# Several Flawed Approaches to Penalized SVDs

*A supplementary note to*

**"The analysis of two-way functional data using two-way regularized singular value decompositions"**

Jianhua Z. Huang, Haipeng Shen, Andreas Buja

**Abstract**

The application of two-way penalization to SVDs is not a trivial matter. In this note, we show that several "natural" approaches to penalized SVDs do not work and explain why so. We present these flawed approaches to spare readers fruitless search in dead ends.

# 1 Outline

Section 2 presents several approaches to penalized SVD through generalization of penalized regressions and then discusses their flaws. Sections 3 and 4 provide some deeper discussion by comparing the criteria in terms of bi-Rayleigh quotients and alternating optimization algorithms, respectively.

# 2 Several approaches to penalized SVD

We consider a $m \times n$ data matrix $\mathbf{X}$ whose indexes both live in structured domains, and our goal is to find best rank-one approximations of $\mathbf{X}$ in a sense that reflects the domain structure. The calculations generalize to any rank-$q$ approximation.

We write rank-one approximations as $\mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ and $\mathbf{v}$ are $n$- and $m$-vectors, respectively. We will not assume that either is normalized, hence they are unique only up to a scale factor that can be shifted between them:

$$\mathbf{u} \mapsto c\mathbf{u}\,, \quad \mathbf{v} \mapsto \mathbf{v}/c\,. \tag{1}$$

It will turn out that these scale transformations play an essential role in the design of a desirable form of regularized SVD. As a reminder, the criterion for unregularized standard SVD is

$$\mathcal{C}_0(\mathbf{u}, \mathbf{v}) \;=\; \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2\,, \tag{2}$$

Jianhua Z. Huang is Professor (Email: jianhua@stat.tamu.edu), Department of Statistics, Texas A&M University, College Station, TX 77843. Haipeng Shen (Email: haipeng@email.unc.edu) is Assistant Professor, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. Andreas Buja (Email: buja@wharton.upenn.edu) is Liem Sioe Liong/First Pacific Company Professor, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104. Jianhua Z. Huang's work was partially supported by NSF grant DMS-0606580, NCI grant CA57030, and Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST). Haipeng Shen's work was partially supported by NSF grant DMS-0606577, CMMI-0800575, and UNC-CH R. J. Reynolds Fund Award for Junior Faculty Development.

where $\|\cdot\|_F^2$ denote the squared Frobenius norm of a matrix. A plausible approach to invariance under (1) would be to require normalization, $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$, and introduce a slope factor $s$ such that $\mathcal{C}_0(\mathbf{u},\mathbf{v}) \;=\; \|\mathbf{X} - s\,\mathbf{u}\mathbf{v}^T\|_F^2$. We prefer, however, to deal with invariance directly because normalization requirements result in extraneous difficulties rather than clarifications.

We will now require $\mathbf{u}$ and $\mathbf{v}$ to satisfy some domain-specific conditions, typically a form of smoothness. Specifically, let $\mathbf{\Omega}_u$ and $\mathbf{\Omega}_v$ be $n \times n$ and $m \times m$ non-negative definite matrices, respectively, such that smaller values of the quadratic forms $\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}$ and $\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}$ are more desirable than larger ones in terms of the respective index domain. These quadratic forms will be chosen to be *roughness penalties* if $\mathbf{u}$ and $\mathbf{v}$ are required to be smooth on their domains. For example, if the index set of $\mathbf{u}$ is $I = \{1, ..., n\}$ corresponding to equi-spaced time points, then a crude roughness penalty is $\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u} = \sum_i (u_{i+1} - 2u_i + u_{i-1})^2$. For future use we recall that the solution of the penalized regression problem

$$\|\mathbf{y} - \mathbf{f}\|^2 \;+\; \mathbf{f}^T\mathbf{\Omega}\,\mathbf{f} \;\;=\;\; \min_{\mathbf{f}} \qquad (\mathbf{f}, \mathbf{y} \in \mathbb{R}^n)$$

is $\mathbf{f} = (\mathbf{I} + \mathbf{\Omega})^{-1}\mathbf{y}$, that is, the matrix $\mathbf{S} = (\mathbf{I} + \mathbf{\Omega})^{-1}$ is the "smoothing matrix" that corresponds to the hat matrix of linear regression. This, however, is not an orthogonal projection but a symmetric matrix with eigenvalues between zero and one (Hastie and Tibshirani, 1990).

