An Inverse Power Method for Nonlinear Eigenproblems with Applications in 1-Spectral Clustering and Sparse PCA

Matthias Hein **Thomas Bühler**
Saarland University, Saarbrücken, Germany
{hein,tb}@cs.uni-saarland.de

Abstract

Many problems in machine learning and statistics can be formulated as (generalized) eigenproblems. In terms of the associated optimization problem, computing linear eigenvectors amounts to finding critical points of a quadratic function subject to quadratic constraints. In this paper we show that a certain class of constrained optimization problems with nonquadratic objective and constraints can be understood as *nonlinear eigenproblems*. We derive a generalization of the inverse power method which is guaranteed to converge to a nonlinear eigenvector. We apply the inverse power method to 1-spectral clustering and sparse PCA which can naturally be formulated as nonlinear eigenproblems. In both applications we achieve state-of-the-art results in terms of solution quality and runtime. Moving beyond the standard eigenproblem should be useful also in many other applications and our inverse power method can be easily adapted to new problems.

1 Introduction

Eigenvalue problems associated to a symmetric and positive semi-definite matrix are quite abundant in machine learning and statistics. However, considering the eigenproblem from a variational point of view using Courant-Fischer-theory, the objective is a ratio of quadratic functions, which is quite restrictive from a modeling perspective. We show in this paper that using a ratio of p-homogeneous functions leads quite naturally to a *nonlinear* eigenvalue problem, associated to a certain nonlinear operator. Clearly, such a generalization is only interesting if certain properties of the standard problem are preserved and efficient algorithms for the computation of nonlinear eigenvectors are available. In this paper we present an efficient generalization of the inverse power method (IPM) to nonlinear eigenvalue problems and study the relation to the standard problem. While our IPM is a general purpose method, we show for two unsupervised learning problems that it can be easily adapted to a particular application.

The first application is spectral clustering [21]. In prior work [5] we proposed p-spectral clustering based on the graph p-Laplacian, a nonlinear operator on graphs which reduces to the standard graph Laplacian for $p = 2$. For p close to one, we obtained much better cuts than standard spectral clustering, at the cost of higher runtime. Using the new IPM, we efficiently compute eigenvectors of the 1-Laplacian for 1-spectral clustering. Similar to the recent work of [20], we improve considerably compared to [5] both in terms of runtime and the achieved Cheeger cuts. However, opposed to the suggested method in [20] our IPM is guaranteed to converge to an eigenvector of the 1-Laplacian.

The second application is sparse Principal Component Analysis (PCA). The motivation for sparse PCA is that the largest PCA component is difficult to interpret as usually all components are nonzero. In order to allow a direct interpretation it is therefore desirable to have only a few features with nonzero components but which still explain most of the variance. This kind of trade-off has been

widely studied in recent years, see [15] and references therein. We show that also sparse PCA has a natural formulation as a nonlinear eigenvalue problem and can be efficiently solved with the IPM.

All proofs had to be omitted due to space restrictions and can be found in the supplementary material.

2 Nonlinear Eigenproblems

The standard eigenproblem for a symmetric matric $A \in \mathbb{R}^{n\times n}$ is of the form

$$Af - \lambda f = 0, \tag{1}$$

where $f \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$. It is a well-known result from linear algebra that for symmetric matrices A, the eigenvectors of A can be characterized as critical points of the functional

$$F_{\text{Standard}}(f) = \frac{\langle f, Af\rangle}{\|f\|_2^2}. \tag{2}$$

The eigenvectors of A can be computed using the Courant-Fischer Min-Max principle. While the ratio of quadratic functions is useful in several applications, it is a severe modeling restriction. This restriction however can be overcome using nonlinear eigenproblems. In this paper we consider functionals F of the form

$$F(f) = \frac{R(f)}{S(f)}, \tag{3}$$

where with $\mathbb{R}_+ = \{x \in \mathbb{R} \,|\, x \geq 0\}$ we assume $R : \mathbb{R}^n \to \mathbb{R}_+$, $S : \mathbb{R}^n \to \mathbb{R}_+$ to be convex, Lipschitz continuous, even and positively p-homogeneous[1] with $p \geq 1$. Moreover, we assume that $S(f) = 0$ if and only if $f = 0$. The condition that R and S are p-homogeneous and even will imply for any eigenvector v that also αv for $\alpha \in \mathbb{R}$ is an eigenvector. It is easy to see that the functional of the standard eigenvalue problem in Equation (2) is a special case of the general functional in (3).

To gain some intuition, let us first consider the case where R and S are differentiable. Then it holds for every critical point f^* of F,

$$\nabla F(f^*) = 0 \qquad \Longleftrightarrow \qquad \nabla R(f^*) - \frac{R(f^*)}{S(f^*)} \cdot \nabla S(f^*) = 0\,.$$

Let $r, s : \mathbb{R}^n \to \mathbb{R}^n$ be the operators defined as $r(f) = \nabla R(f)$, $s(f) = \nabla S(f)$ and $\lambda^* = \frac{R(f^*)}{S(f^*)}$, we see that every critical point f^* of F satisfies the *nonlinear eigenproblem*

$$r(f^*) - \lambda^* s(f^*) = 0, \tag{4}$$

which is in general a system of nonlinear equations, as r and s are nonlinear operators. If R and S are both quadratic, r and s are linear operators and one gets back the standard eigenproblem (1).

Before we proceed to the general nondifferentiable case, we have to introduce some important concepts from nonsmooth analysis. Note that F is in general nonconvex and nondifferentiable. In the following we denote by $\partial F(f)$ the *generalized gradient* of F at f according to Clarke [9],

$$\partial F(f) = \{\xi \in \mathbb{R}^n \;\big|\; F^0(f, v) \;\geq\; \langle \xi, v\rangle\,, \quad \text{for all } v \in \mathbb{R}^n\},$$

where $F^0(f, v) = \lim_{g\to f,\, t\to 0} \sup \frac{F(g+tv)-F(g)}{t}$. In the case where F is convex, ∂F is the subdifferential of F and $F^0(f, v)$ the directional derivative for each $v \in \mathbb{R}^n$. A characterization of critical points of nonsmooth functionals is as follows.

Definition 2.1 ([7]) *A point $f \in \mathbb{R}^n$ is a critical point of F, if $0 \in \partial F$.*

This generalizes the well-known fact that the gradient of a differentiable function vanishes at a critical point. We now show that the nonlinear eigenproblem (4) is a necessary condition for a critical point and in some cases even sufficient. A useful tool is the generalized Euler identity.

Theorem 2.1 ([22]) *Let $R : \mathbb{R}^n \to \mathbb{R}$ be a positively p-homogeneous and convex continuous function. Then, for each $x \in \mathbb{R}^n$ and $r^* \in \partial R(x)$ it holds that $\langle x, r^*\rangle = p\, R(x)$.*

[1] A function $G : \mathbb{R}^n \to \mathbb{R}$ is positively homogeneous of degree p if $G(\gamma x) = \gamma^p G(x)$ for all $\gamma \geq 0$.

The next theorem characterizes the relation between nonlinear eigenvectors and critical points of F.

Theorem 2.2 *Suppose that R, S fulfill the stated conditions. Then a necessary condition for f^* being a critical point of F is*

$$0 \in \partial R(f^*) - \lambda^* \, \partial S(f^*), \qquad \textit{where} \qquad \lambda^* = \frac{R(f^*)}{S(f^*)}. \tag{5}$$

If S is continuously differentiable at f^*, then this is also sufficient.

Proof: Let f^* fulfill the general nonlinear eigenproblem in (5), where $r^* \in \partial R(f^*), s^* \in \partial S(f^*)$, such that $r^* - \lambda^* \, s^* = 0$. Then by Theorem 2.1,

$$0 = \langle f^*, r^* \rangle - \lambda^* \, \langle f^*, s^* \rangle = p \, R(f^*) - p \, \lambda^* \, S(f^*),$$

and thus $\lambda^* = R(f^*)/S(f^*)$. As R, S are Lipschitz continuous, we have, see Prop. 2.3.14 in [9],

$$\partial\Big(\frac{R}{S}\Big)(f) \;\subseteq\; \frac{S(f)\,\partial R(f) - R(f)\,\partial S(f)}{S(f)^2}. \tag{6}$$

Thus if f^* is a critical point, that is $0 \in \partial F(f^*)$, then $0 \in \partial R(f^*) - \frac{R(f^*)}{S(f^*)} \, \partial S(f^*)$ given that $f^* \neq 0$. Moreover, by Prop. 2.3.14 in [9] we have equality in (6), if S is continuously differentiable at f^* and thus (5) implies that f^* is a critical point of F. □

Finally, the definition of the associated nonlinear operators in the nonsmooth case is a bit tricky as r and s can be set-valued. However, as we assume R and S to be Lipschitz, the set where R and S are nondifferentiable has measure zero and thus r and s are single-valued almost everywhere.

