Statistica Sinica **18**(2008), 299-311

IMPROVING DIMENSION REDUCTION VIA CONTOUR-PROJECTION

Hansheng Wang[1], Liqiang Ni[2] and Chih-Ling Tsai[1,3]

[1]*Peking University,* [2]*University of Central Florida and* [3]*University of California-Davis*

Abstract: Most sufficient dimension reduction methods hinge on the existence of finite moments of the predictor vector, a characteristic which is not necessarily warranted for every elliptically contoured distribution as commonly encountered in practice. Hence, we propose a *contour-projection* approach, which projects the elliptically distributed predictor vector onto a unit contour, which shares the same shape as the predictor density contour. As a result, the projected vector has finite moments of any order. Furthermore, *contour-projection* yields a hybrid predictor vector, which encompasses both the direction and length of the original predictor vector. Therefore, it naturally leads to a substantial improvement on many existing dimension reduction methods (e.g., sliced inverse regression and sliced average variance estimation) when the predictor vector has a heavy-tailed distribution. Numerical studies confirm our theoretical findings.

Key words and phrases: Contour-projection, dimension reduction, linearity condition, sliced average variance estimation, sliced inverse regression.

1. Introduction

High-dimensional data frequently encountered in applications pose a serious challenge for parameter estimation and statistical inference, especially in the presence of unknown nonlinear structures. One plausible solution is to reduce the dimension without going through a parametric or nonparametric model fitting process. To this end, Li (1991) introduced seminal work on *sliced inverse regression* (SIR). Since then, various dimension reduction methods have been proposed (e.g., *sliced average variance estimation* (SAVE, Cook and Weisberg (1991)), *principal Hessian directions* (pHd, Li (1992)), and most recently *contour regression* (CR, Li, Zha and Chiaromonte (2005)). See Cook and Ni (2005) for a recent review.

A common feature of the aforementioned dimension reduction methods is that they are model-free, with assumptions on the marginal distribution of the predictor $x \in \mathcal{R}^p$ instead of on the conditional distribution of the response $y|x$. To exploit this information via the inverse regression approach, one often assumes that the predictor is elliptically distributed. However, this assumption does not guarantee that the predictor has the finite first or second moments required for the inverse regression or other dimension reduction methods (e.g., SIR, SAVE, and pHd). Even if those moments do exist, the heavy-tailed predictor may still seriously deteriorate the performance of existing methods. The heavy-tailed distribution has been observed in many natural phenomena, including financial, physical, and sociological variables. Resolving this challenge task is important to further facilitate the use of existing dimension reduction methods.

Observations generated from heavy-tailed distributions are sometimes viewed as outliers (see Hampel, Ronchetti, Rousseeuw and Stahel (1986)). Therefore, one could explore the robust version of SIR for dimension reductions. For example, Gather, Hilker and Becker (2002) discussed the outlier sensitivity of SIR and Prendergast (2005) proposed the detection of influential observations for SIR via influence functions. However, none of them studied dimension reductions for heavy-tailed distributions. This motivated us to propose a *contour-projection* (CP) procedure. Specifically, the CP approach projects the predictor vector onto a unit contour that shares the same shape as the original density contour. The resulting CP-predictor encompasses the direction and length of the original predictor vector and has finite moments of any order. The CP approach enables us to improve parameter estimations for heavy-tailed predictors. To utilize our findings, we focus on two widely used dimension reduction methods, SIR and SAVE.

The rest of this paper is organized as follows. Section 2 presents the basic idea of the *contour-projection* approach and its properties. The utilization of the *contour-projection* for dimension reduction is addressed in Section 3. Simulation studies are carried out in Section 4, and the results show that CP performs well for heavy-tailed predictors. An example is given to illustrate the usefulness of CP. We conclude the paper with a brief discussion in Section 5.

2. Contour-Projection

Let $\mathcal{S}$ denote a linear subspace of $\mathcal{R}^p$ with, for example, $\mathcal{S}(A)$ the linear space spanned by the columns of matrix A. To capture the dependence between y and x, it is assumed that there is $p \times d$ matrix β, for some integer $d > 0$, such that

$$y \perp\!\!\!\perp x | \beta^\top x, \tag{2.1}$$

where " $\perp\!\!\!\perp$ " denotes independence. In other words, $\beta^\top x$ summarizes all information x has about y. We refer to $\mathcal{S}(\beta)$ as the *sufficient dimension reduction subspace* (SDR, Cook (1996)). It is easy to see that any subspace $\mathcal{S} \supset \mathcal{S}(\beta)$ is also a SDR subspace. Therefore, it is only of interest to infer the "smallest" SDR subspace, referred to as the *central subspace* and denoted by $\mathcal{S}_{y|x}$. In the rest of article, we assume that $\mathcal{S}_{y|x}$ exists with a basis being β as defined in (2.1). The notion of *central subspace* is very helpful in exploratory analysis and data visualization. When p is large, sufficient summary plots based on the *central subspace* with dimension $d < p$ can be very informative for statistical modeling. We take $p > 1$ in the rest of the paper.

