DFG-Schwerpunktprogramm 1324

The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited

M. Hein, S. Setzer, L. Jost, S. Rangapuram

Preprint 153

Edited by

AG Numerik/Optimierung
Fachbereich 12 - Mathematik und Informatik
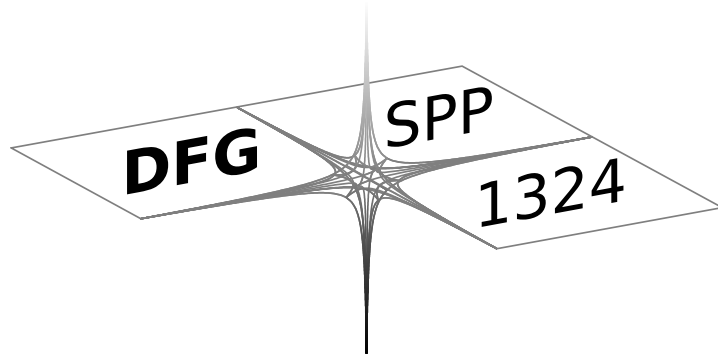Philipps-Universität Marburg
Hans-Meerwein-Str.
35032 Marburg

DFG-Schwerpunktprogramm 1324

The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited

M. Hein, S. Setzer, L. Jost, S. Rangapuram

Preprint 153

The consecutive numbering of the publications is determined by their chronological order.

The aim of this preprint series is to make new research rapidly available for scientific discussion. Therefore, the responsibility for the contents is solely due to the authors. The publications will be distributed by the authors.

The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited

Matthias Hein, Simon Setzer, Leonardo Jost and Syama Sundar Rangapuram
Department of Computer Science
Saarland University

Abstract

Hypergraphs allow to encode higher-order relationships in data and are thus a very flexible modeling tool. Current learning methods are either based on approximations of the hypergraphs via graphs or on tensor methods which are only applicable under special conditions. In this paper, we present a new learning framework on hypergraphs which fully uses the hypergraph structure. The key element is a family of regularization functionals based on the total variation on hypergraphs.

1 Introduction

Graph-based learning is by now well established in machine learning and is the standard way to deal with data that encode pairwise relationships. Hypergraphs are a natural extension of graphs which allow to model also higher-order relations in data. It has been recognized in several application areas such as computer vision [1, 2], bioinformatics [3, 4] and information retrieval [5, 6] that such higher-order relations are available and help to improve the learning performance.

Current approaches in hypergraph-based learning can be divided into two categories. The first one uses tensor methods for clustering as the higher-order extension of matrix (spectral) methods for graphs [7, 8, 9]. While tensor methods are mathematically quite appealing, they are limited to so-called k-uniform hypergraphs, that is, each hyperedge contains exactly k vertices. Thus, they are not able to model mixed higher-order relationships. The second main approach can deal with arbitrary hypergraphs [10, 11]. The basic idea of this line of work is to approximate the hypergraph via a standard weighted graph. In a second step, one then uses methods developed for graph-based clustering and semi-supervised learning. The two main ways of approximating the hypergraph by a standard graph are the clique and the star expansion which were compared in [12]. One can summarize their results by stating that there exists no approximation which fully encodes the hypergraph structure. Earlier it has been proven that an exact representation of the hypergraph via a graph retaining its cut properties is impossible [13].

In this paper, we overcome the limitations of both existing approaches. We note that both for clustering and semi-supervised learning the key element, either explicitly or implicitly, is the cut functional. Our aim is to directly work with the cut defined on the hypergraph. We discuss in detail the differences of this hypergraph cut and the cut functionals induced by the clique and star expansion in Section 2.1. Then, in Section 2.2, we introduce the total variation on a hypergraph as the Lovasz extension of the hypergraph cut. Based on this, we propose a family of regularization functionals which interpolates between the total variation and a regularization functional enforcing smoother functions on the hypergraph corresponding to Laplacian-type regularization on graphs. We show a model for semi-supervised learning based on these regularizers in Section 3. Furthermore, in the line of recent research on tight relaxations of balanced graph cuts [14, 15, 16, 17], we show in Section 4 that there exists a tight relaxation of the normalized hypergraph cut. In both applications, a new type of convex optimization problems has to be solved for which we derive first-order gradient methods in Section 5. The main ingredients of these algorithms are proximal mappings for which

we provide a novel algorithm and analyze its complexity. In the experimental section 6, we show that fully incorporating hypergraph structure is beneficial in semi-supervised learning and clustering on hypergraphs. *All proofs are moved to the supplementary material.*

2   The Total Variation on Hypergraphs

A large class of graph-based algorithms in semi-supervised learning and clustering is based either explicitly or implicitly on the cut. Thus, we discuss first in Section 2.1 the hypergraph cut and the corresponding approximations of [11, 12]. In Section 2.2, we introduce in analogy to the total variation on graphs, the total variation on hypergraphs as the Lovasz extension of the hypergraph cut as well as a corresponding family of regularization functionals.

2.1   Hypergraphs, Graphs and Cuts

Hypergraphs allow to model relations which are not only pairwise as in graphs but involve multiple vertices. In this paper, we consider weighted undirected hypergraphs $H = (V, E, w)$ where V is the vertex set with $|V| = n$ and E the set of hyperedges with $|E| = m$. Each hyperedge $e \in E$ corresponds to a subset of vertices, i.e., to an element of 2^V. The vector $w \in \mathbb{R}^m$ contains for each hyperedge e its non-negative weight w_e. In the following, we use the letter H also for the incidence matrix $H \in \mathbb{R}^{|V| \times |E|}$ which is for $i \in V$ and $e \in E$,

$$H_{i,e} = \begin{cases} 1 & \text{if } i \in e, \\ 0 & \text{else.} \end{cases}$$

The degree of a vertex $i \in V$ is defined as $d_i = \sum_{e \in E} w_e H_{i,e}$ and the cardinality of an edge e can be written as $|e| = \sum_{j \in V} H_{j,e}$. We would like to emphasize that we do *not* impose the restriction that the hypergraph is k-uniform, i.e., that each hyperedge contains *exactly* k vertices.

The considered class of hypergraphs contains the set of undirected, weighted graphs which is equivalent to the set of 2-uniform hypergraphs. The motivation for the total variation on hypergraphs comes from the correspondence between the cut on a graph and the total variation functional. Thus, we recall the definition of the cut on weighted graphs $G = (V, W)$ with weight matrix W. Let $\overline{C} = V \backslash C$ denote the complement of C in V. Then, for a partition $(C, \overline{C})$, the cut is defined as

$$\text{cut}_G(C, \overline{C}) = \sum\nolimits_{i,j \,:\, i \in C, j \in \overline{C}} w_{ij}.$$

This standard definition of the cut carries over naturally to a hypergraph H

$$\text{cut}_H(C, \overline{C}) = \sum\nolimits_{\substack{e \in E: \\ e \cap C \neq \emptyset,\ e \cap \overline{C} \neq \emptyset}} w_e. \tag{1}$$

Thus, the cut functional on a hypergraph is just the sum of the weights of the hyperedges which have vertices both in C and $\overline{C}$. It is not biased towards a particular way the hyperedge is cut, that is, how many vertices of the hyperedge are in C resp. $\overline{C}$. This emphasizes that the vertices in a hyperedge belong together and we penalize every cut of a hyperedge with the same value.

In order to handle hypergraphs with existing methods developed for graphs, the focus in previous works [11, 12] has been on transforming the hypergraph into a graph. In [11], they suggest to use the *clique expansion* (CE), i.e., every hyperedge $e \in H$ is replaced with a fully connected subgraph where every edge in this subgraph has weight $\frac{w_e}{|e|}$. This leads to the cut functional cut_{CE},

$$\text{cut}_{CE}(C, \overline{C}) := \sum\nolimits_{\substack{e \in E: \\ e \cap C \neq \emptyset,\ e \cap \overline{C} \neq \emptyset}} \frac{w_e}{|e|} |e \cap C|\, |e \cap \overline{C}|. \tag{2}$$

Note that in contrast to the hypergraph cut (1), the value of cut_{CE} depends on the way each hyperedge is cut since the term $|e \cap C|\, |e \cap \overline{C}|$ makes the weights dependent on the partition. In particular, the smallest weight is attained if only a single vertex is split off, whereas the largest weight is attained if the partition of the hyperedge is most balanced. In comparison to the hypergraph cut, this leads to a bias towards cuts that favor to split off single vertices from a hyperedge which in our point of view is an undesired property for most applications. We illustrate this with an example in Figure 1, where the minimum hypergraph cut (cut_H) leads to a balanced partition, whereas the minimum
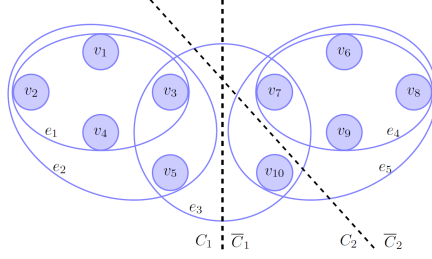
Figure 1: Minimum hypergraph cut cut_H vs. minimum cut of the clique expansion cut_{CE}: For edge weights $w_1 = w_4 = 10$, $w_2 = w_5 = 0.1$ and $w_3 = 0.6$ the minimum hypergraph cut is $(C_1, \overline{C}_1)$ which is perfectly balanced. Although cutting one hyperedge more and being unbalanced, $(C_2, \overline{C}_2)$ is the optimal cut for the clique expansion approximation.

clique expansion cut (cut_{CE}) not only cuts an additional hyperedge but is also unbalanced. This is due to its bias towards splitting off single nodes of a hyperedge. Another argument against the clique expansion is computational complexity. For large hyperedges the clique expansion leads to (almost) fully connected graphs which makes computations slow and is prohibitive for large hypergraphs.

We omit the discussion of the *star graph* approximation of hypergraphs discussed in [12] as it is shown there that the star graph expansion is very similar to the clique expansion. Instead, we want to recall the result of Ihler et al. [13] which states that in general there exists no graph with the same vertex set V which has for every partition $(C, \overline{C})$ the same cut value as the hypergraph cut.

Finally, note that for weighted 3-uniform hypergraphs it is always possible to find a corresponding graph such that any cut of the graph is equal to the corresponding cut of the hypergraph.

Proposition 2.1. *Suppose $H = (V, E, w)$ is a weighted 3-uniform hypergraph. Then, $W \in \mathbb{R}^{|V| \times |V|}$ defined as $W = \frac{1}{2} H \text{diag}(w) H^T$ defines the weight matrix of a graph $G = (V, W)$ where each cut of G has the same value as the corresponding hypergraph cut of H.*

Proof. The cut value of a partition $(C, \overline{C})$ of G is given as

$$\text{cut}_G(C, \overline{C}) = \frac{1}{2} \sum_{e \in E} |e \cap C| |e \cap \overline{C}| w_e.$$

The product $|e \cap C| |e \cap \overline{C}|$ takes the values 2 if e is cut by C and zero otherwise. Because of the factor $\frac{1}{2}$, we thus get equivalence to the hypergraph cut. □

2.2 The Total Variation on Hypergraphs

In this section, we define the total variation on hypergraphs. The key technical element is the Lovasz extension which extends a set function, seen as a mapping on 2^V, to a function on $\mathbb{R}^{|V|}$.