3 The inverse power method for nonlinear Eigenproblems

A standard technique to obtain the smallest eigenvalue of a positive semi-definite symmetric matrix A is the inverse power method [12]. Its main building block is the fact that the iterative scheme

$$A f^{k+1} = f^k \tag{7}$$

converges to the smallest eigenvector of A. Transforming (7) into the optimization problem

$$f^{k+1} = \arg\min_u \frac{1}{2} \, \langle u, A \, u \rangle - \big\langle u, f^k \big\rangle \tag{8}$$

is the motivation for the general IPM. The direct generalization tries to solve

$$0 \in r(f^{k+1}) - s(f^k) \quad \text{ or equivalently } \quad f^{k+1} = \arg\min_u R(u) - \big\langle u, s(f^k) \big\rangle \,, \tag{9}$$

where $r(f) \in \partial R(f)$ and $s(f) \in \partial S(f)$. For $p > 1$ this leads directly to Algorithm 2, however for $p = 1$ the direct generalization fails. In particular, the ball constraint has to be introduced in Algorithm 1 as the objective in the optimization problem (9) is otherwise unbounded from below. (Note that the 2-norm is only chosen for algorithmic convenience). Moreover, the introduction of λ_k in Algorithm 1 is necessary to guarantee descent whereas in Algorithm 2 it would just yield a rescaled solution of the problem in the inner loop (called inner problem in the following).

For both methods we show convergence to a solution of (4), which by Theorem 2.2 is a necessary condition for a critical point of F and often also sufficient. Interestingly, both applications are naturally formulated as 1-homogeneous problems so that we use in both cases Algorithm 1. Nevertheless, we state the second algorithm for completeness. Note that we cannot guarantee convergence to the smallest eigenvector even though our experiments suggest that we often do so. However, as the method is fast one can afford to run it multiple times with different initializations and use the eigenvector with smallest eigenvalue.

The inner optimization problem is convex for both algorithms. In turns out that both for 1-spectral clustering and sparse PCA the inner problem can be solved very efficiently, for sparse PCA it has even a closed form solution. While we do not yet have results about convergence speed, empirical observation shows that one usually converges quite quickly to an eigenvector.

Algorithm 1 Computing a nonlinear eigenvector for convex positively p-homogeneous functions R and S with $p = 1$

Initialization: $f^0 = \text{random with } \left\|f^0\right\| = 1$, $\lambda^0 = F(f^0)$
repeat
  $f^{k+1} = \arg\min_{\|u\|_2 \leq 1} \left\{R(u) - \lambda^k \left\langle u, s(f^k)\right\rangle\right\}$ where $s(f^k) \in \partial S(f^k)$
  $\lambda^{k+1} = R(f^{k+1})/S(f^{k+1})$
until $\frac{\left|\lambda^{k+1} - \lambda^k\right|}{\lambda^k} < \epsilon$
Output: eigenvalue λ^{k+1} and eigenvector f^{k+1}.

Algorithm 2 Computing a nonlinear eigenvector for convex positively p-homogeneous functions R and S with $p > 1$

Initialization: $f^0 = \text{random}$, $\lambda^0 = F(f^0)$
repeat
  $g^{k+1} = \arg\min_{u} \left\{R(u) - \left\langle u, s(f^k)\right\rangle\right\}$ where $s(f^k) \in \partial S(f^k)$
  $f^{k+1} = g^{k+1}/S(g^{k+1})^{1/p}$
  $\lambda^{k+1} = R(f^{k+1})/S(f^{k+1})$
until $\frac{\left|\lambda^{k+1} - \lambda^k\right|}{\lambda^k} < \epsilon$
Output: eigenvalue λ^{k+1} and eigenvector f^{k+1}.

To our best knowledge both suggested methods have not been considered before. In [4] they propose an inverse power method specially tailored towards the continuous p-Laplacian for $p > 1$, which can be seen as a special case of Algorithm 2. In [15] a generalized power method has been proposed which will be discussed in Section 5. Finally, both methods can be easily adapted to compute the largest nonlinear eigenvalue, which however we have to omit due to space constraints.

Lemma 3.1 *The sequences f^k produced by Alg. 1 and 2 satisfy $F(f^k) > F(f^{k+1})$ for all $k \geq 0$ or the sequences terminate.*

Theorem 3.1 *The sequences f^k produced by Algorithms 1 and 2 converge to an eigenvector f^* with eigenvalue $\lambda^* \in \left[0, F(f^0)\right]$ in the sense that it solves the nonlinear eigenproblem* (5)*. If S is continuously differentiable at f^*, then F has a critical point at f^*.*

Throughout the proofs, we use the notation $\Phi_{f^k}(u) = R(u) - \lambda^k \left\langle u, s(f^k)\right\rangle$ and $\Psi_{f^k}(u) = R(u) - \left\langle u, s(f^k)\right\rangle$ for the objectives of the inner problems in Algorithms 1 & 2, respectively.

Proof of Lemma 3.1 for Algorithm 1: First note that the optimal value of the inner problem is non-positive as $\Phi_{f^k}(0) = 0$. Moreover, as Φ_{f^k} is 1-homogeneous, the minimum of Φ_{f^k} is always attained at the boundary of the constraint set. Thus any f^k fulfills $\left\|f^k\right\|_2^2 = 1$ and thus is feasible, and

$$\min_{\|f\|_2^2 \leq 1} \Phi_{f^k}(f) \leq \Phi_{f^k}(f^k) = R(f^k) - \lambda^k \left\langle f^k, s(f^k)\right\rangle = R(f^k) - F(f^k) \cdot S(f^k) = 0\,,$$

where we used $\left\langle f^k, s(f^k)\right\rangle = S(f^k)$ from Theorem 2.1. If the optimal value is zero, then f^k is a possible minimizer and the sequence terminates and f^k is an eigenvector see proof of Theorem 3.1 for Algorithm 1. Otherwise the optimal value is negative and at the optimal point f^{k+1} we get $R(f^{k+1}) < \lambda^k \left\langle f^{k+1}, s(f^k)\right\rangle$. The definition of the subdifferential $s(f^k)$ together with the 1-homogeneity of S yields

$$S(f^{k+1}) \geq S(f^k) + \left\langle f^{k+1} - f^k, s(f^k)\right\rangle = \left\langle f^{k+1}, s(f^k)\right\rangle\,,$$

and finally $F(f^{k+1}) = \frac{R(f^{k+1})}{S(f^{k+1})} < \lambda^k = F(f^k)$. □

Proof of Theorem 3.1 for Algorithm 1: By Lemma 3.1 the sequence $F(f^k)$ is monotonically decreasing. By assumption S and R are nonnegative and hence F is bounded below by zero. Thus

we have convergence towards a limit

$$\lambda^* = \lim_{k\to\infty} F(f^k) \, .$$

Note that $\left\|f^k\right\|_2^2 \leq 1$ for every k, thus the sequence f^k is contained in a compact set, which implies that there exists a subsequence f^{k_j} converging to some element f^*. As the sequence $F(f^{k_j})$ is a subsequence of a convergent sequence, it has to converge towards the same limit, hence also

$$\lim_{j\to\infty} F(f^{k_j}) = \lambda^* \, .$$

As shown before, the objective of the inner optimization problem is nonpositive at the optimal point. Assume now that $\min_{\|f\|_2^2\leq 1} \Phi_{f^*}(f) < 0$. Then the vector $f^{**} = \operatorname*{arg\,min}_{\|f\|_2^2\leq 1} \Phi_{f^*}(f)$ satisfies

$$R(f^{**}) < \lambda^* \left\langle f^{**}, s(f^*)\right\rangle = \lambda^* \left(S(f^*) + \left\langle f^{**} - f^*, s(f^*)\right\rangle\right) \leq \lambda^* S(f^{**}) \, ,$$

where we used the definition of the subdifferential and the 1-homogeneity of S. Hence

$$F(f^{**}) < \lambda^* = F(f^*) \, ,$$

which is a contradiction to the fact that the sequence $F(f^k)$ has converged to λ^*. Thus we must have $\min_{\|f\|_2^2\leq 1} \Phi_{f^*}(f) = 0$, i.e. the function $\Phi_{f^*}(f)$ is nonnegative in the unit ball. Using the fact that for any $\alpha \geq 0$,

$$\Phi_{f^*}(\alpha f) = \alpha \Phi_{f^*} \, ,$$

we can even conclude that the function $\Phi_{f^*}(f)$ is nonnegative everywhere, and thus $\min_f \Phi_{f^*}(f) = 0$. Note that $\Phi_{f^*}(f^*) = 0$, which implies that f^* is a global minimizer of Φ_{f^*}, and hence

$$0 \in \partial \Phi_{f^*}(f^*) = \partial R(f^*) - \lambda^* \partial S(f^*) \, ,$$

which implies that f^* is an eigenvector with eigenvalue λ^*. Note that this argument was independent of the choice of the subsequence, thus every convergent subsequence converges to an eigenvector with the same eigenvalue λ^*. Clearly we have $\lambda^* \leq F(f^0)$. □

The following lemma is useful in the convergence proof of Algorithm 2.

Lemma 3.2 *Let R be a convex, positively p-homogeneous function with $p \geq 1$. Then for any $x \in \mathbb{R}^n$, $t \geq 0$ and any $r^* \in \partial R(x)$ we have $t^{p-1} r^* \in \partial R(tx)$.*

Proof: Using the definition of the subgradient, we have for any $y \in \mathbb{R}^n$ and any $t \geq 0$,

$$t^p R(y) \geq t^p R(x) + t^p \left\langle r^*, y - x\right\rangle \, .$$

Using the p-homogeneity of R, we can rewrite this as

$$R(ty) \geq R(tx) + \left\langle t^{p-1} r^*, ty - tx\right\rangle \, ,$$

which implies $t^{p-1} r^* \in \partial R(tx)$. □

The following Proposition generalizes a result by Zarantonello [23].