We assume that the predictor x follows an elliptically contoured (EC) distribution with density (see Muirhead (1982))

$$f_{\mu,\Sigma}(x) = |\Sigma^{-\frac{1}{2}}| f\big(||x-\mu||^2_{\Sigma}\big), \tag{2.2}$$

where $\mu \in \mathcal{R}^{p\times 1}$ and $\Sigma \in \mathcal{R}^{p\times p}$ are the parameters. Here Σ is a positive definite matrix with $tr(\Sigma) = p$ for identifiability, and $||t||^2_{\Sigma} = t^\top \Sigma^{-1} t$ is the corresponding Mahalanobis distance. The elliptically contoured distribution is the most common assumption for the predictor in sufficient dimension reduction literature, partly because it ensures that a so-called *linearity condition* holds for the basis of *central subspace*: $E(x|\beta^\top x)$ is a linear function of $\beta^\top x$ (Li (1991)). The linearity condition connects the *central subspace* with inverse regression, which is one of the main venues for sufficient dimension reduction. It is worth noting that the EC condition implies the linearity condition when second-order moments exist. In addition, the EC condition can still be satisfied when the second-order moment of the predictor does not exist. Moreover, even if the actual distribution departs from (2.2), it can be improved by either re-weighting methods (Cook and Nachtsheim (1994)) or through the multivariate Box-Cox transformation (Quiroz, Nakamura and Pérez (1996)). A classical vs. robust Mahalanobis distance plot (DD plot) can serve as a diagnostic tool for elliptical symmetry (Olive (2002)).

The idea of the *contour-projection* is simple. Given μ and Σ, $\mathcal{C} = \{x : ||x-\mu||^2_{\Sigma} = 1\}$ defines a unit contour, that shares the same shape as the predictor density in (2.2). For any predictor x, the CP procedure constructs a new predictor $\tilde{x} = (x-\mu)/r$, by projecting x onto the unit contour $\mathcal{C}$, where $r = \|x-\mu\|_{\Sigma}$. The resulting $\tilde{x}$ is bounded with finite moments of any order. Lemma 1 gives some important properties of $\tilde{x}$ and r.

Lemma 1. *Assume x has the density function* (2.2). *Then* (1) $E(\tilde{x}) = 0$ *and* $cov(\tilde{x}) = p^{-1}\Sigma$; (2) *if* $\xi \in \mathcal{R}^{p\times q}$, $q < p$, *is any orthogonal matrix such that* $\xi^\top \xi = I_q$, *we have* $E(\tilde{x}|\xi^\top \tilde{x}) = A\xi^\top \tilde{x}$, *where* $A = \Sigma\xi(\xi^\top\Sigma\xi)^{-1}$; (3) $\tilde{x}$ *and* r *are independent.*

Proof. (1) Define $u = \Sigma^{-1/2}(x-\mu) = (u_1, \ldots, u_p)^\top$, with density given by $f(||u||^2)$. Then, $\tilde{x} = \Sigma^{1/2}u/||u||$. Because the u_i's are exchangeable, we have $\mathrm{E}(u_i|||u||) = 0$, $\mathrm{E}(u_i^2|||u||) = \mathrm{E}(u_j^2|||u||)$, and $\mathrm{E}(u_iu_j|||u||) = 0$ for the given $||u||$

when $i \neq j$. These results together with $trace\{\mathrm{E}(uu^\top/||u||^2)\} = 1$ imply that $\mathrm{E}(\tilde{x}) = \Sigma^{1/2}\mathrm{E}\,(u/||u||) = 0$ and $\mathrm{E}\{\tilde{x}\tilde{x}^\top\} = \Sigma^{1/2}\mathrm{E}\,(uu^\top/||u||^2)\,\Sigma^{1/2} = p^{-1}\Sigma$.
(2) Let $\gamma = \Sigma^{1/2}\xi$, $P_\gamma = \gamma(\gamma^\top\gamma)^{-1}\gamma^\top$ be the orthogonal projection onto the subspace spanned by γ, and $Q_\gamma = \mathrm{I}_p - P_\gamma$. We then have $\mathrm{E}(Q_\gamma u|P_\gamma u, \|u\|) = 0$ and

$$\begin{aligned}\mathrm{E}\Big(\frac{u}{\|u\|}\Big|P_\gamma\frac{u}{\|u\|}\Big) &= \mathrm{E}\Big\{(P_\gamma + Q_\gamma)\frac{u}{\|u\|}\Big|P_\gamma\frac{u}{\|u\|}\Big\}\\ &= P_\gamma\frac{u}{\|u\|} + \mathrm{E}\Big\{Q_\gamma\frac{u}{\|u\|}\Big|P_\gamma\frac{u}{\|u\|}\Big\}\\ &= P_\gamma\frac{u}{\|u\|} + \mathrm{E}\Big\{\|u\|^{-1}\mathrm{E}\big(Q_\gamma u\big|P_\gamma u, \|u\|\big)\Big|P_\gamma\frac{u}{\|u\|}\Big\}\\ &= P_\gamma\frac{u}{\|u\|}.\end{aligned}$$

