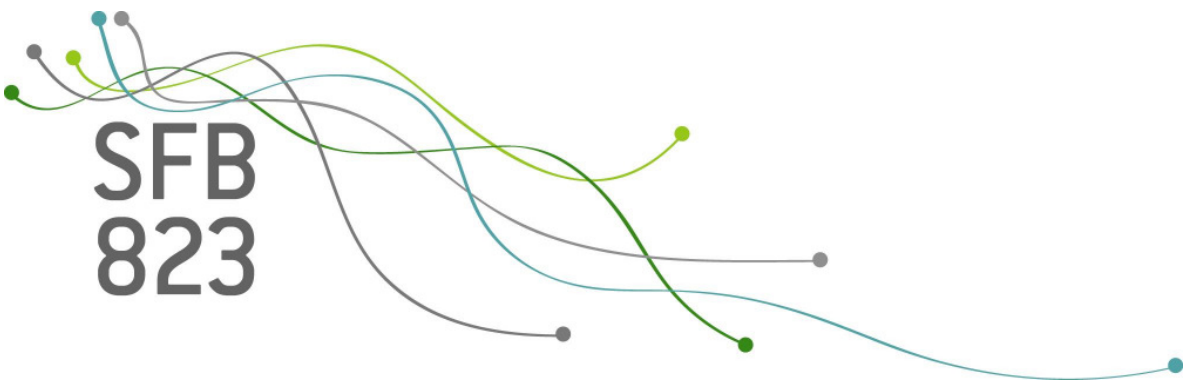
SFB 823

Optimal designs for indirect regression

Stefanie Biedermann, Nicolai Bissantz, Holger Dette, Edmund Jones

Nr. 35/2010

Discussion Paper



Optimal Designs for Indirect Regression

Stefanie Biedermann
University of Southampton
School of Mathematics
Highfield SO17 1BJ, UK
email: s.biedermann@soton.ac.uk

Nicolai Bissantz
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: nicolai.bissantz@rub.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
email: holger.dette@ruhr-uni-bochum.de

Edmund Jones
University of Bristol
Department of Mathematics
Clifton BS8 1TW, UK
email: edmund.jones@bristol.ac.uk

September 8, 2010

Abstract

In many real life applications, it is impossible to observe the feature of interest directly. For example, scientists in Materials Science may be interested in detecting cracks inside objects, not visible from the outside. Similarly, non-invasive medical imaging techniques such as Positrone Emission Tomography rely on indirect observations to reconstruct an image of the patient's internal organs. In this paper, we investigate optimal designs for such indirect regression problems. We determine designs minimizing the integrated mean squared error of estimates of the regression function obtained by Tikhonov or spectral cut-off regularization. We use the optimal designs as benchmarks to investigate the efficiency of the uniform design commonly used in applications. Several examples are discussed to illustrate the results, in most of which the uniform design or a simple modification thereof is demonstrated to be very efficient for the estimation of the regression function. Our designs provide guidelines to scientists regarding the experimental conditions at which the indirect observations should be taken in order to obtain an accurate estimate for the object of interest.

Keywords and Phrases: Indirect Regression, Optimal Design, Uniform Design, Integrated Mean Squared Error Criterion, Tikhonov Regularization, Spectral Cut-off Regularization, Radon Transform.

1 Introduction

Indirect or inverse problems arise in numerous applications such as deconvolution problems [c.f. Fan (1991); Johnstone et al. (2004)], positron emission and X-ray tomography [Johnstone and Silverman (1990); Cavalier (2000, 2001)], Wicksell's problem [Groeneboom and Jongbloed (1995)] and the heat equation [Mair and Ruymgaart (1996)]. The main difference to "classical" inference is that in these models the unknown density or regression function of interest m cannot be observed directly. Such problems have been investigated intensively in the last decades, where most of the work focused on the construction of estimators of m and the determination of their convergence properties with respect to the L^2-risk assuming that m belongs to a certain smoothness class [cf. e.g. Mair and Ruymgaart (1996); Cavalier and Tsybakov (2002)] or their pointwise properties [c.f. Fan (1991); Cavalier (2000); Donoho and Low (1992); Bissantz and Birke (2009)]. In many application areas, e.g. magnetic resonance imaging (MRI), Positron Emission Tomography (PET) or fluorescence microscopy, the data are sampled using a uniform design [see e.g. Shepp and Vardi (1982)].
It is well known in direct regression problems that an optimal design can improve the efficiency of statistical inference substantially and there exists an extensive literature on this subject [see Pukelsheim (2006) or Randall et al. (2007)]. Most authors concentrate on the construction of optimal designs for efficient parameter estimation, where various estimation methods have been considered. Optimal designs for parametric regression models minimizing (integrated) mean squared error criteria have been discussed in Studden (1977), Spruill (1987), Dette and O'Brien (1999) or Broniatowski and Celant (2007) among others. Designs minimizing the integrated mean squared error of a nonparametric estimate in direct regression models have been investigated by Mueller (1984) and Cheng et al. (1998) among others, while Chan (1992) considered optimal designs for variance estimation. More recent work discussed the construction of sequential optimal designs in this context [see Park (2000); Park and Faraway (1998) or Efromovich (2008)].
On the other hand optimal design problems for indirect regression models have found much less attention in the literature so far. Experimental designs have mainly been considered from an empirical point of view in the context of (geo-)physical problems. Among other approaches, Maurer et al. (2000) proposed statistical criteria for the selection of an experimental design for electromagnetic geophysical surveys, while Curtis (1999) modified standard optimality criteria to improve the invertibility of the information matrix. Haber et al. (2008) and Horesh et al. (2010) discussed numerical methods for the determination of optimal designs with respect to different optimality criteria which take into account both the bias and stochastic variability of the estimate. Applications of optimal designs have been discussed for borehole tomography and impedance tomography. Moreover, Stark (2008) focused on the Backus-Gilbert resolution approach controlling the mean squared error (MSE). Van den Berg et al. (2003) applied Bayesian experimental design techniques to an amplitude versus offset experiment. While most authors concentrate on a matrix-vector representation of operator, model and data, there exists no systematic investigation of optimal design problems for

indirect regression models. In particular, there has been no investigation of optimal design problems for estimation techniques in ill-posed problems, which use the singular value decompositions of the operator K and its inverse to construct a series estimator for the unknown regression function.
The present paper tries to fill this gap and is devoted to the construction of optimal designs minimizing the integrated mean squared error of the indirect regression estimator, which is constructed by estimating the coefficients in the singular value decomposition of the corresponding operator. In Section 2.1 we introduce the necessary notation for estimating m by a singular value decomposition in the indirect regression model (1) defined below. In particular, we discuss two regularization schemes (Tikhonov and spectral cut-off regularization) and derive explicit expressions for the integrated mean squared error. Section 2.2 is devoted to the solution of the optimal design problems and the optimal design density is found explicitly. Since the optimal designs depend on the unknown regression function and regularization parameter, they require a certain amount of prior knowledge for implementation. We use the optimal designs as benchmarks against which candidate designs can be assessed. In Section 3 we illustrate our approach through several examples with a one-dimensional predictor. The robustness of optimal designs with respect to model misspecifications is investigated, and an assessment of the commonly used uniform design is provided. Examples with a two-dimensional predictor are considered in Section 4. In particular, we discuss optimal designs for the Radon transform, which is widely used in modeling of Positron Emission Tomography [Johnstone and Silverman (1990); Cavalier (2000, 2001)], and demonstrate that in most situations the uniform design or a straightforward modification thereof is close to the optimal designs in terms of performance. Finally some technical details are given in the Appendix.