Definition 2.1. *Let $\hat{S} : 2^V \to \mathbb{R}$ be a set function with $\hat{S}(\emptyset) = 0$. Let $f \in \mathbb{R}^{|V|}$, let V be ordered such that $f_1 \leq f_2 \leq \ldots \leq f_n$ and define $C_i = \{j \in V \,|\, j > i\}$. Then, the **Lovasz extension** $S : \mathbb{R}^{|V|} \to \mathbb{R}$ of $\hat{S}$ is given by*

$$S(f) = \sum_{i=1}^{n} f_i \Big(\hat{S}(C_{i-1}) - \hat{S}(C_i) \Big) = \sum_{i=1}^{n-1} \hat{S}(C_i)(f_{i+1} - f_i) + f_1 \hat{S}(V).$$

Note that for the characteristic function of a set $C \subset V$, we have $S(\mathbf{1}_C) = \hat{S}(C)$.

It is well-known that the Lovasz extension S is a convex function if and only if $\hat{S}$ is submodular [18]. For graphs $G = (V, W)$, the total variation on graphs is defined as the Lovasz extension of the graph cut [18] given as $TV_G : \mathbb{R}^{|V|} \to \mathbb{R}$, $TV_G(f) = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} |f_i - f_j|$.

Proposition 2.2. *The **total variation** $\text{TV}_H : \mathbb{R}^{|V|} \to \mathbb{R}$ **on a hypergraph** $H = (V, E, w)$ defined as the Lovasz extension of the hypergraph cut, $\hat{S}(C) = \text{cut}_H(C, \overline{C})$, is a convex function given by*

$$\text{TV}_H(f) = \sum_{e \in E} w_e \Big(\max_{i \in e} f_i - \min_{j \in e} f_j \Big) = \sum_{e \in E} w_e \max_{i,j \in e} |f_i - f_j|.$$

Proof. Using $C_{i-1} = C_i \cup \{i\}$ and $\overline{C_i} = \overline{C_{i-1}} \cup \{i\}$ the Lovasz extension can be written as

$$\mathrm{TV}_H(f) = \sum_{i=1}^n f_i \Big(\mathrm{cut}(C_{i-1}, \overline{C_{i-1}}) - \mathrm{cut}(C_i, \overline{C_i}) \Big) = \sum_{i=1}^n f_i \Big(\mathrm{cut}(\{i\}, \overline{C_{i-1}}) - \mathrm{cut}(C_i, \{i\}) \Big)$$

$$= \sum_{i=1}^n f_i \Big(\sum_{\substack{e \in E, i \in e \\ e \cap \{1,\dots,i-1\} \neq \emptyset}} w_e - \sum_{\substack{e \in E, i \in e \\ e \cap \{i+1,\dots,n\} \neq \emptyset}} w_e \Big) = \sum_{e \in E} w_e \Big(\max_{i \in e} f_i - \min_{j \in e} f_j \Big).$$

It is easy to see that the Lovasz extension of the hypergraph cut is a convex function. Since the maximum of convex functions is convex, $-\min_{i \in e} f_i = \max_{i \in e} f_i$ and the hyperedge weights are non-negative, we have a non-negative combination of convex functions which is convex. Alternatively, one could use that the hypergraph cut is submodular and the Lovasz extension of every submodular set function is convex. $\square$

Note that the total variation of a hypergraph cut reduces to the total variation on graphs if H is 2-uniform (standard graph). There is an interesting relation of the total variation on hypergraphs to sparsity inducing group norms. Namely, defining for each edge $e \in E$ the difference operator $D_e : \mathbb{R}^{|V|} \to \mathbb{R}^{|V| \times |V|}$ by $(D_e f)_{ij} = f_i - f_j$ if $i, j \in e$ and 0 otherwise, TV_H can be written as, $\mathrm{TV}_H(f) = \sum_{e \in E} w_e \|D_e f\|_\infty$, which can be seen as inducing group sparse structure on the gradient level. The groups are the hyperedges and thus are typically overlapping. This could lead potentially to extensions of the elastic net on graphs to hypergraphs.

It is known that using the total variation on graphs as a regularization functional in semi-supervised learning (SSL) leads to very spiky solutions for small numbers of labeled points. Thus, one would like to have regularization functionals enforcing more smoothness of the solutions. For graphs this is achieved by using the family of regularization functionals $\Omega_{G,p} : \mathbb{R}^{|V|} \to \mathbb{R}$,

$$\Omega_{G,p}(f) = \frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|^p.$$

For $p = 2$ we get regularization functional of the (combinatorial) graph Laplacian which is the basis of a large class of methods for SSL on graphs. In analogy to graphs, we define the corresponding family of regularizers on hypergraphs.

Definition 2.2. *The regularization functionals $\Omega_{H,p} : \mathbb{R}^{|V|} \to \mathbb{R}$ for a hypergraph $H = (V, E, w)$ are defined for $p \geq 1$ as*

$$\Omega_{H,p}(f) = \sum_{e \in E} w_e \Big(\max_{i \in e} f_i - \min_{j \in e} f_j \Big)^p.$$

Lemma 2.1. *The functionals $\Omega_{H,p} : \mathbb{R}^{|V|} \to \mathbb{R}$ are convex.*

Proof. The p-th power of positive, convex functions for $p \geq 1$ is convex as

$$\big(f(\lambda x + (1-\lambda) y) \big)^p \leq \big(\lambda f(x) + (1-\lambda) f(y) \big)^p \leq \lambda f(x)^p + (1-\lambda) f(y)^p$$

where the last inequality follows from the convexity of x^p on $\mathbb{R}_+$. Thus, the p-th power of $\max_{i \in e} f_i - \min_{j \in e} f_j$ is convex. $\square$

Note that $\Omega_{H,1}(f) = \mathrm{TV}_H(f)$. If H is a graph, $\Omega_{H,2}$ reduces to the Laplacian regularization $\Omega_{G,2}$. Note that for characteristic functions of sets, $f = \mathbf{1}_C$, it holds $\Omega_{H,p}(\mathbf{1}_C) = \mathrm{cut}_H(C, \overline{C})$. Thus, the difference between the hypergraph cut and its approximations such as clique and star expansion carries over to $\Omega_{H,p}$ and $\Omega_{G_{CE},p}$.

3   Semi-supervised Learning

With the regularization functionals derived in the last section, we can immediately write down a formulation for two-class semi-supervised learning on hypergraphs similar to the well-known approaches of [19, 20]. Given the label set L we construct the vector $Y \in \mathbb{R}^n$ with $Y_i = 0$ if $i \notin L$

and Y_i equal to the label in $\{-1, 1\}$ if $i \in L$. We propose to solve

$$f^* = \underset{f \in \mathbb{R}^{|V|}}{\arg\min} \frac{1}{2} \|f - Y\|_2^2 + \lambda \, \Omega_{H,p}(f), \tag{3}$$

where $\lambda > 0$ is the regularization parameter. In Section 5, we discuss how this convex optimization problem can be solved efficiently for the case $p = 1$ and $p = 2$. Note, that other loss functions than the squared loss could be used. However, the regularizer aims at contracting the function and we use the label set $\{-1, 1\}$ so that $f^* \in [-1, 1]^{|V|}$. However, on the interval $[-1, 1]$ the squared loss behaves very similar to other margin-based loss functions.

In general, we recommend to use $p = 2$ as it corresponds to Laplacian-type regularization for graphs which is known to work well. For graphs $p = 1$ is known to produce spiky solutions for small numbers of labeled points. This is due to the effect that cutting "out" the labeled points leads to a much smaller cut than, e.g., producing a balanced partition. However, in the case where one has only a small number of hyperedges this effect is much smaller and we will see in the experiments that $p = 1$ also leads to reasonable solutions.

4   Balanced Hypergraph Cuts

In Section 2.1, we discussed the difference between the hypergraph cut (1) and the graph cut of the clique expansion (2) of the hypergraph and gave a simple example in Figure 1 where these cuts yield quite different results. Clearly, this difference carries over to the famous normalized cut criterion introduced in [21, 22] for clustering of graphs with applications in image segmentation. For a hypergraph the ratio resp. normalized cut can be formulated as

$$\mathrm{RCut}(C, \overline{C}) = \frac{\mathrm{cut}_H(C, \overline{C})}{|C||\overline{C}|}, \quad \mathrm{NCut}(C, \overline{C}) = \frac{\mathrm{cut}_H(C, \overline{C})}{\mathrm{vol}(C) \, \mathrm{vol}(\overline{C})},$$

which incorporate different balancing criteria. Note, that in contrast to the normalized cut for graphs the normalized hypergraph cut allows *no* relaxation into a linear eigenproblem (spectral relaxation).

Thus, we follow a recent line of research [14, 15, 16, 17] where it has been shown that the standard spectral relaxation of the normalized cut used in spectral clustering [22] is loose and that a tight relaxation can be formulated in terms of a nonlinear eigenproblem. Although nonlinear eigenproblems are non-convex, one can compute nonlinear eigenvectors quite efficiently at the price of loosing global optimality. However, it has been shown that the potentially non-optimal solutions of the tight relaxation, outperform in practice the globally optimal solution of the loose relaxation, often by large margin. In this section, we extend their approach to hypergraphs and consider general balanced hypergraph cuts $\mathrm{Bcut}(C, \overline{C})$ of the form, $\mathrm{Bcut}(C, \overline{C}) = \frac{\mathrm{cut}_H(C, \overline{C})}{\hat{S}(C)}$, where $\hat{S} : 2^V \to \mathbb{R}_+$ is a non-negative, symmetric set function (that is $\hat{S}(C) = \hat{S}(\overline{C})$). For the normalized cut one has $\hat{S}(C) = \mathrm{vol}(C) \, \mathrm{vol}(\overline{C})$ whereas for the Cheeger cut one has $\hat{S}(C) = \min\{\mathrm{vol}\, C, \mathrm{vol}\, \overline{C}\}$. Other examples of balancing functions can be found in [16]. Our following result shows that the balanced hypergraph cut also has a tight relaxation into a continuous nonlinear eigenproblem [14].

