T y^{(r)})_{r+1:2r-k} \|_2^2. \end{aligned} \tag{31}$$

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 $\mathcal{E}^{1:r+1}$ 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:

$$W^{(r)} = \begin{pmatrix} R^{(r)} & 0 \\ 0 & 1 \\ 0 & 0 \\ A_{N+r-k+1,(1:r+1)} \end{pmatrix}.$$

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 2×2 submatrix $(G_{jj}, G_{jm}; G_{mj}, G_{mm})$ replaced by a planar rotation matrix; that is,

$$G(j, m, \theta) = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & \cos(\theta) & \cdots & \sin(\theta) & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -\sin(\theta) & \cdots & \cos(\theta) & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}, \tag{32}$$

where θ 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

$$\cos(\Theta_{mj}) = b/\rho, \quad \sin(\Theta_{mj}) = a/\rho \tag{33}$$

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

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

$$q^{(r)} = (Q^{(r)})^T y^{(r)}.$$

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

$$q^{(r+1)} = G^{(r+1)} \begin{pmatrix} q_{1:r}^{(r)} \\ f_{r+1} \\ q_{r+1:2r-k}^{(r)} \\ 0 \end{pmatrix}, \tag{34}$$

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

$$G^{(r+1)} = \prod_{j=1}^{k+1} G\left(r - k + j, 2(r + 1) - k, \Theta_{N+r+1-k,r-k+j}\right), \tag{35}$$

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

$$\|q_{r+1:2r+1-k}^{(r+1)}\|_2^2 = \|q_{r+1:2r-k}^{(r)}\|_2^2 + (q_{2(r+1)-k}^{(r+1)})^2. \tag{36}$$

Therefore, the error update is given by

$$\mathcal{E}^{1:r+1} = \mathcal{E}^{1:r} + (q_{2(r+1)-k}^{(r+1)})^2. \tag{37}$$

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

The errors $\mathcal{E}^{l:r}$ can be updated in the same fashion by applying the above procedure to the data $\tilde{f} = (f_l, \ldots f_r)$. An important practical aspect is that the recurrence coefficients Θ_{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 $\mathcal{E}^{l:r}$.

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 ∇^k the approximation errors $\mathcal{E}^{1:r}$ are exact equal 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 Sect. 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 $\mathcal{E}^{l':r}$ has not been computed yet. In order to overcome this issue without 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 $\mathcal{E}^{l:r}$. 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
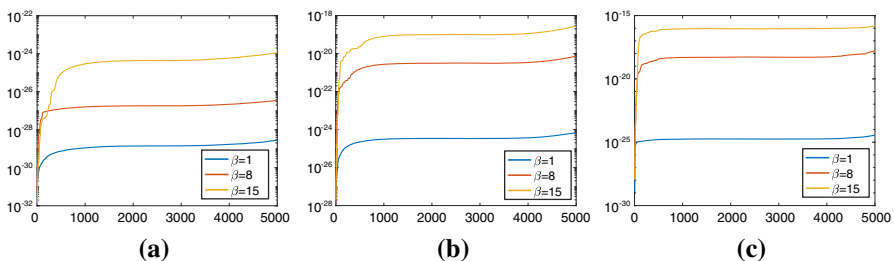

Fig. 3 Approximation errors $\mathcal{E}^{1:r}$, $r = 1, \ldots, N$, for **a** a linear polynomial and order $k = 2$, for **b** a quadratic polynomial and order $k = 3$, and for **c** a cubic polynomial and order $k = 4$. The theoretical approximation errors are equal to zero; the approximation errors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

the pruning strategies we effectively decrease the total number of error updates (37) that have to be performed; see Sect. 4.2 for a numerical study.

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

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

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 $\mathcal{E}^{r-k:r}$, $\mathcal{E}^{r-k-1:r}$, ..., $\mathcal{E}^{1:r}$ 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 $\mathcal{O}(k^2 N)$ 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 $\mathcal{O}(k)$. It follows that computing the errors for all $\mathcal{O}(N^2)$ intervals sums up to $\mathcal{O}(kN^2)$. The reconstruction step from a partition is in $\mathcal{O}(kN)$ 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 $\mathcal{O}(N^2)$. □

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

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

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

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

As in Sect. 3.2, we describe the method for the prototypical case $l = 1$. Furthermore, we assume that $r > k$ since otherwise $\mathcal{E}^{1:r} = 0$. 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

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

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 $B_{r+1,j}$ with the pivot element $R_{j,j}^{(r)}$. Now assume that we have computed $q^{(r)}$ and $\mathcal{E}^{1:r}$. Then, $q^{(r+1)}$ can be expressed by the recurrence relation

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

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

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

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 $q^{(k)}$, we obtain by (40)

$$\mathcal{E}^{1:r+1} = \|q_{k+1:r+1}^{(r+1)}\|_2^2 = \|q_{k+1:r}^{(r)}\|_2^2 + (q^{(r+1)})_{r+1}^2 = \mathcal{E}^{1:r} + (q^{(r+1)})_{r+1}^2. \tag{43}$$

Remark 7 For the higher order Potts problems, there are also closed formulae for the evaluation of the errors $\mathcal{E}^{l:r}$ which one might consider to use directly. Such formulae
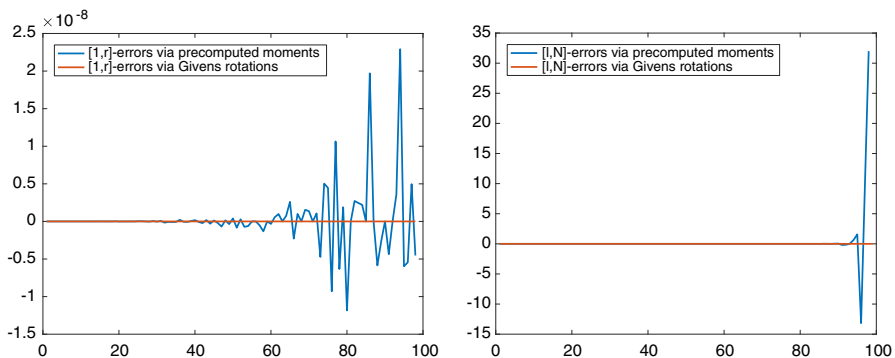

Fig. 4 Approximation errors for a parabolic signal for the higher order Potts model of order $k = 3$. The graphs show the results based on precomputed moments (blue) and based on the proposed scheme (red). The true approximation errors are all equal to zero. Left: errors $\mathcal{E}^{1:r}$ for $r = 1, \ldots, N$. The computation based on precomputed moments is distorted beyond machine precision and gives even negative values. Right: errors $\mathcal{E}^{l:N}$ for $l = 1, \ldots, N - 1$. 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 (colour figure online)

are derived in [26,41] 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 $\mathcal{E}^{l:r}$ can then be computed in $\mathcal{O}(1)$ 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 precomputation of moments is prone to numerical instability. This is illustrated by the following experiment (cf. Fig. 4). We consider the parabolic signal $f_n = n^2/100$, $n = 0, ..., N$ where $N = 100$. The true approximation errors for the higher order Potts model of order $k = 3$ are given by $\mathcal{E}^{l:r} = 0$ for all l, r with $1 \leq l \leq r \leq N$. Figure 4 shows that the results for $\mathcal{E}^{1:r}$ are distorted when using the approach based on the precomputation of moments, in particular if r is close to N. The errors $\mathcal{E}^{l:N}$ 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 $\mathcal{G}_{[\mathcal{I}]}(f)$ defined in (19) and omit the brackets and simply write $\mathcal{G}_{\mathcal{I}}(f)$.