Therefore, $\mathrm{E}\{\tilde{x}|\Sigma^{1/2}\xi(\xi^\top\Sigma\xi)^{-1}\xi^\top\tilde{x}\} = \Sigma\xi(\xi^\top\Sigma\xi)^{-1}\xi^\top\tilde{x}$. Consequently, we have $\mathrm{E}(\tilde{x}|\xi^\top\tilde{x}) = A\xi^\top\tilde{x}$ since $\Sigma^{1/2}\xi(\xi^\top\Sigma\xi)^{-1} \in \mathcal{R}^{p\times q}$ is a full column rank matrix.
(3) This is a direct result of Theorem 1 in Cambanis, Huang and Simons (1981).

Remark 1. Lemma 1 shows that Li's (1991) linearity condition holds for the CP predictor. It also suggests that $\tilde{x}$ itself may contribute useful information on the *central subspace*, while the role of the length of the predictor x can be secondary (see Section 3 for detailed illustrations).

In practice, both μ and Σ are unknown and need to be estimated from the data. To this end, (Tyler (1987, p.245)) proposed simultaneous M-estimators for μ and Σ that coincide with those from a simple iterative estimating procedure motivated by Lemma 1. Specifically, let $\mu^{(m)}$ and $\Sigma^{(m)}$ be the estimates obtained after the m'th iteration. Because $\mathrm{E}(\tilde{x}) = \mathrm{E}((x-\mu)/(||x-\mu||_\Sigma)) = 0$ suggests that, $\mu^{(m+1)}$ can be obtained, based on $(\mu^{(m)}, \Sigma^{(m)})$, as

$$\mu^{(m+1)} = \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{||x_i - \mu^{(m)}||_{\Sigma^{(m)}}}\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{||x_i - \mu^{(m)}||_{\Sigma^{(m)}}}\Big). \tag{2.3}$$

Subsequently, using the fact that $\mathrm{cov}(\tilde{x}) = p^{-1}\Sigma$, $\Sigma^{(m+1)}$ can be obtained by first computing

$$\Sigma^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \mu^{(m)})(x_i - \mu^{(m)})^\top}{\|x_i - \mu^{(m)}\|^2_{\Sigma^{(m)}}}, \tag{2.4}$$

and then re-scaling $\Sigma^{(m+1)}$ so that its trace equals p.

It seems natural to obtain parameter estimators of μ and Σ by iterating (2.3) and (2.4) until the algorithm converges. However, Tyler (1987) pointed out that the convergence and consistency of simultaneous fixed-point solutions is still an open question. Under some regularity conditions, Tyler (1987) showed that if $\mu^{(m)}$ is a consistent estimate of μ, the fixed point solution of (2.4) uniquely exists and produces a consistent estimate of Σ (see Frahm (2004) for detailed discussions). This motivates us to propose a hybrid estimator that combines the fixed-point solution and one-step reweighting algorithm given below.

1. Let $\hat{\mu}_0$ be the componentwise median and $\hat{\Sigma}_0$ the sample covariance matrix.
2. Let $\mu^{(m)} = \hat{\mu}_0$ in (2.4). The fixed point solution is $\hat{\Sigma}_1$.
3. Plug $(\hat{\mu}_0, \hat{\Sigma}_1)$ into (2.3) to obtain $\hat{\mu}_1$.
4. Let $\mu^{(m)} = \hat{\mu}_1$ in (2.4). The fixed point solution is $\hat{\Sigma}_2$.
5. Plug $(\hat{\mu}_1, \hat{\Sigma}_2)$ into (2.3) to obtain $\hat{\mu}_2$.

The above procedure not only avoids the convergence issue of (2.3) but also facilitates the computation. Furthermore, the consistent estimate $\hat{\mu}_0$ of μ leads to the fact that $\hat{\mu}_2$ and $\hat{\Sigma}_2$ are consistent. They perform well in simulation studies (see Section 4).

Remark 2. In multivariate distributions, there is considerable literature on robust location and dispersion estimators (e.g., minimum covariance determinant (MCD) and minimum volume ellipsoid (MVE) methods, see Maronna and Zamar (2002)). In addition, the function *cov.rob* in the package *MASS* (Venables and Ripley (2006)) of R provides robust estimators of location and dispersion parameters. Once those estimators' theoretical and computational issues have been resolved, they may also be employed to estimate μ and Σ in *contour-projection.*

3. Dimension reduction via contour-projection

To make *contour-projection* an effective SDR technique for regression, we build up a connection between the dimension reduction subspaces before and after *contour-projection* via the following lemma.