Proposition 3.1 *Let $R : \mathbb{R}^n \to \mathbb{R}$ be a convex, continuous and positively p-homogeneous and even functional and $\partial R(f)$ its subdifferential at f. Then it holds for any $f, g \in R^n$ and $r(f) \in \partial R(f)$,*

$$|\langle r(f), g\rangle| \leq \langle r(f), f\rangle^{1-\frac{1}{p}} \langle r(g), g\rangle^{\frac{1}{p}} = p \cdot R(f)^{1-\frac{1}{p}} R(g)^{\frac{1}{p}} \, .$$

Proof: First observe that for any k points $x_0, \ldots x_{k-1} \in \mathbb{R}^n$, the subdifferential inequality yields

$$\begin{aligned} R(x_l) &\geq R(x_{l-1}) + \langle r(x_{l-1}), x_l - x_{l-1}\rangle \, , \; \forall 1 \leq l \leq k-1 \\ R(x_0) &\geq R(x_{k-1}) + \langle r(x_{k-1}), x_0 - x_{k-1}\rangle \, , \end{aligned}$$

and hence, by summing up,

$$\langle r(x_0), x_1 - x_0\rangle + \cdots + \langle r(x_{k-2}), x_{k-1} - x_{k-2}\rangle + \langle r(x_{k-1}), x_0 - x_{k-1}\rangle \leq 0 \, . \qquad (10)$$

Let now $f, g \in \mathbb{R}^n$, and $r(f) \in \partial R(f), r(g) \in \partial R(g)$. We construct a set of $2m$ points $x_0, \dots x_{2m-1}$ in $\mathbb{R}^n$, where $m \in \mathbb{N}$, as follows:

$$x_i = \begin{cases} \frac{i+1}{m} f & , \, 0 \leq i \leq m-1 \\ \frac{2m-1-i}{m} g & , \, m \leq i \leq 2m-1 \end{cases} .$$

By Lemma 3.2 for all $i \in \{0, \dots 2m-1\}$ there exists an $r^*(x_i) \in \partial R(x_i)$ s.t.

$$r^*(x_i) = \begin{cases} \left(\frac{i+1}{m}\right)^{p-1} r(f) & , \, 0 \leq i \leq m-1 \\ \left(\frac{2m-1-i}{m}\right)^{p-1} r(g) & , \, m \leq i \leq 2m-1 \end{cases} .$$

Eq. (10) now yields

$$\sum_{i=0}^{m-2} \left\langle \left(\frac{i+1}{m}\right)^{p-1} r(f), \frac{1}{m} f \right\rangle + \left\langle r(f), \frac{m-1}{m} g - f \right\rangle$$
$$- \sum_{i=m}^{2m-2} \left\langle \left(\frac{2m-1-i}{m}\right)^{p-1} r(g), \frac{1}{m} g \right\rangle + \left\langle 0 \cdot r(g), \frac{1}{m} f - 0 \cdot g \right\rangle \leq 0$$

which simplifies to

$$\left(\frac{1}{m} \sum_{j=1}^{m-1} \left(\frac{j}{m}\right)^{p-1} - 1\right) \langle r(f), f \rangle - \left(\frac{1}{m} \sum_{j=1}^{m-1} \left(\frac{j}{m}\right)^{p-1}\right) \langle r(g), g \rangle + \frac{m-1}{m} \langle r(f), g \rangle \leq 0 . \tag{11}$$

By letting $m \to \infty$ we obtain for the two sums

$$\lim_{m \to \infty} \left(\frac{1}{m} \sum_{j=1}^{m-1} \left(\frac{j}{m}\right)^{p-1}\right) = \lim_{m \to \infty} \left(\frac{1}{m} \sum_{j=\frac{1}{m}, \frac{2}{m}, \dots}^{\frac{m-1}{m}} j^{p-1}\right) = \int_0^1 j^{p-1} dj = \frac{1}{p} .$$

Hence in total in the limit $m \to \infty$ Eq. (11) becomes

$$\langle r(f), g \rangle - \left(1 - \frac{1}{p}\right) \langle r(f), f \rangle - \frac{1}{p} \langle r(g), g \rangle \leq 0 .$$

As the above inequality holds for all $f, g \in \mathbb{R}^n$, clearly we can now perform the substitution $f \to t^{-1} f, g \to t^{p-1} g, r(f) \to t^{-(p-1)} r(f), r(g) \to t^{(p-1)^2} r(g)$, where $t \in \mathbb{R}^+$, which gives

$$\langle r(f), g \rangle - \left(1 - \frac{1}{p}\right) t^{-p} \langle r(f), f \rangle - \frac{1}{p} t^{p(p-1)} \langle r(g), g \rangle \leq 0 . \tag{12}$$

A local optimum with respect to t of the left side satisfies the necessary condition

$$\begin{aligned} 0 = & \, (p-1) \, t^{-p-1} \langle r(f), f \rangle - (p-1) t^{p^2-p-1} \langle r(g), g \rangle \\ & = t^{-p-1} (p-1) \left(\langle r(f), f \rangle - t^{p^2} \langle r(g), g \rangle \right) , \end{aligned}$$

which implies that

$$t^p = \left(\frac{\langle r(f), f \rangle}{\langle r(g), g \rangle} \right)^{\frac{1}{p}} .$$

Plugging this into (12) yields

$$\begin{aligned} 0 \geq & \langle r(f), g \rangle - \left(1 - \frac{1}{p}\right) \langle r(g), g \rangle^{\frac{1}{p}} \langle r(f), f \rangle^{1 - \frac{1}{p}} - \frac{1}{p} \langle r(f), f \rangle^{1 - \frac{1}{p}} \langle r(g), g \rangle^{\frac{1}{p}} \\ & = \langle r(f), g \rangle - \langle r(f), f \rangle^{1 - \frac{1}{p}} \langle r(g), g \rangle^{\frac{1}{p}} . \end{aligned}$$

By the homogeneity of R we then have

$$\langle r(f), g \rangle \leq \langle r(f), f \rangle^{1 - \frac{1}{p}} \langle r(g), g \rangle^{\frac{1}{p}} = p \cdot R(f)^{1 - \frac{1}{p}} R(g)^{\frac{1}{p}} .$$

Finally, note that we can replace the left side by its absolute value since replacing g with $-g$ yields

$$\langle r(f), -g\rangle \leq p \cdot R(f)^{1-\frac{1}{p}} R(-g)^{\frac{1}{p}} = p \cdot R(f)^{1-\frac{1}{p}} R(g)^{\frac{1}{p}} ,$$

where we used the fact that R is even. □

Proof of Lemma 3.1 for Algorithm 2: Note that as $R(u) \geq 0$, the minimum of the objective of the inner problem is attained for some u with $\langle u, s(f^k)\rangle > 0$. Choose u such that $\langle u, s(f^k)\rangle > 0$. Then we minimize Ψ_{f^k} on the ray tu, $t \geq 0$. We have

$$\Psi_{f^k}(tu) = R(tu) - \langle t\, u, s(f^k)\rangle = t^p\, R(u) - t\, \langle u, s(f^k)\rangle$$

and hence

$$\frac{\partial}{\partial t}\Psi_{f^k}(tu) = p\, t^{p-1} R(u) - \langle u, s(f^k)\rangle$$

and thus the minimum is attained at $t^*(u) = \Big(\frac{\langle u, s(f^k)\rangle}{p\, R(u)}\Big)^{\frac{1}{p-1}} > 0$ and

$$\Psi_{f^k}(t^*(u)u) = t^*(u)^p R(u) - t^*(u)\, \langle u, s(f^k)\rangle = (1-p)\Big(\frac{\langle u, s(f^k)\rangle^p}{p^p\, R(u)}\Big)^{\frac{1}{p-1}}.$$

Assume there exists u that satisfies $\Psi_{f^k}(u) < \Psi_{f^k}(\hat{f})$ where $\hat{f} = F(f^k)^{\frac{1}{1-p}}\, f^k$. Hence, also $\Psi_{f^k}(t^*(u)u) < \Psi_{f^k}(\hat{f})$, which implies

$$\begin{aligned}(1-p)\Big(\frac{\langle u, s(f^k)\rangle^p}{p^p\, R(u)}\Big)^{\frac{1}{p-1}} &< F(f^k)^{\frac{p}{1-p}} R(f^k) - F(f^k)^{\frac{1}{1-p}}\, \langle f^k, s(f^k)\rangle \\ &= F(f^k)^{\frac{1}{1-p}}(1-p) ,\end{aligned}$$

where we used the fact that $\langle f^k, s(f^k)\rangle = pS(f^k)$ and $S(f^k) = 1$. Rearranging, we obtain

$$F(f^k) > \frac{p^p R(u)}{\langle u, s(f^k)\rangle^p} .$$

Using the Hölder-type inequality of Proposition 3.1 and $S(f^k) = 1$, we obtain

$$\langle u, s(f^k)\rangle \leq pS(f^k)^{1-\frac{1}{p}} S(u)^{\frac{1}{p}} = pS(u)^{\frac{1}{p}},$$

which gives $F(f^k) > F(u)$. Let now f^* be the minimizer of Ψ_{f^k}. Then f^* satisfies $\Psi_{f^k}(f^*) \leq \Psi_{f^k}(\hat{f})$. If equality holds then $\hat{f} = F(f^k)^{\frac{1}{1-p}}\, f^k$ is a minimizer of the inner problem and the sequence terminates. In this case f^k is an eigenvector, see proof of Theorem 3.1 for Algorithm 2. Otherwise $\Psi_{f^k}(f^*) < \Psi_{f^k}(\hat{f})$ and thus $u = f^*$ fulfills the above assumption and we get $F(f^k) > F(f^*)$, as claimed. □

