Online Supplemental Material

A. Proofs

Proof of Theorem 3.2. For $f_i \in \mathcal{K}$ define

$$\nu_D^\top = (f_1(\theta_D).\dots, f_N(\theta_D)),$$

and the matrix $\mathbf{K} = \{\langle f_i, f_j\rangle_{\mathcal{K}}\}$; recall $C(f_i, f_j) = \langle f_i, f_j\rangle_{\mathcal{K}}$. Using Proposition 3 of Hall et al. (2013), we then need only to show that

$$(\nu_D - \nu_{D'})^\top \mathbf{K}^+ (\nu_D - \nu_{D'}) \leq \|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2,$$

where $+$ denotes the Moore-Penrose generalized inverse. We take a common strategy to such problems by showing that the left hand side can be expressed as $\|P(\theta_D - \theta_{D'})\|_{\mathcal{H}}^2$, where P is a projection operator. Recall that we can move between $\mathcal{K}$ and $\mathcal{H}$ via the transformation $h = CT_h$ for $h \in \mathcal{H}$ and $T_h \in \mathcal{K}$. Define the operator, $P_1 : \mathbb{B} \to \operatorname{span}\{f_1, \dots, f_N\} \subset \mathcal{K}$ as

$$P_1(x) = \sum_{i=1}^N f_i \sum_{j=1}^N (K^+)_{ij} f_j(x)$$

and its analog into $\mathcal{H}$, $P : \mathbb{B} \to \operatorname{span}\{C(f_1), \dots, C(f_N)\}$:

$$P(x) = C(P_1(x)) = \sum_{i=1}^N C(f_i) \sum_{j=1}^N (K^+)_{ij} f_j(x).$$

Notice that while P_1 maps elements of $\mathbb{B}$ to $\mathcal{K} \supset B^*$, P_2 maps elements of $\mathbb{B}$ into the Cameron-Martin space, $\mathcal{H} \subset \mathbb{B}$. By the reproducing property, there exists $T_{\theta_D - \theta_{D'}} \in \mathcal{K}$ such that

$$\begin{aligned}
&\langle T_{\theta_D - \theta_{D'}}, P_1(\theta_D - \theta_{D'})\rangle_{\mathcal{K}} \\
&= \sum_{i=1}^N \langle f_i, T_{\theta_D - \theta_{D'}}\rangle_{\mathcal{K}} \sum_{j=1}^N (K^+)_{ij} f_j(\theta_D - \theta_{D'}) \\
&= \sum_{i=1}^N f_i(\theta_D - \theta_{D'}) \sum_{j=1}^N (K^+)_{ij} f_j(\theta_D - \theta_{D'}) \\
&= (\nu_D - \nu_{D'})^\top \mathbf{K}^+ (\nu_D - \nu_{D'}).
\end{aligned}$$

Moving to $\mathcal{H}$ we have

$$\langle T_{\theta_D - \theta_{D'}}, P_1(\theta_D - \theta_{D'})\rangle_{\mathcal{K}} = \langle \theta_D - \theta_{D'}, P(\theta_D - \theta_{D'})\rangle_{\mathcal{H}}.$$

If we show that P is a projection operator over $\mathcal{H}$, i.e., symmetric and idempotent, we will have the desired bound.

First, P is **idempotent** by direct verification:

$$\begin{aligned}
P^2(x) &= \sum_{i=1}^N C(f_i) \sum_{j=1}^N (K^+)_{ij} f_j \left(\sum_{k=1}^N C(f_k) \sum_{l=1}^N (K^+)_{kl} f_l(x) \right) \\
&= \sum_{i=1}^N C(f_i) \sum_{j=1}^N (K^+)_{ij} \sum_{k=1}^N C(f_k, f_j) \sum_{l=1}^N (K^+)_{kl} f_l(x) \\
&= \sum_{i=1}^N C(f_i) \sum_{l=1}^N (K^+)_{il} f_l(x) = P(x).
\end{aligned}$$

Second, we show P is **symmetric** with respect to the $\mathcal{H}$ inner product by making repeated use of the reproducing property:

$$\begin{aligned}
\langle P(x), y\rangle_{\mathcal{H}} &= \langle P_1(x), T_y\rangle_{\mathcal{K}} \\
&= \sum_{i=1}^N f_i(y) \sum_{j=1}^N (K^{-1})_{ij} f_j(x) \\
&= \langle T_x, P_1(y)\rangle_{\mathcal{K}} = \langle x, P(y)\rangle_{\mathcal{H}}.
\end{aligned}$$

Hence P is a projection operator from $\mathbb{H}$ to $\mathbb{K}$, and the claim of the theorem holds. □