The goal is to balance the penalties $\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}$ and $\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}$ against goodness-of-fit as measured by the residual sum of squares $\|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2$. Following practices in regression, it would be natural to try to achieve such a balance by minimizing the sum

$$\mathcal{C}_1(\mathbf{u},\mathbf{v}) \;\;=\;\; \|\mathbf{X} \;-\; \mathbf{u}\mathbf{v}^T\|_F^2 \;+\; \mathbf{u}^T\mathbf{\Omega}_u\mathbf{u} \;+\; \mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}\;. \tag{3}$$

This criterion has, however, the undesirable property that it is not invariant under scale transformations (1). While the goodness-of-fit criterion $\|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2$ remains unchanged, the penalization terms change to $c^2\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}$ and $c^{-2}\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}$, respectively. It appears therefore that this additive combination of the penalties imposes specific scales on $\mathbf{u}$ and $\mathbf{v}$ relative to each other, while the goodness-of-fit criterion has nothing to say about these relative scales. Indeed, we will see in Sections 3 below that minimization of (3) forces the two penalties to attain identical values.

If the obvious approach is deficient, what would be a desirable way to balance goodness-of-fit and penalties? A heuristic pointer can be found by expanding the goodness-of-fit criterion $\mathcal{C}_0$ with some trace algebra as follows:

$$\mathcal{C}_0(\mathbf{u},\mathbf{v}) \;\;=\;\; \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2 \;\;=\;\; \|\mathbf{X}\|_F^2 \;-\; 2\,\mathbf{u}^T\mathbf{X}\mathbf{v} \;+\; \|\mathbf{u}\|^2\|\mathbf{v}\|^2\;. \tag{4}$$

Of note is that the rightmost term is bi-quadratic. This matters because the penalties act as modifiers of this term and should be of comparable functional form. It is therefore natural to search for bi-quadratic forms of combined penalties, and the simplest form would be the product as opposed to the sum of the penalties:

$$\mathcal{C}_2(\mathbf{u},\mathbf{v}) \;\;=\;\; \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2 \;+\; \mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}\;. \tag{5}$$

This criterion, while satisfying invariance under the scale transformations (1), has another deficiency: it does not specialize to a one-way regularization method in $\mathbf{v}$ when $\mathbf{\Omega}_u = 0$, say. (This specialization requirement would appear to be satisfied by $\mathcal{C}_1$, but appearances can be misleading, as is the case here; see Section 3).

The search for criteria that satisfy both requirements lead us to the following proposal:

$$\mathcal{C}_3(\mathbf{u}, \mathbf{v}) \;=\; \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2 \;+\; \mathbf{u}^T\mathbf{\Omega}_u\mathbf{u} \cdot \|\mathbf{v}\|^2 \;+\; \|\mathbf{u}\|^2 \cdot \mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} \;. \tag{6}$$

The idea is to combine smoothing in $\mathbf{u}$ (caused by $\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}$) with shrinkage in $\mathbf{v}$ (caused by $\|\mathbf{v}\|^2$), and vice versa. The criterion is invariant under (1) and it specializes to a version of one-way regularized SVD when one of the penalities vanishes (Huang, Shen and Buja, 2008). We will see, however, that this criterion has a coupling problem (as do $\mathcal{C}_1$ and $\mathcal{C}_2$): the natural alternating algorithm that optimizes $\mathbf{u}$ and $\mathbf{v}$ in turn amounts to alternating smoothing where the amount of smoothing for $\mathbf{u}$ depends on $\mathbf{v}$, and vice versa (Section 4). While this "defect" may make only a weak heuristic argument compared to the scale invariance and specialization requirements, it takes on more weight with hindsight once the conceptually cleanest solution is found.