Proof of Theorem 3.1 for Algorithm 2: Note that as $F(f) \geq 0$, the sequence $F(f^k)$ is bounded from below, and by Lemma 3.1 it is monotonically decreasing and thus converges to some $\lambda^* \in [0, F(f^0)]$. Moreover, $S(f^k) = 1$ for all k. As S is continuous it attains its minimum m on the unit sphere in $\mathbb{R}^n$. By assumption $m > 0$. We obtain

$$1 = S(f^k) = S\Big(\frac{f^k}{\|f^k\|_2}\, \|f^k\|_2\Big) \geq m\, \|f^k\|_2^p, \quad \Longrightarrow \quad \|f^k\|_2 \leq \Big(\frac{1}{m}\Big)^{\frac{1}{p}}.$$

Thus the sequence f^k is bounded and there exists a convergent subsequence f^{k_j}. Clearly, $\lim_{j\to\infty} F(f^{k_j}) = \lim_{k\to\infty} F(f^k) = \lambda^*$. Let now $f^* = \lim_{j\to\infty} f^{k_j}$, and suppose that there exists $u \in \mathbb{R}^n$ with $\Psi_{f^*}(u) < \Psi_{f^*}(\hat{f})$ where $\hat{f} = F(f^*)^{\frac{1}{1-p}} f^*$. Then, analogously to the proof of Lemma 3.1, one can conclude that $F(u) < F(f^*) = \lambda^*$ which contradicts the fact that $F(f^{k_j})$ has as its limit λ^*. Thus $\hat{f}$ is a minimizer of Ψ_{f^*}, which implies

$$\begin{aligned}0 \in \partial\Psi_{f^*}\big(F(f^*)^{\frac{1}{1-p}} f^*\big) &= \partial R\big(F(f^*)^{\frac{1}{1-p}} f^*\big) - s(f^*) = \Big(F(f^*)^{\frac{1}{1-p}}\Big)^{p-1} \partial R(f^*) - s(f^*) \\ &= \frac{1}{F(f^*)}\Big(\partial R(f^*) - F(f^*)s(f^*)\Big),\end{aligned}$$

so that f^* is an eigenvector with eigenvalue λ^*. As this argument was independent of the subsequence, any convergent subsequence of f^k converges towards an eigenvector with eigenvalue λ^*. □

Practical implementation: By the proof of Lemma 3.1, descent in F is not only guaranteed for the optimal solution of the inner problem, but for any vector u which has inner objective value $\Phi_{f^k}(u) < 0 = \Phi_{f^k}(f^k)$ for Alg. 1 and $\Psi_{f^k}(u) < \Psi_{f^k}(F(f^k)^{\frac{1}{1-p}} f^k)$ in the case of Alg. 2. This has two important practical implications. First, for the convergence of the IPM, it is sufficient to use a vector u satisfying the above conditions instead of the optimal solution of the inner problem. In particular, in an early stage where one is far away from the limit, it makes no sense to invest much effort to solve the inner problem accurately. Second, if the inner problem is solved by a descent method, a good initialization for the inner problem at step $k+1$ is given by f^k in the case of Alg. 1 and $F(f^k)^{\frac{1}{1-p}} f^k$ in the case of Alg. 2 as descent in F is guaranteed after one step.

4 Application 1: 1-spectral clustering and Cheeger cuts

Spectral clustering is a graph-based clustering method (see [21] for an overview) based on a relaxation of the NP-hard problem of finding the optimal balanced cut of an undirected graph. The spectral relaxation has as its solution the second eigenvector of the graph Laplacian and the final partition is found by optimal thresholding. While usually spectral clustering is understood as relaxation of the so called ratio/normalized cut, it can be equally seen as relaxation of the ratio/normalized Cheeger cut, see [5]. Given a weighted undirected graph with vertex set V and weight matrix W, the ratio Cheeger cut (RCC) of a partition $(C, \overline{C})$, where $C \subset V$ and $\overline{C} = V \backslash C$, is defined as

$$\mathrm{RCC}(C, \overline{C}) := \frac{\mathrm{cut}(C, \overline{C})}{\min\{|C|, |\overline{C}|\}}, \qquad \text{where} \qquad \mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij},$$

where we assume in the following that the graph is connected. Due to limited space the normalized version is omitted, but the proposed IPM can be adapted to this case. In [5] we proposed p-spectral clustering, a generalization of spectral clustering based on the second eigenvector of the nonlinear graph p-Laplacian (the graph Laplacian is recovered for $p = 2$). The main motivation was the relation between the optimal Cheeger cut $h_{\mathrm{RCC}} = \min_{C \subset V} \mathrm{RCC}(C, \overline{C})$ and the Cheeger cut h^*_{RCC} obtained by optimal thresholding the second eigenvector of the p-Laplacian, see [5, 8],

$$\forall\, p > 1, \qquad \frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \;\leq\; \frac{h^*_{\mathrm{RCC}}}{\max_{i \in V} d_i} \;\leq\; p \left(\frac{h_{\mathrm{RCC}}}{\max_{i \in V} d_i} \right)^{\frac{1}{p}},$$

where $d_i = \sum_{i \in V} w_{ij}$ denotes the degree of vertex i. While the inequality is quite loose for spectral clustering ($p = 2$), it becomes tight for $p \to 1$. Indeed in [5] much better cuts than standard spectral clustering were obtained, at the expense of higher runtime. In [20] the idea was taken up and they considered directly the variational characterization of the ratio Cheeger cut, see also [8],

$$h_{\mathrm{RCC}} = \min_{f \text{ nonconstant}} \frac{\frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|}{\|f - \mathrm{median}(f) \mathbf{1}\|_1} = \min_{\substack{f \text{ nonconstant} \\ \mathrm{median(f)}=0}} \frac{\frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|}{\|f\|_1}. \quad (13)$$

In [20] they proposed a minimization scheme based on the Split Bregman method [11]. Their method produces comparable cuts to the ones in [5], while being computationally much more efficient. However, they could not provide any convergence guarantee about their method.

In this paper we consider the functional associated to the 1-Laplacian Δ_1,

$$F_1(f) = \frac{\frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|}{\|f\|_1} = \frac{\langle f, \Delta_1 f \rangle}{\|f\|_1}, \quad (14)$$

where

$$(\Delta_1 f)_i = \left\{ \sum_{j=1}^n w_{ij} u_{ij} \,|\, u_{ij} = -u_{ji}, u_{ij} \in \mathrm{sign}(f_i - f_j) \right\} \text{ and } \mathrm{sign}(x) = \begin{cases} -1, & x < 0, \\ [-1, 1], & x = 0, \\ 1, & x > 0. \end{cases}$$

and study its associated nonlinear eigenproblem $0 \in \Delta_1 f - \lambda\, \mathrm{sign}(f)$.

Proposition 4.1 *Any non-constant eigenvector f^* of the 1-Laplacian has median zero. Moreover, let λ_2 be the second eigenvalue of the 1-Laplacian, then if G is connected it holds $\lambda_2 = h_{\mathrm{RCC}}$.*

Proof: The subdifferential of the enumerator of F_1 can be computed as

$$\partial\big(\frac{1}{2}\sum_{i,j=1}^{n} w_{ij}|f_i - f_j|\big)_i = \Big\{ \sum_{j=1}^{n} w_{ij}u_{ij} \,|\, u_{ij} = -u_{ji}, u_{ij} \in \operatorname{sign}(f_i - f_j)\Big\},$$

where we use the set-valued mapping

$$\operatorname{sign}(x) = \begin{cases} -1, & x < 0, \\ [-1,1], & x = 0, \\ 1, & x > 0. \end{cases}$$

Moreover, the subdifferential of the denominator of F_1 is

$$\partial \left\|f\right\|_1 = \operatorname{sign}(f).$$

Note that, assuming that the graph is connected, any non-constant eigenvector f^* must have $\lambda^* > 0$. Thus if f^* is an eigenvector of the 1-Laplacian, there must exist u_{ij} with $u_{ij} = -u_{ji}$ and $u_{ij} \in \operatorname{sign}(f_i^* - f_j^*)$ and α_i with $\alpha_i \in \operatorname{sign}(f_i^*)$ such that

$$0 = \sum_{j=1}^{n} w_{ij}u_{ij} - \lambda^*\alpha_i.$$

Summing over i yields due to the anti-symmetry of u_{ij}, $\sum_i \alpha_i = |f_+^*| - |f_-^*| + \sum_{f_i^*=0} \alpha_i = 0$, where $|f_+^*|, |f_-^*|$ are the cardinalities of the positive and negative part of f^* and $|f_0^*|$ is the number of components with value zero. Thus we get

$$\big||f_+^*| - |f_-^*|\big| \leq |f_0^*|,$$

which implies with $|f_+^*|+|f_-^*|+|f_0^*| = |V|$ that $|f_+^*| \leq \frac{|V|}{2}$ and $|f_-^*| \leq \frac{|V|}{2}$. Thus the median of f^* is zero if $|V|$ is odd. If $|V|$ is even, the median is non-unique and is contained in $[\max f_-^*, \min f_+^*]$ which contains zero.