Theorem 4.1. *Let $H = (V, E, w)$ be a finite, weighted hypergraph and $S : \mathbb{R}^{|V|} \to \mathbb{R}$ be the Lovász extension of the symmetric, non-negative set function $\hat{S} : 2^V \to \mathbb{R}$. Then, it holds that*

$$\min_{f \in \mathbb{R}^{|V|}} \frac{\sum_{e \in E} w_e \big(\max_{i \in e} f_i - \min_{j \in e} f_j \big)}{S(f)} = \min_{C \subset V} \frac{\mathrm{cut}_H(C, \overline{C})}{\hat{S}(C)}.$$

Further, let $f \in \mathbb{R}^{|V|}$ and define $C_t := \{i \in V \mid f_i > t\}$. Then,

$$\min_{t \in \mathbb{R}} \frac{\mathrm{cut}_H(C_t, \overline{C_t})}{\hat{S}(C_t)} \leq \frac{\sum_{e \in E} w_e \big(\max_{i \in e} f_i - \min_{j \in e} f_j \big)}{S(f)}.$$

Proof. By Prop. 2.2 the Lovász extension of $\mathrm{cut}_H(C, \overline{C})$ is given by $\sum_{e \in E} w_e \big(\max_{i \in e} f_i - \min_{j \in e} f_j \big)$.

Noting that both $\mathrm{cut}_H(C, \overline{C})$ and $\hat{S}(C)$ vanish on the full set V, the proof then follows from the recent result [17], which shows in this case the equivalence between the set problem and the continuous problem written in terms of the Lovász extensions. □

5

The last part of the theorem shows that "optimal thresholding" among all level sets of $f \in \mathbb{R}^{|V|}$ can only lead to a better or equally balanced hypergraph cut.

The question remains how to minimize the ratio $Q(f) = \frac{\text{TV}_H(f)}{S(f)}$. As discussed in [16], every Lovász extension S can be written as a difference of convex positively 1-homogeneous functions[1] $S = S_1 - S_2$. Moreover, as shown in Prop. 2.2 the total variation TV_H is convex. Thus, we have to minimize a non-negative ratio of a convex and a d.c. function[2]. For this we use the RatioDCA algorithm of [16], which for completeness is shown in Algorithm 1. The main part of RatioDCA is

---
Algorithm 1 RatioDCA – Minimization of a non-negative ratio of 1-homogeneous d.c. functions
---
1: **Input:** $Q = \frac{R_1 - R_2}{S_1 - S_2}$
2: **Initialization:** $f^0 = $ random with $\|f^0\| = 1$, $\lambda^0 = Q(f^0)$
3: **repeat**
4:   $s_1(f^k) \in \partial S_1(f^k)$, $r_2(f^k) \in \partial R_2(f^k)$
5:   $f^{k+1} = \underset{\|u\|_2 \leq 1}{\arg \min} \left\{ R_1(u) - \langle u, r_2(f^k) \rangle + \lambda^k \left(S_2(u) - \langle u, s_1(f^k) \rangle \right) \right\}$
6:   $\lambda^{k+1} = (R_1(f^{k+1}) - R_2(f^{k+1}))/(S_1(f^{k+1}) - S_2(f^{k+1}))$
7: **until** $\frac{|\lambda^{k+1} - \lambda^k|}{\lambda^k} < \epsilon$
8: **Output:** eigenvalue λ^{k+1} and eigenvector f^{k+1}.

---

the inner problem which is a convex optimization problem. In our case $R_1 = TV_H, R_2 = 0$ and the inner problem reads

$$\min_{\|u\|_2 \leq 1} \{\text{TV}_H(u) + \lambda^k \left(S_2(u) - \langle u, s_1(f^k) \rangle \right)\}. \tag{4}$$

While the strategy for general balancing functions is discussed in [16], we restrict ourselves for simplicity to submodular balancing functions, in which case S is convex and thus $S_2 = 0$. Note that the balancing functions of ratio/normalized cut and Cheeger cut are submodular. It turns out that the inner problem is very similar to the semi-supervised learning formulation (3). The efficient solution of both problems is discussed in the next section.

5 Algorithms for the Total Variation on Hypergraphs

The problem (3) we want to solve for semi-supervised learning and the inner problem (4) of RatioDCA have a common structure. They are the sum of convex functionals where one of them is the novel regularizer $\Omega_{H,p}$. We propose to solve these problems using a primal-dual algorithm, denoted PDHG in this paper, which was proposed in [23, 24]. Its main idea is to iteratively solve for each convex term in the objective function a so-called proximal problem. Solving the proximal problem w.r.t. a mapping $g : \mathbb{R}^n \to \mathbb{R}$ and a vector $\tilde{x} \in \mathbb{R}^n$ means to compute the proximal map prox_g defined by

$$\text{prox}_g(\tilde{x}) = \underset{x \in \mathbb{R}^n}{\arg \min} \{\frac{1}{2}\|x - \tilde{x}\|_2^2 + g(x)\}.$$

The main idea here is that often these proximal problems can be solved efficiently leading to a fast convergence of the overall algorithm. In order to point out the common structure of PDHG for both (3) and the inner problems of Algorithm 1, we first consider a general optimization problem of the form

$$\min_{f \in \mathbb{R}^n} \{G(f) + F(Kf)\}, \tag{5}$$

where $K \in \mathbb{R}^{m,n}$ and $G : \mathbb{R}^n \to \mathbb{R}, F : \mathbb{R}^m \to \mathbb{R}$ are lower-semicontinuous convex functions. Recall that the conjugate function of G^* of G is defined as

$$G^*(x) = \sup_{f \in \mathbb{R}^n} \{\langle x, f \rangle - G(f)\}$$

and similarly for F^*. In terms of these conjugate functions, we can write the dual problem of (5) as

$$-\min_{\alpha \in \mathbb{R}^m} \{G^*(-K^\mathsf{T}\alpha) + F^*(\alpha)\}. \tag{6}$$

---
[1]A function $f : \mathbb{R}^d \to \mathbb{R}$ is positively 1-homogeneous if $\forall \alpha > 0, f(\alpha x) = \alpha f(x)$.
[2]A function is called d.c. if it can be written as a difference of convex functions

The PDHG algorithm for (5) has the following general form. For convergence proofs we refer to [23, 24].

Algorithm 2 PDHG

1: **Initialization:** $f^{(0)} = \bar{f}^{(0)} = 0$, $\theta \in [0,1]$, $\sigma, \tau > 0$ with $\sigma\tau < 1/\|K\|_2^2$
2: **repeat**
3:     $\alpha^{(k)} = \text{prox}_{\sigma F^*}(\alpha^{(k)} + \sigma K \bar{f}^{(k)})$
4:     $f^{(k+1)} = \text{prox}_{\tau G(f)}(f^{(k)} - \tau K^{\mathsf{T}}(\alpha^{(k)}))$
5:     $\bar{f}^{(k+1)} = f^{(k+1)} + \theta(f^{(k+1)} - f^{(k)})$
6: **until** relative duality gap $< \epsilon$
7: **Output:** $f^{(k+1)}$.

We will now apply this general setting to the convex optimization problems arising in this paper. First, the following Table 1 shows how one can choose G in (5) in order to solve (3) and (4), provides the solutions of the corresponding proximal problems, and gives the conjugate functions. Note that we write the constraint in the inner problem of RatioDCA via the indicator function $\iota_{\|\cdot\|_2 \leq 1}$ defined by $\iota_{\|\cdot\|_2 \leq 1}(x) = 0$, if $\|x\|_2 \leq 1$ and $+\infty$ otherwise. Clearly, both proximal problems have an explicit solution.

$G(f) = \frac{1}{2}\|f - Y\|_2^2$	$G(f) = -\langle s_1(f^k), f\rangle + \iota_{\|\cdot\|_2 \leq 1}(f)$
$\text{prox}_{\tau G(f)}(\tilde{x}) = \frac{1}{1+\tau}(\tilde{x} + \tau Y)$	$\text{prox}_{\tau G(f)}(\tilde{x}) = \frac{\tilde{x} + \tau s_1(f^k)}{\max\{1, \|\tilde{x} + \tau s_1(f^k)\|_2\}}$
$G^*(x) = \frac{1}{2}\|x + Y\|_2^2 - \frac{1}{2}\|Y\|_2^2$	$G^*(x) = \|x + s_1(f^k)\|_2$

Table 1: Data-dependent terms of the SSL functional (3) (left) and the inner problem of RatioDCA (4) (right) with respective proximal map and conjugate.

Second, we discuss the choice of F and K to incorporate $\Omega_{H,p}$.

PDHG algorithm for $\Omega_{H,1}$. Let m_e denote the number of vertices in hyperedge $e \in E$. The main idea is to write

$$\lambda\Omega_{H,1}(f) = F(Kf) := \sum_{e \in E}(F_{(e,1)}(K_e f) + F_{(e,2)}(K_e f)), \tag{7}$$

where the rows of the matrices $K_e \in \mathbb{R}^{m_e,n}$ are the i-th standard unit vectors for $i \in e$ and the functionals $F_{(e,j)} : \mathbb{R}^{m_e} \to \mathbb{R}$ are defined as

$$F_{(e,1)}(\alpha^{(e,1)}) = \lambda w_e \max(\alpha^{(e,1)}), \quad F_{(e,2)}(\alpha^{(e,2)}) = -\lambda w_e \min(\alpha^{(e,2)}).$$

The primal problem has thus the form

$$\min_{f \in \mathbb{R}^n}\{G(f) + \sum_{e \in E}(F_{(e,1)}(K_e f) + F_{(e,2)}(K_e f))\}.$$

In contrast to the function G, we need in the PDHG algorithm the proximal maps for the conjugate functions of $F_{(e,j)}$. They are given by

$$F_{(e,1)}^* = \iota_{S_{\lambda w_e}}, \quad F_{(e,2)}^* = \iota_{-S_{\lambda w_e}},$$

where $S_{\lambda w_e} = \{x \in \mathbb{R}^{m_e} : \sum_{i=1}^{m_e} x_i = \lambda w_e, x_i \geq 0\}$ is the scaled simplex in $\mathbb{R}^{m_e}$. By (6) the dual problem has the form

$$-\min_{\alpha^{(e,1)},\alpha^{(e,2)}}\{G^*(-\sum_{e \in E} K_e^{\mathsf{T}}(\alpha^{(e,1)} + \alpha^{(e,2)})) + \sum_{e \in E}(\iota_{S_{\lambda w_e}^e}(\alpha^{(e,1)}) + \iota_{-S_{\lambda w_e}^e}(\alpha^{(e,2)}))\},$$

where G^* is given as in Table 1. The solutions of the proximal problems for $F_{(e,1)}^*$ and $F_{(e,1)}^*$ are simply the orthogonal projections onto these simplexes written here as $P_{S_{\lambda w_e}^e}$ and $P_{-S_{\lambda w_e}^e}$, respectively. These projections can be performed in linear time, cf., [25].