Proof of Theorem 3.3. We aim to show that for any measurable subset $A \subset \mathbb{B}$ we have

$$P_D(A) \leq e^\epsilon P_{D'}(A) + \delta,$$

where P_D denotes the measure of $\tilde{\theta}_D$, which is Gaussian with mean θ_D and covariance $\sigma^2 C$. Recall the global sensitivity for the functional case is

$$\Delta^2 = \sup_{D \sim D'} \|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2.$$

The density of $\tilde{\theta}_D$ wrt σZ is

$$\exp\left\{ -\frac{1}{2\sigma^2} (\|\theta_D\|_{\mathcal{H}}^2 - 2T_D(x)) \right\},$$

where for simplicity we denote $T_D = T_{\theta_D}$. We equivalently aim to show that

$$\begin{aligned}
P_D(A) &= \int_A dP_D(x) = \int_A \frac{dP_D}{dP_{D'}}(x) dP_{D'}(x) \\
&\leq e^\epsilon \int_A dP_{D'}(x) + \delta.
\end{aligned}$$

We can express

$$\begin{aligned}
&\frac{dP_D}{dP_{D'}}(x) = \frac{dP_D}{dQ}(x) / \frac{dP_{D'}}{dQ}(x) \\
&= \exp\left\{ -\frac{1}{2\sigma^2} (\|\theta_D\|_{\mathcal{H}}^2 - \|\theta_{D'}\|_{\mathcal{H}}^2 - 2(T_D - T_{D'})(x)) \right\}.
\end{aligned}$$

Expand

$$\begin{aligned}&\|\theta_{D'}\|_{\mathcal{H}}^2 = \|\theta_{D'} - \theta_D + \theta_D\|_{\mathcal{H}}^2\\ &= \|\theta_{D'} - \theta_D\|_{\mathcal{H}}^2 + \|\theta_D\|_{\mathcal{H}}^2 - 2\langle \theta_D - \theta_{D'}, \theta_D\rangle_{\mathcal{H}},\end{aligned}$$

and recall that we can write $\langle x, y\rangle_{\mathcal{H}} = T_x(y)$. So we have

$$\begin{aligned}&\frac{dP_D}{dP_{D'}}(x) =\\ &\exp\left\{-\frac{1}{2\sigma^2}(-\|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2 - 2(T_D - T_{D'})(x - \theta_D))\right\}.\end{aligned}$$

Decompose $\mathbb{B} = \mathcal{H}_1 \bigcup \mathcal{H}_2$ where for $x \in \mathcal{H}_1$ we have $\dfrac{dP_D}{dP_{D'}}(x) \leq e^\epsilon$ and for $x \in \mathcal{H}_2$ we have $\dfrac{dP_D}{dP_{D'}}(x) > e^\epsilon$. Then trivially we have that

$$P_D(A) = P_D(A \cap \mathcal{H}_1) + P_D(A \cap \mathcal{H}_2).$$

Using the definition of $\mathcal{H}_1$ we have that

$$\begin{aligned}&P_D(A \cap \mathcal{H}_1) = \int_{A\cap\mathcal{H}_1} \frac{dP_D}{dP_{D'}}(x)\frac{dP_{D'}}{dQ}(x)\, dQ(x)\\ &\leq e^\epsilon \int_{A\cap\mathcal{H}_1} \frac{dP_{D'}}{dQ}(x)\, dQ(x) \leq e^\epsilon P_{D'}(A).\end{aligned}$$

The proof will be complete if we can show that

$$P_D(A \cap \mathcal{H}_2) \leq \delta.$$

This is equivalent to showing that

$$\begin{aligned}P\Bigg(&-\frac{1}{2\sigma^2}(-\|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2\\ &- 2(T_D - T_{D'})(\tilde{\theta}_D - \theta_D)) \geq \epsilon\Bigg) \leq \delta.\end{aligned}$$

Recall that $\tilde{\theta}_D = \theta_D + \sigma Z$, where $Z \sim \mathcal{N}_{\mathbb{B}}(0, C)$. The event above can equivalently be stated as

$$\begin{aligned}&-\frac{1}{2\sigma^2}(-\|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2 - 2(T_D - T_{D'})(\tilde{\theta}_D - \theta_D)) \geq \epsilon\\ &\Leftrightarrow (T_D - T_{D'})(Z) \geq \sigma\left[\epsilon - \frac{1}{2\sigma^2}\|\theta_{D'} - \theta_D\|_{\mathcal{H}}^2\right].\end{aligned}$$