If the graph is connected, the only eigenvector corresponding to the first eigenvalue $\lambda_1 = 0$ of the 1-Laplacian is the constant one. As all non-constant eigenvectors have median zero, it follows with Equation 13 that $\lambda_2 \geq h_{\mathrm{RCC}}$. For the other direction, we have to use the algorithm we present in the following and some subsequent results. By Lemma 4.2 there exists a vector $f^* = \mathbf{1}_C$ with $|C| \leq \big|\overline{C}\big|$ such that $F_1(f^*) = h_{\mathrm{RCC}}$. Obviously, f^* is non-constant and has median zero and thus can be used as initial point f^0 for Algorithm 3. By Lemma 4.1 starting with $f^0 = f^*$ the sequence either terminates and the current iterate f^0 is an eigenvector or one finds a f^1 with $F_1(f^1) < F_1(f^0)$, where f^1 has median zero. Suppose that there exists such a f^1, then $F_1(f^1) < F_1(f^0) = \min_{\substack{f \text{ nonconstant} \\ \text{median(f)}=0}} F_1(f)$ which is a contradiction. Therefore the sequence has to terminate and thus by the argument in the proof of Theorem 4.1 the corresponding iterate is an eigenvector. Thus we get $h_{\mathrm{RCC}} \geq \lambda_2$ and thus with $\lambda_2 \geq h_{\mathrm{RCC}}$ we arrive at the desired result. □

For the computation of the second eigenvector we have to modify the IPM which is discussed in the next section.

4.1 Modification of the IPM for computing the second eigenvector of the 1-Laplacian

The direct minimization of (14) would be compatible with the IPM, but the global minimizer is the first eigenvector which is constant. For computing the second eigenvector note that, unlike in the case $p = 2$, we cannot simply project on the space orthogonal to the constant eigenvector, since mutual orthogonality of the eigenvectors does not hold in the nonlinear case.

Algorithm 3 is a modification of Algorithm 1 which computes a nonconstant eigenvector of the 1-Laplacian. The notation $|f_+^{k+1}|, |f_-^{k+1}|$ and $|f_0^{k+1}|$ refers to the cardinality of positive, negative and zero elements, respectively. Note that Algorithm 1 requires in each step the computation of *some* subgradient $s(f^k) \in \partial S(f^k)$, whereas in Algorithm 3 the subgradient v^k has to satisfy $\langle v^k, \mathbf{1}\rangle = 0$. This condition ensures that the inner objective is invariant under addition of a constant and thus not affected by the subtraction of the median. Opposite to [20] we can prove convergence to a nonconstant eigenvector of the 1-Laplacian. However, we cannot guarantee convergence to the *second* eigenvector. Thus we recommend to use multiple random initializations and use the result which achieves the best ratio Cheeger cut.

Algorithm 3 Computing a nonconstant 1-eigenvector of the graph 1-Laplacian

Input: weight matrix W
Initialization: nonconstant f^0 with $\operatorname{median}(f^0) = 0$ and $\left\|f^0\right\|_1 = 1$, accuracy ϵ
repeat
  $g^{k+1} = \operatorname*{arg\,min}_{\|f\|_2^2 \leq 1} \left\{ \frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j| - \lambda^k \left\langle f, v^k \right\rangle \right\}$
  $f^{k+1} = g^{k+1} - \operatorname{median}\left(g^{k+1}\right)$
  $v_i^{k+1} = \begin{cases} \operatorname{sign}(f_i^{k+1}), & \text{if } f_i^{k+1} \neq 0, \\ -\frac{|f_+^{k+1}| - |f_-^{k+1}|}{|f_0^{k+1}|}, & \text{if } f_i^{k+1} = 0. \end{cases}$,
  $\lambda^{k+1} = F_1(f^{k+1})$
until $\frac{\left|\lambda^{k+1} - \lambda^k\right|}{\lambda^k} < \epsilon$

Lemma 4.1 *The sequence f^k produced by Algorithm 3 satisfies $F_1(f^k) > F_1(f^{k+1})$ for all $k \geq 0$ or the sequence terminates.*

Proof: Note that, analogously to the proof of Lemma 3.1, we can conclude that the inner objective is nonpositive at the optimum, where the sequence terminates if the optimal value is zero as the previous f^k is among the minimizers of the inner problem. Now observe that the objective of the inner optimization problem is invariant under addition of a constant. This follows from the fact that we always have $\left\langle v^k, \mathbf{1} \right\rangle = 0$, which can be easily verified. Hence, with $R(f) = \frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|$, we get

$$R(f^{k+1}) - \lambda^k \left\langle f^{k+1}, v^k \right\rangle = R(g^{k+1}) - \lambda^k \left\langle g^{k+1}, v^k \right\rangle < 0 \,.$$

Dividing both sides by $\left\|f^{k+1}\right\|_1$ yields

$$\frac{R(f^{k+1})}{\|f^{k+1}\|_1} - \lambda^k \frac{\left\langle f^{k+1}, v^k \right\rangle}{\|f^{k+1}\|_1} < 0 \,,$$

and with $\left\langle f^{k+1}, v^k \right\rangle \leq \left\|f^{k+1}\right\|_1 \left\|v^k\right\|_\infty = \left\|f^{k+1}\right\|_1$, the result follows. □

Theorem 4.1 *The sequence f^k produced by Algorithm 3 converges to an eigenvector f^* of the 1-Laplacian with eigenvalue $\lambda^* \in \left[h_{\mathrm{RCC}}, F_1(f^0)\right)$. Furthermore, $F_1(f^k) > F_1(f^{k+1})$ for all $k \geq 0$ or the sequence terminates.*

Proof: Note that every constant vector u_0 satisfies $\Phi_{f^k}(u_0) = 0$ as $\left\langle v^k, \mathbf{1} \right\rangle = 0$. The minimizer of Φ_{f^k} is either negative or the sequence terminates in which case the previous non-constant g^k is a minimizer. In any case g^{k+1} cannot be constant and in turn f^{k+1} is nonconstant and has median zero. Thus for all k,

$$F_1(f^k) = \frac{\frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i^k - f_j^k|}{\|f^k\|_1} = \frac{\frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i^k - f_j^k|}{\|f^k - \operatorname{median}(f^k)\mathbf{1}\|_1} \geq h_{\mathrm{RCC}},$$

where we use that the median of f^k is zero. Thus $F_1(f^k)$ is lower-bounded by h_{RCC}. Note that $h_{\mathrm{RCC}} \leq \lambda_2$. We can conclude now analogously to Theorem 3.1 that the sequence $F_1(f^k)$ converges to some limit

$$\lambda^* = \lim_{k \to \infty} F_1(f^k) \geq h_{\mathrm{RCC}} \,.$$

As in Theorem 3.1 the compactness of the set containing the sequence g^k implies the existence of a convergent subsequence g^{k_j}, and using the fact that subtracting the median is continuous we have $\lim_{j \to \infty} f^{k_j} = g^* - \operatorname{median}(g^*)\mathbf{1} =: f^*$. The proof now proceeds analogously to Theorem 3.1. □

4.2 Quality guarantee for 1-spectral clustering

Even though we cannot guarantee that we obtain the optimal ratio Cheeger cut, we can guarantee that 1-spectral clustering always leads to a ratio Cheeger cut at least as good as the one found by

standard spectral clustering. Let $(C_f^*, \overline{C_f^*})$ be the partition of V obtained by optimal thresholding of f, where $C_f^* = \arg\min_t \mathrm{RCC}(C_f^t, \overline{C_f^t})$, and for $t \in \mathbb{R}$, $C_f^t = \{i \in V \mid f_i > t\}$. Furthermore, $\mathbf{1}_C$ denotes the vector which is 1 on C and 0 else.