2 Indirect regression

2.1 Model specification and mean squared error

We focus on the indirect regression model with random design, i.e. we suppose that we have N independent pairs of observations $(X_1, Y_1), \ldots, (X_N, Y_N)$ available from the model

$$Y_k = (Km)(X_k) + \varepsilon_k, \tag{1}$$

where K is a bounded linear operator between L^2-spaces $L^2(\mu_1)$ and $L^2(\mu_2)$, which is compact and injective. Here μ_1 and μ_2 are probability measures on the corresponding Borel σ-fields of the sets $\mathcal{X}_1$ and $\mathcal{X}_2$ with Lebesgue densities w_φ and w_ψ, respectively. The random design points X_k have a μ_2-density, say h, defined on the design space $\mathcal{X}_2 \subset \mathbb{R}^d$ which has a non empty interior. The ε_k's are independent identically distributed errors, independent of the X_k's, such that

$$E[Y_i|X_i = x] = (Km)(x),\ \mathrm{Var}(Y_i|X_i = x) = \sigma^2(x), \qquad i = 1, \ldots, N.$$

Here x denotes the predictor, and m and σ^2 are the regression and the variance function, respectively. The object of interest is the regression function $m : \mathcal{X}_1 \to \mathbb{R}$, an element of $L^2(\mu_1)$, which is only observable in the form (Km), i.e. after application of the operator K. For the regression function m we obtain the Fourier expansion

$$m = \sum_{j=1}^{\infty} a_j \varphi_j \tag{2}$$

with coefficients $a_j = \langle m, \varphi_j \rangle_{\mu_1}$, where $\{\varphi_j |\ j \in \mathbb{N}\} \subset L^2(\mu_1)$ is an orthonormal system which is part of the singular system $\{\lambda_j, \varphi_j, \psi_j\}$ of the operator K, i.e.

$$\lambda_j \psi_j = K\varphi_j, \quad \langle \varphi_j, \varphi_i \rangle_{\mu_1} = \delta_{ij}, \quad \langle \psi_j, \psi_i \rangle_{\mu_2} = \delta_{ij}, \quad i, j \in \mathbb{N}.$$

$\langle \cdot, \cdot \rangle_{\mu_1}$ and $\langle \cdot, \cdot \rangle_{\mu_2}$ denote the corresponding inner products on $L^2(\mu_1)$ and $L^2(\mu_2)$, respectively, and $\lambda_1, \lambda_2, \ldots$ are the eigenvalues of K. Similarly, the function $(Km) \in L^2(\mu_2)$ has an expansion of the form

$$Km = \sum_{j=1}^{\infty} b_j \psi_j = \sum_{j=1}^{\infty} a_j K\varphi_j = \sum_{j=1}^{\infty} \lambda_j a_j \psi_j,$$

where the Fourier coefficients b_j are given by the inner product $b_j = \langle Km, \psi_j \rangle_{\mu_2}$. A natural estimator for the coefficient b_j is

$$\hat{b}_j \ = \ \frac{1}{N} \sum_{i=1}^{N} \frac{\psi_j(X_i)}{h(X_i)} Y_i. \tag{3}$$

It is easy to see that this estimator is unbiased for b_j, i.e.

$$E[\hat{b}_j] = E[E[\frac{\psi_j(X_1)}{h(X_1)} Y_1 | X_1]] = \int_{\mathcal{X}_2} \psi_j(x) Km(x)\, d\mu_2(x) = b_j,$$

and also, unlike the least squares estimator, avoids the inversion of possibly highdimensional and ill-conditioned matrices. The estimator of the regression function m is now constructed from the expansion in (2) by an appropriate regularization. For the sake of definiteness we restrict ourselves to the Tikhonov and the spectral cut-off regularization [Engl et al. (1996)]. For the Tikhonov regularization we fix a parameter $\alpha > 0$ and define

$$\hat{m}_\alpha = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} \hat{b}_j \varphi_j \tag{4}$$

as an estimator of the regression function m. Throughout this paper we call this the Tikhonov estimator. The second estimator is obtained by truncating the expansion (2) at some index $M \in \mathbb{N}$, yielding

$$\hat{m}_M = \sum_{j=1}^{M} \frac{\hat{b}_j}{\lambda_j} \ \varphi_j, \tag{5}$$

and is therefore called spectral cut-off estimator. In the following theorem we specify the integrated mean squared error $\mathrm{IMSE}(\hat{m}) = \int_{\mathcal{X}_1} \mathrm{MSE}(\hat{m}(z))\, d\mu_1(z)$ of the two estimators. Throughout this paper we assume that the parameters of regularization satisfy $M \to \infty$ or $\alpha \to 0$ with increasing sample size $N \to \infty$.

Theorem 1 *If the assumptions specified in this section are satisfied, then the integrated mean squared error of the Tikhonov estimator* (4) *is given by*

$$\begin{aligned} IMSE(\hat{m}_\alpha) = \Phi_T(h,\alpha) \;&=\; \frac{1}{N}\int_{\mathcal{X}_2} \frac{g_\alpha(x)\{\sigma^2(x)+(Km)^2(x)\}}{h(x)}\, d\mu_2(x) \\ &+\; \alpha^2 \sum_{j=1}^{\infty} \frac{a_j^2}{(\lambda_j^2+\alpha)^2} - \frac{1}{N}\sum_{j=1}^{\infty} \frac{\lambda_j^4 a_j^2}{(\lambda_j^2+\alpha)^2}, \end{aligned} \tag{6}$$

where the function g_α is defined by

$$g_\alpha(x) = \sum_{j=1}^{\infty} \frac{\lambda_j^2}{(\lambda_j^2+\alpha)^2}\; \psi_j^2(x)\,. \tag{7}$$

For the spectral cut-off estimator (5) *we obtain*

$$\begin{aligned} IMSE(\hat{m}_M) = \Phi_C(h,M) \;&=\; \frac{1}{N}\int_{\mathcal{X}_2} \frac{g_M(x)\{\sigma^2(x)+(Km)^2(x)\}}{h(x)} d\mu_2(x) \\ &+\; \sum_{j=M+1}^{\infty} \frac{b_j^2}{\lambda_j^2} - \frac{1}{N}\sum_{j=1}^{M} \frac{b_j^2}{\lambda_j^2}, \end{aligned} \tag{8}$$

where the function g_M is defined by

$$g_M(x) = \sum_{j=1}^{M} \frac{\psi_j^2(x)}{\lambda_j^2}\,. \tag{9}$$

2.2 Optimal designs

In this section we will determine designs which minimize the integrated mean squared error of the estimators $\hat{m}_\alpha$ or $\hat{m}_M$, corresponding to Tikhonov and spectral cut-off regularization, respectively. This criterion depends on the parameter of regularization, the design density h and the functions m and σ^2. We will assume that m and σ^2 are known and determine the optimal design density, which corresponds to the concept of locally optimal designs [see Chernoff (1953)]. As a consequence, the designs derived here require some preliminary knowledge about the regression curve in the specific problem under investigation. On the other hand the important application of our results consists in the fact that the derived optimal designs serve as a benchmark for the commonly used designs. In particular we use the optimal designs to demonstrate that in many cases the popular uniform allocation is extremely efficient with respect to the integrated mean squared error criterion.

Moreover, the optimal designs determined in this section can be used in more advanced sequential design procedures as considered by Park (2000); Park and Faraway (1998) or Efromovich (2008) in the case of direct nonparametric regression.
While for fixed m and σ^2 the optimal design density can be found explicitly, the parameter of regularization usually has to be determined numerically from experimental data. The following result specifies the optimal design density.

Theorem 2

(1) For fixed $\alpha > 0$ the optimal design density minimizing the function $\Phi_T(h, \alpha)$ defined in (6) *is given by*

$$h^*_\alpha(x) = \frac{\sqrt{g_\alpha(x)}\sqrt{\sigma^2(x) + (Km)^2(x)}}{\int_{\mathcal{X}_2} \sqrt{g_\alpha(t)}\sqrt{\sigma^2(t) + (Km)^2(t)}\, d\mu_2(t)}, \tag{10}$$

where the function g_α is defined by (7).

(2) For fixed $M \in \mathbb{N}$, the optimal design density minimizing the function $\Phi_C(h, M)$ defined in (8) *is given by*

$$h^*_M(x) = \frac{\sqrt{g_M(x)}\sqrt{\sigma^2(x) + (Km)^2(x)}}{\int_{\mathcal{X}_2} \sqrt{g_M(t)}\sqrt{\sigma^2(t) + (Km)^2(t)}\, d\mu_2(t)}, \tag{11}$$

where the function g_M is defined by (9).