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

$$\mathcal{G}_{\mathcal{I}'}(f) < \mathcal{G}_{\mathcal{I}}(f) - \varepsilon \quad \text{for all} \quad \mathcal{I} \notin [\mathcal{I}'] \tag{44}$$

then there is an euclidean δ-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 $u = u_{g, \mathcal{I}^*}$ of the problem $(\mathcal{P}_{k,\beta,\gamma})$, and the corresponding partition $\mathcal{I}^*$ fulfills $[\mathcal{I}^*] = [\mathcal{I}']$. We may choose δ by

$$\delta := \min\left(\frac{\varepsilon}{2(2\beta)^{2k}\left(\|f\| + 1/2\right)}, \frac{1}{2}\right). \tag{45}$$

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

$$\bar{\mathcal{G}}_{\mathcal{I}}(f) := \beta^{2k} \|L_{\mathcal{I}} S_{\mathcal{I}} f\|_2^2 + \|S_{\mathcal{I}} f - f\|_2^2 \tag{46}$$

which are the main parts of the $\mathcal{G}_{\mathcal{I}}$ given by (19). Each $\bar{\mathcal{G}}_{\mathcal{I}}$ may be represented w.r.t. the euclidean standard scalar product $\langle \cdot, \cdot \rangle$ via a symmetric matrix $A_{\mathcal{I}}$ as $\bar{\mathcal{G}}_{\mathcal{I}}(x) = \langle A_{\mathcal{I}} x, x \rangle$. The operator norm of $A_{\mathcal{I}}$ equals the norm of the corresponding bilinear form which in turn, since the $\bar{\mathcal{G}}_{\mathcal{I}}$ are positive (semi-definite), corresponds to

$$\|A_{\mathcal{I}}\| = \sup_{x:\|x\|=1} \bar{\mathcal{G}}_{\mathcal{I}}(x). \tag{47}$$

We first let δ' be defined by

$$\delta' := \min\left(\frac{1}{2}, \frac{\varepsilon}{2\max_{\mathcal{I}}\|A_{\mathcal{I}}\|\,(\|f\| + 1/2)}\right). \tag{48}$$

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

$$
\begin{aligned}
\mathcal{G}_{\mathcal{I}'}(g) = \bar{\mathcal{G}}_{\mathcal{I}'}(g) + \gamma|[\mathcal{I}']| &= \bar{\mathcal{G}}_{\mathcal{I}'}(f) + \bar{\mathcal{G}}_{\mathcal{I}'}(f - g) - 2\langle A'_{\mathcal{I}}f, f - g\rangle + \gamma|[\mathcal{I}']| \\
&< \bar{\mathcal{G}}_{\mathcal{I}}(f) - \varepsilon + \delta'^2\|A_{\mathcal{I}'}\| + \delta'\|A_{\mathcal{I}'}\|\|f\| + \gamma|[\mathcal{I}]| \\
&= \bar{\mathcal{G}}_{\mathcal{I}}(g) + \bar{\mathcal{G}}_{\mathcal{I}}(f - g) + 2\langle A_{\mathcal{I}}f, f - g\rangle \\
&\quad - \varepsilon + \delta'^2\|A_{\mathcal{I}'}\| + \delta'\|A_{\mathcal{I}'}\|\|f\| + \gamma|[\mathcal{I}]| \\
&\leq \bar{\mathcal{G}}_{\mathcal{I}}(g) - \varepsilon + 2\delta'\max_{\mathcal{I}}\|A_{\mathcal{I}}\|\,(\|f\| + \delta') + \gamma|[\mathcal{I}]| \\
&\leq \bar{\mathcal{G}}_{\mathcal{I}}(g) - \varepsilon + \varepsilon + \gamma|[\mathcal{I}]| = \bar{\mathcal{G}}_{\mathcal{I}}(g) + \gamma|[\mathcal{I}]| = \mathcal{G}_{\mathcal{I}}(g).
\end{aligned} \tag{49}
$$

For the first inequality, we applied (44) for $\mathcal{G}_{\mathcal{I}'}(f)$ 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 $\max_{\mathcal{I}}\|A_{\mathcal{I}}\|$ using basic spectral theory for self-adjoint bounded operators. Since $A_{\mathcal{I}}$ is the matrix representing the bilinear form corresponding to $\bar{\mathcal{G}}_{\mathcal{I}}$, we may estimate using (46) that

$$\|A_{\mathcal{I}}\| \leq \beta^{2k}\|L_{\mathcal{I}}^T L_{\mathcal{I}}\|\,\|S_{\mathcal{I}}\|^2 + \|I - S_{\mathcal{I}}\|^2, \quad \text{for any partition } \mathcal{I}, \tag{50}$$

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

$$\|A_{\mathcal{I}}\| \leq \beta^{2k}2^{2k} + 1, \quad \text{for any partition } \mathcal{I}. \tag{51}$$

We now can 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 $\mathcal{I}'$ which shows the assertion. $\square$

Next, we need a backward stability result for the QR algorithm [28]. 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 $\tilde{S}_{l:r}$ needed for computing the $\mathcal{E}^{l:r}$ is backward stable, i.e., given data $f_{l:r}$ living on the subinterval $l:r$, there is a perturbation $\tilde{f}_{l:r}$ of $f_{l:r}$ such that*

$$\tilde{S}_{l:r}(f_{l:r}) = S_{l:r}(\tilde{f}_{l:r}) \quad with \quad \|\tilde{f}_{l:r} - f_{l:r}\| \le \delta_{l:r}, \tag{52}$$

where $\delta_{l:r}$ depends on the machine precision τ and on the norm $\|f_{l:r}\|$ via

$$\delta_{l:r} \le 6\tau\sqrt{r-l+1} \cdot \left(\frac{9(r-l+1)-5}{4} - k\right)(1+6\tau)^{3(r-l)-k}\|f_{l:r}\|. \tag{53}$$

Here, the QR algorithm is understood as in the analysis setup of [28,71].

Proof If $r - l < k$, then $\tilde{S}_{l:r}(f_{l:r}) = S_{l:r}(f_{l:r}) = 0$. So we may assume $r - l \ge k$. Recall that calculating $\mathcal{E}^{l:r}$ corresponds to computing the residual vector of the least squares problem with system matrix $A \in \mathbb{R}^{2(r-l+1)-k \times (r-l+1)}$ from (29) and data $f_{l:r}$. In [28] it is shown that

$$\tilde{R} = \bar{Q}^T(A + \Delta A), \quad \|\Delta A\|_F \le \mu_{l:r}(\tau)\|A\|_F,$$
$$\mu_{l:r}(\tau) = 6\tau\sqrt{r-l+1} \cdot \left(\frac{9(r-l+1)-5}{4} - k\right)(1+6\tau)^{3(r-l)-k},$$

where $\tilde{R}$ is the computed upper triangular matrix by means of Givens rotations and note that $\bar{Q}^T$ is the orthogonal matrix that is the product of exact Givens rotations we apply. Analogously, for the data vector it is shown in [28] that

$$\widetilde{Q^T f_{l:r}} = \bar{Q}^T(f_{l:r} + \Delta f_{l:r}), \quad \|\Delta f_{l:r}\| \le \mu_{l:r}(\tau)\|f_{l:r}\|,$$

hence $\tilde{S}_{l:r}(f_{l:r}) = S_{l:r}(f_{l:r} + \Delta f_{l:r})$ which implies (53). $\qquad\square$