Lemma 4.2 *Let $C, \overline{C}$ be a partitioning of the vertex set V, and assume that $|C| \leq \left|\overline{C}\right|$. Then for any vector $f \in \mathbb{R}^n$ of the form $f = \alpha \mathbf{1}_C$, where $\alpha \in \mathbb{R}$, it holds that $F_1(f) = \mathrm{RCC}(C, \overline{C})$.*

Proof: As F_1 is scale invariant, we can without loss of generality assume that $\alpha = 1$. Then we have

$$\begin{aligned} F_1(f) &= \frac{\frac{1}{2}\sum_{i,j=1}^n w_{ij}|f_i - f_j|}{\|f\|_1} = \frac{\frac{1}{2}\sum_{i\in C, j\notin C} w_{ij} + \frac{1}{2}\sum_{i\notin C, j\in C} w_{ij}}{\sum_{i\in C} 1} \\ &= \frac{\mathrm{cut}(C, \overline{C})}{|C|} = \frac{\mathrm{cut}(C, \overline{C}}{\min\left\{|C|, \left|\overline{C}\right|\right\}} = \mathrm{RCC}(C, \overline{C}) \,. \end{aligned}$$

□

Lemma 4.3 *Let $f \in \mathbb{R}^n$ with $\mathrm{median}(f) = 0$, and $C = \arg\min\{|C_f^*|, |\overline{C_f^*}|\}$. Then the vector $f^* = \mathbf{1}_C$ satisfies $F_1(f) \geq F_1(f^*)$.*

Proof: Denote by $f^+ : V \to \mathbb{R}$ the function $f_i^+ := \max\{0, f_i\}$, and analogously, let $f^- := \max\{0, -f_i\}$. Then we have

$$R(f) = \frac{1}{2}\sum_{i,j} w_{ij}\,|f_i - f_j| = \frac{1}{2}\sum_{i,j} w_{ij}\left|f_i^+ - f_i^- - f_j^+ + f_j^-\right| \,. \tag{15}$$

Note that we always have $\left|f_i^+ - f_i^- - f_j^+ + f_j^-\right| = \left|f_i^+ - f_j^+\right| + \left|f_i^- - f_j^-\right|$, which can easily be verified by performing a case distinction over the signs of f_i and f_j. Eq. (15) can now be written as

$$R(f) = \frac{1}{2}\sum_{i,j} w_{ij}\left|f_i^+ - f_j^+\right| + \frac{1}{2}\sum_{i,j} w_{ij}\left|f_i^- - f_j^-\right| = R(f^+) + R(f^-) \,.$$

Using the fact that $\|f\|_1$ can be decomposed as $\|f^+\|_1 + \|f^-\|_1$, we obtain

$$\frac{R(f)}{\|f\|_1} = \frac{R(f^+) + R(f^-)}{\|f^+\|_1 + \|f^-\|_1} \geq \min\left\{\frac{R(f^+)}{\|f^+\|_1}, \frac{R(f^-)}{\|f^-\|_1}\right\} \,. \tag{16}$$

The last inequality follows from the fact that we always have for $a, b, c, d > 0$,

$$\frac{a+b}{c+d} \geq \min\left\{\frac{a}{c}, \frac{b}{d}\right\} \,,$$

which can be easily shown by contradiction. Let wlog $\min\left\{\frac{a}{c}, \frac{b}{d}\right\} = \frac{a}{c}$, and assume that $\frac{a+b}{c+d} < \frac{a}{c}$. This implies $\frac{a}{c} > \frac{b}{d}$, which is a contradiction to $\frac{a}{c} \leq \frac{b}{d}$. Note that $\mathrm{median}(f) = 0$, hence we have

$$0 \in \arg\min_c \sum_{i\in V} |f_i - c| \,,$$

which implies that $0 \in \partial \sum_{i\in V} |f_i|$ and hence there exist coefficients $|\alpha_i| \leq 1$ such that

$$0 = \sum_{f_i \neq 0} \mathrm{sign}(f_i) + \sum_{f_i = 0} \alpha_i \,,$$

which is equivalent to $\Big|\,|\{i, f_i > 0\}| - |\{i, f_i < 0\}|\,\Big| \leq |\{i, f_i = 0\}|$. This inequality implies that $|\{i, f_i > 0\}| \leq \frac{|V|}{2}$ and $|\{i, f_i < 0\}| \leq \frac{|V|}{2}$. We now rewrite $R(f^+)$ as follows:

$$R(f^+) = \frac{1}{2}\sum_{f_i^+ > f_j^+} w_{ij}\left(f_i^+ - f_j^+\right) = \sum_{f_i^+ > f_j^+} w_{ij}\int_{f_j^+}^{f_i^+} 1dt = \int_0^\infty \sum_{f_i^+ > t \geq f_j^+} w_{ij}\; dt \,.$$

Note that for $t \geq 0$,

$$\sum_{f_i^+ > t \geq f_j^+} w_{ij} = \mathrm{cut}(C_f^t, \overline{C_f^t}) = \frac{\mathrm{cut}(C_f^t, \overline{C_f^t})}{\min\left\{\left|C_f^t\right|, \left|\overline{C_f^t}\right|\right\}} \cdot \left|C_f^t\right| \geq \mathrm{RCC}(C_f^*, \overline{C_f^*}) \cdot \left|C_f^t\right|,$$

where in the second step we used that $\left|C_f^t\right| \leq |\{i, f_i > 0\}| \leq \frac{|V|}{2}$. Hence we have

$$\begin{aligned} R(f^+) &\geq \int_0^\infty \mathrm{RCC}(C_f^*, \overline{C_f^*}) \cdot \left|C_f^t\right| dt = \mathrm{RCC}(C_f^*, \overline{C_f^*}) \int_0^\infty \sum_{f_i > t} 1 dt \\ &= \mathrm{RCC}(C_f^*, \overline{C_f^*}) \sum_{f_i > 0} \int_0^{f_i} 1 dt = \mathrm{RCC}(C_f^*, \overline{C_f^*}) \left\|f^+\right\|_1 . \end{aligned}$$

Hence it holds that $F_1(f^+) \geq \mathrm{RCC}(C_f^*, \overline{C_f^*})$, and analogously one shows that $F_1(f^-) \geq \mathrm{RCC}(C_{-f}^*, \overline{C_{-f}^*})$. Note that $\mathrm{RCC}(C_f^*, \overline{C_f^*}) = \mathrm{RCC}(C_{-f}^*, \overline{C_{-f}^*}) = F_1(f^*)$, by Lemma 4.2. Combining this with Eq. (16) yields the result. □

Theorem 4.2 *Let u denote the second eigenvector of the standard graph Laplacian, and f denote the result of Algorithm 3 after initializing with the vector $\frac{1}{|C|}\mathbf{1}_C$, where $C = \arg\min\{|C_u^*|, |\overline{C_u^*}|\}$. Then $\mathrm{RCC}(C_u^*, \overline{C_u^*}) \geq \mathrm{RCC}(C_f^*, \overline{C_f^*})$.*

Proof: Using Lemma 4.1 and 4.2 , we have the following chain of inequalities:

$$\mathrm{RCC}(C_u^*, \overline{C_u^*}) \overset{4.2}{=} F_1\left(\frac{1}{|C|}\mathbf{1}_C\right) = F_1(\mathbf{1}_C) \overset{4.1}{\geq} F_1(f).$$

With $C_1 := \arg\min\{|C_f^*|, |\overline{C_f^*}|\}$, we obtain by Lemma 4.3 and 4.2:

$$F_1(f) \overset{4.3}{\geq} F_1\left(\mathbf{1}_{C_1}\right) \overset{4.2}{=} \mathrm{RCC}(C_f^*, \overline{C_f^*}) .$$

□

4.3 Solution of the inner problem

The inner problem is convex, thus a solution can be computed by any standard method for solving convex nonsmooth programs, e.g. subgradient methods [3]. However, in this particular case we can exploit the structure of the problem and use the equivalent dual formulation of the inner problem.

Lemma 4.4 *Let $E \subset V \times V$ denote the set of edges and $A : \mathbb{R}^E \to \mathbb{R}^V$ be defined as $(A\alpha)_i = \sum_{j \,|\, (i,j) \in E} w_{ij}\alpha_{ij}$. The inner problem is equivalent to*

$$\min_{\{\alpha \in \mathbb{R}^E \,|\, \|\alpha\|_\infty \leq 1,\ \alpha_{ij} = -\alpha_{ji}\}} \Psi(\alpha) := \left\|A\alpha - F(f^k)v^k\right\|_2^2 .$$

The Lipschitz constant of the gradient of Ψ is upper bounded by $2\max_r \sum_{s=1}^n w_{rs}^2$.

Proof: First, we note that

$$\frac{1}{2}\sum_{i,j=1}^n w_{ij}|u_i - u_j| = \max_{\{\beta \in \mathbb{R}^E \,|\, \|\beta\|_\infty \leq 1\}} \frac{1}{2}\sum_{(i,j)\in E} w_{ij}(u_i - u_j)\beta_{ij} .$$

Introducing the new variable $\alpha_{ij} = \frac{1}{2}(\beta_{ij} - \beta_{ji})$, this can be rewritten as

$$\max_{\{\alpha \in \mathbb{R}^E \,|\, \|\alpha\|_\infty \leq 1,\ \alpha_{ij} = -\alpha_{ji}\}} \sum_{(i,j)\in E} w_{ij}\alpha_{ij}u_i = \max_{\{\alpha \in \mathbb{R}^E \,|\, \|\alpha\|_\infty \leq 1,\ \alpha_{ij} = -\alpha_{ji}\}} \langle u, A\alpha\rangle ,$$

where we have introduced the notation $(A\alpha)_i = \sum_{j \mid (i,j)\in E} w_{ij}\alpha_{ij}$. Both u and α are constrained to lie in non-empty compact, convex sets, and thus we can reformulate the inner objective by the standard min-max-theorem (see e.g. Corollary 37.3.2. in [17]) as follows:

$$\begin{aligned}
&\min_{\|u\|_2\leq 1} \max_{\{\alpha\in\mathbb{R}^E \mid \|\alpha\|_\infty\leq 1,\, \alpha_{ij}=-\alpha_{ji}\}} \langle u, A\alpha\rangle - F(f^k)\left\langle u, v^k\right\rangle \\
&= \max_{\{\alpha\in\mathbb{R}^E \mid \|\alpha\|_\infty\leq 1,\, \alpha_{ij}=-\alpha_{ji}\}} \min_{\|u\|_2\leq 1} \left\langle u, A\alpha - F(f^k)v^k\right\rangle \\
&= \max_{\{\alpha\in\mathbb{R}^E \mid \|\alpha\|_\infty\leq 1,\, \alpha_{ij}=-\alpha_{ji}\}} -\left\|A\alpha - F(f^k)v^k\right\|_2 .
\end{aligned}$$