Lemma 2. *Suppose ξ is a basis of $\mathcal{S}_{y|x}$. Then, $y \perp\!\!\!\perp \tilde{x} | \xi^\top \tilde{x}$.*

Proof. Let F be a generic cumulative distribution function. Suppose ξ is a basis of $\mathcal{S}_{y|x}$. Then $y \perp\!\!\!\perp x | \xi^\top x$, which is equivalent to $y \perp\!\!\!\perp (\tilde{x}, r) | \xi^\top x$. Moreover, $y \perp\!\!\!\perp (\tilde{x}, r) | (\xi^\top \tilde{x}, r)$, which leads to $y \perp\!\!\!\perp \tilde{x} | (\xi^\top \tilde{x}, r)$. Hence, for any scalar a,

$$F_{y|\tilde{x}}(a) = \mathrm{E}_r[F_{y|(\tilde{x},r)}(a)] = \mathrm{E}_r[F_{y|(\xi^\top \tilde{x},r)}(a)] = F_{y|\xi^\top \tilde{x}}(a).$$

Consequently, $y \perp\!\!\!\perp \tilde{x} | \xi^\top \tilde{x}$ (compare Cook (1998, pp. 63, 64 and 106)), and the proof is completed.

Remark 3. Lemma 2 indicates that one may apply the CP predictor $\tilde{x}$ to gain information of the *central subspace* $\mathcal{S}_{y|x}$. To further elucidate this finding, we consider the multi-index model

$$y = g(\beta^\top x, \epsilon), \tag{3.1}$$

where $g(\cdot)$ is some unknown function and ϵ is a random noise independent of x. Because $x = \mu + r\tilde{x}$, (3.1) can be expressed as

$$y = \tilde{g}(\beta^\top \tilde{x}, r, \epsilon), \tag{3.2}$$

for some unknown function $\tilde{g}$, where (ϵ, r) is independent of $\tilde{x}$. Therefore, it is possible to find β via (3.2) rather than (3.1).

Remark 4. For the sake of simplicity, one can exclude r from (3.2) to compute β. However, this may lead to an identification problem. For example, assume that x follows a p-dimensional standard normal distribution with $\Sigma = I$, $p > 2$, and

$$y = \sum_{i=2}^{p} x_i^2 + \epsilon = r^2 \sum_{i=2}^{p} \tilde{x}_i^2 + \epsilon = r^2 - x_1^2 + \epsilon = r^2(1 - \tilde{x}_1^2) + \epsilon, \tag{3.3}$$

where ϵ is an independent noise. As a result, both $\mathcal{S}(e_1)$ and its orthogonal complement $\mathcal{S}_{y|x}$ are dimension reduction subspaces in the $\tilde{x}$ scale, where $e_1 \in \mathcal{R}^p$ is a vector with first component 1, and the others 0. Therefore, we recommend including r in the analysis and also comparing the results from dimension reduction methods with and without contour projection.

We explore applications of SIR and SAVE on dimension reduction via the CP predictor. In regression analysis, Li (1991) proposed a sliced inverse regression approach to dimension reduction, which implicitly assumes that the first two moments of x exist. In addition, SIR requires the linearity condition: $\mathrm{E}(x|P_\beta x)$ is a linear function of $P_\beta x$, where β is a basis of $\mathcal{S}_{y|x}$. To find the *central subspace*, Li (1991) considered the standardized predictor z with zero mean and identity covariance matrix. Under the linearity condition, $\mathrm{E}(z|y) \in \mathcal{S}_{y|z}$, thus $M_{sir} = \mathrm{cov}(\mathrm{E}(z|y)) \subseteq \mathcal{S}_{y|z}$. When y is discrete, it is straightforward to obtain a sample version of M_{sir}. If y is continuous, we discretize y first by partitioning the range of y to a few slices. If $dim(\mathcal{S}_{y|z}) = d$, then the d eigenvectors corresponding to the largest d eigenvalues of $\hat{M}_{sir}$ constitute an estimated basis of $\mathcal{S}_{y|z}$. Then we transform the *central subspace* back to the original x-scale via $\mathcal{S}_{y|x} = \mathrm{cov}(x)^{-1/2}\mathcal{S}_{y|z}$.

In practice, however, an elliptically contoured predictor may not have a finite second moment even though it satisfies the linearity condition. In this case, SIR is not applicable. This motivated us to employ *contour-projection* since the CP predictor has finite moments of any order. Based on $\tilde{x}$ and assuming β is an orthogonal basis, we then apply the linearity condition given in Lemma 1 to obtain

$$\mathrm{E}(\tilde{x}|y) = \mathrm{E}[\mathrm{E}(\tilde{x}|(\beta^\top \tilde{x}, y))|y] = \mathrm{E}[\mathrm{E}(\tilde{x}|\beta^\top \tilde{x})|y] = \Sigma\beta(\beta^\top\Sigma\beta)^{-1}\mathrm{E}[\beta^\top \tilde{x}|y].$$

Under (2.2), $\Sigma^{-1}\mathrm{E}(\tilde{x}|y) \in \mathcal{S}_{y|x}$, which suggests a direct application of SIR on $\tilde{x}$ is warranted, and this is referred to as *contour-projected* SIR (CP-SIR). Accordingly, the non-identifiability of $\mathcal{S}_{y|\tilde{x}}$ does not affect the validity of CP-SIR.