Corollary 10 *We consider bounded data $f \in \mathbb{R}^N$, $\|f\| < C$. For any partition $\mathcal{I}$ considered in the proposed algorithm for the higher order Potts and Mumford–Shah problem, there is a perturbation $\tilde{f}$ of f such that*

$$\tilde{S}_{\mathcal{I}}(f) = S_{\mathcal{I}}(\tilde{f}) \quad where \quad \|\tilde{f} - f\| < \delta_{\mathcal{I}}(\tau), \tag{54}$$

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

$$\delta_{\mathcal{I}}(\tau)^2 \le 36C^2\tau^2 \sum_i (r_i - l_i + 1)\left(\frac{9(r_i - l_i + 1)-5}{4} - k\right)^2 (1+6\tau)^{6(r-l)-2k}. \tag{55}$$

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

For the formulation of the next statement, we use the notation $\tilde{\mathcal{G}}_\mathcal{I}$ to denote the algorithm to compute the energy $\mathcal{G}_\mathcal{I}$ given by (19). Further, we use the notation $g(\tau)$ to bound the approximation error between $\tilde{\mathcal{G}}_\mathcal{I}$ and $\mathcal{G}_\mathcal{I}$ for all $\mathcal{I}$ in dependence of the precision τ.

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

$$\delta^*(\tau) = \max_{\mathcal{I}} \delta_\mathcal{I}(\tau) \le 6C\tau N^{\frac{3}{2}} \left(\frac{9N-5}{4} - k \right)(1+6\tau)^{3(N-1)-k} \qquad (56)$$

for $\delta_\mathcal{I}(\tau)$ in Corollary 10 and assume that τ is small enough such that $\delta^*(\tau) < \delta/2$ with δ given by (45) and such that $g(\tau) \le \varepsilon/4$ with ε given in (44). Then, the higher order Potts and Mumford–Shah problem $(\mathcal{P}_{k,\beta,\gamma})$ has a unique minimizer u_f, 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

$$\tilde{u}_f = u_{\tilde{f}} \quad \text{where} \quad \|\tilde{f} - f\| < \delta^*(\tau). \qquad (57)$$

Here, $\tilde{u}_f$ is the result produced by the proposed algorithm for data f and $u_{\tilde{f}}$ 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 ε and the precision τ without using the δ's sufficient for the assumptions of Proposition 11 to hold is given by

$$(1+6\tau)^{3N-k}\tau < \frac{\varepsilon}{12CN(2\beta)^{2k}(C+\frac{1}{2})\left(\frac{9N-5}{4} - k\right)} \quad \text{if } \varepsilon \le (2\beta)^{2k}\left(C+\frac{1}{2}\right), \qquad (58)$$

$$(1+6\tau)^{3N-k}\tau < \frac{1}{12CN\left(\frac{9N-5}{4} - k\right)} \quad \text{if } \varepsilon > (2\beta)^{2k}\left(C+\frac{1}{2}\right), \qquad (59)$$

$$\mu_{[1,N]}(\tau) < \frac{1}{2}\left(\frac{4C^2N+\varepsilon}{C^2N} \right)^{\frac{1}{2}} - 1 \qquad (60)$$

w.r.t. $\mu_{[1,N]}(\tau)$ 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 $g(\tau) \le \varepsilon/4$ since: for any admissible $l : r$ we have

$$\left| \|\widetilde{Q^T f_{l:r}}\|^2 - \|Q^T f_{l:r}\|^2 \right| \leq \left(\|Q^T f_{l:r}\| + \|\widetilde{Q^T f_{l:r}}\| \right) \left| \|Q^T f_{l:r}\| - \|\widetilde{Q^T f_{l:r}}\| \right|$$

$$\leq \left(C + \|\widetilde{Q^T f_{l:r}}\| \right) \|Q^T f_{l:r} - \widetilde{Q^T f_{l:r}}\|$$

and

$$\|\widetilde{Q^T f_{l:r}}\| \leq \|Q^T f_{l:r}\| + \|Q^T f_{l:r} - \widetilde{Q^T f_{l:r}}\|$$

$$\leq C + \|Q^T f_{l:r} - \widetilde{Q^T f_{l:r}}\| \leq C + \mu_{[1,N]}C.$$

Combining both yields

$$g(\tau) = \left| \tilde{\mathcal{G}}_{\mathcal{I}}(f) - \mathcal{G}_{\mathcal{I}}(f) \right| \leq \sum_{l:r \in \mathcal{I}} \left| \tilde{\mathcal{E}}^{l:r} - \mathcal{E}^{l:r} \right|$$

$$\leq \sum_{l:r \in \mathcal{I}} \left| \|\widetilde{Q^T f_{l:r}}\|^2 - \|Q^T f_{l:r}\|^2 \right| \leq NC^2 \mu_{[1,N]}(\tau) \left(2 + \mu_{[1,N]}(\tau) \right).$$

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

$$\|\tilde{S}_{l:r}(f_{l:r}) - S_{l:r}(f_{l:r})\| = \|S_{l:r}(\tilde{f}_{l:r}) - S_{l:r}(f_{l:r})\|$$

$$\leq \|S_{l:r}\| \|\tilde{f}_{l:r} - f_{l:r}\| \leq \|\tilde{f}_{l:r} - f_{l:r}\| < \delta_{l:r}. \qquad (61)$$

For the second before last inequality, we used that $\|S_{l:r}\| \leq 1$ which we have shown in the proof of Lemma 8. In consequence, summing over all intervals of $\mathcal{I}$ of length at least $k + 1$, we obtain from (61) that

$$\|\tilde{S}_{\mathcal{I}}(f) - S_{\mathcal{I}}(f)\| = \|S_{\mathcal{I}}(\tilde{f}_{\mathcal{I}}) - S_{\mathcal{I}}(f)\| \leq \|\tilde{f}_{\mathcal{I}} - f\| < \delta_{\mathcal{I}}(\tau). \qquad (62)$$

for any partition $\mathcal{I}$ which is considered by the algorithm. Then, using the notation $\tilde{\mathcal{G}}_{\mathcal{I}}$ for the algorithmic variant of $\mathcal{G}_{\mathcal{I}}$, we have (with the notation as in Lemma 8) that

$$\tilde{\mathcal{G}}_{\mathcal{I}'}(\tilde{f}_{\mathcal{I}'}) \leq \mathcal{G}_{\mathcal{I}'}(\tilde{f}_{\mathcal{I}'}) + |\tilde{\mathcal{G}}_{\mathcal{I}'}(\tilde{f}_{\mathcal{I}'}) - \mathcal{G}_{\mathcal{I}'}(\tilde{f}_{\mathcal{I}'})|$$

$$\leq \mathcal{G}_{\mathcal{I}'}(\tilde{f}_{\mathcal{I}'}) + g(\tau)$$

$$< \mathcal{G}_{\mathcal{I}}(f) - \varepsilon + \delta_{\mathcal{I}'}(\tau)^2 \|A_{\mathcal{I}'}\| + \delta_{\mathcal{I}'}(\tau) \|A_{\mathcal{I}'}\| \|f\| + g(\tau)$$

$$\leq \mathcal{G}_{\mathcal{I}}(\tilde{f}_{\mathcal{I}}) - \varepsilon + 2\delta^*(\tau) \max_{\mathcal{I}} \|A_{\mathcal{I}}\| \left(\|f\| + \delta^*(\tau) \right) + g(\tau)$$

$$\leq \mathcal{G}_{\mathcal{I}}(\tilde{f}_{\mathcal{I}}) - \varepsilon + \varepsilon/2 + g(\tau)$$
$$\leq \tilde{\mathcal{G}}_{\mathcal{I}}(\tilde{f}_{\mathcal{I}}) - \varepsilon + \varepsilon/2 + 2g(\tau) \leq \tilde{\mathcal{G}}_{\mathcal{I}}(\tilde{f}_{\mathcal{I}}). \tag{63}$$