This solution turns out to be the most complex of all, the sum of both types of bi-quadratic penalties tried so far:

$$\mathcal{C}_4(\mathbf{u}, \mathbf{v}) \;=\; \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2 \;+\; \mathbf{u}^T\mathbf{\Omega}_1\mathbf{u} \cdot \|\mathbf{v}\|^2 \;+\; \|\mathbf{u}\|^2 \cdot \mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} \;+\; \mathbf{u}^T\mathbf{\Omega}_1\mathbf{u} \cdot \mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} \;. \tag{7}$$

While this criterion is derived from axiomatic argument in our paper (Huang, Shen and Buja, 2009), one can get a glimpse of the reason why this succeeds by substituting the expansion (4) of the goodness-of-fit criterion in (7) and simplifying the algebra:

$$\mathcal{C}_4(\mathbf{u}, \mathbf{v}) \;=\; \|\mathbf{X}\|_F^2 \;-\; 2\,\mathbf{u}^T\mathbf{X}\mathbf{v} \;+\; \mathbf{u}^T(\mathbf{I} + \mathbf{\Omega}_u)\mathbf{u} \cdot \mathbf{v}^T(\mathbf{I} + \mathbf{\Omega}_v)\mathbf{v} \;. \tag{8}$$

This kind of factorization is absent from the previous combined penalties. The matrices $\mathbf{I} + \mathbf{\Omega}_u$ and $\mathbf{I} + \mathbf{\Omega}_v$ are the inverses of the smoother matrices $\mathbf{S}_u$ and $\mathbf{S}_v$, a fact that will result in an intuitive alternating algorithm (Section 4). Comparing the expanded form (4) of $\mathcal{C}_0$ with (8), we see that the Euclidean squared norms $\|\mathbf{u}\|^2$ and $\|\mathbf{v}\|^2$ are replaced by the quadratic forms $\mathbf{u}^T(\mathbf{I} + \mathbf{\Omega}_u)\mathbf{u}$ and $\mathbf{v}^T(\mathbf{I} + \mathbf{\Omega}_v)\mathbf{v}$, respectively. This fact hints that hierarchies of regularized singular vectors based on $\mathcal{C}_4$ will be orthogonal in the sense that $\mathbf{u}_1^T(\mathbf{I} + \mathbf{\Omega}_u)\mathbf{u}_2 = 0$ and $\mathbf{v}_1^T(\mathbf{I} + \mathbf{\Omega}_v)\mathbf{v}_2 = 0$, as opposed to $\mathbf{u}_1^T\mathbf{u}_2 = 0$ and $\mathbf{v}_1^T\mathbf{v}_2 = 0$.

# 3 Comparison of criteria in terms of bi-Rayleigh quotients

In this section we compare the penalized LS criteria in terms of scale-invariant ratios that are a form of bi-Rayleigh quotients, and in the next subsection in terms of stationary equations that suggests alternating algorithms.

$$\begin{array}{rcll}
\mathcal{C}_0(\mathbf{u},\mathbf{v}) & = & \|\mathbf{X}-\mathbf{u}\mathbf{v}^T\|_F^2 : & \\
\min_{s,t}\ \mathcal{C}_0(s\mathbf{u},t\mathbf{v}) & = & \|\mathbf{X}\|_F^2 - \frac{(\mathbf{u}^T\mathbf{X}\mathbf{v})^2}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2} & \\
\mathbf{u} & = & c_v\,\mathbf{X}\mathbf{v}\,, & c_v = \|\mathbf{v}\|^{-2} \\
\mathbf{v} & = & c_u\,\mathbf{X}^T\mathbf{u}\,, & c_u = \|\mathbf{u}\|^{-2}
\end{array}$$

$$\begin{array}{rcll}
\mathcal{C}_1(\mathbf{u},\mathbf{v}) & = & \|\mathbf{X}-\mathbf{u}\mathbf{v}^T\|_F^2 + \mathbf{u}^T\mathbf{\Omega}_u\mathbf{u} + \mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} : & \\
\min_{s,t}\ \mathcal{C}_1(s\mathbf{u},t\mathbf{v}) & = & \|\mathbf{X}\|^2 - \frac{\left(\mathbf{u}^T\mathbf{X}\mathbf{v}-\left(\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}\right)^{1/2}\right)^2}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2} & \\
\mathbf{u} & = & c_v\,\mathbf{S}_u(\lambda_v)\,\mathbf{X}\mathbf{v}\,, & \lambda_v = c_v = \|\mathbf{v}\|^{-2} \\
\mathbf{v} & = & c_u\,\mathbf{S}_v(\lambda_u)\,\mathbf{X}^T\mathbf{u}\,, & \lambda_u = c_u = \|\mathbf{u}\|^{-2}
\end{array}$$