Algorithm 3 PDHG for $\Omega_{H,1}$

1: **Initialization:** $f^{(0)} = \bar{f}^{(0)} = 0$, $\theta \in [0,1]$, $\sigma, \tau > 0$ with $\sigma\tau < 1/(2\max_{i=1,\ldots,n}\{c_i\})$
2: **repeat**
3:    $\alpha^{(e,1)(k+1)} = P_{S^e_{\lambda w_e}}(\alpha^{(e,1)(k)} + \sigma K_e \bar{f}^{(k)})$, $e \in E$
4:    $\alpha^{(e,2)(k+1)} = P_{-S^e_{\lambda w_e}}(\alpha^{(e,2)(k)} + \sigma K_e \bar{f}^{(k)})$, $e \in E$
5:    $f^{(k+1)} = \text{prox}_{\tau G}(f^{(k)} - \tau \sum_{e \in E} K_e^{\mathrm{T}}(\alpha^{(e,1)(k+1)} + \alpha^{(e,2)(k+1)}))$
6:    $\bar{f}^{(k+1)} = f^{(k+1)} + \theta(f^{(k+1)} - f^{(k)})$
7: **until** relative duality gap $< \epsilon$
8: **Output:** $f^{(k+1)}$.

Using the proximal mappings we have presented so far, we obtain Algorithm 3. In line 1, c_i is the number of hyperedges the vertex i lies in. This bound on the product of the step sizes can be derived as follows

$$\|K\|_2^2 = \|K^{\mathrm{T}}K\|_2 = 2\|\sum_{e \in E} K_e^{\mathrm{T}} K_e\|_2 = 2\max_{i=1,\ldots,n}\{c_i\}.$$

It is important to point out here that the algorithm decouples the problem in the sense that in every iteration we solve subproblems which treat the functionals $G, F_{(e,1)}, F_{(e,2)}$ separately and thus can be solved in an efficient way.

PDHG algorithm for $\Omega_{H,2}$. We define G and K_e as above. Moreover, we set

$$F_e(\alpha^e) = \lambda w_e \underbrace{(\max(\alpha^e) - \min(\alpha^e))^2}_{=:h_e(\alpha^e)}. \tag{8}$$

Hence, the primal problem can be written as

$$\min_{f \in \mathbb{R}^n}\{G(f) + \sum_{e \in E} F_e(K_e f)\}.$$

To formulate the dual problem we need the conjugate of F_e. To this end, we first derive the conjugate function of h_e defined in (8), i.e.,

$$h_e^*(\alpha^e) = \sup_{\phi \in \mathbb{R}^{m_e}} \{\langle \alpha^e, \phi \rangle - (\max(\phi) - \min(\phi))^2\}.$$

Lemma 5.1. *Let $\alpha^e \in \mathbb{R}^{m_e}$ and $t_+ = \sum_{i:\alpha_i^e > 0} \alpha_i^e$ and $t_- = \sum_{i:\alpha_i^e < 0} \alpha_i^e$. It holds that*

$$h_e^*(\alpha^e) = \begin{cases} \frac{1}{4}t_+^2 & \text{if } \langle \alpha^e, \mathbf{1} \rangle = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Using the decomposition, $\phi = \psi + \gamma \mathbf{1}$, where $\langle \psi, \mathbf{1} \rangle = 0$ and $\gamma \in \mathbb{R}$, we can write

$$\langle \alpha^e, \phi \rangle - (\max(\phi) - \min(\phi))^2 = \gamma \langle \alpha^e, \mathbf{1} \rangle + \langle \alpha^e, \psi \rangle - (\max(\psi) - \min(\psi))^2.$$

Thus for $\langle \alpha^e, \mathbf{1} \rangle \neq 0$, we have $h_e^*(\alpha^e) = \infty$. Now we consider the case where $\langle \alpha^e, \mathbf{1} \rangle = 0$. We write $I_- = \{i : \alpha_i^e < 0\}$ and $I_+ = \{i : \alpha_i^e > 0\}$ and define $t_+ = \sum_{i \in I_+} \alpha_i^e$ and $t_- = \sum_{i \in I_-} \alpha_i^e$. Note that $\langle \alpha^e, \mathbf{1} \rangle = 0$ implies $t_+ = -t_-$. Let us assume $a = \max(\phi)$ and $b = \min(\phi)$ are fixed. To maximize $\langle \alpha^e, \phi \rangle - (\max(\phi) - \min(\phi))^2$ it is clearly best to choose $\phi_i = a$ for $i \in I_-$ and $\phi_i = b$ for $i \in I_+$. Consequently,

$$\langle \alpha^e, \phi \rangle - (\max(\phi) - \min(\phi))^2 = t_+(b - a) - (b - a)^2. \tag{9}$$

We maximize the gap $\Delta = b - a$ for the objective $m(\Delta) = t_+\Delta - \Delta^2$ and obtain the maximizer as $\Delta = \frac{t_+}{2}$. Thus we have $h_e^*(\alpha^e) = \frac{t_+^2}{4}$ if $\langle \alpha^e, \mathbf{1} \rangle \neq 0$. $\qquad\square$

With $t_+ = \sum_{i:\alpha_i^e > 0} \alpha_i^e$ and $t_- = \sum_{i:\alpha_i^e < 0} \alpha_i^e$ we thus get

$$F_e^*(\alpha^e) = \lambda w_e h^*\left(\frac{\alpha^e}{\lambda w_e}\right) = \begin{cases} \frac{1}{4\lambda w_e}t_+^2 & \text{if } t_+ = -t_-, \\ +\infty & \text{otherwise.} \end{cases} \tag{10}$$

8

So, we obtain the dual problem

$$-\min_{\alpha^e}\{G^*(-\sum_{e\in E} K_e^{\mathrm{T}}\alpha^e) + \sum_{e\in E}\frac{1}{4\lambda w_e}(t_+^e)^2 + \sum_{e\in E}\iota_{\{0\}}(t_+^e + t_-^e)\},$$

where $t_+^e = \sum_{i:\alpha_i^e>0}\alpha_i^e$ and $t_-^e = \sum_{i:\alpha_i^e<0}\alpha_i^e$.

As we have seen in (10), the conjugate functions F_e^* are not indicator functions and we thus solve the corresponding proximal problems via proximal problems for F_e. More specifically, we exploit the fact that

$$\mathrm{prox}_{\sigma F_e^*}(\tilde{\alpha}^e) = \tilde{\alpha}^e - \mathrm{prox}_{\frac{1}{\sigma}F_e}(\tilde{\alpha}^e), \tag{11}$$

see [26, Lemma 2.10], and use the following novel result concerning the proximal problem on the right-hand side of (11).

Proposition 5.1. *For any $\sigma > 0$ and any $\tilde{\alpha}^e \in \mathbb{R}^{m_e}$ the proximum*

$$\mathrm{prox}_{\frac{1}{\sigma}F_e}(\tilde{\alpha}^e) = \arg\min_{\alpha^e \in \mathbb{R}^{m_e}}\{\frac{1}{2}\|\alpha^e - \tilde{\alpha}^e\|_2^2 + \frac{1}{\sigma}\lambda w_e(\max(\alpha^e) - \min(\alpha^e))^2\}$$

can be computed with $\mathcal{O}(m_e \log m_e)$ arithmetic operations.

We will now derive such an algorithm. To simplify the notation, we consider instead of $\frac{1}{\sigma}F_e$ the function $h : \mathbb{R}^m \to \mathbb{R}$ defined by

$$h(\alpha) = (\max(\alpha) - \min(\alpha))^2$$

and show that $\mathrm{prox}_{\mu h}(\alpha)$, $\mu > 0$, can be computed with $\mathcal{O}(m \log m)$ arithmetic operations.

Let us fix $\alpha \in \mathbb{R}^m$. For every pair $r, s \in [\min(\alpha), \max(\alpha)]$ with $r \geq s$, we define $\alpha^{(r,s)}$ by

$$\alpha_i^{(r,s)} = \begin{cases} r & \text{if } \alpha_i \geq r \\ \alpha_i & \text{if } \alpha_i \in (r, s) \\ s & \text{if } \alpha_i \leq s \end{cases} \tag{12}$$

Clearly, if $r = \max(\mathrm{prox}_{\mu h}(\alpha))$ and $s = \min(\mathrm{prox}_{\mu h}(\alpha))$ then $\alpha^{(r,s)} = \mathrm{prox}_{\mu h}(\alpha)$. Hence, the above definition allows us to write the proximal problem in terms of the variables r, s since for

$$(r, s) = \arg\min_{\tilde{r},\tilde{s}}\{\underbrace{\frac{1}{2}\|\alpha^{(\tilde{r},\tilde{s})} - \alpha\|_2^2}_{=:E_1(\tilde{r},\tilde{s})} + \underbrace{\mu(\tilde{r} - \tilde{s})^2}_{=:E_2(\tilde{r},\tilde{s})}\} \tag{13}$$

we have

$$\mathrm{prox}_{\mu h}(\alpha) = \alpha^{(r,s)}.$$

Our goal is now to find a minimizer of (13). To this end, we first order α in an increasing order which can be done in $\mathcal{O}(m \log m)$ arithmetic operations. W.l.o.g. we assume here that the components of α are pairwise different. Moreover, we introduce the following notation. For $r, s \in [\alpha_1, \alpha_m]$ there exist unique $p, q \in \{1, \ldots, m\}$ characterized by $\alpha_{m-p+1} = \min\{\alpha_i | \alpha_i \geq r\}$ and $\alpha_q = \max\{\alpha_i | \alpha_i \leq s\}$. Thus, the directional partial derivatives w.r.t. r and s are given by

$$\frac{\partial E_1}{\partial r^-}(r, s) = \sum_{i=m-p+1}^{m}(\alpha_i - r), \qquad \frac{\partial E_1}{\partial s^+}(r, s) = \sum_{i=1}^{q}(s - \alpha_i). \tag{14}$$

They tell us how much we increase E_1 by decreasing r and increasing s, respectively. On the other hand both of these changes lead to a decrease in the energy E_2. More precisely, it holds that

$$\frac{\partial E_2}{\partial r^-}(r, s) = \frac{\partial E_2}{\partial s^+}(r, s) = 2\mu(s - r). \tag{15}$$

Thus, the main ideas behind our algorithm are as follows. Starting with $r = \max(\alpha)$ and $s = \min(\alpha)$, we decrease r and increase s keeping the two partial derivatives of (14) equal. We stop when the sum of the partial derivatives vanishes. So, the optimal r, s are characterized by the system

$$\sum_{i=m-p+1}^{m}(\alpha_i - r) = \sum_{i=1}^{q}(s - \alpha_i), \tag{16}$$

$$\sum_{i=m-p+1}^{m}(\alpha_i - r) + 2\mu(s - r) = 0. \tag{17}$$

9

We will now generate a sequence of pairs $r^{(k)}, s^{(k)}$ satisfying $r^{(k)} \geq s^{(k)}$ and (16) for each k. The corresponding indices needed to calculate the partial derivatives will be denoted by $p^{(k)}, q^{(k)}$. The main procedure is described in the next lemma.