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 $\tilde{f}'_{\mathcal{I}}$ 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 $\tilde{f}_{\mathcal{I}}$ of f. The second before last and last inequality follow by our assumptions made on τ. Together, (63) tells us that the solution $\tilde{u}_f$ computed by the proposed algorithm has partition $\mathcal{I}'$ and

$$\tilde{u}_f = \tilde{S}_{\mathcal{I}'}(f). \tag{64}$$

Using again Corollary 10, we have

$$\tilde{S}_{\mathcal{I}'}(f) = S_{\mathcal{I}'}(\tilde{f}) \quad \text{for} \quad \|\tilde{f} - f\| < \delta_{\mathcal{I}'}(\tau) \leq \delta^*(\tau), \tag{65}$$

with the perturbation $\tilde{f}$ of f. We have that $\|\tilde{f} - f\| < \delta^*(\tau) < \delta$ with δ 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 $u_{\tilde{f}}$ agrees with the optimal solution for the partition $\mathcal{I}'$ which we have denoted by $u_{\tilde{f},\mathcal{I}'} = S_{\mathcal{I}}(\tilde{f})$, i.e.,

$$u_{\tilde{f}} = u_{\tilde{f},\mathcal{I}'} = S_{\mathcal{I}'}(\tilde{f}). \tag{66}$$

Combined with (64) and (65), this shows (57) which completes the proof.   □

Lemma 13 *We consider a nonzero quadratic form H in a ball of radius C in $\mathbb{R}^N$. Then, the Lebesgue measure λ of the set $H_{\varepsilon,c} = \{x : \|x\| \leq C, c - \varepsilon < H(x) < c + \varepsilon\}$ fulfills*

$$\lambda(H_{\varepsilon,c}) \leq 2\sqrt{\frac{\varepsilon}{\|A\|}} \, C^{N-1} \tag{67}$$

where $\|A\|$ denotes the spectral norm of the representing matrix A of H.

Proof Without loss of generality, we may use a orthogonal transformation of the coordinate system to represent H by $H(x) = \sum_i \alpha_i x_i^2$ with the eigenvalues α_i of the corresponding representing matrix of the quadratic form. We sort the α_i by modulus, i.e., $|\alpha_1| \geq |\alpha_2| \geq \dots$. With repect 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 α_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 $H_{\varepsilon,c}$. To this end, we notice that, for

fixed $x_2, \ldots, x_N$, we may estimate the univariate Lebesgue measure λ^1 of the section $\{x_1 :\leq C, c - \varepsilon < H(x_1, x_2, \ldots, x_N) < c + \varepsilon\}$

$$\lambda^1 \left(c + \sum_{i=2}^{N} \frac{\alpha_i}{|\alpha_1|} x_i^2 - \varepsilon < \mathrm{sign}(\alpha_1)\, x_1^2 < c + \sum_{i=2}^{N} \frac{\alpha_i}{|\alpha_1|} x_i^2 + \varepsilon \right) \leq 2\sqrt{\frac{\varepsilon}{|\alpha_1|}}. \quad (68)$$

(Notice that 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. □

Theorem 14 *Let $\varepsilon > 0$ be given and assume that the precision τ 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 $2\binom{\sigma_{N,k}}{2}\sqrt{\frac{\varepsilon}{\sup_{\mathcal{I}} \|A_{\mathcal{I}}\|}}\, C^{N-1}$, $A_{\mathcal{I}}$ 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 $(\mathcal{P}_{k,\beta,\gamma})$ is backward stable in the sense that*

$$\tilde{u}_f = u_{\tilde{f}} \quad \text{where} \quad \|\tilde{f} - f\| < \delta^*(\tau), \quad (69)$$

where $\delta^*(\tau)$ is given by (56). Here, $\tilde{u}_f$ is the result produced by the proposed algorithm for data f and $u_{\tilde{f}}$ is the (unique) solution of the higher order Potts and Mumford–Shah problem $(\mathcal{P}_{k,\beta,\gamma})$ 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 $2\binom{\sigma_{N,k}}{2}\sqrt{\frac{\varepsilon}{\sup_{\mathcal{I}} \|A_{\mathcal{I}}\|}}\, C^{N-1}$. We choose two different partitions $\mathcal{I}, \mathcal{I}'$ with $[\mathcal{I}] \neq [\mathcal{I}']$, i.e., their equivalence classes do not agree, and consider the corresponding quadratic forms $\mathcal{G}_{\mathcal{I}}, \mathcal{G}_{\mathcal{I}'}$. Their difference $\mathcal{G}_{\mathcal{I}} - \mathcal{G}_{\mathcal{I}'}$ is again a quadratic form (plus a constant). By Lemma 13, the set where $\mathcal{G}_{\mathcal{I}}$ and $\mathcal{G}_{\mathcal{I}'}$ are closer than ε has Lebesgue measure $2\sqrt{\frac{\varepsilon}{|\alpha_1|}}\, C^{N-1}$. Iterating this for all $\binom{\sigma_{N,k}}{2}$ different bilinear forms $\mathcal{G}_{\mathcal{I}} - \mathcal{G}_{\mathcal{I}'}$ 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. □

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 proposed solvers for the higher order Mumford–Shah and Potts models in C++ with wrappers to Matlab using mex-files. The pseudocode is given as Algorithm 1 in the appendix. All experiments were conducted on a desktop computer with 3.1 GHz Intel Core i5-2400 processor and 8 GB RAM.

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 [19,45]. We corrupt the signals by additive zero mean Gaussian noise with variance σ^2. We let the noise level η be given by $\eta = \sigma N/\|g\|_1$, where g denotes the clean signal. To obtain a meaningful comparison of the models' potentials, we determined parameters β and γ such that the result u^* has the best relative ℓ_2-error ε_{rel}, given by $\varepsilon_{\text{rel}} = \|u^* - g\|_2/\|g\|_2$. (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 $\mathcal{I}^*$. A commonly used measure for segmentation quality is the Rand index [52] which we briefly explain. The Rand index R_{ind} of two partitions $\mathcal{I}, \mathcal{I}'$ is given by $R_{\text{ind}}(\mathcal{I}, \mathcal{I}') = \binom{N}{2} \sum_{\{i, j: 1 \leq i < j \leq N\}} t_{ij}$ where t_{ij} is equal to one if there are $I \in \mathcal{I}$ and $I' \in \mathcal{I}'$ 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 $\mathcal{I}$ and $\mathcal{I}'$ 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 β leads to stronger smoothing on the segments, and that a high parameter γ leads to less segments.

The first signal is a sinusoidal with two steps (Fig. 5). We observe that the first order model requires choosing a relatively small β parameter to avoid the gradient limit effect. 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. The reconstruction quality starts decaying from order $k = 6$ on which can be attributed to overfitting.