In the last decade, SIR has been used successfully in many applications. However, it fails in a strictly symmetric case, e.g., $y = (\beta^\top z)^2 + \epsilon$, where (z, ϵ) are multivariate independent standard normal. To this end, Cook and Weisberg (1991) proposed SAVE, which may recover some of the information overlooked by SIR. Let η be a basis of $\mathcal{S}_{y|z}$. Then, for the elliptically distributed predictor, Cook and Weisberg (1991) found that

$$w_y \mathrm{~I}_p - \mathrm{cov}(z|y) = P_\eta[w_y \mathrm{~I}_p - \mathrm{cov}(z|y)]P_\eta,$$

where the scalar w_y is a function of y that depends on the distribution of x. Since $\mathrm{E}(w_y) = 1$, w_y varies for different values of y about 1. Empirical results indicate that a SAVE kernel matrix $M_{save} = \mathrm{E}[(\mathrm{~I}_p - \mathrm{cov}(z|y))^2]$ can be utilized to estimate $\mathcal{S}_{y|x}$ in a way similar to the procedure for M_{sir} (see Cook and Weisberg (1991)).

To employ the CP predictor $\tilde{x}$, take $v = \Sigma^{-1/2}\tilde{x}$ and let $\gamma = \Sigma^{1/2}\beta$ denote a basis for a dimension reduction subspace of the regression of y on v. Since v is uniformly distributed on a unit hypersphere, we have $\mathrm{E}(v|y) \in \mathcal{S}(\gamma)$, $\mathrm{E}(v|P_\gamma v) = P_\gamma v$, and $\mathrm{cov}(v|P_\gamma v) = (1 - \|P_\gamma v\|^2)Q_\gamma/(p-d)$. It is easy to see that $\mathrm{cov}(v|y) = u_y Q_\gamma + P_\gamma \mathrm{cov}(v|y) P_\gamma$, or $u_y \mathrm{~I}_p - \mathrm{cov}(v|y) = P_\gamma[u_y \mathrm{~I}_p - \mathrm{cov}(v|y)]P_\gamma$, where $u_y = \mathrm{E}(1 - \|P_\gamma v\|^2|y)/(p-d)$ and $\mathrm{E}(u_y) = 1/p$. This motivates us to adopt the SAVE approach to estimate the *central subspace* via a kernel matrix $\mathrm{E}[(\mathrm{~I}_p/p - \mathrm{cov}(v|y))^2]$, which is equivalent to applying SAVE to $\tilde{x}$. We call this procedure *contour-projected* SAVE (CP-SAVE).

4. Simulations and an Example

In this section, we employ the hybrid estimation procedure introduced in Section 2 to estimate location and dispersion parameters. Furthermore, we adopt Weisberg's (2002) *dr* package in *R* to implement SIR and SAVE and to make comparisons.

4.1. Comparison of estimates

Consider the two models given by

$$\text{Model I:} \quad y = \frac{x_1}{0.5 + (x_2 + 1.5)^2} + 0.2\epsilon,$$

$$\text{Model II:} \quad y = (x_1 + 0.5)^2 + x_2 + 0.2\epsilon,$$

where ϵ is a standard normal distribution. For each model, we generated 500 data sets with the sample sizes n = 100, 200, 300, and 400. Furthermore, $x = (x_1, \ldots, x_{10})' = w/\sqrt{v_{df}/df}$, where $w \in \mathcal{R}^{10}$ is a standard multivariate normal distribution and v_{df} is a chi-square distribution with degrees of freedom df (df=1, 3, 5, ∞). As a result, x follows a multivariate t distribution (see Lange, Little, and Taylor (1989)) and we have situations where the tail of the predictor is extremely heavy so that moments do not exist ($df = 1$); the predictor has a heavy tail with finite first order moment ($df = 3$); the predictor has a heavy tail with finite second order moment ($df = 5$); and the tail of the predictor is not heavy at all ($df = \infty$ corresponds to a multivariate normal distribution of x_i).

The *central subspace* associated with both models is given by $\mathcal{S}(\beta)$, where $\beta = \{e_1, e_2\}$ and $e_l \in \mathcal{R}^p$ is a vector with l-th component 1, and the others 0. For each data set, SIR, CP-SIR, SAVE, and CP-SAVE were applied with the number of slices set at 5. Without loss of generality, let $\hat{\beta}$ denote the estimated basis obtained from any of the four inverse regression approaches. The estimation accuracy of the *central subspace* is measured by $\Delta = \|P_{\hat{\beta}} - P_\beta\|$ (see Li, Zha and Chiaromonte (2005)), where $\|\cdot\|$ is the maximum singular value of a matrix.