3 Deconvolution with a one-dimensional predictor

In this section we focus on deconvolution problems of periodic functions in $L^2[0, 1]$ which are symmetric around 0.5 (in the following denoted by $L^2_s[0, 1]$), i.e. we consider the convolution operator

$$(Km)(x) = \Psi * m(x) = \int_0^1 \Psi(x - t)m(t)\, dt,$$

with $m \in L^2_s[0, 1]$ the (unknown) function of interest, and $\Psi \in L^2_s[0, 1]$ the (known) convolution function. In this case the operator K is self-adjoint with eigenvalues $\lambda_j = \int_0^1 \Psi(t)\varphi_j(t)\, dt$, $j \geq 1$, and eigenfunctions $\varphi_j(x) = \psi_j(x) = \sqrt{2}\cos(2(j - 1)\pi x)$ for $j \geq 2$ and $\varphi_1(x) = \psi_1(x) = 1$. The measures μ_1 and μ_2 are the Lebesgue measure on the interval $[0, 1]$.
In the subsequent examples, we assume that the eigenvalues λ_j of the operator K are given by $\lambda_j = 1/j^{1+\delta}$ and the coefficients a_j in the Fourier expansion of the function m are also given by $a_j = 1/j^{1+\delta}$ for some $\delta > 0$. Here, the larger δ is, the smoother is Ψ and, in consequence, the

smoother is the operator K. The functions g_α and g_M appearing in the optimal densities $h^*_\alpha(x)$ and $h^*_M(x)$ defined by (10) and (11), respectively, simplify to

$$g_\alpha(x) = \frac{1}{(1+\alpha)^2} + 2\sum_{j=2}^{\infty} \frac{j^{2(1+\delta)}}{(1+j^{2(1+\delta)}\alpha)^2} \cos^2(2(j-1)\pi x)$$

$$\text{and} \quad g_M(x) = 1 + 2\sum_{j=2}^{M} j^{2(1+\delta)} \cos^2(2(j-1)\pi x). \tag{12}$$

We distinguish two cases in the following discussion corresponding to homo- and heteroscedastic data.

3.1 Homoscedasticity

For $\delta = 1$, $\sigma^2 = 1$ and various values of the regularization parameter, the optimal design densities are depicted in Figure 1. It is interesting to note that the optimal design densities for the Tikhonov estimator appear to be less oscillating compared to the optimal densities for spectral cut-off estimation. On the other hand, both cases yield designs with a similar form as the uniform design except in neighborhoods of the points $0, 0.5$ and 1.

In what follows, we will use the optimal designs as benchmarks and investigate the performance of the commonly used uniform allocation $h_U(x) \equiv 1$. For brevity we restrict ourselves to spectral cut-off regularization; Tikhonov regularization yields similar conclusions.

We seek values for M that balance the contributions of the bias and the variance in the integrated mean squared error. A simple calculation yields for the integrated squared bias in (8)

$$\sum_{j=M+1}^{\infty} a_j^2 = \sum_{j=M+1}^{\infty} \frac{1}{j^{2(1+\delta)}} = \frac{1}{(2\delta+1)M^{2\delta+1}} + o(M^{-2\delta-1}).$$

On the other hand, the integral of the function g_M defined in (12) with respect to the Lebesgue measure is of order $M^{2\delta+3}$ and so M has to be chosen proportionally to $N^{1/4(1+\delta)}$. Therefore we consider the choice

$$M = \left\lfloor c\Big(\frac{N}{\tau^2}\Big)^{1/4(1+\delta)} \right\rfloor + 1 \tag{13}$$

for different values of the constant c, where $\tau^2 = \int_0^1 (\sigma^2(x) + (Km)^2(x))dx$.

We investigate two examples, namely $a_j = \lambda_j = j^{-2}$ and $a_j = \lambda_j = j^{-1.25}$. In Table 1 we show the efficiencies of the uniform design h_U with respect to the optimal design minimizing the integrated mean squared error, i.e.

$$\text{eff}(h_U, M) = \frac{\Phi_C(h^*_M, M)}{\Phi_C(h_U, M)}.$$

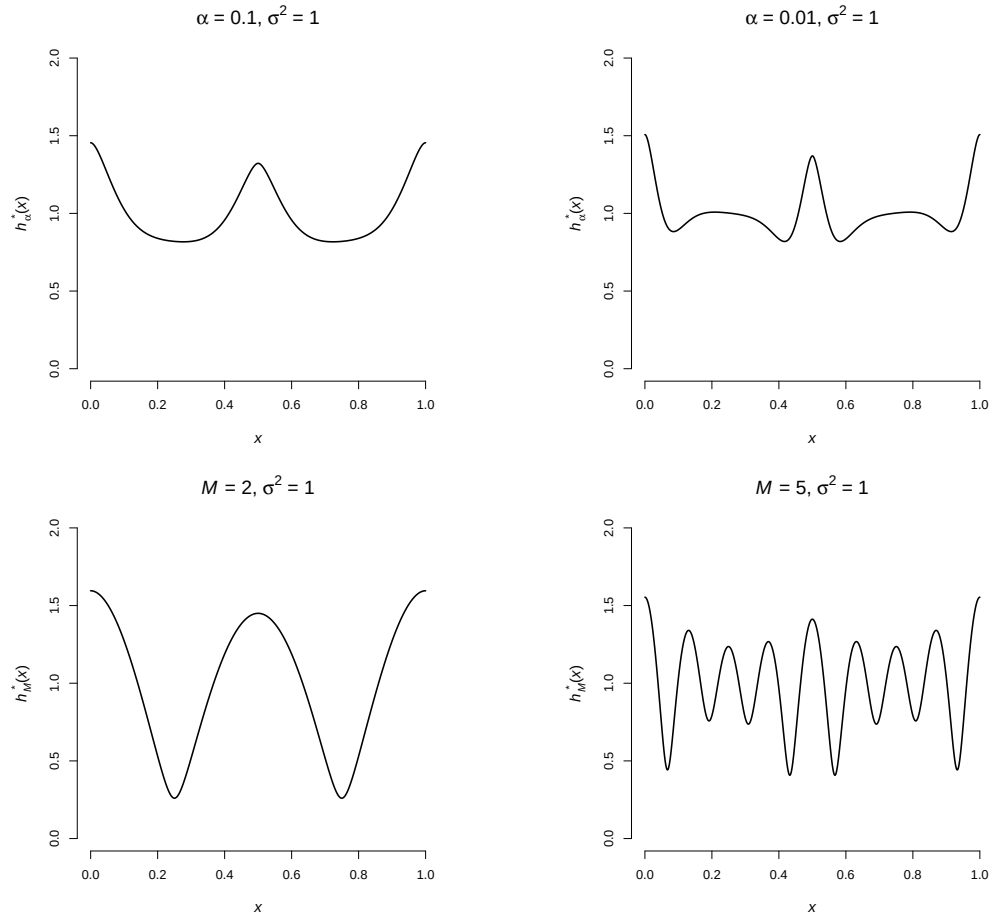


Figure 1: *The optimal densities $h^*_\alpha(x)$ and $h^*_M(x)$ minimizing the integrated mean squared error of the Tikhonov estimator and the spectral cut-off estimator, respectively, for $\delta = 1, \sigma^2 = 1$ and some selected values of the regularization parameters α and M. Top Left: $h^*_\alpha(x)$ for $\alpha = 0.1$, Top Right: $h^*_\alpha(x)$ for $\alpha = 0.01$, Bottom Left: $h^*_M(x)$ for $M = 2$, Bottom Right: $h^*_M(x)$ for $M = 5$.*

We observe that the uniform design is rather efficient for both examples across all scenarios (at least 83.9% for $\delta = 1$ and 87.6% for $\delta = 0.25$). For the situation of faster decay of coefficients a_j we observe slightly larger advantages of the optimal design. Similarly, for small sample sizes or if the value of the constant c used to determine M is 1 the improvement through using the optimal design can be more substantial. The influence of the size of σ^2 appears to be negligible.
In practice, the values for a_j, $j = 1, 2, \ldots$, σ^2 and M are not known prior to the experiment, and so the optimal design densities are locally optimal. To assess the robustness of locally optimal designs under model misspecifications, we find 8 locally optimal designs and compare them across these 8 scenarios. The uniform design is also included in this study. We assume $\lambda_j = j^{-2}$, $N = 100$, and specify $a_j = j^{-2}$ or $j^{-1.25}$ $(j = 1, 2, \ldots)$, $\sigma^2 = 1$ or 0.25 and $M = 2$ or 5. The efficiencies of the 9 designs under consideration are given in Table 2.