The second example is a piecewise constant signal (Fig. 6). As the signal has no variation on the segments, the experiment confirms the intuition that large elasticity parameters – mostly $\beta = \infty$ – are preferable here. 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 leads to very good segmentation results up to order $k = 5$ and good reconstructions up to order $k = 3$, as well.

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 Fig. 8. The data are time-averaged (hourly) wind speeds at the summit of high-

---

[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.
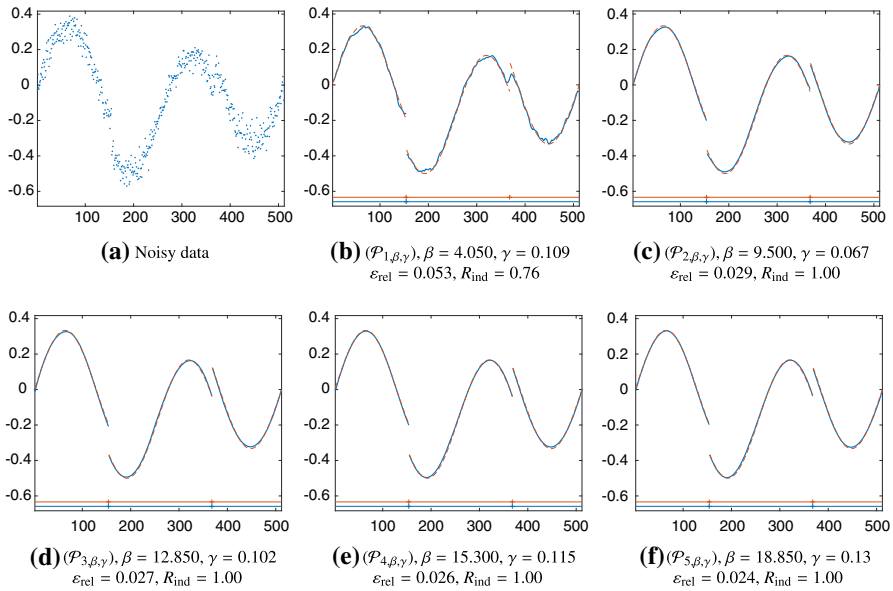
Fig. 5 Reconstructions of "Heavy Sine"-signal from noisy data. **a** Data corrupted by Gaussian noise of level 0.2. **b–f** Reconstructions using higher order Mumford–Shah and Potts models
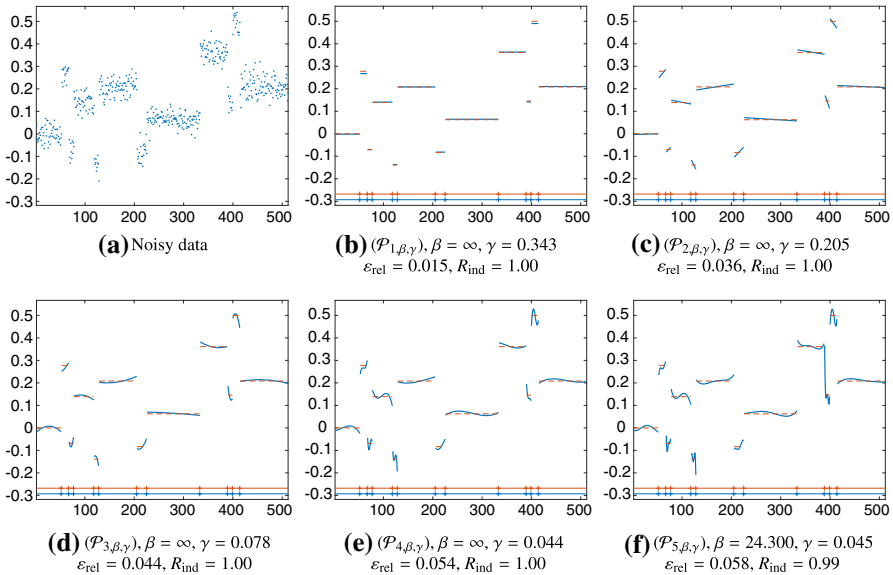

Fig. 6 Reconstructions of "Blocks"-signal from noisy data. **a** Data corrupted by Gaussian noise of level 0.2. **b–f** Reconstructions using higher order Mumford–Shah and Potts models
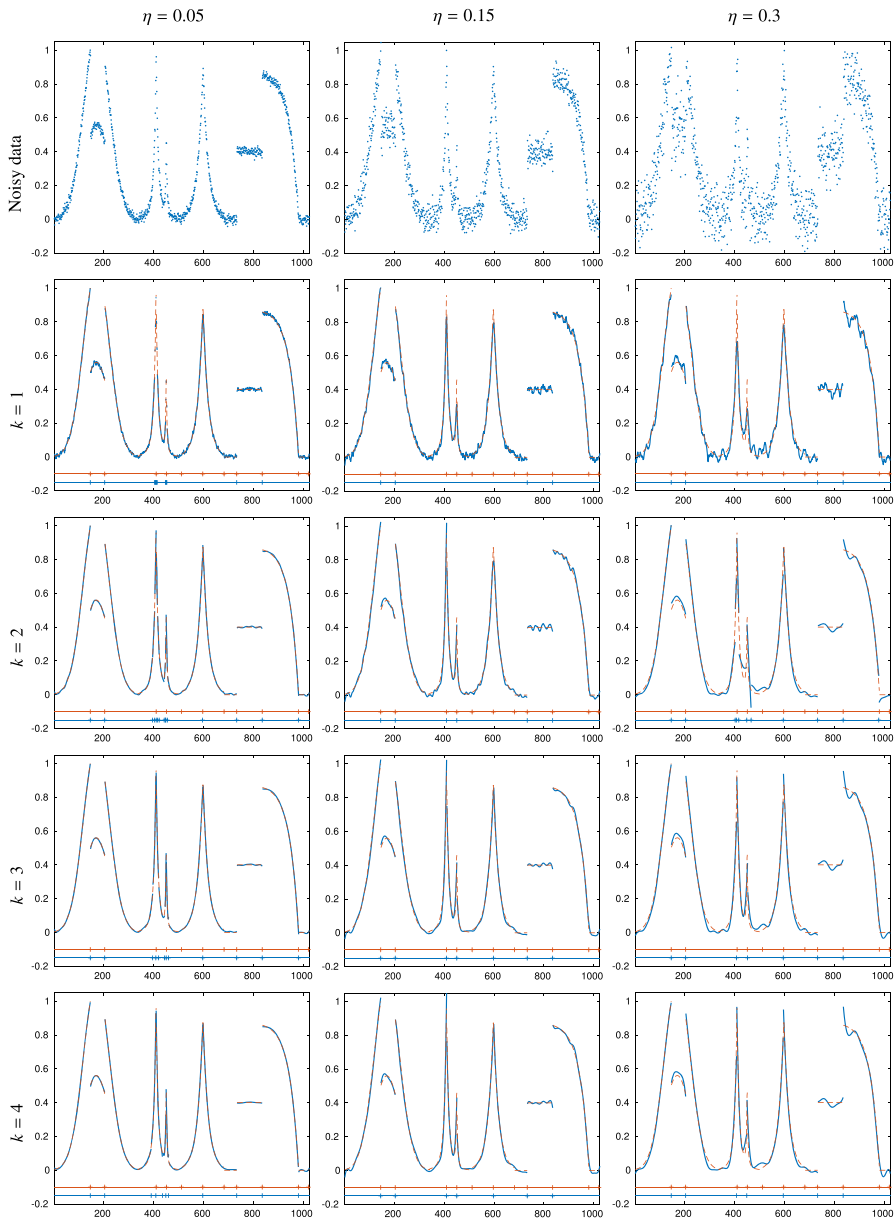
Fig. 7 Smoothing of piecewise defined signal of increasing noise level η (top row). We observe that the segmentation quality gets higher and that the noise is smoothed out better on the segments when using higher order Mumford–Shah models
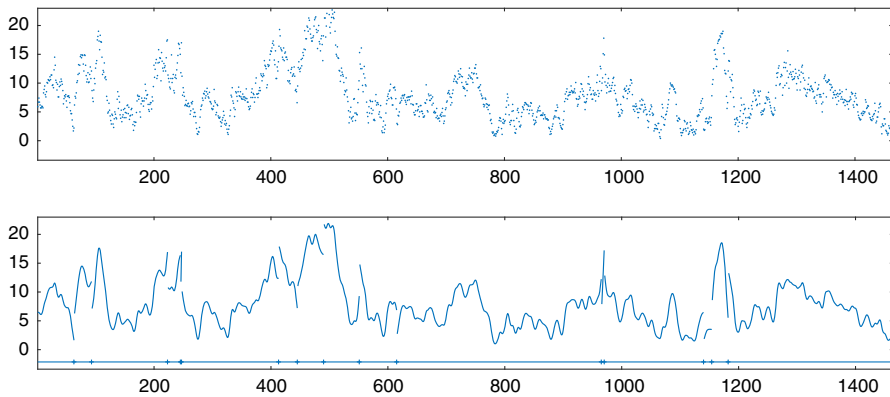
Fig. 8 Top: hourly averaged windspeeds (m/s) at the summit of the Zugspitze from November to December 2016. Bottom: result of higher order Mumford–Shah model ($\mathcal{P}_{2;2;15}$)