Lemma 5.2. *Assume $r^{(k)} \in (\alpha_{m-p^{(k)}}, \alpha_{m-p^{(k)}+1}]$ and $s^{(k)} \in [\alpha_{q^{(k)}}, \alpha_{q^{(k)}+1})$ and property* (16) *holds for $(r^{(k)}, s^{(k)})$. Then, we can either choose*

$$r^{(k+1)} = r^{(k)} - \frac{q^{(k)}}{p^{(k)}}(s^{(k+1)} - s^{(k)}) \text{ and } s^{(k+1)} = \alpha_{q^{(k)}+1} \tag{18}$$

or

$$r^{(k+1)} = \alpha_{m-p^{(k)}} \text{ and } s^{(k+1)} = s^{(k)} + \frac{p^{(k)}}{q^{(k)}}(r^{(k)} - r^{(k+1)}) \tag{19}$$

such that $r^{(k+1)} \in [\alpha_{m-p^{(k)}}, \alpha_{m-p^{(k)}+1})$, $s^{(k+1)} \in (\alpha_{q^{(k)}}, \alpha_{q^{(k)}+1}]$ and (16) *holds true for $(r^{(k+1)}, s^{(k+1)})$.*

Proof. Property (16) for $(r^{(k+1)}, s^{(k+1)})$ means that

$$\sum_{i=m-p^{(k)}+1}^{m} (\alpha_i - r^{(k+1)}) = \sum_{i=1}^{q^{(k)}} (s^{(k+1)} - \alpha_i). \tag{20}$$

Since by assumption (16) holds for $(r^{(k)}, s^{(k)})$, equation (20) is equivalent to

$$p^{(k)}(r^{(k+1)} - r^{(k)}) = q^{(k)}(s^{(k)} - s^{(k+1)}).$$

If we set $(r^{(k+1)}, s^{(k+1)})$ according to (18) but $r^{(k+1)} < \alpha_{m-p^{(k)}}$. Then we get

$$r^{(k)} - \frac{q^{(k)}}{p^{(k)}}(\alpha_{q^{(k)}+1} - s^{(k)}) < \alpha_{m-p^{(k)}}$$

$$\Rightarrow s^{(k)} + \frac{p^{(k)}}{q^{(k)}}(r^{(k)} - \alpha_{m-p^{(k)}}) < \alpha_{q^{(k)}+1},$$

i.e., we can choose $r^{(k+1)}, s^{(k+1)}$ according to (19) and vice versa. $\quad\square$

After each computation of a new pair $(r^{(k+1)}, s^{(k+1)})$ we check if the left-hand side of (17) is smaller than zero (note that initially the left-hand side of (17) is negative and it is increasing for every iteration). If this is not the case, we found the intervals where the optimal values r and s lie in. Restricted to this domain the functional $E_1 + E_2$ is a differentiable. Hence, we can compute r, s as follows.

Lemma 5.3. *Assume that the optimal r, s of* (13) *fulfill $r \in [\alpha_{m-p}, \alpha_{m-p+1}]$ and $s \in [\alpha_q, \alpha_{q+1}]$. Then, it holds that*

$$s = \left(q + 2\mu - \frac{(2\mu)^2}{\sum_{i=m-p+1}^{m} \alpha_i + 2\mu}\right)^{-1} \left(\frac{2\mu}{p + 2\mu} \sum_{i=m-p+1}^{m} \alpha_i + \sum_{i=1}^{q} \alpha_i\right)$$

$$r = \frac{1}{2\mu}\left((q + 2\mu)s - \sum_{i=1}^{q} \alpha_i\right).$$

Proof. When restricted to $[\alpha_i, \alpha_{i+1}] \times [\alpha_j, \alpha_{j+1}]$, the function $(r, s) \mapsto E_1(r, s) + E_2(r, s)$ is a quadratic function in (r, s). We can thus simply set the gradient to zero and solve the corresponding system of linear equations which yields the above result. $\quad\square$

In conclusion, we obtain the following algorithm. Note that after the sorting, the algorithm takes in the order of m steps to compute the proximum which proves Proposition 5.1.

Hence, the corresponding PDHG algorithm can be formulated as follows.

Algorithm 4 – Solution of the proximal problem $\text{prox}_{\mu h}(\alpha)$

1: Sort $\alpha \in \mathbb{R}^m$ in increasing order.
2: **Initialization:** $r^{(0)} = \max(\alpha), s^{(0)} = \min(\alpha)$
3: **while** $\frac{\partial E_1}{\partial r^-}(r^{(k)}, s^{(k)}) < 2\mu(r^{(k)} - s^{(k)})$ and $q^{(k)} + 1 \leq m - p^{(k)}$ **do**
4:    Find $(r^{(k+1)}, s^{(k+1)})$ according to Lemma 5.2.
5: **end while**
6: Compute r, s as described in Lemma 5.3.
7: **Output:** After restoring the original order, set

$$(\text{prox}_{\mu h}(\alpha))_i = \begin{cases} r & \text{if } \alpha_i \geq r, \\ \alpha_i & \text{if } \alpha_i \in (r, s), \quad \text{for } i = 1, \dots, m. \\ s & \text{if } \alpha_i \leq s, \end{cases}$$

Algorithm 5 PDHG for $\Omega_{H,2}$

1: **Initialization:** $f^{(0)} = \bar{f}^{(0)} = 0, \theta \in [0,1], \sigma, \tau > 0$ with $\sigma\tau < 1/\max_{i=1,\dots,n}\{c_i\}$
2: **repeat**
3:    $\alpha^{e(k+1)} = \alpha^{e(k)} + \sigma K_e \bar{f}^{(k)} - \text{prox}_{\frac{1}{\sigma}F_e}(\alpha^{e(k)} + \sigma K_e \bar{f}^{(k)}), \quad e \in E$
4:    $f^{(k+1)} = \text{prox}_{\tau G}(f^{(k)} - \tau \sum_{e \in E} K_e^{\mathrm{T}}(\alpha^{e(k+1)}))$
5:    $\bar{f}^{(k+1)} = f^{(k+1)} + \theta(f^{(k+1)} - f^{(k)})$
6: **until** relative duality gap $< \epsilon$
7: **Output:** $f^{(k+1)}$.

We solve the subproblems in line 3 via Algorithm 4. Note that the bound on the step sizes is now doubled, i.e., less restrictive since we have defined for each hyperedge one functional F_e and not two as for $p = 1$, i.e.,

$$\|K\|_2^2 = \|K^{\mathrm{T}}K\|_2 = \|\sum_{e \in E} K_e^{\mathrm{T}} K_e\|_2 = \max_{i=1,\dots,n}\{c_i\}.$$

6   Experiments

As our literature review revealed, in practice the method of Zhou et al [11] is most often used for clustering and semi-supervised learning on hypergraphs. We compare to them on the selection of UCI datasets which are summarized in Table 2. Zoo, Mushrooms and 20Newsgroups[3] have been used also in [11] and contain only categorical features. As in [11], a hyperedge of weight one is created by all data points which have same the value of a categorical feature. This is repeated for all possible values of every categorical feature. For covertype a subset of the features is numerical. These features are turned into categorical features by binning them into 10 bins of equal size. We create two datasets each with two classes (4,5 and 6,7) of the original dataset.

---

[3]This is a modified version by Sam Roweis of the original 20 newsgroups dataset available at `http://www.cs.nyu.edu/~roweis/data/20news_w100.mat`.

Prop. \ Dataset	Zoo	Mushrooms	Covertype (4,5)	Covertype (6,7)	20Newsgroups
Number of classes	7	2	2	2	4
$|V|$	101	8124	12240	37877	16242
$|E|$	42	112	104	123	100
$\sum_{e \in E}|e|$	1717	170604	146880	454522	65451
$|E|$ of Clique Exp.	10201	65999376	143008092	1348219153	53284642

Table 2: Datasets used for SSL and clustering. Note that the clique expansion leads for all datasets to a graph which is close to being fully connected as all datasets contain large hyperedges. For covertype (6,7) the weight matrix needs over 10GB of memory, the original hypergraph only 4MB.

Semi-supervised Learning. In [11], they suggest to use a regularizer induced by the normalized Laplacian L_{CE} arising from the clique expansion

$$L_{CE} = \mathbb{I} - D_{CE}^{-\frac{1}{2}} H W' H^T D_{CE}^{-\frac{1}{2}},$$

where D_{CE} is a diagonal matrix with entries $d_{EC}(i) = \sum_{e \in E} H_{i,e} \frac{w_e}{|e|}$ and $W' \in \mathbb{R}^{|E| \times |E|}$ is a diagonal matrix with entries $w'(e) = w_e/|e|$. The semi-supervised learning problem can then be formulated with $\lambda > 0$ as

$$\arg\min_{f \in \mathbb{R}^{|V|}} \{ \|f - Y\|_2^2 + \lambda \langle f, L_{CE} f \rangle \}.$$

The advantage of this formulation is that the solution can be found via a linear system. However, as Table 2 indicates the obvious downside is that L_{CE} is a potentially very dense matrix and thus one needs in the worst case $|V|^2$ memory and $O(|V|^3)$ computations. This is in contrast to our method which needs $2 \sum_{e \in E} |e| + |V|$ memory. For the largest example (covertype 6,7), where the clique expansion fails due to memory problems, our method takes 30-100s (depending on λ). We stop our method for all experiments when we achieve a relative duality gap of 10^{-6}.

In the experiments we do 10 trials for different numbers of labeled points. We choose the regularization parameter for both methods from the set 10^{-k}, where $k = \{0, 1, 2, 3, 4, 5, 6\}$ via 5-fold cross validation. The resulting errors and standard deviations can be found in the following table (first row lists the no. of labeled points).