Table 1: *Efficiency of the uniform design for different sample sizes, variances and choices of the regularization parameter M. The value of M, determined by (13), is shown in brackets.*

$\delta = 1$									
N	$\sigma^2 = 0.25$			$\sigma^2 = 1$			$\sigma^2 = 4$		
	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
25	.889 (1)	.839 (2)	.889 (3)	.890 (1)	.845 (2)	.891 (3)	.891 (1)	.849 (2)	.893 (3)
100	.911 (1)	.850 (2)	.911 (4)	.905 (1)	.851 (2)	.913 (4)	.898 (1)	.852 (2)	.893 (3)
1000	.916 (2)	.895 (3)	.926 (5)	.901 (2)	.895 (3)	.928 (5)	.941 (1)	.877 (2)	.915 (4)
$\delta = 0.25$									
N	$\sigma^2 = 0.25$			$\sigma^2 = 1$			$\sigma^2 = 4$		
	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
25	.934 (1)	.876 (2)	.905 (4)	.939 (1)	.885 (2)	.923 (4)	.942 (1)	.889 (2)	.936 (4)
100	.939 (2)	.908 (3)	.920 (5)	.933 (2)	.918 (3)	.936 (5)	.961 (1)	.924 (3)	.947 (5)
1000	.991 (2)	.951 (4)	.942 (8)	.989 (2)	.952 (4)	.952 (7)	.979 (2)	.948 (4)	.961 (7)

Table 2: *Efficiencies of the 9 designs under investigation for 8 different scenarios with $N = 100$. $h^*(a_j, \sigma^2, M)$ is the locally optimal design for the given selection of (a_j, σ^2, M), and h_U is the uniform design.*

design $\setminus$ *scenario*	$a_j = j^{-2}$				$a_j = j^{-1.25}$			
	$\sigma^2 = 0.25$		$\sigma^2 = 1$		$\sigma^2 = 0.25$		$\sigma^2 = 1$	
	$M = 2$	$M = 5$	$M = 2$	$M = 5$	$M = 2$	$M = 5$	$M = 2$	$M = 5$
$h^*(j^{-2}, 0.25, 2)$	1	.681	1	.679	.999	.690	1	.685
$h^*(j^{-2}, 0.25, 5)$	.743	1	.740	1	.830	.999	.805	1
$h^*(j^{-2}, 1, 2)$	1	.683	1	.681	.998	.692	1	.687
$h^*(j^{-2}, 1, 5)$	.740	1	.739	1	.827	.997	.804	.999
$h^*(j^{-1.25}, 0.25, 2)$	.998	.673	.996	.670	1	.683	.999	.677
$h^*(j^{-1.25}, 0.25, 5)$	.747	.999	.743	.997	.835	1	.809	.999
$h^*(j^{-1.25}, 1, 2)$	1	.678	.999	.676	.999	.688	1	.682
$h^*(j^{-1.25}, 1, 5)$	.745	1	.742	.999	.831	.999	.807	1
h_U	.850	.926	.851	.928	.900	.920	.889	.925

Note that all off-diagonal elements equal to 1 result from rounding to three decimal places. We see from Table 2 that the uniform design is most robust among its competitors with a minimal efficiency of 85% across all scenarios. For the locally optimal designs we observe an alternating pattern of very high and relatively low efficiencies. These imply that misspecifications of the coefficients a_j and the variance σ^2 hardly affect the efficiency of the locally optimal designs whereas the misspecification of M can lead to poor design performance. Following this up, we found that the optimal designs for the same M but different a_j and σ^2 are very similar, which explains their similar performance. We further note that optimal designs for $M = 5$ are slightly more robust than those for $M = 2$. From the bottom panel of Figure 1 we see that $h_5^*(x)$ despite its oscillating form resembles a uniform density more closely than $h_2^*(x)$.

3.2 Heteroscedasticity - Poisson distribution

In many applications of inverse problems, e.g. tomography, the data are counts. In such situations the assumption of constant variance is not realistic and a popular distributional assumption is that of a Poisson distribution where we have $\mathrm{Var}(Y_i|X_i = x) = \mathrm{E}[Y_i|X_i = x] = (Km)(x)$. Therefore it is of considerable interest to compare the results of the previous section with the corresponding situation in the heteroscedastic case to assess if the uniform design will also do well in this situation. Again, we restrict ourselves to the case of spectral cut-off regularization and consider the situation discussed in the previous paragraph, that is $\lambda_j = a_j = 1/j^{1+\delta}$, where $\delta > 0$. The optimal design density is obtained from (11) with $\sigma^2(x) = (Km)(x) = 1 + \sqrt{2}\sum_{j=2}^{\infty} j^{-2(1+\delta)} \cos(2\pi(j-1)x)$. The resulting densities are directly comparable with those depicted in the bottom panel of Figure 1, but not shown here since there are no substantial differences.

In Table 3 we present the corresponding efficiencies of the uniform design, where the parameter of regularization M again is chosen by the rule of thumb in (13), with $\tau^2 = \int_0^1 [Km(x) + (Km)^2(x)]dx$. A comparison with Table 3 shows that in the case of heteroscedasticity the uniform design is similarly efficient as for homoscedasticity.

Table 3: *Efficiency of the uniform design in the Poisson model for different sample sizes and various choices of the regularization parameter M. The value of M is shown in brackets.*

N	$\delta = 0.25$			$\delta = 1$		
	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
25	.927 (1)	.870 (2)	.909 (4)	.888 (1)	.842 (2)	.890 (3)
100	.924 (2)	.905 (3)	.924 (5)	.903 (1)	.849 (2)	.912 (4)
1000	.987 (2)	.943 (4)	.941 (7)	.899 (2)	.894 (3)	.926 (5)

We investigate a further example corresponding to a sudden change of signal over a certain period.

The function $m(z)$ is given by

$$m(z) \;=\; 2\; I_{[\frac{1}{4},\frac{3}{4}]}(z) + 1, \tag{14}$$

which yields for the coefficients in the Fourier expansion $a_1 = 2$,

$$a_j \;=\; \int_0^1 m(z)\varphi_j(z)dz = \frac{2\sqrt{2}(-1)^{j/2}}{\pi(j-1)}, \;\; \text{if} \;\; j \geq 2, \; j \text{ even}$$

and $a_j = 0$ otherwise. We consider three different functions with which $m(z)$ is convoluted, resulting in eigenvalues $\lambda_j = a_j, j^{-1.25}$ or j^{-2}, respectively, for $j = 1, 2, \ldots$. Figure 2 shows the optimal densities for the choice $\lambda_j = a_j$ and different values of M. These designs look considerably different from those found for the previous examples, which is due to the different form of the function $Km(x)$.

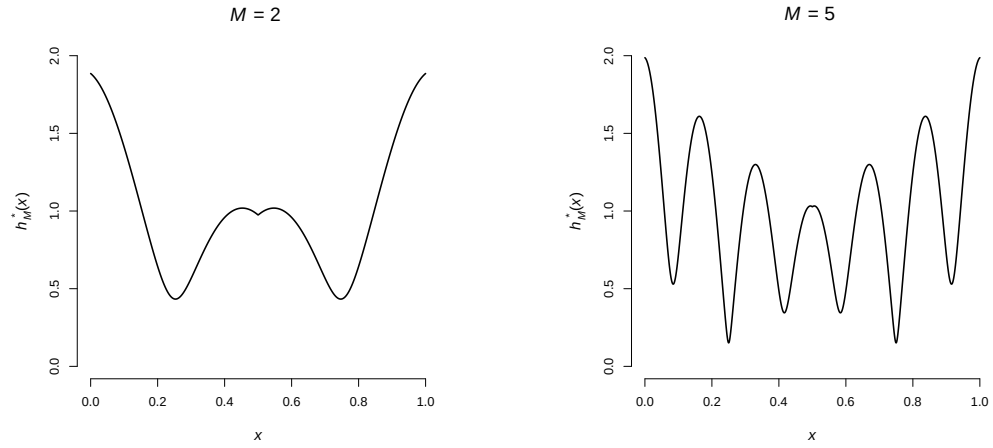


Figure 2: *The optimal density $h_M^*(x)$ minimizing the integrated mean squared error of the spectral cut-off estimator in the case of heteroscedasticity for the step function (14) with $\lambda_j = a_j$. Left: $M = 2$, Right: $M = 5$.*

Again, we use the optimal designs as benchmarks to assess the performance of the uniform design. To find values for M through (13) we compute the order of the integrated squared bias as $\sum_{j=M+1}^{\infty} a_j^2 = (8/\pi^2) \sum_{j=M+1,\, j \text{ even}}^{\infty} (j-1)^{-2} = O(1/M)$. The order of the integrated variance is $O(M^{2\delta+3}/N)$ for $\delta = 0, 0.25$ and 1, respectively, depending on the choice of eigenvalues.
The efficiencies of the uniform design for various scenarios are given in Table 4. As before, the uniform design is doing remarkably well.