However $(T_D - T_{D'})(Z)$ is a (real) normal random variable with mean zero and variance $\|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2 \leq \Delta^2$. So, if $Y \sim \mathcal{N}(0, 1)$ then we have that

$$\begin{aligned}&P\left(-\frac{1}{2\sigma^2}(-\|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2 - 2(T_D - T_{D'})(\tilde{\theta}_D - \theta_D)) \geq \epsilon\right)\\ &\leq P\left(\Delta Y \geq \sigma\left[\epsilon - \frac{1}{2\sigma^2}\|\theta_D - \theta_{D'}\|_{\mathcal{H}}^2\right]\right)\\ &\leq P\left(Y \geq \frac{\sigma}{\Delta}\left[\epsilon - \frac{\Delta^2}{2\sigma^2}\right]\right)\\ &= P\left(Y \geq \sqrt{2\log(2/\delta)} - \frac{\epsilon}{2\sqrt{2\log(2/\delta)}}\right)\\ &\leq P\left(Y \geq \sqrt{2\log(2/\delta)} - \frac{1}{2\sqrt{2\log(2/\delta)}}\right) \leq \delta\end{aligned}$$

as long as $\epsilon \leq 1$ (Hall et al., 2013).

□

A.1. Derivation of RKHS Estimate

Recall that

$$g(m) = \frac{1}{N}\sum_{n=1}^{N} \|X_n - m\|_{\mathbb{H}}^2 + \phi\|m\|_\eta^2.$$

Without loss of generality, we may drop any terms not involving m and write

$$\begin{aligned}g(m) &= -2\langle \bar{X}, m\rangle_{\mathbb{H}} + \|m\|_{\mathbb{H}}^2 + \phi\|m\|_\eta^2\\ &= -2\langle \bar{X}, m\rangle_{\mathbb{H}} + \langle m, m\rangle_{\mathbb{H}} + \phi\langle m, C^{-\eta}m\rangle_{\mathbb{H}}.\end{aligned}$$

Since we are working with a Hilbert space, it can be identified with its own dual. We transfer everything over to the Cameron-Martin Space of C^η, call it $\mathcal{H}_\eta$, which contains $\mathcal{H}$:

$$\begin{aligned}g(m) &= -2\langle \bar{X}, C^\eta m\rangle_{\mathcal{H}}\\ &+ \langle m, C^\eta m\rangle_{\mathcal{H}} + \phi\langle m, m\rangle_{\mathcal{H}}.\end{aligned}$$

We then have that

$$g'(m) = -2C^\eta\bar{X} + 2C^\eta m + 2\phi m.$$

Setting the above equal to zero we have that

$$C^\eta\bar{X} = C^\eta\hat{\mu} + \phi\hat{\mu}. \tag{5}$$

or

$$\hat{\mu} = (C^\eta + \phi I)^{-1}C^\eta(\bar{X}).$$

Since (λ_j, v_j) are the eigenvalues/eigenfunctions of C and $X_i = \sum_{j=1}^\infty x_{ij}v_j$ then we have

$$\begin{aligned}\hat{\mu} &= \sum_{j=1}^\infty \langle\hat{\mu}, v_j\rangle_{\mathbb{H}} v_j = \sum_{j=1}^\infty \frac{\lambda_j^\eta}{\lambda_j^\eta + \phi}\langle\bar{X}, v_j\rangle_{\mathbb{H}} v_j\\ &= \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^\infty \frac{\lambda_j^\eta}{\lambda_j^\eta + \phi} x_{ij}v_j.\end{aligned}$$

Proof of Theorem 4.1. The upper bound for Δ_n^2 is derived as following:

$$\begin{aligned}\Delta_n^2 &= \sup_{D\sim D'} \|\hat{\mu}_D - \hat{\mu}_{D'}\|_{\mathcal{H}}^2\\ &= \sup_{D\sim D'} \|\frac{1}{N}\sum_{j=1}^\infty \frac{\lambda_j^\eta}{(\lambda_j^\eta + \phi)}(x_{1j} - x_{1'j})v_j\|_{\mathcal{H}}^2\\ &\leq \frac{1}{N^2}\sup_j \frac{\lambda_j^{2\eta-1}}{(\lambda_j^\eta + \phi)^2}\sup_{D\sim D'}\sum_{j=1}^\infty \langle X_1 - X_1', v_j\rangle_{\mathbb{H}}^2\\ &= \frac{1}{N^2}\sup_j \frac{\lambda_j^{2\eta-1}}{(\lambda_j^\eta + \phi)^2}\sup_{D\sim D'}\|X_1 - X_1'\|_{\mathbb{H}}^2\\ &\leq \frac{4\tau^2}{N^2}\sup_j \frac{\lambda_j^{2\eta-1}}{(\lambda_j^\eta + \phi)^2}.\end{aligned}$$