Zoo	20	25	30	35	40	45	50
Zhou et al.	35.1 ± 17.2	30.3 ± 7.9	40.7 ± 14.2	29.7 ± 8.8	32.9 ± 16.8	27.6 ± 10.8	25.3 ± 14.4
$\Omega_{H,1}$	2.9 ± 3.0	$\mathbf{1.4 \pm 2.2}$	$\mathbf{2.2 \pm 2.1}$	$\mathbf{0.7 \pm 1.0}$	$\mathbf{0.7 \pm 1.5}$	$\mathbf{0.9 \pm 1.4}$	1.9 ± 3.0
$\Omega_{H,2}$	$\mathbf{2.3 \pm 1.9}$	1.5 ± 2.4	2.9 ± 2.3	0.9 ± 1.4	0.8 ± 1.7	1.2 ± 1.8	$\mathbf{1.6 \pm 2.9}$

Mushr.	20	40	60	80	100	120	160	200
Zhou et al.	$\mathbf{15.5 \pm 12.8}$	10.9 ± 4.4	9.5 ± 2.7	10.3 ± 2.0	9.0 ± 4.5	8.8 ± 1.4	8.8 ± 2.3	9.3 ± 1.0
$\Omega_{H,1}$	19.5 ± 10.5	10.8 ± 3.7	$\mathbf{7.4 \pm 3.8}$	5.6 ± 1.9	5.7 ± 2.2	5.4 ± 2.4	5.6 ± 3.8	5.6 ± 3.8
$\Omega_{H,2}$	18.4 ± 7.4	$\mathbf{9.8 \pm 4.5}$	9.9 ± 5.5	6.4 ± 2.7	6.3 ± 2.5	$\mathbf{4.5 \pm 1.8}$	$\mathbf{4.4 \pm 2.1}$	$\mathbf{3.0 \pm 0.6}$

covert45	20	40	60	80	100	120	160	200
Zhou et al.	$\mathbf{18.9 \pm 4.6}$	18.3 ± 5.2	17.2 ± 6.7	16.6 ± 6.4	17.6 ± 5.2	18.4 ± 5.1	19.2 ± 4.0	20.4 ± 2.9
$\Omega_{H,1}$	21.4 ± 0.9	17.6 ± 2.6	12.6 ± 4.3	7.6 ± 3.5	6.2 ± 3.8	4.5 ± 3.6	2.6 ± 1.6	1.5 ± 1.3
$\Omega_{H,2}$	20.7 ± 2.0	$\mathbf{16.1 \pm 4.1}$	$\mathbf{10.9 \pm 4.9}$	$\mathbf{5.9 \pm 3.7}$	$\mathbf{4.6 \pm 3.4}$	$\mathbf{3.3 \pm 3.1}$	$\mathbf{2.2 \pm 1.8}$	$\mathbf{1.0 \pm 1.1}$

covert67	20	40	60	80	100	120	160	200
$\Omega_{H,1}$	40.6 ± 8.9	6.4 ± 10.4	3.6 ± 3.2	3.3 ± 2.5	1.8 ± 0.8	1.3 ± 0.9	0.9 ± 0.4	1.2 ± 0.9
$\Omega_{H,2}$	$\mathbf{25.2 \pm 18.3}$	$\mathbf{4.3 \pm 9.6}$	$\mathbf{2.1 \pm 2.0}$	$\mathbf{2.2 \pm 1.4}$	$\mathbf{1.4 \pm 1.1}$	$\mathbf{1.0 \pm 0.8}$	$\mathbf{0.7 \pm 0.4}$	$\mathbf{1.1 \pm 0.8}$

20news	20	40	60	80	100	120	160	200
Zhou et al.	$\mathbf{45.5 \pm 7.5}$	$\mathbf{34.4 \pm 3.1}$	$\mathbf{31.5 \pm 1.4}$	$\mathbf{29.8 \pm 4.0}$	$\mathbf{27.0 \pm 1.3}$	$\mathbf{27.3 \pm 1.5}$	$\mathbf{25.7 \pm 1.4}$	$\mathbf{25.0 \pm 1.3}$
$\Omega_{H,1}$	65.7 ± 6.1	61.4 ± 6.1	53.2 ± 5.7	46.2 ± 3.7	42.4 ± 3.3	40.9 ± 3.2	36.1 ± 1.5	34.7 ± 3.6
$\Omega_{H,2}$	55.0 ± 4.8	48.0 ± 6.0	45.0 ± 5.9	40.3 ± 3.0	38.3 ± 2.7	38.1 ± 2.6	35.0 ± 2.8	34.1 ± 2.0

We observe that our SSL methods based on $\Omega_{H,p}$, $p = 1, 2$ outperform consistently the clique expansion technique of Zhou et al [11] on all datasets except 20newsgroups[4]. However, 20newsgroups is a very difficult dataset as only 10,267 out of the 16,242 data points are different which leads to a minimum possible error of 9.6%. A method based on pairwise interaction such as the clique expansion can better deal with such label noise as the hyperedges for this dataset are very large and thus accumulate the label noise. On all other datasets we observe that incorporating hypergraph structure leads to much better results. As expected our squared TV functional ($p = 2$) outperforms slightly the total variation on hypergraph ($p = 1$) even though the difference is small. However, as $\Omega_{H,2}$ reduces to the standard regularization based on the graph Laplacian, which is known to work well, we recommend this for general SSL on hypergraphs.

Clustering. We use the normalized hypergraph cut as clustering objective. For more than two clusters we recursively partition the hypergraph until the desired number of clusters is reached. For comparison we use the normalized spectral clustering approach based on the Laplacian L_{CE} [11](clique expansion). The first part (first 6 columns) of the following table shows the clustering errors of both methods as well as the normalized cuts achieved by these methods on the hypergraph and on the graph resulting from the clique expansion. Moreover, we show results (last 4 columns) which are obtained based on a kNN graph (unit weights) which is built based on the Hamming distance (note that we have categorical features) in order to check if the hypergraph modeling of the

---

[4]Note that the results on 20newsgroups differ from [11] and communication with the authors of [11] could not clarify the difference.

problem is actually useful compared to a standard similarity based graph construction. The number k is chosen as the smallest number for which the graph becomes connected and we compare results of normalized 1-spectral clustering [14] and the standard spectral clustering [22]. Note that the employed hypergraph construction has no free parameter.

Dataset	Clustering Error %		Hypergraph Ncut		Graph(CE) Ncut		Clustering Error %		kNN-Graph Ncut	
	Ours	[11]	Ours	[11]	Ours	[11]	[14]	[22]	[14]	[22]
Mushrooms	10.98	32.25	0.0011	0.0013	0.6991	0.7053	48.2	48.2	1e-4	1e-4
Zoo	16.83	15.84	0.6739	0.6784	5.1315	5.1703	5.94	5.94	1.636	1.636
20-newsgroup	47.77	33.20	0.0176	0.0303	2.3846	1.8492	66.38	66.38	0.1031	0.1034
covertype (4,5)	22.44	22.44	0.0018	0.0022	0.7400	0.6691	22.44	22.44	0.0152	0.02182
covertype (6,7)	8.16	-	8.18e-4	-	0.6882	-	45.85	45.85	0.0041	0.0041

First, we observe that optimizing the normalized cuts on hypergraphs yields better or similar results in terms of clustering errors compared to the clique expansion (except for 20-newsgroup for the same reason given in the previous paragraph). The improvement is significant in case of Mushrooms while for Zoo our clustering error is slightly higher. However, we always achieve smaller normalized hypergraph cuts. Moreover, our method sometimes has even smaller cuts on the graphs resulting from the clique expansion, although it does not directly optimize this objective in contrast to [11]. Again, we could not run the method of [11] on covertype (6,7) since the weight matrix is very dense. Second, the comparison to a standard graph-based approach where the similarity structure is obtained using the Hamming distance on the categorical feature shows that using hypergraph structure is indeed useful. Nevertheless, we think that there is room for improvement regarding the construction of the hypergraph.

Acknowledgments

References

[1] Y. Huang, Q. Liu, and D. Metaxas. Video object segmentation by hypergraph cut. In *CVPR*, pages 1738 – 1745, 2009.

[2] P. Ochs and T. Brox. Higher order motion models and spectral clustering. In *CVPR*, pages 614–621, 2012.

[3] S. Klamt, U.-U. Haus, and F. Theis. Hypergraphs and cellular networks. *PLoS Computational Biology*, 5:e1000385, 2009.

[4] Z. Tian, T. Hwang, and R. Kuang. A hypergraph-based learning algorithm for classifying gene expression and arraycgh data with prior knowledge. *Bioinformatics*, 25:2831–2838, 2009.

[5] D. Gibson, J. Kleinberg, and P. Raghavan. Clustering categorical data: an approach based on dynamical systems. *VLDB Journal*, 8:222–236, 2000.

[6] J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, L. Zhang, and X. He. Music recommendation by unified hypergraph: Combining social media information and music content. In *Proc. of the Int. Conf. on Multimedia (MM)*, pages 391–400, 2010.

[7] A. Shashua, R. Zass, and T. Hazan. Multi-way clustering using super-symmetric non-negative tensor factorization. In *ECCV*, pages 595–608, 2006.

[8] S. Rota Bulo and M. Pellilo. A game-theoretic approach to hypergraph clustering. In *NIPS*, pages 1571– 1579, 2009.

[9] M. Leordeanu and C. Sminchisescu. Efficient hypergraph clustering. In *AISTATS*, pages 676–684, 2012.

[10] S. Agarwal, J. Lim, L. Zelnik-Manor, P. Petrona, D. J. Kriegman, and S. Belongie. Beyond pairwise clustering. In *CVPR*, pages 838–845, 2005.

[11] D. Zhou, J. Huang, and B. Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *NIPS*, pages 1601–1608, 2006.

[12] S. Agarwal, K. Branson, and S. Belongie. Higher order learning with graphs. In *ICML*, pages 17–24, 2006.

[13] E. Ihler, D. Wagner, and F. Wagner. Modeling hypergraphs by graphs with the same mincut properties. *Information Processing Letters*, 45:171–175, 1993.

[14] M. Hein and T. Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA. In *NIPS*, pages 847–855, 2010.

[15] A. Szlam and X. Bresson. Total variation and Cheeger cuts. In *ICML*, pages 1039–1046, 2010.

[16] M. Hein and S. Setzer. Beyond spectral clustering - tight relaxations of balanced graph cuts. In *NIPS*, pages 2366–2374, 2011.

[17] T. Bühler, S. Rangapuram, S. Setzer, and M. Hein. Constrained fractional set programs and their application in local clustering and community detection. In *ICML*, pages 624–632, 2013.

[18] F. Bach. Learning with submodular functions: A convex optimization perspective. *CoRR*, abs/1111.6453, 2011.

[19] M. Belkin and P. Niyogi. Semi-supervised learning on manifolds. *Machine Learning*, 56:209–239, 2004.

[20] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf. Learning with local and global consistency. In *NIPS*, volume 16, pages 321–328, 2004.

[21] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 22(8):888–905, 2000.

[22] U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007.

[23] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[24] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. of Math. Imaging and Vision*, 40:120–145, 2011.

[25] K. Kiwiel. On Linear-Time algorithms for the continuous quadratic knapsack problem. *J. Opt. Theory Appl.*, 134(3):549–554, 2007.

[26] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.

[29] E. Faou, V. Gradinaru, and C. Lubich. Computing Semi-classical Quantum Dynamics with Hagedorn Wavepackets. Preprint 29, DFG-SPP 1324, November 2009.

[30] D. Conte and C. Lubich. An Error Analysis of the Multi-configuration Time-dependent Hartree Method of Quantum Dynamics. Preprint 30, DFG-SPP 1324, November 2009.

[31] C. E. Powell and E. Ullmann. Preconditioning Stochastic Galerkin Saddle Point Problems. Preprint 31, DFG-SPP 1324, November 2009.

[32] O. G. Ernst and E. Ullmann. Stochastic Galerkin Matrices. Preprint 32, DFG-SPP 1324, November 2009.

[33] F. Lindner and R. L. Schilling. Weak Order for the Discretization of the Stochastic Heat Equation Driven by Impulsive Noise. Preprint 33, DFG-SPP 1324, November 2009.

[34] L. Kämmerer and S. Kunis. On the Stability of the Hyperbolic Cross Discrete Fourier Transform. Preprint 34, DFG-SPP 1324, December 2009.