4 Two-dimensional indirect regression problems

In this section we investigate optimal design problems for two dimensional indirect regression problems. Throughout this section $\mathbf{x}, \mathbf{z} \in \mathbb{R}^2$ denote two dimensional variables. Referring to four partic-

Table 4: *Efficiency of the uniform design for the regression function defined in (14) for different sample sizes and various choices of the parameter M given in brackets.*

N	$\lambda_j = a_j$			$\delta = 0.25$			$\delta = 1$		
	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
25	.994 (1)	.852 (2)	.852 (3)	.937 (1)	.861 (2)	.901 (3)	.899 (1)	.841 (2)	.886 (3)
100	.998 (1)	.898 (2)	.898 (3)	.969 (1)	.896 (2)	.918 (4)	.935 (1)	.860 (2)	.907 (4)
1000	.979 (2)	.979 (3)	.885 (6)	.975 (2)	.935 (4)	.941 (7)	.943 (2)	.913 (3)	.922 (5)

ular applications given below, we assume that the bases of the underlying L^2-spaces are subspaces of the complex valued functions and that the corresponding bases are indexed by two parameters, such that the singular value decompositions of the functions m and Km are given by

$$m(\mathbf{z}) = \sum_{q=0}^{\infty} \sum_{p} a_{pq} \varphi_{pq}(\mathbf{z}), \quad (Km)(\mathbf{x}) = \sum_{q=0}^{\infty} \sum_{p} b_{pq} \psi_{pq}(\mathbf{x}),$$

respectively, where the range of the second index p *is finite* and depends on the parameter q. The functions φ_{pq} and ψ_{pq} are known orthonormal bases of the L^2-spaces, that is

$$\int\int \varphi_{pq}(\mathbf{z}) \overline{\varphi_{rs}(\mathbf{z})}\, d\mu_1(\mathbf{z}) \;=\; \int\int \psi_{pq}(\mathbf{x}) \overline{\psi_{rs}(\mathbf{x})}\, d\mu_2(\mathbf{x}) \;=\; \delta_{pr}\delta_{qs},$$

where $\overline{\varphi}$ denotes the complex conjugate of the function φ and δ_{ij} is the Kronecker delta. The singular values λ_{pq} satisfy $K\varphi_{pq} = \lambda_{pq}\psi_{pq}$ and $\lambda_{pq} a_{pq} = b_{pq}$.

For brevity we restrict ourselves to the case of spectral cut-off regularization and consider the estimators

$$\hat{b}_{pq} \;=\; \frac{1}{N} \sum_{i=1}^{N} \frac{\overline{\psi_{pq}(\mathbf{X_i})}}{h(\mathbf{X_i})} Y_i, \quad \hat{m}(\mathbf{z}) \;=\; \sum_{q=0}^{M} \sum_{p} \frac{\hat{b}_{pq}}{\lambda_{pq}} \varphi_{pq}(\mathbf{z})$$

for the coefficients b_{pq} and the regression function m, respectively. From Theorem 1 we obtain for the integrated mean squared error

$$\begin{aligned} \mathrm{IMSE}(h, M) \;=\; & \frac{1}{N} \int\int \frac{g_M(\mathbf{x}) \{\sigma^2(\mathbf{x}) + (Km)^2(\mathbf{x})\}}{h(\mathbf{x})}\, d\mu_2(\mathbf{x}) \\ & + \sum_{q=M+1}^{\infty} \sum_{p} \frac{|b_{pq}|^2}{\lambda_{pq}^2} - \frac{1}{N} \sum_{q=0}^{M} \sum_{p} \frac{|b_{pq}|^2}{\lambda_{pq}^2}, \end{aligned}$$

where the function g_M is defined by

$$g_M(\mathbf{x}) = \sum_{q=0}^{M} \sum_{p} \frac{|\psi_{pq}(\mathbf{x})|^2}{\lambda_{pq}^2} \tag{15}$$

and $|b_{pq}|^2 = b_{pq}\overline{b_{pq}}$ is the squared complex modulus. The optimal density minimizing the integrated mean squared error is obtained from equation (11) in Theorem 2.

4.1 Optimal design for the Radon transform

As a special case of the situation discussed in the previous paragraph we consider the Radon transform, which appears e.g. in the modeling of Positron Emission Tomography (PET) experiments [e.g. Johnstone and Silverman (1990), Cavalier (2000)]. PET is concerned with the estimation of the density of positron emission due to a radioactively labeled metabolite which was injected into a patient's body. In the two-dimensional case, which we consider here, the aim is to recover the density of emission in a slice through the patient's body. In this case the Radon transform $\mathcal{R}$ represents the line integrals through the emission density in the body, taken along all possible lines through the slice. Hence, $\mathcal{R}$ is an injective integral operator mapping a function in the space of observations (often called brain space) $L^2(B, \mu_B)$ of emission densities in the patient's body to the detector space $L^2(D, \mu_D)$. In what follows we assume B to be the unit circle, parametrized by polar coordinates (r, ϑ), and in a similar way D to be parametrized by the angle $\phi \in [0, 2\pi)$ of the detected line through the patient's body, and its impact parameter $s \in [0, 1]$.
In our subsequent analysis we model the PET data as noisy discrete observations in the indirect regression model (1), where $m(r, \vartheta)$ is the emission density in the patient's body, which is to be recovered from the observations, and the operator $K = \mathcal{R}$ is

$$\mathcal{R}m(s, \phi) = \frac{1}{2\sqrt{1-s^2}} \int_{-\sqrt{1-s^2}}^{\sqrt{1-s^2}} m\left(s\cos(\phi) - t\sin(\phi), s\sin(\phi) + t\cos(\phi)\right) dt. \tag{16}$$

Unlike for the deconvolution problems considered in Section 3 the system of basis functions considered here is not orthogonal with respect to the Lebesgue measure. We briefly discuss the singular value decomposition of $\mathcal{R}$, which is required for the subsequent computations. The Lebesgue densities of the measures μ_B and μ_D corresponding to the L^2-spaces $L^2(B, \mu_B)$ and $L^2(D, \mu_D)$ are given by $w_\varphi(r, \vartheta) = r/\pi$ for $0 \le r \le 1,\ 0 \le \vartheta < 2\pi$ and $w_\psi(s, \phi) = 2(1-s^2)^{1/2}/\pi^2$ for $0 \le s \le 1,\ 0 \le \phi < 2\pi$. The orthonormal system of basis functions $\{\varphi_{pq}\}$ of the brain space is defined by the Zernike polynomials $\varphi_{p,q}(r, \vartheta) = \sqrt{q+1} \cdot Z_q^{|p|}(r) e^{ip\vartheta}$, $q = 0, 1, 2, \ldots$, $p = -q, -q+2, \ldots, q$, where $Z_m^k(r)$ denotes a polynomial of degree m [see Zernike (1934)] and is defined as

$$Z_m^k(r) = \sum_{j=0}^{(m-k)/2} \frac{(-1)^j (m-j)!}{j!((m+k)/2 - j)!((m-k)/2 - j)!} r^{m-2j}$$

if $m - k$ is even and $Z_m^k(r) = 0$ if $m - k$ is odd. Similarly, the associated basis functions of the detector space are given by $\psi_{pq}(s, \phi) = U_q(s) e^{ip\phi}$, $q = 0, 1, 2, \ldots$, $p = -q, -q+2, \ldots, q$, where $U_q(\cos(\kappa)) = \sin\left((q+1)\kappa\right) / \sin(\kappa)$ is the qth Chebyshev polynomial of the second kind [see Szegö

(1975)]. Finally, the singular values of the operator $\mathcal{R}$ are given by $\lambda_{pq} = (q+1)^{-1/2}$ for every $(p,q) \in \{q \in \mathbb{N}_0; p = -q, -q+2, \ldots, q\}$, and $\mathcal{R}\varphi_{pq} = \lambda_{pq}\psi_{pq}$. For further details see Johnstone and Silverman (1990), who studied the PET problem in a density estimation framework. In what follows we will derive the optimal design density for the Radon transform.