$$\begin{array}{rclll}
\mathcal{C}_2(\mathbf{u},\mathbf{v}) & = & \|\mathbf{X}-\mathbf{u}\mathbf{v}^T\|_F^2 + \mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} : & & \\
\min_{s,t}\ \mathcal{C}_2(s\mathbf{u},t\mathbf{v}) & = & \|\mathbf{X}\|^2 - \frac{(\mathbf{u}^T\mathbf{X}\mathbf{v})^2}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2+\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}} & & \\
\mathbf{u} & = & c_v\,\mathbf{S}_u(\lambda_v)\,\mathbf{X}\mathbf{v}\,, & \lambda_v = \mathcal{R}_v(\mathbf{v})\,, & c_v = \|\mathbf{v}\|^{-2} \\
\mathbf{v} & = & c_u\,\mathbf{S}_v(\lambda_u)\,\mathbf{X}^T\mathbf{u}\,, & \lambda_u = \mathcal{R}_u(\mathbf{u})\,, & c_u = \|\mathbf{u}\|^{-2}
\end{array}$$

$$\begin{array}{rclll}
\mathcal{C}_3(\mathbf{u},\mathbf{v}) & = & \|\mathbf{X}-\mathbf{u}\mathbf{v}^T\|_F^2 + \mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}\cdot\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} : & & \\
\min_{s,t}\ \mathcal{C}_3(s\mathbf{u},t\mathbf{v}) & = & \|\mathbf{X}\|^2 - \frac{(\mathbf{u}^T\mathbf{X}\mathbf{v})^2}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2+\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}\cdot\|\mathbf{v}\|^2+\|\mathbf{u}\|^2\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}} & & \\
\mathbf{u} & = & c_v\,\mathbf{S}_u(\lambda_v)\,\mathbf{X}\mathbf{v}\,, & \lambda_v = (1+\mathcal{R}_v(\mathbf{v}))^{-1}, & c_v = \|\mathbf{v}\|^{-2}\,\lambda_v \\
\mathbf{v} & = & c_u\,\mathbf{S}_v(\lambda_u)\,\mathbf{X}^T\mathbf{u}\,, & \lambda_u = (1+\mathcal{R}_u(\mathbf{u}))^{-1}, & c_u = \|\mathbf{u}\|^{-2}\,\lambda_v
\end{array}$$

$$\begin{array}{rcll}
\mathcal{C}_4(\mathbf{u},\mathbf{v}) & = & \|\mathbf{X}-\mathbf{u}\mathbf{v}^T\|_F^2 + \mathbf{u}^T\mathbf{\Omega}_1\mathbf{u}\cdot\|\mathbf{v}\|^2 + \|\mathbf{u}\|^2\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} + \mathbf{u}^T\mathbf{\Omega}_1\mathbf{u}\cdot\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v} : & \\
\min_{s,t}\ \mathcal{C}_4(s\mathbf{u},t\mathbf{v}) & = & \|\mathbf{X}\|^2 - \frac{(\mathbf{u}^T\mathbf{X}\mathbf{v})^2}{\mathbf{u}^T(\mathbf{I}+\mathbf{\Omega}_u)\mathbf{u}\cdot\mathbf{v}^T(\mathbf{I}+\mathbf{\Omega}_v)\mathbf{v}} & \\
\mathbf{u} & = & c_v\,\mathbf{S}_u(1)\,\mathbf{X}\mathbf{v}\,, & c_v = \|\mathbf{v}\|^{-2}(1+\mathcal{R}_v(\mathbf{v}))^{-1} \\
\mathbf{v} & = & c_u\,\mathbf{S}_v(1)\,\mathbf{X}^T\mathbf{u}\,, & c_u = \|\mathbf{u}\|^{-2}(1+\mathcal{R}_u(\mathbf{u}))^{-1}
\end{array}$$

Table 1: The five (penalized) LS criteria, their bi-Rayleigh quotients, and the stationary equations for alternating algorithms; $\mathbf{S}_{\mathbf{u/v}}$ are the smoother matrices and $\mathcal{R}_{\mathbf{u/v}}$ the plain Rayleigh quotients of $\mathbf{\Omega}_{u/v}$ [see (9) and (10)].