[35] P. Cerejeiras, M. Ferreira, U. Kähler, and G. Teschke. Inversion of the noisy Radon transform on $SO(3)$ by Gabor frames and sparse recovery principles. Preprint 35, DFG-SPP 1324, January 2010.

[36] T. Jahnke and T. Udrescu. Solving Chemical Master Equations by Adaptive Wavelet Compression. Preprint 36, DFG-SPP 1324, January 2010.

[37] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Irregular Shearlet Frames: Geometry and Approximation Properties. Preprint 37, DFG-SPP 1324, February 2010.

[38] G. Kutyniok and W.-Q Lim. Compactly Supported Shearlets are Optimally Sparse. Preprint 38, DFG-SPP 1324, February 2010.

[39] M. Hansen and W. Sickel. Best m-Term Approximation and Tensor Products of Sobolev and Besov Spaces – the Case of Non-compact Embeddings. Preprint 39, DFG-SPP 1324, March 2010.

[40] B. Niu, F.J. Hickernell, T. Müller-Gronbach, and K. Ritter. Deterministic Multilevel Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 40, DFG-SPP 1324, March 2010.

[41] P. Kittipoom, G. Kutyniok, and W.-Q Lim. Construction of Compactly Supported Shearlet Frames. Preprint 41, DFG-SPP 1324, March 2010.

[42] C. Bender and J. Steiner. Error Criteria for Numerical Solutions of Backward SDEs. Preprint 42, DFG-SPP 1324, April 2010.

[43] L. Grasedyck. Polynomial Approximation in Hierarchical Tucker Format by Vector-Tensorization. Preprint 43, DFG-SPP 1324, April 2010.

[44] M. Hansen und W. Sickel. Best m-Term Approximation and Sobolev-Besov Spaces of Dominating Mixed Smoothness - the Case of Compact Embeddings. Preprint 44, DFG-SPP 1324, April 2010.

[45] P. Binev, W. Dahmen, and P. Lamby. Fast High-Dimensional Approximation with Sparse Occupancy Trees. Preprint 45, DFG-SPP 1324, May 2010.

[46] J. Ballani and L. Grasedyck. A Projection Method to Solve Linear Systems in Tensor Format. Preprint 46, DFG-SPP 1324, May 2010.

[47] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, and P. Wojtaszczyk. Convergence Rates for Greedy Algorithms in Reduced Basis Methods. Preprint 47, DFG-SPP 1324, May 2010.

[48] S. Kestler and K. Urban. Adaptive Wavelet Methods on Unbounded Domains. Preprint 48, DFG-SPP 1324, June 2010.

[49] H. Yserentant. The Mixed Regularity of Electronic Wave Functions Multiplied by Explicit Correlation Factors. Preprint 49, DFG-SPP 1324, June 2010.

[50] H. Yserentant. On the Complexity of the Electronic Schrödinger Equation. Preprint 50, DFG-SPP 1324, June 2010.

[51] M. Guillemard and A. Iske. Curvature Analysis of Frequency Modulated Manifolds in Dimensionality Reduction. Preprint 51, DFG-SPP 1324, June 2010.

[52] E. Herrholz and G. Teschke. Compressive Sensing Principles and Iterative Sparse Recovery for Inverse and Ill-Posed Problems. Preprint 52, DFG-SPP 1324, July 2010.

[53] L. Kämmerer, S. Kunis, and D. Potts. Interpolation Lattices for Hyperbolic Cross Trigonometric Polynomials. Preprint 53, DFG-SPP 1324, July 2010.

[54] G. Kutyniok and W.-Q Lim. Shearlets on Bounded Domains. Preprint 54, DFG-SPP 1324, July 2010.

[55] A. Zeiser. Wavelet Approximation in Weighted Sobolev Spaces of Mixed Order with Applications to the Electronic Schrödinger Equation. Preprint 55, DFG-SPP 1324, July 2010.

[56] G. Kutyniok, J. Lemvig, and W.-Q Lim. Compactly Supported Shearlets. Preprint 56, DFG-SPP 1324, July 2010.

[57] A. Zeiser. On the Optimality of the Inexact Inverse Iteration Coupled with Adaptive Finite Element Methods. Preprint 57, DFG-SPP 1324, July 2010.

[58] S. Jokar. Sparse Recovery and Kronecker Products. Preprint 58, DFG-SPP 1324, August 2010.

[59] T. Aboiyar, E. H. Georgoulis, and A. Iske. Adaptive ADER Methods Using Kernel-Based Polyharmonic Spline WENO Reconstruction. Preprint 59, DFG-SPP 1324, August 2010.

[60] O. G. Ernst, A. Mugler, H.-J. Starkloff, and E. Ullmann. On the Convergence of Generalized Polynomial Chaos Expansions. Preprint 60, DFG-SPP 1324, August 2010.

[61] S. Holtz, T. Rohwedder, and R. Schneider. On Manifolds of Tensors of Fixed TT-Rank. Preprint 61, DFG-SPP 1324, September 2010.

[62] J. Ballani, L. Grasedyck, and M. Kluge. Black Box Approximation of Tensors in Hierarchical Tucker Format. Preprint 62, DFG-SPP 1324, October 2010.

[63] M. Hansen. On Tensor Products of Quasi-Banach Spaces. Preprint 63, DFG-SPP 1324, October 2010.

[64] S. Dahlke, G. Steidl, and G. Teschke. Shearlet Coorbit Spaces: Compactly Supported Analyzing Shearlets, Traces and Embeddings. Preprint 64, DFG-SPP 1324, October 2010.

[65] W. Hackbusch. Tensorisation of Vectors and their Efficient Convolution. Preprint 65, DFG-SPP 1324, November 2010.

[66] P. A. Cioica, S. Dahlke, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Spatial Besov Regularity for Stochastic Partial Differential Equations on Lipschitz Domains. Preprint 66, DFG-SPP 1324, November 2010.

[67] E. Novak and H. Woźniakowski. On the Power of Function Values for the Approximation Problem in Various Settings. Preprint 67, DFG-SPP 1324, November 2010.

[68] A. Hinrichs, E. Novak, and H. Woźniakowski. The Curse of Dimensionality for Monotone and Convex Functions of Many Variables. Preprint 68, DFG-SPP 1324, November 2010.

[69] G. Kutyniok and W.-Q Lim. Image Separation Using Shearlets. Preprint 69, DFG-SPP 1324, November 2010.

[70] B. Jin and P. Maass. An Analysis of Electrical Impedance Tomography with Applications to Tikhonov Regularization. Preprint 70, DFG-SPP 1324, December 2010.

[71] S. Holtz, T. Rohwedder, and R. Schneider. The Alternating Linear Scheme for Tensor Optimisation in the TT Format. Preprint 71, DFG-SPP 1324, December 2010.

[72] T. Müller-Gronbach and K. Ritter. A Local Refinement Strategy for Constructive Quantization of Scalar SDEs. Preprint 72, DFG-SPP 1324, December 2010.

[73] T. Rohwedder and R. Schneider. An Analysis for the DIIS Acceleration Method used in Quantum Chemistry Calculations. Preprint 73, DFG-SPP 1324, December 2010.

[74] C. Bender and J. Steiner. Least-Squares Monte Carlo for Backward SDEs. Preprint 74, DFG-SPP 1324, December 2010.

[75] C. Bender. Primal and Dual Pricing of Multiple Exercise Options in Continuous Time. Preprint 75, DFG-SPP 1324, December 2010.

[76] H. Harbrecht, M. Peters, and R. Schneider. On the Low-rank Approximation by the Pivoted Cholesky Decomposition. Preprint 76, DFG-SPP 1324, December 2010.

[77] P. A. Cioica, S. Dahlke, N. Döhring, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R. L. Schilling. Adaptive Wavelet Methods for Elliptic Stochastic Partial Differential Equations. Preprint 77, DFG-SPP 1324, January 2011.

[78] G. Plonka, S. Tenorth, and A. Iske. Optimal Representation of Piecewise Hölder Smooth Bivariate Functions by the Easy Path Wavelet Transform. Preprint 78, DFG-SPP 1324, January 2011.

[79] A. Mugler and H.-J. Starkloff. On Elliptic Partial Differential Equations with Random Coefficients. Preprint 79, DFG-SPP 1324, January 2011.

[80] T. Müller-Gronbach, K. Ritter, and L. Yaroslavtseva. A Derandomization of the Euler Scheme for Scalar Stochastic Differential Equations. Preprint 80, DFG-SPP 1324, January 2011.

[81] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. Preprint 81, DFG-SPP 1324, January 2011.

[82] K. Grella and C. Schwab. Sparse Tensor Spherical Harmonics Approximation in Radiative Transfer. Preprint 82, DFG-SPP 1324, January 2011.

[83] D.A. Lorenz, S. Schiffler, and D. Trede. Beyond Convergence Rates: Exact Inversion With Tikhonov Regularization With Sparsity Constraints. Preprint 83, DFG-SPP 1324, January 2011.

[84] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. Preprint 84, DFG-SPP 1324, January 2011.

[85] S. Dahlke and W. Sickel. On Besov Regularity of Solutions to Nonlinear Elliptic Partial Differential Equations. Preprint 85, DFG-SPP 1324, January 2011.

[86] S. Dahlke, U. Friedrich, P. Maass, T. Raasch, and R.A. Ressel. An adaptive wavelet method for parameter identification problems in parabolic partial differential equations. Preprint 86, DFG-SPP 1324, January 2011.

[87] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and Variational Stabilization for Convection-Diffusion Equations. Preprint 87, DFG-SPP 1324, January 2011.

[88] T. Jahnke. On Reduced Models for the Chemical Master Equation. Preprint 88, DFG-SPP 1324, January 2011.

[89] P. Binev, W. Dahmen, R. DeVore, P. Lamby, D. Savu, and R. Sharpley. Compressed Sensing and Electron Microscopy. Preprint 89, DFG-SPP 1324, March 2011.

[90] P. Binev, F. Blanco-Silva, D. Blom, W. Dahmen, P. Lamby, R. Sharpley, and T. Vogt. High Quality Image Formation by Nonlocal Means Applied to High-Angle Annular Dark Field Scanning Transmission Electron Microscopy (HAADF-STEM). Preprint 90, DFG-SPP 1324, March 2011.

[91] R. A. Ressel. A Parameter Identification Problem for a Nonlinear Parabolic Differential Equation. Preprint 91, DFG-SPP 1324, May 2011.

[92] G. Kutyniok. Data Separation by Sparse Representations. Preprint 92, DFG-SPP 1324, May 2011.

[93] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to Compressed Sensing. Preprint 93, DFG-SPP 1324, May 2011.

[94] H.-C. Kreusler and H. Yserentant. The Mixed Regularity of Electronic Wave Functions in Fractional Order and Weighted Sobolev Spaces. Preprint 94, DFG-SPP 1324, June 2011.

[95] E. Ullmann, H. C. Elman, and O. G. Ernst. Efficient Iterative Solvers for Stochastic Galerkin Discretizations of Log-Transformed Random Diffusion Problems. Preprint 95, DFG-SPP 1324, June 2011.