Note that the function g_M defined in (15) does not depend on the variable ϕ, that is

$$g_M(s) = g_M(s,\phi) = \sum_{q=0}^{M}\sum_{p}(q+1)U_q^2(s) = \sum_{q=0}^{M}(q+1)^2U_q^2(s). \tag{17}$$

It follows from Theorem 2 that the optimal density is given by

$$h_M^*(s,\phi) = \frac{\pi^2}{2}\,\frac{\sqrt{\sigma^2(s,\phi) + (\mathcal{R}m)^2(s,\phi)}\,\sqrt{g_M(s)}}{\int_0^1\int_0^{2\pi}\sqrt{\sigma^2(t,\rho) + (\mathcal{R}m)^2(t,\rho)}\,\sqrt{1-t^2}\,\sqrt{g_M(t)}\,d\rho dt},$$

where the function $g_M(s)$ is defined in (17). In what follows, we investigate the performance of the uniform design with constant density $h_U(s,\phi) \equiv 1$ on $[0,1]\times[0,2\pi]$ in four examples.

4.2 Specific Examples

We consider: Two objects positioned in the center of the scan field, a solid disc and a polar rose, the latter representing an object with cracks appearing in several places as e.g. observed in materials science; a disc shifted to the right of the scan field; a double disc having positive mass throughout the scan field. A schematic of a slice of each example object, embedded in the detector ring, is shown in Figure 3.

For each slice of a solid disc of radius $r_0 < 1$, positioned in the middle of the scan field, we obtain $m(r,\theta) = 1$ if $0 \le r \le r_0, 0 \le \theta \le 2\pi$, and $m(r,\theta) = 0$ otherwise.

Since the observations in tomography applications are usually photon counts, we assume the observations $Y|(S,\Phi) = (s,\phi)$ come from a Poisson distribution with parameter

$$\mathcal{R}m(s,\phi) = \sigma^2(s,\phi) = \sqrt{r_0^2 - s^2}/\sqrt{1-s^2}\,I_{[0,r_0]}(s).$$

Since $\mathcal{R}m(s,\phi)$ does not depend on ϕ the optimal design density $h_M^*(s,\phi)$ simplifies to

$$h_M^*(s,\phi) \;=\; \frac{\pi}{4}\;\frac{\sqrt{g_M(s)}\sqrt{\frac{\sqrt{r_0^2-s^2}}{\sqrt{1-s^2}} + \frac{r_0^2-s^2}{1-s^2}}}{\int_0^{r_0}\sqrt{g_M(t)}\sqrt{\frac{\sqrt{r_0^2-t^2}}{\sqrt{1-t^2}} + \frac{r_0^2-t^2}{1-t^2}}\,dt} \quad \text{if } \; 0 \le s \le r_0,\; 0 \le \phi \le 2\pi$$

and $h_M^*(s,\phi) = 0$ otherwise. Obviously, this design would be useless for objects that extend beyond distance r_0 from the center.

For a polar rose with 8 petals and choosing 0.5 for the maximal extension from the center, each slice is described by

$$m(r,\theta) \;=\; 1 \quad \text{if } \; 0 \le r \le 0.5\,|\cos(4\theta)|,\; 0 \le \theta \le 2\pi \tag{18}$$

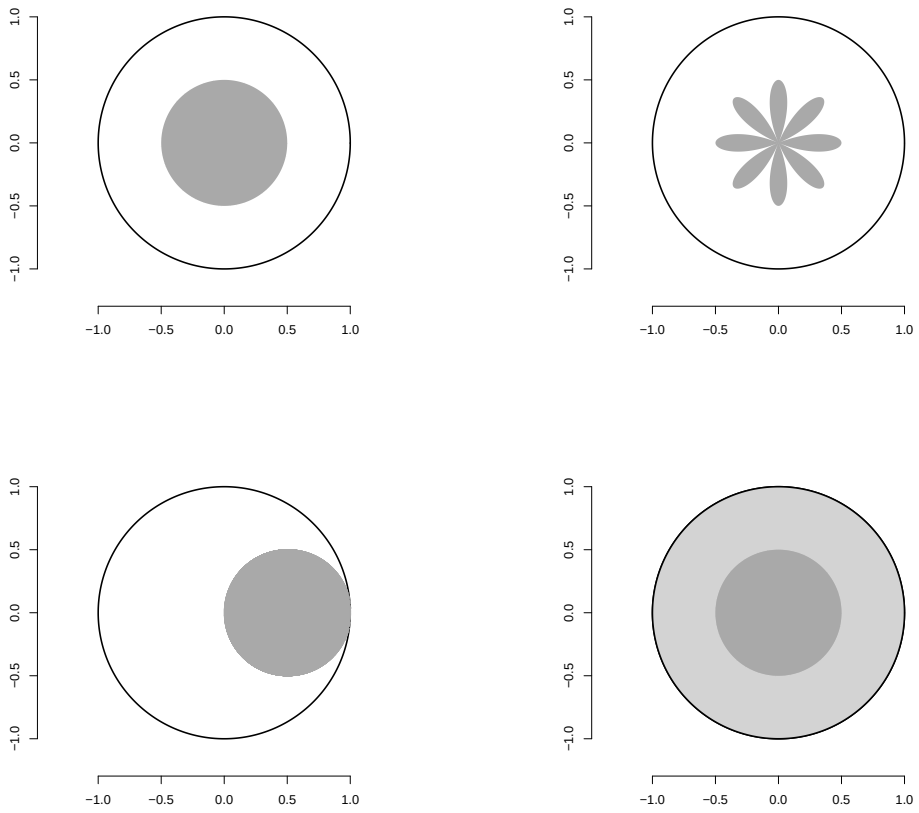


Figure 3: *Schematic of slices of the example objects. Top left: Disc with radius 0.5 positioned in the center of the detector ring. Top right: Polar rose with 8 petals positioned in the center of the detector ring. Bottom left: Disc of radius 0.5 positioned on the right hand side of the detector ring. Bottom right: Double disc with higher density towards its center.*

and $m(r, \theta) = 0$ otherwise. The graphs in Figure 4 show the optimal design density $h^*_M(s, \phi)$ for the centered disc with radius $r_0 = 0.5$ and the polar rose for different values of M. For both objects, the densities are zero for $s > 0.5$.

Numerical calculations suggest that the integrated squared bias is approximately of order M^{-1} while the integrated variance is of order M^3/N. To obtain a balance of orders we consider the choice $M = \lfloor c(N/\tau^2)^{0.25} \rfloor + 1$ for the parameter in the spectral cut-off estimator, where $\tau^2 = \int_0^1 \int_0^{2\pi} (\mathcal{R}m(s, \phi) + (\mathcal{R}m)^2(s, \phi))\, d\mu_D(s, \phi)$.

In the left panel of Table 5 we show the efficiencies of the uniform design h_U for scanning the centered disc for various values of N and M, while the efficiencies for scanning the polar rose defined in (18) are displayed in the right panel. These are reasonably good when M is small, i.e. when the bias dominates the IMSE, but rather poor for larger values of the regularization parameter.