In the last step we have used that the solution of the minimization of the linear function over the Euclidean unit ball is given by

$$u^* = -\frac{A\alpha - F(f^k)v^k}{\|A\alpha - F(f^k)v^k\|_2},$$

if $\left\|A\alpha - F(f^k)v^k\right\| \neq 0$ and otherwise u^* is an arbitrary element of the Euclidean unit ball. Transforming the maximization problem into a minimization problem finishes the proof of the first statement. Regarding the Lipschitz constant, a straightforward computation shows that

$$(\nabla\Psi(\alpha))_{rs} = 2w_{rs}\Big(\sum_{j\mid(r,j)\in E} w_{rj}\alpha_{rj} - F(f^k)v_r^k\Big).$$

Thus,

$$\begin{aligned}
\|\nabla\Psi(\alpha) - \nabla\Psi(\beta)\|^2 &= 4\sum_{(r,s)\in E} w_{rs}^2\Big(\sum_{j\mid(r,j)\in E} w_{rj}(\alpha_{rj}-\beta_{rj})\Big)^2 \\
&\leq 4\sum_{(r,s)\in E} w_{rs}^2\Big(\sum_{j\mid(r,j)\in E} w_{rj}^2 \sum_{i\mid(r,i)\in E}(\alpha_{ri}-\beta_{ri})^2\Big) \\
&= 4\sum_{r=1}^n\Big(\sum_{s\mid(r,s)\in E} w_{rs}^2\Big)^2 \sum_{i\mid(r,i)\in E}(\alpha_{ri}-\beta_{ri})^2 \\
&\leq 4\Big(\max_r \sum_{s=1}^n w_{rs}^2\Big)^2 \sum_{(r,i)\in E}(\alpha_{ri}-\beta_{ri})^2.
\end{aligned}$$

□

Compared to the primal problem, the objective of the dual problem is smooth. Moreover, it can be efficiently solved using FISTA ([2]), a two-step subgradient method with guaranteed convergence rate $O(\frac{1}{k^2})$ where k is the number of steps. The only input of FISTA is an upper bound on the Lipschitz constant of the gradient of the objective. FISTA provides a good solution in a few steps which guarantees descent in functional (13) and thus makes the modified IPM very fast. The resulting Algorithm is shown in Alg. 4.

5 Application 2: Sparse PCA

Principal Component Analysis (PCA) is a standard technique for dimensionality reduction and data analysis [13]. PCA finds the k-dimensional subspace of maximal variance in the data. For $k = 1$, given a data matrix $X \in \mathbb{R}^{n\times p}$ where each column has mean 0, in PCA one computes

$$f^* = \arg\max_{f\in\mathbb{R}^p} \frac{\left\langle f, X^TXf\right\rangle}{\|f\|_2^2}, \tag{17}$$

where the maximizer f^* is the largest eigenvector of the covariance matrix $\Sigma = X^TX \in \mathbb{R}^{p\times p}$. The interpretation of the PCA component f^* is difficult as usually all components are nonzero. In sparse PCA one wants to get a small number of features which still capture most of the variance. For instance, in the case of gene expression data one would like the principal components to consist only of a few significant genes, making it easy to interpret by a human. Thus one needs to enforce sparsity of the PCA component, which yields a trade-off between explained variance and sparsity.

Algorithm 4 Solution of the dual inner problem with FISTA

Input: Lipschitz-constant L of $\nabla\Psi$,
Initialization: $t^1 = 1$, $\alpha^1 \in \mathbb{R}^E$,
repeat

$$\begin{aligned}\beta_{rs}^{t+1} &= \alpha_{rs}^t - \frac{1}{L}\nabla\Psi(\alpha^t)_{rs} \\ &= \alpha_{rs}^t - \frac{2}{L}w_{rs}\big(\sum_{j\,|\,(r,j)\in E} w_{rj}\alpha_{rj}^t - F(f^k)v_r^k\big)\end{aligned}$$

$t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$,
$\alpha_{rs}^{t+1} = \beta_{rs}^{t+1} + \frac{t_k-1}{t_{k+1}}\Big(\beta_{rs}^{t+1} - \beta_{rs}^t\Big)$.
until stop if gap between original and dual problem is smaller than ϵ

While standard PCA leads to an eigenproblem, adding a constraint on the cardinality, i.e. the number of nonzero coefficients, makes the problem NP-hard. The first approaches performed simple thresholding of the principal components which was shown to be misleading [6]. Since then several methods have been proposed, mainly based on penalizing the L_1 norm of the principal components, including SCoTLASS [14] and SPCA [24]. D'Aspremont et al.[10] focused on the L_0-constrained formulation and proposed a greedy algorithm to compute a full set of good candidate solutions up to a specified target sparsity, and derived sufficient conditions for a vector to be globally optimal. Moghaddam et al. [16] used branch and bound to compute optimal solutions for small problem instances. Other approaches include D.C. [19] and EM-based methods [18]. Recently, Journee et al. [15] proposed two single unit (computation of one component only) and two block (simultaneous computation of multiple components) methods based on L_0-penalization and L_1-penalization.

Problem (17) is equivalent to

$$f^* = \operatorname*{arg\,min}_{f\in\mathbb{R}^p} \frac{\|f\|_2^2}{\langle f, \Sigma f\rangle} = \operatorname*{arg\,min}_{f\in\mathbb{R}^p} \frac{\|f\|_2}{\|Xf\|_2} \,.$$

In order to enforce sparsity we use instead of the L_2-norm a convex combination of an L_1 norm and L_2 norm in the enumerator, which yields the functional

$$F(f) = \frac{(1-\alpha)\,\|f\|_2 + \alpha\,\|f\|_1}{\|Xf\|_2} \,, \tag{18}$$

with sparsity controlling parameter $\alpha \in [0,1]$. Standard PCA is recovered for $\alpha = 0$, whereas $\alpha = 1$ yields the sparsest non-trivial solution: the component with the maximal variance. One easily sees that the formulation (18) fits in our general framework, as both enumerator and denominator are 1-homogeneous functions. The inner problem of the IPM becomes

$$g^{k+1} = \operatorname*{arg\,min}_{\|f\|_2\le 1} (1-\alpha)\,\|f\|_2 + \alpha\,\|f\|_1 - \lambda^k\left\langle f, \mu^k\right\rangle \,, \quad \text{where} \quad \mu^k = \frac{\Sigma f^k}{\sqrt{\langle f^k, \Sigma f^k\rangle}}. \tag{19}$$

This problem has a closed form solution. In the following we use the notation $x_+ = \max\{0, x\}$.

Lemma 5.1 *The convex optimization problem* (19) *has the analytical solution*

$$g_i^{k+1} = \frac{1}{s}\operatorname{sign}(\mu_i^k)\big(\lambda^k\left|\mu_i^k\right| - \alpha\big)_+, \quad \textit{where} \quad s = \sqrt{\textstyle\sum_{i=1}^n (\lambda^k|\mu_i^k| - \alpha)_+^2} \,.$$

Proof: We note that the objective is positively 1-homogenous and that the optimum is either zero by plugging in the previous iterate or negative in which case the optimum is attained at the boundary. Thus wlog we can assume that at the optimum $\|f\|_2 = 1$. Thus the problem reduces to

$$\min_{\|f\|_2\le 1} \alpha\,\|f\|_1 - \lambda^k\left\langle f, \mu^k\right\rangle .$$

First, we derive an equivalent "dual" problem, noting

$$\alpha \|f\|_1 - \lambda^k \langle \mu^k, f \rangle = \max_{\|v\|_\infty \leq 1} \langle f, \alpha v - \lambda^k \mu^k \rangle \ .$$

Using the fact that the objective is convex in f and concave in v and the feasible set is compact, we obtain by the min-max equality:

$$\begin{aligned} \min_{\|f\|_2 \leq 1} \ \max_{\|v\|_\infty \leq 1} \langle f, \alpha v - \lambda^k \mu^k \rangle &= \max_{\|v\|_\infty \leq 1} \ \min_{\|f\|_2 \leq 1} \langle f, \alpha v - \lambda^k \mu^k \rangle \\ &= \max_{\|v\|_\infty \leq 1} - \left\| \alpha v - \lambda^k \mu^k \right\|_2 \ . \end{aligned}$$

The objective of the dual problem is separable in v and the constraints of v as well. Thus each component can be optimized separately which gives

$$v_i = \operatorname{sign}(\mu_i^k) \min \left\{ 1, \frac{\lambda^k |\mu_i^k|}{\alpha} \right\}.$$

Using that $f^* = (-\alpha v + \lambda^k \mu^k) / \left\| \lambda^k \mu^k - \alpha v \right\|_2$, we get the solution

$$f_i = \frac{\operatorname{sign}(\mu_i^k)(\lambda^k |\mu_i^k| - \alpha)_+}{\sqrt{\sum_{i=1}^n (\lambda^k |\mu_i^k| - \alpha)_+^2}}.$$

□

As s is just a scaling factor, we can omit it and obtain the simple and efficient scheme to compute sparse principal components shown in Algorithm 5. While the derivation is quite different from [15], the resulting algorithms are very similar. The subtle difference is that in our formulation the thresholding parameter of the inner problem depends on the current eigenvalue estimate whereas it is fixed in [15]. Empirically, this leads to the fact that we need slightly less iterations to converge.