We can also derive a simpler bound by examining the function

$$f(x) = \frac{x^{2\eta-1}}{(x^\eta + \phi)^2}, \qquad x \geq 0,$$

and where it attains its maximum. Taking the derivative we have $f'(x) = 0$ if and only if

$$\begin{aligned}
&(x^\eta + \phi)^2(2\eta - 1)x^{2\eta-2} - x^{2\eta-1}2\eta x^{\eta-1}(x^\eta + \phi) = 0 \\
&(x^\eta + \phi)(2\eta - 1) - 2\eta x^\eta = 0 \\
&x = (\phi(2\eta - 1))^{1/\eta}.
\end{aligned}$$

Taking a second derivative shows that this is where the maximum occurs. We then have that

$$f(x) \leq \frac{(\phi(2\eta-1))^{2-1/\eta}}{(\phi(2\eta-1)+\phi)^2} = \phi^{-1/\eta}\frac{(2\eta-1)^{2-1/\eta}}{4\eta^2}$$

Thus, we can also use the bound

$$\frac{4\tau^2}{N^2}\sup_j \frac{\lambda_j^{2\eta-1}}{(\lambda_j^\eta + \phi)^2} \leq \frac{\tau^2}{N^2\phi^{1/\eta}}\frac{(2\eta-1)^{2-1/\eta}}{\eta^2}.$$

For $\eta = 1$, the bound becomes $\tau^2 N^{-2}\phi^{-1}$, while another calculus argument shows that regardless of η, one will always have

$$\frac{\tau^2}{N^2\phi^{1/\eta}}\frac{(2\eta-1)^{2-1/\eta}}{\eta^2} \leq \frac{4\tau^2}{N^2\phi^{1/\eta}},$$

as desired.

□

B. Extension of Empirical Study

In this section we review the impact of different parameters on the utility of sanitized releases introduced in Section 5. For the RKHS, $\mathcal{H}$, we would consider four popular kernels:

$$\begin{aligned}
C_1(t,s) &= \exp\left\{\frac{-|t-s|^2}{\rho}\right\} \qquad (6)\\
C_2(t,s) &= \left(1 + \frac{\sqrt{5}|t-s|}{\rho} + \frac{5(t-s)^2}{3\rho^2}\right)\exp\left\{\frac{-\sqrt{5}|t-s|}{\rho}\right\}\\
C_3(t,s) &= \left(1 + \frac{\sqrt{3}|t-s|}{\rho}\right)\exp\left\{\frac{-\sqrt{3}|t-s|}{\rho}\right\}\\
C_4(t,s) &= \exp\left\{\frac{-|t-s|}{\rho}\right\}.
\end{aligned}$$

the first is also known as the Gaussian or squared exponential kernel and the last is also known as the exponential, Laplacian, or Ornstein-Uhlenbeck kernel.

Recall the all parameters discussed in Section 5 will be fixed in all scenarios, except for the one where they are explicitly varied to consider their effect.

The scenario 1 was discussed in Section 5.

SCENARIO 2: VARYING KERNEL RANGE PARAMETER ρ

Here all defaults are used except the range parameter for the noise and RKHS (which are taken to be the same in all settings) that ranges from 0.002 to 2. The results are presented in Figure 4. We see very similar patterns to Scenario 1, where increasing ρ increases the smoothing of both the estimate and its privacy enhanced version. However, increasing ρ smooths more than it shrinks and there is still a non-negligible difference between the two estimates for larger values, (e.g., $\rho = 0.2$). Practically, both ρ and ϕ should be chosen together for the best performance, which we will explore further in Section 6.

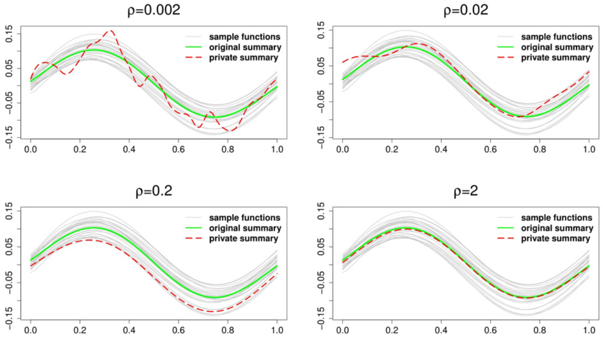

Figure 4. Original and private RKHS smoothing mean with Gaussian Kernel (C_1) for different values of kernel range parameter ρ

SCENARIO 3: VARYING THE KERNEL FUNCTION $c(t, s)$

Here we consider the four different kernels given in (6) for both the noise and RKHS kernel (which are taken to be the same). The results are summarized in Figure 5. All kernels give roughly the same pattern, however, C_1 produces curves which are infinitely differentiable, while the exponential kernel produces curves that are nowhere differentiable (they follow an Ornstein-Uhlenbeck process). The two Matérn covariances give paths that have either one (C_3) or two (C_2) derivatives. Since the underlying function to be estimated is already very smooth, the kernel does not have a substantial impact. However, for more irregular shapes, this choice can play a substantial role on the efficiency of the resulting RKHS estimate.