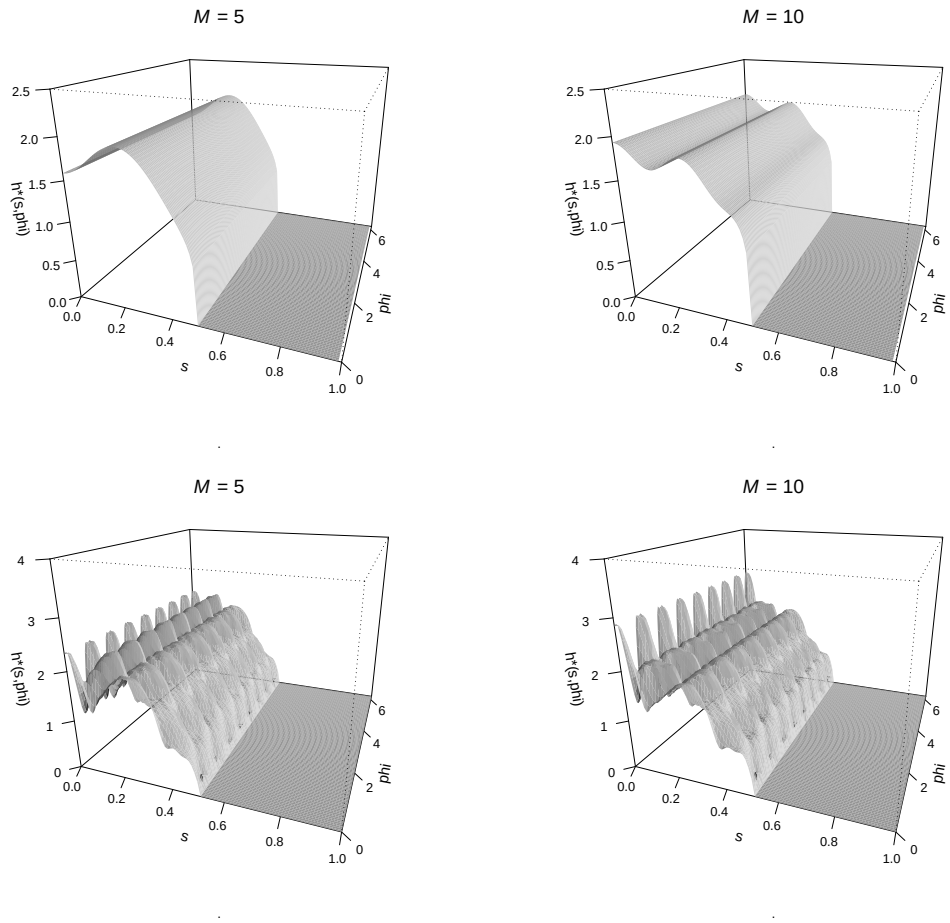


Figure 4: *Plots of selected optimal densities $h^*_M(s,\phi)$ for scanning a centered disc and a polar rose for different values of M. Top left: Centered disc, $M = 5$, Top right: Centered disc, $M = 10$, Bottom left: Polar rose, $M = 5$, Bottom right: Polar rose, $M = 10$.*

Table 5: *Efficiency of the uniform design h_U for estimating a disc and a polar rose in the middle of the scan field, respectively, for different sample sizes and various choices of the parameter M used in the spectral cut-off regularization. The values of M are given in brackets.*

N	centered disc			polar rose		
	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
25	.751 (2)	.696 (3)	.607 (6)	.830 (2)	.691 (4)	.632 (8)
100	.833 (3)	.658 (5)	.611 (9)	.910 (3)	.725 (6)	.646 (11)
1000	.915 (4)	.733 (8)	.620 (15)	.950 (5)	.842 (9)	.679 (18)
10000	.962 (7)	.801 (13)	.623 (26)	.981 (8)	.901 (16)	.661 (32)

For calculating the optimal density we used the assumption that we know the exact shape of the object to be scanned. In some applications, e.g. when looking for interior cracks in an object in materials science, information on the outer shape and position of the object may well be available. Using that the objects do not extend more than 0.5 units from the center of the detector circle, it seems reasonable to consider the uniform design with constant density $h_{U,0.5}(s,\phi) \equiv \pi/(\sqrt{0.75}+2\arcsin(0.5)) \approx 1.642$ on $[0,0.5]\times[0,2\pi]$. The efficiencies of this design show a considerable improvement compared with the uniform design on the larger space: Across the same scenarios as in Table 5, the minimal efficiency of $h_{U,0.5}$ is 96.3% and 91.2%, respectively, for estimating the centered disc and the polar rose.

We next consider the scanning of a solid disc with radius r_0, but this time the object is not located in the center of the scan field. For the choice $r_0 = 0.5$ for the radius and $(0.5, 0)$ for the center of the object, we obtain for its density

$$m(r,\theta) \;=\; 1 \quad \text{if} \;\; 0 \le r \le \cos(\theta),\; 0 \le \theta \le 2\pi$$

and $m(r,\theta) = 0$ otherwise.

As an example of an object which has positive density everywhere in the scan field we consider two nested discs of different density. A slice of this double disc is described by

$$m(r,\theta) = \begin{cases} 1 & \text{if} \;\; 0 \le r_0,\; 0 \le \theta \le 2\pi \\ 0.5 & \text{if} \;\; r_0 < r \le 1,\; 0 \le \theta \le 2\pi, \end{cases}$$

i.e. the density of the object is higher towards the center.

The optimal densities for scanning the shifted disc and the double disc are depicted in Figure 5 for different values of the regularization parameter M. Unlike the previous examples, for the shifted disc the area with zero density depends on both s and ϕ. For the double disc, the optimal densities increase with s as $s \to 1$.

In the left panel of Table 6 we show the efficiencies of the uniform design h_U for estimating the shifted disc for various values of N and M while the efficiencies for estimating the double disc are displayed in the right panel. For the shifted disc, the uniform design only does well in situations where the regularization parameter M is small, i.e. where the integrated squared bias dominates the IMSE. Unlike in the situation of Example 1, where this problem could be fixed by reducing the domain of the uniform design accordingly, there is no obvious way around this issue in this case. The double disc can be estimated reasonably well using the uniform design.

5 Conclusions

This is the first paper to provide a systematic approach to optimal design for indirect regression problems. We have focused on the derivation of designs leading to an efficient estimation of the

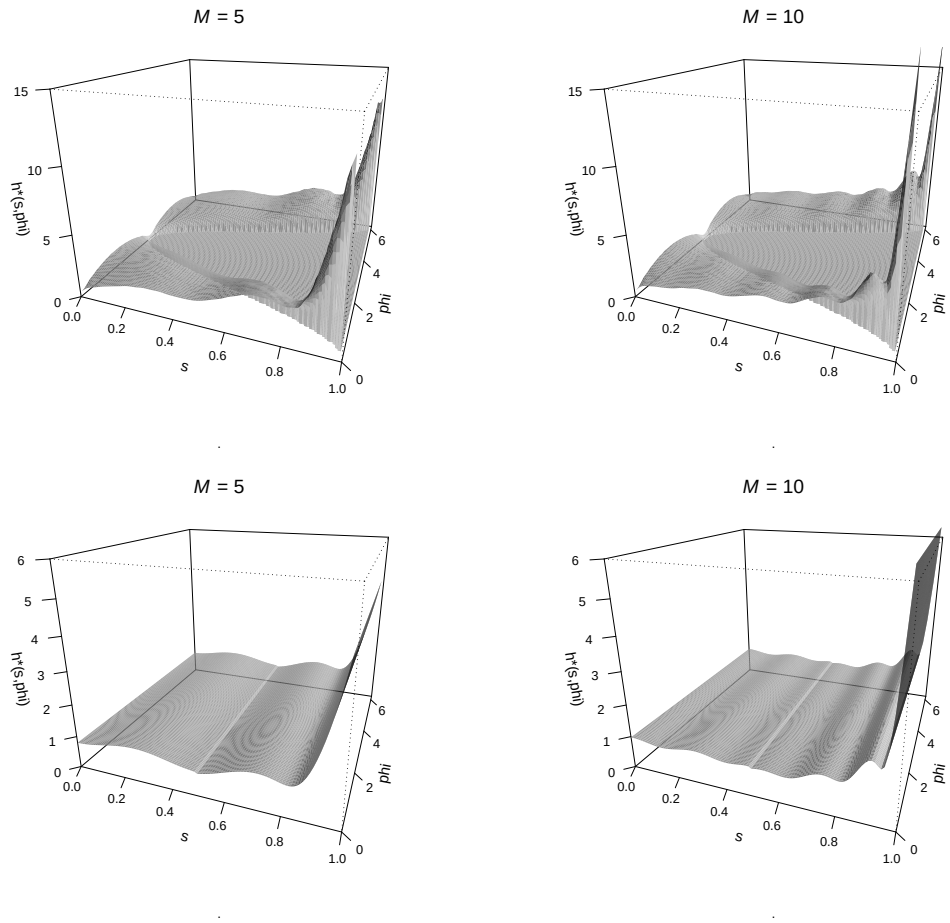


Figure 5: *Plots of selected optimal densities* $h^*_M(s,\phi)$ *for scanning a shifted disc and a double disc, respectively, for different values of* M*. Top left: Shifted disc* $M = 5$*, Top right: Shifted disc,* $M = 10$*, Bottom left: Double disc,* $M = 5$*, Bottom right: Double disc,* $M = 10$*.*

Table 6: *Efficiency of the uniform design* h_U *for estimating a non-centered disc and a double disc in the middle of the scan field, respectively, for different sample sizes and various choices of the parameter* M *used in the spectral cut-off regularization. The values of* M *are given in brackets.*

N	shifted disc			double disc		
	$c = 0.5$	$c = 1$	$c = 2$	$c = 0.5$	$c = 1$	$c = 2$
25	.679 (2)	.568 (3)	.541 (6)	.856 (2)	.860 (3)	.863 (5)
100	.693 (3)	.581 (5)	.543 (9)	.873 (2)	.866 (4)	.866 (7)
1000	.864 (4)	.644 (8)	.554 (15)	.920 (3)	.873 (6)	.866 (12)
10000	.923 (7)	.702 (13)	.559 (26)	.937 (5)	.879 (10)	.867 (20)

unknown regression function m. Using the singular value decomposition of the operator K, an expression for the integrated mean squared error of a natural series estimator was derived. Designs minimizing this expression were found explicitly. These designs serve as benchmarks for commonly used designs in indirect regression. Moreover they can be used in more advanced sequential design procedures as considered by Park (2000); Park and Faraway (1998) or Efromovich (2008) in the case of direct nonparametric regression. In this paper we worked in the first named direction and investigated the efficiency of the uniform design in several situations of practical interest. It was demonstrated that the uniform design is performing efficiently under most scenarios. In particular, the uniform design is rather robust with respect to the choice of the regularization parameter.