For Model I, Table 1 reports the average Δ for estimates obtained from SIR and CP-SIR, respectively, in 500 realizations. Although SIR deteriorates with heavier tails, CP-SIR holds its accuracy. Because the constant 1.5 in the denominator of Model I is relatively large, the results from SAVE methods are less satisfactory than those in Table 1. In this situation, however, the unreported results showed that CP-SAVE is superior to SAVE.

Table 1. The average of Δ from 500 realizations for model I when the predictors are generated from multivariate t distributions.

sample	$df = 1$		$df = 3$		$df = 5$		$df = \infty$	
size	sir	cp-sir	sir	cp-sir	sir	cp-sir	sir	cp-sir
100	0.914	0.691	0.655	0.558	0.591	0.550	0.519	0.523
200	0.909	0.492	0.526	0.391	0.425	0.379	0.355	0.355
300	0.914	0.413	0.459	0.316	0.364	0.309	0.287	0.289
400	0.904	0.361	0.427	0.280	0.316	0.264	0.246	0.244

As discussed in Section 3, SAVE is usually more comprehensive than SIR in the estimation of the *central subspace*. For example, if the constant 1.5 in Model I is closer to 0, SIR is less capable of detecting the direction associated with x_2, while SAVE still can do this. Model II is another example where SAVE can be sharper than SIR. Table 2 reports the average Δ for estimates obtained from

SAVE and CP-SAVE for Model II, which corroborates the benefits of *contour-projection*. Next, we show the effect of contour projection on parameter inference.

Table 2. The average of Δ from 500 realizations for model II when the predictors are generated from multivariate t distributions.

sample	$df = 1$		$df = 3$		$df = 5$		$df = \infty$	
size	save	cp-save	save	cp-save	save	cp-save	save	cp-save
100	0.949	0.956	0.942	0.938	0.943	0.937	0.940	0.937
200	0.939	0.822	0.928	0.689	0.900	0.660	0.573	0.613
300	0.941	0.627	0.912	0.448	0.832	0.426	0.360	0.406
400	0.946	0.461	0.912	0.313	0.807	0.294	0.283	0.298

4.2. Comparison of inferences

Suppose that $\hat{\eta}_i$ is the i-th eigenvector of $\hat{M}_{sir}$ with corresponding eigenvalue $0 < \hat{\lambda}_i < 1$. Then $\hat{\beta}_i = \text{cov}(x)^{-1/2}\hat{\eta}_i$ is the estimate of β_i (the i-th column of β) in the *central subspace* $\mathcal{S}_{y|x}$. Under the null hypothesis, $H_0 : e_l^\top \beta_i = 0$, Chen and Li (1998) showed that the test statistic $n(e_l^\top \hat{\beta}_i)^2(\hat{\lambda}_i/(1-\hat{\lambda}_i))[e_l^\top \widehat{\text{cov}}(x)^{-1} e_l]^{-1}$ follows an asymptotic chi-squared distribution with 1 degree of freedom. The above result can be analogously applied to CP-SIR.

To make comparisons, we revisit Model I. For the sake of simplicity, we only consider x to be the multivariate t distribution with $df = 3$ and sample size $n = 300$. Under the null hypothesis $H_0 : e_3^\top \beta_1 = 0$, the p-values should be uniformly distributed. Figure 1 depicts the uniform quantile-quantile plot of 500 p-values obtained from SIR and CP-SIR, respectively. Apparently, the CP-SIR performs better than SIR. When x has $df = 1$, the contrast is more conspicuous than that of $df = 3$. But we do not show that here.

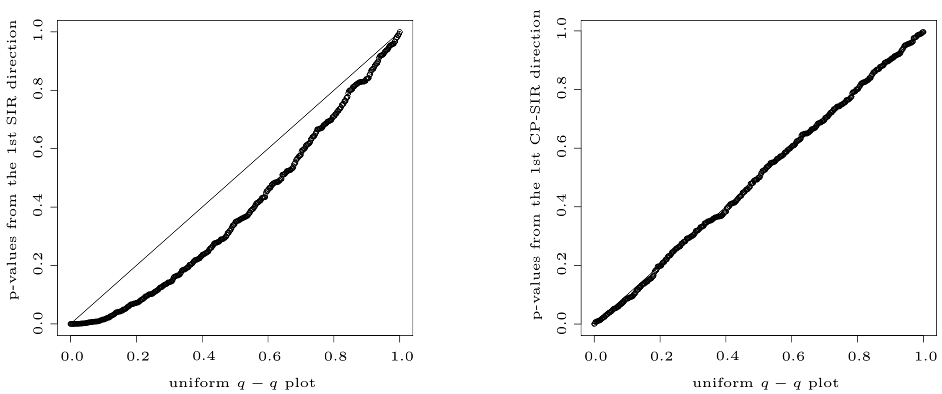

Figure 1. Uniform quantile plots of p-values for testing $H_0 : e_3^\top \beta_1 = 0$ obtained from the SIR and CP-SIR approaches with 5 slices.