We first show how one arrives at "bi-Rayleigh quotients", here so called because they are the analogs of the usual Rayleigh quotients in eigen problems adapted to singular value problems. One arrives at them by minimizing scalar factors in quadratic loss functions. We exemplify with unpenalized LS (4) where we introduce two scale factors, one for $\mathbf{u}$ and $\mathbf{v}$ each:

$$\begin{aligned}\mathcal{C}_0(s\mathbf{u}, t\mathbf{v}) &= \|\mathbf{X} - st\,\mathbf{u}\mathbf{v}^T\|_F^2 \\ &= \|\mathbf{X}\|^2 \,-\, 2\,st\,\mathbf{u}^T\mathbf{X}\mathbf{v} + s^2t^2\,\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \\ &= \|\mathbf{X}\|^2 \,-\, 2\,r\,\mathbf{u}^T\mathbf{X}\mathbf{v} + r^2\,\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \ ,\end{aligned}$$

where $r = st$. As a simple quadratic in $r$ of the form $C^2 - 2Br + A^2r^2$, the minimum is attained at $r = B/A^2$ and the value of the minimum is $C^2 - B^2/A^2$. In this case, we obtain

$$\min_{s,t} \mathcal{C}_0(s\mathbf{u}, t\mathbf{v}) \;=\; \|\mathbf{X}\|_F^2 \,-\, \frac{(\mathbf{u}^T\mathbf{X}\mathbf{v})^2}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2} \ ,$$

We call the rightmost term a "bi-Rayleigh quotient" as it is the ratio of two bi-quadratics, in analogy to the usual Rayleigh quotient which is the ratio of two quadratics. Maximization of the bi-Rayleigh quotient is equivalent to minimization of the LS criterion. The stationary solutions of a bi-Rayleigh quotient are the pairs of left- and right-singular vectors of a singular value problem, just as the stationary solutions of a plain Rayleigh quotient are the eigenvectors of an eigen problem.

The above exercise of scale optimization can be as easily executed for the penalized criteria $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ of Section 2. Their quadratic functions $C^2 - 2Br + A^2r^2$ only differ in the coefficient $A^2$ which in each case is the sum of $\|\mathbf{u}\|^2\|\mathbf{v}\|^2$ and the penalties. The results are shown on the second line of each box of Table 1.

The naive penalized LS criterion $\mathcal{C}_1(\mathbf{u}, \mathbf{v}) = \|\mathbf{X} \,-\, \mathbf{u}\mathbf{v}^T\|_F^2 \,+\, \mathbf{u}^T\boldsymbol{\Omega}_u\mathbf{u} \,+\, \mathbf{v}^T\boldsymbol{\Omega}_v\mathbf{v}$ requires separate treatment as it is the only one that is not scale invariant, so that the slopes $s$ and $t$ do not coalesce into a product $r$. Here are the stationary equations:

$$\begin{aligned}\frac{\partial}{\partial s}\mathcal{C}_1(s\mathbf{u}, t\mathbf{v}) &= \,-\, 2t\mathbf{u}^T\mathbf{X}\mathbf{v} \,+\, 2st^2\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \,+\, 2s\mathbf{u}^T\boldsymbol{\Omega}_u\mathbf{u} \;=\; 0 \\ \frac{\partial}{\partial t}\mathcal{C}_1(s\mathbf{u}, t\mathbf{v}) &= \,-\, 2s\mathbf{u}^T\mathbf{X}\mathbf{v} \,+\, 2s^2t\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \,+\, 2t\mathbf{v}^T\boldsymbol{\Omega}_v\mathbf{v} \;=\; 0\end{aligned}$$