Acknowledgements The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) and the BMBF (project 'INVERS' 03BIPAH4).

A Appendix: Proofs

A.1 Proof of Theorem 1

We restrict ourselves to the spectral cut-off estimator. The arguments for the Tikhonov estimator (4) are similar and therefore omitted for brevity. First note that the bias of the spectral cut-off estimator $\hat{m}_M$ is given by

$$E[\hat{m}_M(z) - m(z)] = \sum_{j=1}^{M} \frac{E[\hat{b}_j - b_j]}{\lambda_j}\varphi_j(z) - \sum_{j=M+1}^{\infty} \frac{b_j}{\lambda_j}\varphi_j(z) = -\sum_{j=M+1}^{\infty} \frac{b_j}{\lambda_j}\varphi_j(z). \tag{19}$$

For the variance of the estimators we have from definition (3) that

$$\begin{aligned} \mathrm{Var}(\hat{b}_j) &= \frac{1}{N}\Big\{\mathrm{Var}(E[Z_{1j}Y_1|X_1]) + E[\mathrm{Var}(Z_{1j}Y_1|X_1)]\Big\} \\ &= \frac{1}{N}\int_{\mathcal{X}_2} \frac{\{\sigma^2(x) + (Km)^2(x)\}\psi_j^2(x)}{h(x)} d\mu_2(x) - \frac{b_j^2}{N}, \end{aligned}$$

where the random variables Z_{ij} are given by $Z_{ij} = \psi_j(X_i)/h(X_i)$. The variance of the spectral cut-off estimator is $\mathrm{Var}(\hat{m}_M(z)) = \sum_{l,k=1}^{M} \mathrm{Cov}(\hat{b}_l, \hat{b}_k)\varphi_l(z)\varphi_k(z)/(\lambda_k\lambda_l)$.
Now note that the functions $\{\varphi_1, \varphi_2, \dots\}$ define an orthonormal basis of $L^2(\mu_1)$, which implies for the integrated variance that

$$\begin{aligned} \int_{\mathcal{X}_1} \mathrm{Var}(\hat{m}_M(x)) d\mu_1(x) &= \sum_{j=1}^{M} \frac{\mathrm{Var}(\hat{b}_j)}{\lambda_j^2} \\ &= \frac{1}{N}\int_{\mathcal{X}_2} \frac{g_M(x)\{\sigma^2(x) + (Km)^2(x)\}}{h(x)} d\mu_2(x) - \frac{1}{N}\sum_{l=1}^{M} \frac{b_l^2}{\lambda_l^2}, \end{aligned}$$

where the function g_M is defined in (9). By a similar argument applied to (19), we obtain for the integrated mean squared error of the estimator the expression (8), which proves the second assertion of Theorem 1. □

A.2 Proof of Theorem 2

Both cases are shown similarly and we restrict ourselves to the case (2) of spectral cut-off regularization. First note that for fixed $M \in \mathbb{N}$ the optimization of the integrated mean squared error (8) reduces to minimization of the expression

$$f(h) = \int_{\mathcal{X}_2} \frac{g_M(x)\{\sigma^2(x) + (Km)^2(x)\}}{h(x)}\, d\mu_2(x)$$

with respect to the design density h. Now Cauchy's inequality yields

$$f(h) \geq \left(\int_{\mathcal{X}_2} \sqrt{g_M(x)}\sqrt{\sigma^2(x) + (Km)^2(x)}\, d\mu_2(x) \right)^2,$$

where there is equality if and only if

$$h_M^*(x) = \frac{\sqrt{g_M(x)}\sqrt{\sigma^2(x) + (Km)^2(x)}}{\int_{\mathcal{X}_2} \sqrt{g_M(t)}\sqrt{\sigma^2(t) + (Km)^2(t)}\, d\mu_2(t)}. \qquad \square$$

References

Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *Journal of Multivariate Analysis*, 100:2364–2375.

Broniatowski, M. and Celant, G. (2007). Optimality and bias of some interpolation and extrapolation designs. *Journal of Statistical Planning and Inference*, 137:858–868.

Cavalier, L. (2000). Efficient estimation of a density in a problem of tomography. *Annals of Statistics*, 28:630–637.

Cavalier, L. (2001). On the problem of local adaptive estimationin tomography. *Bernoulli*, 7:63–78.

Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123:323–354.

Chan, L.-Y. (1992). Optimal design for estimation of variance in nonparametric regression using first order differences. *Biometrika*, 78:926–929.

Cheng, M.-Y., Hall, P., and Titterington, D. (1998). Optimal design for curve estimation by local linear smoothing. *Bernoulli*, 4:3–14.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24:586–602.

Curtis, A. (1999). Optimal experiment design: cross-borehole tomographic examples. *Geophysical Journal International*, 136:637–650.

Dette, H. and O'Brien, T. (1999). Optimality criteria for regression models based on predicted variance. *Biometrika*, 86:93–106.

Donoho, D. L. and Low, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Annals of Statistics*, 20:944–970.

Efromovich, S. (2008). Optimal sequential design in a controlled non-parametric regression. *Scandinavian Journal of Statistics*, 35:266–285.

Engl, H. W., Hanke, M., and Neubauer, A. (1996). *Regularization of Inverse Problems.* Kluwer Academic, Dordrecht, Boston, London.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, 19:1257–1272.

Groeneboom, P. and Jongbloed, G. (1995). Isotonic estimation and rates of convergence in wicksell's problem. *Annals of Statistics*, 23:1518–1542.

Haber, E., Horesh, L., and Tenorio, L. (2008). Numerical methods for experimental design of large-scale linear ill-posed inverse problems. *Inverse Problems*, 24:055012.

Horesh, L., Haber, E., and Tenorio, L. (2010). Optimal experimental design for the large-scale nonlinear ill-posed problem of impedance imaging. Technical report.

Johnstone, I. M., Kerkyacharian, G., Picard, D., and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society Series B*, 66:547–573.

Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Annals of Statistics*, 18:251–280.

Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in hilbert scales. *SIAM Journal on Applied Mathematics*, 56:1424–1444.

Maurer, H., Boerner, D., and Curtis, A. (2000). Design strategies for electromagnetic geophysical surveys. *Inverse Problems*, 16:1097–1117.

Mueller, H. (1984). Optimal designs for nonparametric kernel regression. *Statistics & Probabaility Letters*, 4:285–290.

Park, D. (2000). Sequential design for local bandwidth response curve estimator. *Journal of Nonparametric Statistics*, 12:593–548.

Park, D. and Faraway, J. (1998). Sequential design for response curve estimation. *Journal of Nonparametric Statistics*, 10:155–164.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Randall, T., Donev, A., and Atkinson, A. C. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Shepp, L. and Vardi, Y. (1982). maximum likelihood reconstruction for emission tomography.

Spruill, M. C. (1987). Optimal designs for interpolation. *Journal of Statistical Planning and Inference*, 16:219–229.

Stark, P. B. (2008). Generalizing resolution. *Inverse Problems*, 24:034014.

Studden, W. J. (1977). Optimal designs for integrated variance in polynomial regression. In *Statistical decision theory and related topics. II (Proc. Sympos., Purdue Univ., Lafayette, Ind., 1976)*, pages 411–420. Academic Press, New York.

Szegö, G. (1975). *Orthogonal Polynomials.* American Mathematical Society, Providence, R.I.

Van den Berg, J., Curtis, A., and Trampert, J. (2003). Optimal nonlinear bayesian experimental design: an application to amplitude versus offset experiments. *Geophysical Journal International*, 155:411–421.

Zernike, F. (1934). Diffraction theory of the cut procedure and its improved form, the phase contrast method (in German). *Physica*, 1:689–704.