Algorithm 5 Sparse PCA

1: **Input:** data matrix X, sparsity controlling parameter α, accuracy ϵ
2: **Initialization:** $f^0 =$ random with $S(f^k) = 1$, $\lambda^0 = F(f^k)$
3: **repeat**
4: $\quad g_i^{k+1} = \operatorname{sign}(\mu_i^k)\left(\lambda^k \left|\mu_i^k\right| - \alpha\right)_+$,
5: $\quad f^{k+1} = \frac{g^{k+1}}{\|X g^{k+1}\|_2}$
6: $\quad \lambda^{k+1} = (1-\alpha) \left\| f^{k+1} \right\|_2 + \alpha \left\| f^{k+1} \right\|_1$
7: $\quad \mu^{k+1} = \frac{\Sigma f^{k+1}}{\|X f^{k+1}\|_2}$
8: **until** $\frac{\left|\lambda^{k+1} - \lambda^k\right|}{\lambda^k} < \epsilon$

6 Experiments

1-Spectral Clustering: We compare our IPM with the total variation (TV) based algorithm by [20], p-spectral clustering with $p = 1.1$ [5] as well as standard spectral clustering with optimal thresholding the second eigenvector of the graph Laplacian ($p = 2$). The graph and the two-moons dataset is constructed as in [5]. The following table shows the average ratio Cheeger cut (RCC) and error (classification as in [5]) for 100 draws of a two-moons dataset with 2000 points. In the case of the IPM, we use the best result of 10 runs with random initializations and one run initialized with the second eigenvector of the unnormalized graph Laplacian. For [20] we initialize once with the second eigenvector of the normalized graph Laplacian as proposed in [20] and 10 times randomly. IPM and the TV-based method yield similar results, slightly better than 1.1-spectral and clearly outperforming standard spectral clustering. In terms of runtime, IPM and [20] are on the same level.

	Inverse Power Method	Szlam & Bresson [20]	1.1-spectral [5]	Standard spectral
Avg. RCC	0.0195 ($\pm$ 0.0015)	0.0195 ($\pm$ 0.0015)	0.0196 ($\pm$ 0.0016)	0.0247 ($\pm$ 0.0016)
Avg. error	0.0462 ($\pm$ 0.0161)	0.0491 ($\pm$ 0.0181)	0.0578 ($\pm$ 0.0285)	0.1685 ($\pm$ 0.0200)

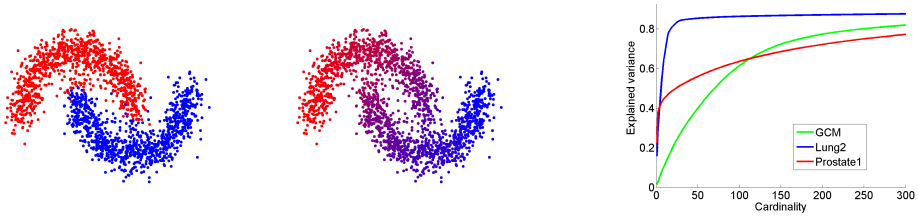

Figure 1: Left and middle: Second eigenvector of the 1-Laplacian and 2-Laplacian, respectively. Right: Relative Variance (relative to maximal possible variance) versus number of non-zero components for the three datasets Lung2, GCM and Prostate1.

Next we perform unnormalized 1-spectral clustering on the full USPS and MNIST-datasets (9298 resp. 70000 points). As clustering criterion we use the multicut version of RCut, given as

$$\mathrm{RCut}(C_1, \ldots, C_K) = \sum_{i=1}^{K} \frac{\mathrm{cut}(C_i, \overline{C_i})}{|C_i|} \ .$$

We successively subdivide clusters until the desired number of clusters ($K = 10$) is reached. In each substep the eigenvector obtained on the subgraph is thresholded such that the multi-cut criterion is minimized. This recursive partitioning scheme is used for all methods. As in the previous experiment, we perform one run initialized with the thresholded second eigenvector of the unnormalized graph Laplacian in the case of the IPM and with the second eigenvector of the normalized graph Laplacian in the case of [20]. In both cases we add 100 runs with random initializations. The next table shows the obtained RCut and errors.

		Inverse Power Method	S.&B. [20]	1.1-spectral [5]	Standard spectral
MNIST	Rcut	0.1507	0.1545	0.1529	0.2252
	Error	0.1244	0.1318	0.1293	0.1883
USPS	Rcut	0.6661	0.6663	0.6676	0.8180
	Error	0.1349	0.1309	0.1308	0.1686

Again the three nonlinear eigenvector methods clearly outperform standard spectral clustering. Note that our method requires additional effort (100 runs) but we get better results. For both datasets our method achieves the best RCut. However, if one wants to do only a single run, by Theorem 4.2 for bi-partitions one achieves a cut at least as good as the one of standard spectral clustering if one initializes with the thresholded 2nd eigenvector of the 2-Laplacian.

Sparse PCA: We evaluate our IPM for sparse PCA on gene expression datasets obtained from [1]. We compare with two recent algorithms: the L_1 based single-unit power algorithm of [15] as well as the EM-based algorithm in [18]. For all considered datasets, the three methods achieve very similar performance in terms of the tradeoff between explained variance and sparsity of the solution, see Fig.1 (Right). In fact the results are so similar that for each dataset, the plots of all three methods coincide in one line. In [15] it also has been observed that the best state-of-the-art algorithms produce the same trade-off curve if one uses the same initialization strategy.

Acknowledgments: This work has been supported by the Excellence Cluster on Multimodal Computing and Interaction at Saarland University.

References

[1] http://www.stat.ucla.edu/~wxl/research/microarray/DBC/index.htm.

[2] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434, 2009.

[3] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[4] R.J. Biezuner, G. Ercole, and E.M. Martins. Computing the first eigenvalue of the p-Laplacian via the inverse power method. *Journal of Functional Analysis*, 257:243–270, 2009.

[5] T. Bühler and M. Hein. Spectral Clustering based on the graph p-Laplacian. In *Proceedings of the 26th International Conference on Machine Learning*, pages 81–88. Omnipress, 2009.

[6] J. Cadima and I.T. Jolliffe. Loading and correlations in the interpretation of principal components. *Journal of Applied Statistics*, 22:203–214, 1995.

[7] K.-C. Chang. Variational methods for non-differentiable functionals and their applications to partial differential equations. *Journal of Mathematical Analysis and Applications*, 80:102–129, 1981.

[8] F.R.K. Chung. *Spectral Graph Theory*. AMS, 1997.

[9] F.H. Clarke. *Optimization and Nonsmooth Analysis*. Wiley New York, 1983.

[10] A. d'Aspremont, F. Bach, and L. El Ghaoui. Optimal solutions for sparse principal component analysis. *Journal of Machine Learning Research*, 9:1269–1294, 2008.

[11] T. Goldstein and S. Osher. The Split Bregman method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[12] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996.

[13] I.T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.

[14] I.T. Jolliffe, N. Trendafilov, and M. Uddin. A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, 12:531–547, 2003.

[15] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre. Generalized Power Method for Sparse Principal Component Analysis. *Journal of Machine Learning Research*, 11:517–553, 2010.

[16] B. Moghaddam, Y. Weiss, and S. Avidan. Spectral bounds for sparse PCA: Exact and greedy algorithms. In *Advances in Neural Information Processing Systems*, pages 915–922. MIT Press, 2006.

[17] R.T. Rockafellar. *Convex analysis*. Princeton University Press, 1970.

[18] C.D. Sigg and J.M. Buhmann. Expectation-maximization for sparse and non-negative PCA. In *Proceedings of the 25th International Conference on Machine Learning*, pages 960–967. ACM, 2008.

[19] B.K. Sriperumbudur, D.A. Torres, and G.R.G. Lanckriet. Sparse eigen methods by D.C. programming. In *Proceedings of the 24th International Conference on Machine Learning*, pages 831–838. ACM, 2007.

[20] A. Szlam and X. Bresson. Total variation and Cheeger cuts. In *Proceedings of the 27th International Conference on Machine Learning*, pages 1039–1046. Omnipress, 2010.

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

[22] F. Yang and Z. Wei. Generalized Euler identity for subdifferentials of homogeneous functions and applications. *Mathematical Analysis and Applications*, 337:516–523, 2008.

[23] E. Zarantonello. The meaning of the Cauchy-Schwartz-Buniakovsky inequality. *Proceedings of the American Mathematical Society*, 59(1), 1976.

[24] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15:265–286, 2006.