Analogously, we conducted the test to assess the null hypothesis, $H_0 : e_3^\top \beta_2 = 0$, and obtained the similar results to those given above. Next, we assess the sensitivity of the contour projection against non-elliptical predictors.

4.3 Non-elliptical predictors

Although SIR and SAVE are not particularly sensitive to the violation of the elliptical assumption, it is still of interest to evaluate the performance of CP under a non-elliptically symmetric distribution. To this end, we consider the same simulation settings as those in Section 4.1, except that the random variable w was generated from a uniform distribution on $[-\sqrt{3}, \sqrt{3}]^{10}$. The simulation results are similar to those in Section 4.1. For example, Table 3 shows that CP-SIR is superior to SIR for heavy-tailed predictors. Here CP is not sensitive to the elliptical assumption.

Table 3. The average of Δ from 500 realizations for model I when the predictors are generated from multivariate uniform distributions.

sample	$df = 1$		$df = 3$		$df = 5$		$df = \infty$	
size	sir	cp-sir	sir	cp-sir	sir	cp-sir	sir	cp-sir
100	0.896	0.581	0.553	0.435	0.443	0.410	0.373	0.393
200	0.895	0.393	0.439	0.298	0.328	0.280	0.254	0.263
300	0.907	0.327	0.396	0.243	0.275	0.226	0.203	0.212
400	0.899	0.288	0.354	0.208	0.241	0.195	0.180	0.186

4.4 An example

To illustrate the effectiveness of the CP approach, we consider the data set created by CCER (Center for China Economic Services). The CCER database is one of most authoritative and widely used stock market databases on the China stock market (*http://www.ccerdata.com/*). The objective of this study is to derive an index that can be used for predicting the firm's next year's earning. To this end, the response variable (y_i) is taken as next year's *return on equity* (ROEy), while the predictor vector (x_i) includes the firm's current year's *return on equity* (ROE,), *log-transformed asset* (ASSET), *profitability margin* (PM), *sales growth rate* (SALES), *leverage level* (LEV), and *asset turnover ratio* (ATO). In addition, the data contain yearly information about firms from 1995 to 2004. The sample size of firms per year ranges from 283 (Year 1995) to 1,172 (Year 2003). Moreover, the kurtosis measures indicate that all predictors have very heavy tailed distributions.

For each of nine years' data sets, we employ both SIR and CP-SIR with the same number of slices (5) as used in the simulation studies. For the sake of illustration, we only reported the most important direction estimate identified by SIR and CP-SIR, respectively, in each year. As a result, we obtained the nine most important direction estimates of SIR and CP-SIR. Figure 2 depicts the absolute value of the estimated coefficient for each of the six predictors across nine years of data sets. We found that the CP-SIR estimates yield much less variabilities than those of SIR estimates (except for the predictor PM). In sum, CP-SIR should be considered when the predictor has a heavy-tailed distribution.