est 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]

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 Fig. 7. In all cases, the signals are corrupted by additive Gaussian noise with noise level $\eta = 0.1$. For

2 The data were collected by German climate data center and are available via ftp at ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate/hourly/wind/historical/ (station id: 02115).

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

Time [s]	$k=1$	2	3	4	Time [s]	$k=1$	2	3	4
$n=1000$	0.0040	0.0049	0.0061	0.0065	$n=1000$	0.0032	0.0038	0.0044	0.0052
4000	0.0183	0.0217	0.0268	0.0293	4000	0.0141	0.0164	0.0191	0.0222
7000	0.0319	0.0382	0.0473	0.0513	7000	0.0247	0.0291	0.0333	0.0395
10000	0.0464	0.0549	0.0682	0.0738	10000	0.0358	0.0417	0.0485	0.0570

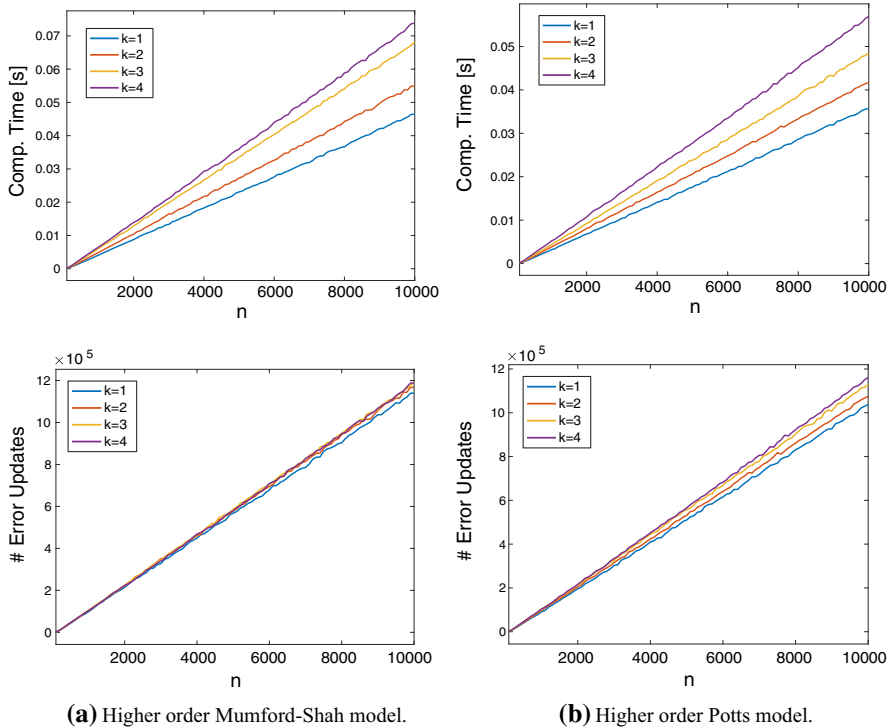**(a)** Higher order Mumford-Shah model.   **(b)** Higher order Potts model.

Fig. 9 Computational costs of the proposed algorithm for the randomly generated piecewise polynomial signals corrupted by Gaussian noise of level $\eta = 0.1$. Computation times for selected lengths N are tabulated (top) and plotted for $N = 100, 200, \ldots, 10{,}000$ (center). We observe that the runtime only grows linearly in N, so much more favorable than the worst case scenario of quadratic growth. This means that the pruning strategies show their full effectiveness for this type of signals, which is reflected by the linear growth in the number of error updates (bottom)

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 Fig. 9. It is an important observation that the runtime and the errors updates exhibit linear growth in the signal length. Thus, the proposed algorithm 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 the proposed algorithm attains its worst case complexity. These results suggest that an increasing number of discontinuities is beneficial for the efficiency of the proposed algorithm.

Table 1 Mean computation times of the proposed algorithm (in seconds) for the signal from Fig. 7 sampled on N points and corrupted by Gaussian noise with noise level 0.1

k	N								
	2^9	2^{10}	2^{11}	2^{12}	2^{13}	2^{14}	2^{15}	2^{16}	2^{17}
(a) Runtime (s) for (higher order) Mumford–Shah solver									
1	0.0025	0.0091	0.0363	0.1132	0.4230	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
(b) Runtime (s) for (higher order) Potts solver									
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

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 Fig. 9 is that the number of discontinuities does not increase with N. Yet, the solver processes signals of size 2^{16} in less than 1 min

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. 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 obtain 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 1 s. This way, the family of higher order Mumford–Shah and Potts models can serve as efficient smoothers for signals with discontinuities.

Appendix: Pseudocode of the proposed solver

A pseudocode for the proposed solver for ($\mathcal{P}_{k,\beta,\gamma}$) is given in Algorithm 1.

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 $(u^*, \mathcal{I}^*)$ of $(\mathcal{P}_{k,\beta,\gamma})$ or $(\mathcal{P}_{k,\infty,\gamma})$

/* Precomputations   */

1 Row-wise transform the matrix $\begin{cases} (\mathcal{P}_{k,\beta,\gamma}) : & A \text{ from (29)} \\ (\mathcal{P}_{k,\infty,\gamma}) : & B \text{ from (39)} \end{cases}$
   to upper triangular form using successive Givens rotations and store the rotation angles in Θ.