Multiplying the first equation with $s$ and the second with $t$, one recognizes immediately that

$$s^2\,\mathbf{u}^T\boldsymbol{\Omega}_u\mathbf{u} \;=\; t^2\,\mathbf{v}^T\boldsymbol{\Omega}_v\mathbf{v} \ ,$$

forcing the two penalties to be equal at the minimum. Using this identity, one can express the combined penalty symmetrically as follows:

$$s^2\,\mathbf{u}^T\boldsymbol{\Omega}_u\mathbf{u} \,+\, t^2\,\mathbf{v}^T\boldsymbol{\Omega}_v\mathbf{v} \;=\; 2st\,\left(\mathbf{u}^T\boldsymbol{\Omega}_u\mathbf{u}\cdot\mathbf{v}^T\boldsymbol{\Omega}_v\mathbf{v}\right)^{1/2} .$$

Thus after scale minimization the combined penalty becomes invariant under scale changes (1). Using this fact one obtains the scale-minimized solution without problem; the result is shown in the second box of Table 1. These calculations show that, counter to appearances, $\mathcal{C}_1$ does not specialize to a form of one-way regularized SVD when one of the penalties vanishes.

Comparing the bi-Rayleigh quotients across the penalized criteria, we see that the naive criterion $\mathcal{C}_1$ is somewhat reminiscent of the regularized PCA proposal by Rice and Silverman (1991) in that it penalizes by subtracting from the numerator rather than adding to the denominator. Criteria $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$ share the property that they penalize the denominator, and in this regard they are all reminiscent of the regularized PCA proposal by Silverman (1996). We also note that only the last, $\mathcal{C}_4$, has a denominator that factors similar to the unpenalized case $\mathcal{C}_0$, providing a heuristic argument in its favor. As a comparison, the numerator of the bi-Rayleigh quotient for Criterion $\mathcal{C}_1$ is not of the simple biquadratic type, and the denominators for $\mathcal{C}_2$ and $\mathcal{C}_3$ do not factorize into a simple product of quadratics.

# 4 Comparison of criteria in terms of alternating algorithms

Table 1 also shows the stationary equations of the (penalized) LS criteria, solved for $\mathbf{u}$ and $\mathbf{v}$, respectively, cast in terms of smoother matrices and (plain) Rayleigh coefficients:

$$\mathbf{S}_u(\lambda) = (\mathbf{I} + \lambda\mathbf{\Omega}_u)^{-1}, \qquad \mathcal{R}_u(\mathbf{u}) = \frac{\mathbf{u}^T\mathbf{\Omega}_u\mathbf{u}}{\|\mathbf{u}\|^2}, \tag{9}$$

$$\mathbf{S}_v(\lambda) = (\mathbf{I} + \lambda\mathbf{\Omega}_v)^{-1}, \qquad \mathcal{R}_v(\mathbf{v}) = \frac{\mathbf{v}^T\mathbf{\Omega}_v\mathbf{v}}{\|\mathbf{v}\|^2}; \tag{10}$$

see the third and fourth lines in each box of Table 1. The bandwidth parameters $\lambda$ are necessary to accommodate criteria $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ whose stationary equations apparently require variable bandwidths whereby the bandwidth for updating $\mathbf{u}$ depends on $\mathbf{v}$, and vice versa. Such coupling of the bandwidths is a heuristic argument against these criteria. Among the penalized criteria, $\mathcal{C}_4$ is the only one in whose stationary equations the bandwidths are fixed. The stationary equations immediately suggest alternating algorithms for $\mathcal{C}_4$: update $\mathbf{u}$ from $\mathbf{v}$, then update $\mathbf{v}$ from $\mathbf{u}$.

# References

Hastie T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Huang, J. Z., Shen, H. and Buja, A. (2008). Functional principal components analysis via penalized rank one approximation. *Electronic Journal of Statistics*, **2**, 678–695.

Huang, J. Z., Shen, H. and Buja, A. (2009). The analysis of two-way functional data using two-way regularized singular value decompositions. *Journal of American Statistical Association*, revised.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B*, **53**, 233–243.

Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, **24**, 1–24,