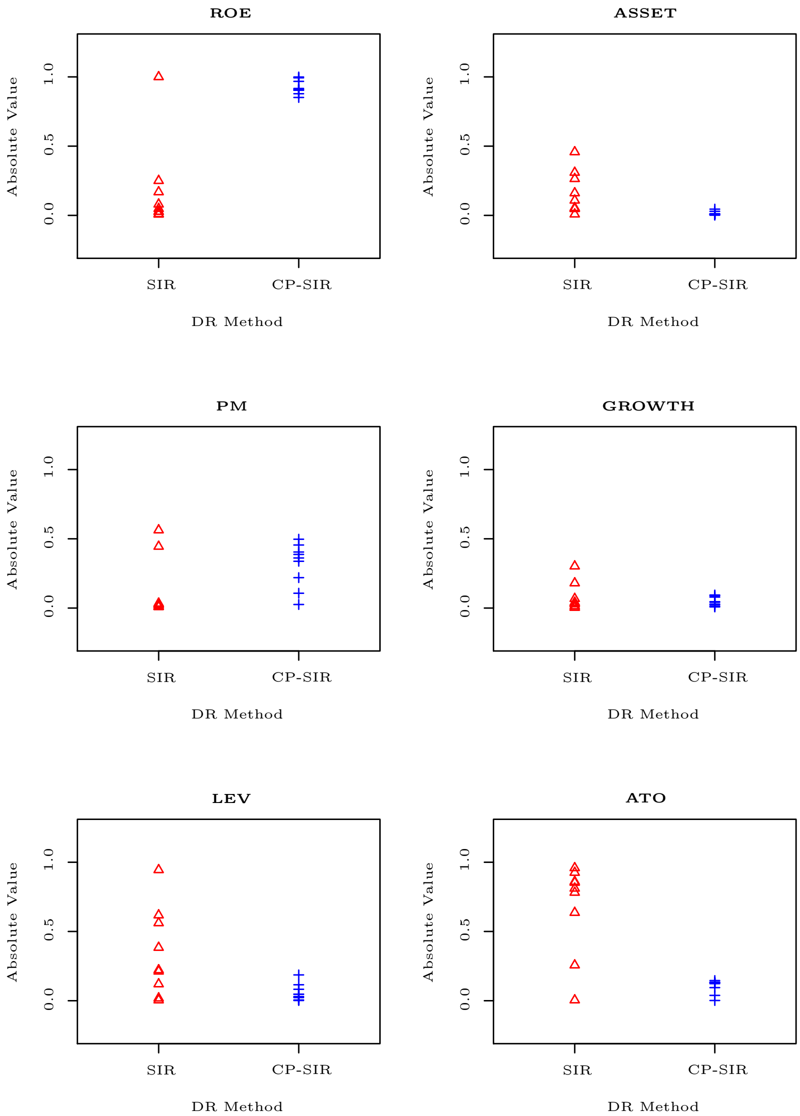

Figure 2. The stability comparison between the associated SIR and CP-SIR estimates of each predictor across nine years' data sets.

5. Conclusion

In this article, we propose a *contour-projection* procedure and then employ it to improve inverse regression methods for dimension reduction. We also present theoretical properties and Monte Carlo results, showing that SIR and SAVE can be improved substantially via the CP approach. However, this improvement is not limited to SIR and SAVE. We may extend the application of CP to other inverse regression approaches if the predictor has an elliptically contoured distribution. Possible gains of CP over recently developed approaches (e.g., Xia et al. (2002) and Yin and Cook (2005)) are under consideration.

Acknowledgements

We are grateful to the Editor, an associate editor, and the referees for their valuable comments and constructive suggestions, which led to a substantial improvement of this manuscript. We are also grateful to Dr. Gabriel Frahm for some helpful discussions. Hansheng Wang's research is partially supported by National Natural Science Foundation of China (Grant 10771006).

References

Cambanis, S., Huang, S. and Simons, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11**, 368-385.

Chen, C. H. and Li, K. C. (1998). Can SIR be as popular as the multiple linear regression? *Statist. Sinica.* **8**, 289-316.

Cook, R. D. (1996). Graphics for regression regressions with a binary response. *J. Amer. Statist. Assoc.* **91**, 983-992.

Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regression Through Graphics.* Wiley, New York.

Cook, R. D. and Nachtsheim, C. J. (1994). Reweighting to achieve elliptically contoured covariates in regression. *J. Amer. Statist. Assoc.* **89**, 592-599.

Cook, R. D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. *J. Amer. Statist. Assoc.* **100**, 410-428.

Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression for dimension reduction". *J. Amer. Statist. Assoc.* **86**, 328-332.

Frahm, G. (2004). Generalized elliptical distributions: theory and applications. Ph.D. thesis, University of Cologne, Faculty of Management, Economics, and Social Sciences, Department of Statistics, Germany. Retrievable from http://kups.ub.uni-koeln.de/volltexte/2004/1319/.

Gather, U., Hilker, T. and Becker, C. (2002). A note on outlier sensitivity of sliced inverse regression. *Statisitcs* **36**, 271-281.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics-The Approach Based on Influence.* Wiley, New York.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755-770.

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using t distribution. *J. Amer. Statist. Assoc.* **84**, 881-896.

Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *Ann. Statist.* **33**, 1580-1616.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**, 316-327.

Li, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87**, 1025-1039.

Maronna, R. A. and Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional data sets. *Technometrics* **44**, 307-317.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

Olive, D. J. (2002). Applications of robust distances for regression. *Technometrics.* **44**, 64-67.

Prendergast, L. A. (2005). Influence functions for sliced inverse regression. *Scand. J. Statist.* **32**, 385-404.

Quiroz, A. J., Nakamura, M. and Pérez, F. J. (1996). Estimation of a multivariate Box-Cox transformation to elliptical symmetry via the empirical characteristic function. *Ann. Inst. Statist. Math.* **48**, 687-709.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15**, 234-251.

Venables, W. N. and Ripley, B. D. (2006). VR bundle for R. http://www.stats.ox.ac.uk/pub/MASS4/Software.html.

Weisberg, S. (2002). Dimension reduction regression in R. *J. Statist. Soft.* **7**, available from http://www.jstatsoft.org.

Xia, Y., Tong, H., Li, W. K. and Zhu, L.-X.(2002). An adaptive estimation of dimension reduction space. *J. Roy. Statist. Soc. Ser. B* **64**, 363-410.

Yin, X. and Cook, R. D. (2005). Direction estimation in single-index regressions. *Biometrika* **92**, 371-384.

Guanghua School of Management, Peking University, Beijing, P. R. China, 100871.

E-mail: hansheng@gsm.pku.edu.cn

Department of Statistics and Actuarial Science, University of Central Florida, Orlando, FL 32816, U.S.A.

E-mail: lni@mail.ucf.edu

Graduate School of Management, University of California, Davis, CA 95616, U.S.A.

Guanghua School of Management, Peking University, Beijing, P. R. China, 100871.

E-mail: cltsai@ucdavis.edu

(Received December 2005; accepted July 2006)