2 Compute $\mathcal{E}^{1:r}$ for all $r = 1, \ldots, N$ with Θ

/* Find optimal changepoints   */

3 Initialize lists $L = [2]$, $R = [2]$, $E = [0]$

4 $J_1 \leftarrow 0$, $P_1^* \leftarrow 0$

5 **for** $r \leftarrow 2, \ldots, N$ **do**

      /* Initialization   */

6     $J_r \leftarrow 0$, $P_r^* \leftarrow \mathcal{E}^{1:r}$

      /* Find optimal P_r^* using (25)   */

7   **for** $i = 1, ..., length \ of \ L$ **do**

8   **while** $r_i < r$ **do**

            /* Update approximation error   */

9         $\begin{cases} (\mathcal{P}_{k,\beta,\gamma}) : & \text{Compute } \mathcal{E}^{l_i:r_i+1} \text{ from } \mathcal{E}^{l_i:r_i} \text{ using the recurrence (34)-(37)} \\ (\mathcal{P}_{k,\infty,\gamma}) : & \text{Compute } \mathcal{E}^{l_i:r_i+1} \text{ from } \mathcal{E}^{l_i:r_i} \text{ using the recurrence (41)-(43)} \end{cases}$

10         $E_i \leftarrow \mathcal{E}^{l_i:r_i+1}$,

11         $r_i \leftarrow r_i + 1$

            /* Pruning (27)   */

12   **if** $P_{l_i-1}^* + E_i \geq P_{r_i}^*$ **then**

13   Delete: l_i from L, r_i from R and E_i from E

14   **go to** 7

15   **end**

16   **end**

17        $b \leftarrow P_{l_i-1}^* + \gamma + E_i$

18   **if** $b \leq P_r^*$ **then**

19         $\mid$ $P_r^* \leftarrow b$, $J_r \leftarrow l - 1$

20   **end**

      /* Pruning (26)   */

21   **If** $E_i + \gamma > P_r^*$ **then break end**

22   **end**

      /* Update lists   */

23   Prepend: $r + 1$ to L, $r + 1$ to R, 0 to E

24 **end**

   /* Recover partition $\mathcal{I}^*$ from segment boundary locations J */

25 $r \leftarrow N$, $\mathcal{I}^* \leftarrow \emptyset$

26 **while** $r > 0$ **do**

27     $\mid$ $l \leftarrow J(r) + 1$, $\mathcal{I}^* \leftarrow \mathcal{I}^* \cup \{(l : r)\}$, $r \leftarrow l - 1$

28 **end**

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

29 **for** $I \in \mathcal{I}^*$ **do**

30     $\begin{cases} (\mathcal{P}_{k,\beta,\gamma}): \text{Solve } u_I^* = \underset{v \in \mathbb{R}^{|I|}}{\operatorname{argmin}} \|v - f_I\|_2^2 + \beta^{2k} \|\nabla^k v\|_2^2 \\ (\mathcal{P}_{k,\infty,\gamma}): \text{Solve } u_I^* = \operatorname{argmin} \|v - f_I\|_2^2 \text{ such that } v \text{ is polynomial of degree} \leq k - 1 \text{ on } I \end{cases}$

31 **end**

References

1. Amat, S., Li, Z., Ruiz, J.: On an new algorithm for function approximation with full accuracy in the presence of discontinuities based on the immersed interface method. J. Sci. Comput. **75**(3), 1500–1534 (2017)

2. Ambrosio, L., Tortorelli, V.M.: Approximation of functional depending on jumps by elliptic functional via γ-convergence. Commun. Pure Appl. Math. **43**(8), 999–1036 (1990)

3. Arandiga, F., Cohen, A., Donat, R., Dyn, N.: Interpolation and approximation of piecewise smooth functions. SIAM J. Numer. Anal. **43**(1), 41–57 (2005)

4. Artina, M., Fornasier, M., Solombrino, F.: Linearly constrained nonsmooth and nonconvex minimization. SIAM J. Optim. **23**(3), 1904–1937 (2013)

5. Auger, I., Lawrence, C.: Algorithms for the optimal identification of segment neighborhoods. Bull. Math. Biol. **51**(1), 39–54 (1989)

6. Bar, L., Sochen, N., Kiryati, N.: Variational pairing of image segmentation and blind restoration. In: ECCV, pp. 166–177. Springer, Berlin (2004)

7. Bellman, R., Roth, R.: Curve fitting by segmented straight lines. J. Am. Stat. Assoc. **64**(327), 1079–1084 (1969)

8. Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press, Cambridge (1987)

9. Blake, A.: Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. IEEE Trans. Pattern Anal. Mach. Intell. **11**(1), 2–12 (1989)

10. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)

11. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM J. Imaging Sci. **3**(3), 492–526 (2010)

12. Bruce, J.: Optimum Quantization. Technical report, Massachusetts Institute of Technology (1965)

13. Carriero, M., Leaci, A., Tomarelli, F.: A survey on the Blake–Zisserman functional. Milan J. Math. **83**(2), 397–420 (2015)

14. Chambolle, A.: Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. SIAM J. Appl. Math. **55**(3), 827–863 (1995)

15. Chambolle, A.: Finite-differences discretizations of the Mumford–Shah functional. ESAIM Math. Model. Numer. Anal. **33**(02), 261–288 (1999)

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

17. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Process. **10**(2), 266–277 (2001)

18. Chen, C., Leng, J., Xu, G.: A general framework of piecewise-polynomial Mumford–Shah model for image segmentation. Int. J. Comput. Math. **94**(10), 1981–1997 (2017)

19. Donoho, D., Johnstone, I.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**, 425–455 (1994)

20. Drobyshev, A., Machka, C., Horsch, M., Seltmann, M., Liebscher, V., de Angelis, M., Beckers, J.: Specificity assessment from fractionation experiments (safe): a novel method to evaluate microarray probe specificity based on hybridisation stringencies. Nucleic Acids Res. **31**(2), 1–10 (2003)

21. Fearnhead, P., Maidstone, R., Letchford, A.: Detecting changes in slope with an L_0 penalty. J. Comput. Graph. Stat. **28**(2), 265–275 (2018)

22. Fornasier, M., Ward, R.: Iterative thresholding meets free-discontinuity problems. Found. Computat. Math. **10**(5), 527–567 (2010)

23. Fornasier, M., March, R., Solombrino, F.: 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)

24. Fortun, D., Storath, M., Rickert, D., Weinmann, A., Unser, M.: Fast piecewise-affine motion estimation without segmentation. IEEE Trans. Image Process. **27**(11), 5612–5624 (2018)

25. Frick, K., Munk, A., Sieling, H.: Multiscale change point inference. J. R. Stat. Soc. Ser. B Stat. Methodol. **76**(3), 495–580 (2014)

26. Friedrich, F., Kempe, A., Liebscher, V., Winkler, G.: Complexity penalized M-estimation: fast computation. J. Comput. Graph. Stat. **17**(1), 201–224 (2008)

27. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. **6**(6), 721–741 (1984)

28. Gentleman, W.: Least squares computations by Givens transformations without square roots. J. Inst. Math. Appl. **12**, 329–336 (1973)

29. Harten, A.: Multiresolution representation of data: a general framework. SIAM J. Numer. Anal. **33**(3), 1205–1256 (1996)

30. Hohm, K., Storath, M., Weinmann, A.: An algorithmic framework for Mumford–Shah regularization of inverse problems in imaging. Inverse Probl. **31**(11), 115011 (2015)

31. Hotz, T., Schutte, O.M., Sieling, H., Polupanow, T., Diederichsen, U., Steinem, C., Munk, A.: Idealizing ion channel recordings by a jump segmentation multiresolution filter. IEEE Trans. NanoBiosci. **12**(4), 376–386 (2013)