[96] S. Kunis and I. Melzer. On the Butterfly Sparse Fourier Transform. Preprint 96, DFG-SPP 1324, June 2011.

[97] T. Rohwedder. The Continuous Coupled Cluster Formulation for the Electronic Schrödinger Equation. Preprint 97, DFG-SPP 1324, June 2011.

[98] T. Rohwedder and R. Schneider. Error Estimates for the Coupled Cluster Method. Preprint 98, DFG-SPP 1324, June 2011.

[99] P. A. Cioica and S. Dahlke. Spatial Besov Regularity for Semilinear Stochastic Partial Differential Equations on Bounded Lipschitz Domains. Preprint 99, DFG-SPP 1324, July 2011.

[100] L. Grasedyck and W. Hackbusch. An Introduction to Hierarchical (H-) Rank and TT-Rank of Tensors with Examples. Preprint 100, DFG-SPP 1324, August 2011.

[101] N. Chegini, S. Dahlke, U. Friedrich, and R. Stevenson. Piecewise Tensor Product Wavelet Bases by Extensions and Approximation Rates. Preprint 101, DFG-SPP 1324, September 2011.

[102] S. Dahlke, P. Oswald, and T. Raasch. A Note on Quarkonial Systems and Multilevel Partition of Unity Methods. Preprint 102, DFG-SPP 1324, September 2011.

[103] A. Uschmajew. Local Convergence of the Alternating Least Squares Algorithm For Canonical Tensor Approximation. Preprint 103, DFG-SPP 1324, September 2011.

[104] S. Kvaal. Multiconfigurational time-dependent Hartree method for describing particle loss due to absorbing boundary conditions. Preprint 104, DFG-SPP 1324, September 2011.

[105] M. Guillemard and A. Iske. On Groupoid C*-Algebras, Persistent Homology and Time-Frequency Analysis. Preprint 105, DFG-SPP 1324, September 2011.

[106] A. Hinrichs, E. Novak, and H. Woźniakowski. Discontinuous information in the worst case and randomized settings. Preprint 106, DFG-SPP 1324, September 2011.

[107] M. Espig, W. Hackbusch, A. Litvinenko, H. Matthies, and E. Zander. Efficient Analysis of High Dimensional Data in Tensor Formats. Preprint 107, DFG-SPP 1324, September 2011.

[108] M. Espig, W. Hackbusch, S. Handschuh, and R. Schneider. Optimization Problems in Contracted Tensor Networks. Preprint 108, DFG-SPP 1324, October 2011.

[109] S. Dereich, T. Müller-Gronbach, and K. Ritter. On the Complexity of Computing Quadrature Formulas for SDEs. Preprint 109, DFG-SPP 1324, October 2011.

[110] D. Belomestny. Solving optimal stopping problems by empirical dual optimization and penalization. Preprint 110, DFG-SPP 1324, November 2011.

[111] D. Belomestny and J. Schoenmakers. Multilevel dual approach for pricing American style derivatives. Preprint 111, DFG-SPP 1324, November 2011.

[112] T. Rohwedder and A. Uschmajew. Local convergence of alternating schemes for optimization of convex problems in the TT format. Preprint 112, DFG-SPP 1324, December 2011.

[113] T. Görner, R. Hielscher, and S. Kunis. Efficient and accurate computation of spherical mean values at scattered center points. Preprint 113, DFG-SPP 1324, December 2011.

[114] Y. Dong, T. Görner, and S. Kunis. An iterative reconstruction scheme for photoacoustic imaging. Preprint 114, DFG-SPP 1324, December 2011.

[115] L. Kämmerer. Reconstructing hyperbolic cross trigonometric polynomials by sampling along generated sets. Preprint 115, DFG-SPP 1324, February 2012.

[116] H. Chen and R. Schneider. Numerical analysis of augmented plane waves methods for full-potential electronic structure calculations. Preprint 116, DFG-SPP 1324, February 2012.

[117] J. Ma, G. Plonka, and M.Y. Hussaini. Compressive Video Sampling with Approximate Message Passing Decoding. Preprint 117, DFG-SPP 1324, February 2012.

[118] D. Heinen and G. Plonka. Wavelet shrinkage on paths for scattered data denoising. Preprint 118, DFG-SPP 1324, February 2012.

[119] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. Preprint 119, DFG-SPP 1324, March 2012.

[120] C. Bender and J. Steiner. A-posteriori estimates for backward SDEs. Preprint 120, DFG-SPP 1324, April 2012.

[121] M. Espig, W. Hackbusch, A. Litvinenkoy, H.G. Matthiesy, and P. Wähnert. Effcient low-rank approximation of the stochastic Galerkin matrix in tensor formats. Preprint 121, DFG-SPP 1324, May 2012.

[122] O. Bokanowski, J. Garcke, M. Griebel, and I. Klompmaker. An adaptive sparse grid semi-Lagrangian scheme for first order Hamilton-Jacobi Bellman equations. Preprint 122, DFG-SPP 1324, June 2012.

[123] A. Mugler and H.-J. Starkloff. On the convergence of the stochastic Galerkin method for random elliptic partial differential equations. Preprint 123, DFG-SPP 1324, June 2012.

[124] P.A. Cioica, S. Dahlke, N. Döhring, U. Friedrich, S. Kinzel, F. Lindner, T. Raasch, K. Ritter, and R.L. Schilling. On the convergence analysis of Rothe's method. Preprint 124, DFG-SPP 1324, July 2012.

[125] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Classification Algorithms using Adaptive Partitioning. Preprint 125, DFG-SPP 1324, July 2012.

[126] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation of hierarchical Tucker and Tensor-Train tensors. Preprint 126, DFG-SPP 1324, July 2012.

[127] M. Kovács, S. Larsson, and K. Urban. On Wavelet-Galerkin methods for semilinear parabolic equations with additive noise. Preprint 127, DFG-SPP 1324, August 2012.

[128] M. Bachmayr, H. Chen, and R. Schneider. Numerical analysis of Gaussian approximations in quantum chemistry. Preprint 128, DFG-SPP 1324, August 2012.

[129] D. Rudolf. Explicit error bounds for Markov chain Monte Carlo. Preprint 129, DFG-SPP 1324, August 2012.

[130] P.A. Cioica, K.-H. Kim, K. Lee, and F. Lindner. On the $L_q(L_p)$-regularity and Besov smoothness of stochastic parabolic equations on bounded Lipschitz domains. Preprint 130, DFG-SPP 1324, December 2012.

[131] M. Hansen. $n-$term Approximation Rates and Besov Regularity for Elliptic PDEs on Polyhedral Domains. Preprint 131, DFG-SPP 1324, December 2012.

[132] R. E. Bank and H. Yserentant. On the H^1 -stability of the L_2 -projection onto finite element spaces. Preprint 132, DFG-SPP 1324, December 2012.

[133] M. Gnewuch, S. Mayer, and K. Ritter. On Weighted Hilbert Spaces and Integration of Functions of Infinitely Many Variables. Preprint 133, DFG-SPP 1324, December 2012.

[134] D. Crisan, J. Diehl, P.K. Friz, and H. Oberhauser. Robust Filtering: Correlated Noise and Multidimensional Observation. Preprint 134, DFG-SPP 1324, January 2013.

[135] Wolfgang Dahmen, Christian Plesken, and Gerrit Welper. Double Greedy Algorithms: Reduced Basis Methods for Transport Dominated Problems. Preprint 135, DFG-SPP 1324, February 2013.

[136] Aicke Hinrichs, Erich Novak, Mario Ullrich, and Henryk Wozniakowski. The Curse of Dimensionality for Numerical Integration of Smooth Functions. Preprint 136, DFG-SPP 1324, February 2013.

[137] Markus Bachmayr, Wolfgang Dahmen, Ronald DeVore, and Lars Grasedyck. Approximation of High-Dimensional Rank One Tensors. Preprint 137, DFG-SPP 1324, March 2013.

[138] Markus Bachmayr and Wolfgang Dahmen. Adaptive Near-Optimal Rank Tensor Approximation for High-Dimensional Operator Equations. Preprint 138, DFG-SPP 1324, April 2013.

[139] Felix Lindner. Singular Behavior of the Solution to the Stochastic Heat Equation on a Polygonal Domain. Preprint 139, DFG-SPP 1324, May 2013.

[140] Stephan Dahlke, Dominik Lellek, Shiu Hong Lui, and Rob Stevenson. Adaptive Wavelet Schwarz Methods for the Navier-Stokes Equation. Preprint 140, DFG-SPP 1324, May 2013.

[141] Jonas Ballani and Lars Grasedyck. Tree Adaptive Approximation in the Hierarchical Tensor Format. Preprint 141, DFG-SPP 1324, June 2013.

[142] Harry Yserentant. A short theory of the Rayleigh-Ritz method. Preprint 142, DFG-SPP 1324, July 2013.

[143] M. Hefter and K. Ritter. On Embeddings of Weighted Tensor Product Hilbert Spaces. Preprint 143, DFG-SPP 1324, August 2013.

[144] M. Altmayer and A. Neuenkirch. Multilevel Monte Carlo Quadrature of Discontinuous Payoffs in the Generalized Heston Model using Malliavin Integration by Parts. Preprint 144, DFG-SPP 1324, August 2013.

[145] L. Kämmerer, D. Potts, and T. Volkmer. Approximation of multivariate functions by trigonometric polynomials based on rank-1 lattice sampling. Preprint 145, DFG-SPP 1324, September 2013.

[146] C. Bender, N. Schweizer, and J. Zhuo. A primal-dual algorithm for BSDEs. Preprint 146, DFG-SPP 1324, October 2013.

[147] D. Rudolf. Hit-and-run for numerical integration. Preprint 147, DFG-SPP 1324, October 2013.

[148] D. Rudolf and M. Ullrich. Positivity of hit-and-run and related algorithms. Preprint 148, DFG-SPP 1324, October 2013.

[149] L. Grasedyck, M. Kluge, and S. Krämer. Alternating Directions Fitting (ADF) of Hierarchical Low Rank Tensors. Preprint 149, DFG-SPP 1324, October 2013.

[150] F. Filbir, S. Kunis, and R. Seyfried. Effective discretization of direct reconstruction schemes for photoacoustic imaging in spherical geometries. Preprint 150, DFG-SPP 1324, November 2013.

[151] E. Novak, M. Ullrich, and H. Woźniakowski. Complexity of Oscillatory Integration for Univariate Sobolev Spaces. Preprint 151, DFG-SPP 1324, November 2013.

[152] A. Hinrichs, E. Novak, and M. Ullrich. A Tractability Result for the Clenshaw Curtis Smolyak Algorithm. Preprint 152, DFG-SPP 1324, November 2013.

[153] M. Hein, S. Setzer, L. Jost, and S. Rangapuram. The Total Variation on Hypergraphs - Learning on Hypergraphs Revisited. Preprint 153, DFG-SPP 1324, November 2013.