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

33. Isack, H., Veksler, O., Sonka, M., Boykov, Y.: Hedgehog shape priors for multi-object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2434–2442 (2016)

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

35. Jiang, M., Maass, P., Page, T.: Regularizing properties of the Mumford–Shah functional for imaging applications. Inverse Probl. **30**(3), 035007 (2014)

36. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C., Ha, T.: Advances in single-molecule fluorescence methods for molecular biology. Annu. Rev. Biochem. **77**, 51–76 (2008)

37. Kiefer, L., Storath, M., Weinmann, A.: Iterative Potts minimization for the recovery of signals with discontinuities from indirect measurements—the multivariate case. arXiv:1812.00862 (2018)

38. Killick, R., Fearnhead, P., Eckley, I.: Optimal detection of changepoints with a linear computational cost. J. Am. Stat. Assoc. **107**(500), 1590–1598 (2012)

39. Kleinberg, J., Tardos, E.: Algorithm Design. Pearson Education India, New Delhi (2006)

40. Kolmogorov, V., Pock, T., Rolinek, M.: Total variation on a tree. Preprint arXiv:1502.07770 (2015)

41. Lemire, D.: A better alternative to piecewise linear time series segmentation. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 545–550. SIAM (2007)

42. Liebscher, V., Winkler, G.: A Potts model for segmentation and jump-detection. In: Proceedings S4G International Conference on Stereology, Spatial Statistics and Stochastic Geometry, Prague June, vol. 21, pp. 185–190. Citeseer (1999)

43. Little, M., Jones, N.: Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory. Proc. R. Soc. A Math. Phys. Eng. Sci. **467**(2135), 3088–3114 (2011a)

44. Little, M., Jones, N.: Generalized methods and solvers for noise removal from piecewise constant signals. II. New methods. Proc. R. Soc. A Math. Phys. Eng. Sci. **467**(2135), 3115–3140 (2011b)

45. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic Press, London (2008)

46. Mumford, D., Shah, J.: Boundary detection by minimizing functionals. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 17, pp. 137–154 (1985)

47. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math. **42**(5), 577–685 (1989)

48. Nord, A.L., Gachon, E., Perez-Carrasco, R., Nirody, J.A., Barducci, A., Berry, R.M., Pedaci, F.: Catch bond drives stator mechanosensitivity in the bacterial flagellar motor. Proc. Natl. Acad. Sci. **114**(49), 12952–12957 (2017)

49. Potts, R.: Some generalized order-disorder transformations. Math. Proc. Camb. Philos. Soc. **48**(1), 106–109 (1952)

50. Ramlau, R., Ring, W.: A Mumford–Shah level-set approach for the inversion and segmentation of X-ray tomography data. J. Comput. Phys. **221**(2), 539–557 (2007)

51. Ramlau, R., Ring, W.: Regularization of ill-posed Mumford–Shah models with perimeter penalization. Inverse Probl. **26**(11), 115001 (2010)

52. Rand, W.: Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. **66**(336), 846 (1971)

53. Rondi, L.: On the regularization of the inverse conductivity problem with discontinuous conductivities. Inverse Probl. Imaging **2**(3), 397–409 (2008)

54. Rondi, L., Santosa, F.: Enhanced electrical impedance tomography via the Mumford–Shah functional. ESAIM Control Optim. Calc. Var. **6**, 517–538 (2001)

55. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D Nonlinear Phenom. **60**(1), 259–268 (1992)
56. Snijders, A., Nowak, N., Segraves, R., et al.: Assembly of microarrays for genome-wide measurement of DNA copy number by CGH. Nat. Genet. **29**, 263–264 (2001)
57. Sowa, Y., Rowe, A., Leake, M., Yakushi, T., Homma, M., Ishijima, A., Berry, R.: Direct observation of steps in rotation of the bacterial flagellar motor. Nature **437**(7060), 916–919 (2005)
58. Storath, M., Weinmann, A., Unser, M.: Jump-penalized least absolute values estimation of scalar or circle-valued signals. Inf. Inference **6**(3), 225–245 (2017)
59. Storath, M., Weinmann, A.: Fast partitioning of vector-valued images. SIAM J. Imaging Sci. **7**(3), 1826–1852 (2014)
60. Strekalovskiy, E., Chambolle, A., Cremers, D.: A convex representation for the vectorial Mumford–Shah functional. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1712–1719 (2012)
61. Strekalovskiy, E., Cremers, D.: Real-time minimization of the piecewise smooth Mumford–Shah functional. In European Conference on Computer Vision, pp. 127–141. Springer, Berlin (2014)
62. Tsai, A., Yezzi, A., Willsky, A.S.: Curve evolution implementation of the mumford-shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Trans. Image Process. **10**(8), 1169–1186 (2001)
63. Veksler, O.: Efficient graph-based energy minimization methods in computer vision. PhD thesis, Cornell University (1999)
64. Wahba, G.: Spline Models for Observational Data. SIAM, Philadelphia (1990)
65. Wang, Y., Xiang, S., Pan, C., Wang, L., Meng, G.: Level set evolution with locally linear classification for image segmentation. Pattern Recognit. **46**(6), 1734–1746 (2013)
66. Wang, Z., Vemuri, B.C.: DTI segmentation using an information theoretic tensor dissimilarity measure. IEEE Trans. Med. Imaging **24**(10), 1267–1277 (2005)
67. Weinmann, A., Storath, M.: Iterative Potts and Blake–Zisserman minimization for the recovery of functions with discontinuities from indirect measurements. Proc. R. Soc. A **471**(2176), 20140638 (2015)
68. Weinmann, A., Storath, M., Demaret, L.: The L^1-Potts functional for robust jump-sparse reconstruction. SIAM J. Numer. Anal. **53**(1), 644–673 (2015)
69. Weinmann, A., Demaret, L., Storath, M.: Mumford–Shah and Potts regularization for manifold-valued data. J. Math. Imaging Vis. **55**(3), 428–445 (2016)
70. Whittaker, E.: On a new method of graduation. Proc. Edinb. Math. Soc. **41**, 63–75 (1922)
71. Wilkinson, J.: The Algebraic Eigenvalue Problem. Monographs on Numerical Analysis. Clarendon Press, Oxford (1965)
72. Winkler, G.: Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction, vol. 27. Springer, Berlin (2003)
73. Winkler, G., Liebscher, V.: Smoothers for discontinuous signals. J. Nonparametr. Stat. **14**(1–2), 203–222 (2002)
74. Winkler, G., Wittich, O., Liebscher, V., Kempe, A.: Don't shed tears over breaks. Jahresbericht der Deutschen Mathematiker-Vereinigung **107**, 57–87 (2005)
75. Wittich, O., Kempe, A., Winkler, G., Liebscher, V.: Complexity penalized least squares estimators: analytical results. Mathematische Nachrichten **281**(4), 582–595 (2008)
76. Yang, J., Li, H.: Dense, accurate optical flow estimation with piecewise parametric model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1019–1027 (2015)
77. Zanetti, M., Bruzzone, L.: Piecewise linear approximation of vector-valued images and curves via second-order variational model. IEEE Trans. Image Process. **26**(9), 4414–4429 (2017)
78. Zanetti, M., Ruggiero, V., Miranda Jr., M.: Numerical minimization of a second-order functional for image segmentation. Commun. Nonlinear Sci. Numer. Simul. **36**, 528–548 (2016)