SCENARIO 4: VARYING THE SMOOTHING PARAMETER OF SAMPLES p

In this setting we vary p from 1.1 to 4, which determines the smoothness of the data, $X_n(t)$. Note that p has to be strictly larger than 1 or the X_i will not be square integrable. Figure 6 summarizes these results. As we can see, the smoothness of

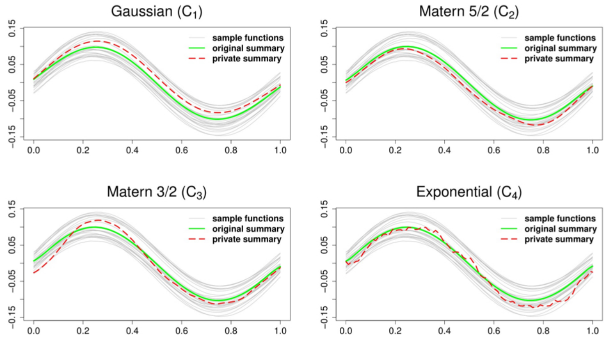

Figure 5. Original and private RKHS smoothing mean for different kernels

the curves has a smaller impact on the utility of the sanitized estimates as compared to other parameters. As the curves become smoother, the global sensitivity decreases implying the need for less noise being added in order to maintain the desired privacy level, and thus resulting in a higher utility for the privacy enhanced curves. However, the smoothness, in terms of derivatives, of the estimates is not affected, as this is determined by the kernel.

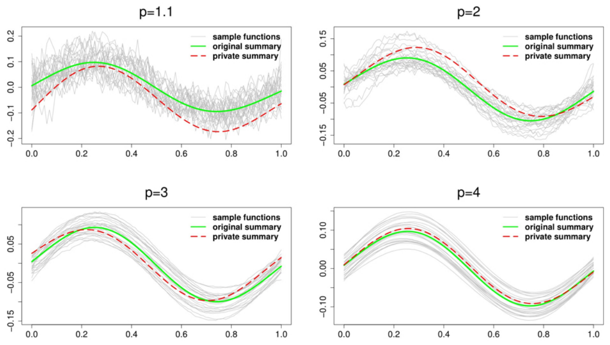

Figure 6. Original and private RKHS smoothing mean with Gaussian Kernel (C_1) for different values of smoothing parameter of samples p

Scenario 5: Varying the privacy parameters (ϵ, δ)

In this setting we vary the privacy parameters, ϵ and δ. Figure 7 present the effects of varying ϵ from 5 to 0.1 while in Figure 8 we vary δ from 0.1 to 10^{-6}. As we decrease the parameters, we are requiring a stricter form of privacy, which is reflected in the plots; recall that $\delta = 0$ will give us the stricter form of DP, ϵ-DP (also called ϵ-DP). As we decrease these values, the overall noise added increases, and we expect larger deviations of the sanitized estimates from the mean. There is less sensitivity in the output to changes in δ than to ϵ. However, as with the previous scenario these parameters play no role in the overall smoothness, in terms of derivatives of the resulting estimates.

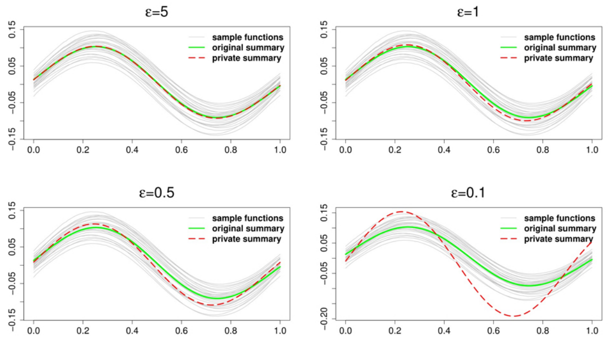

Figure 7. Original and private RKHS smoothing mean with Gaussian Kernel (C_1) for different values of privacy level parameter ϵ when $\delta = 1$

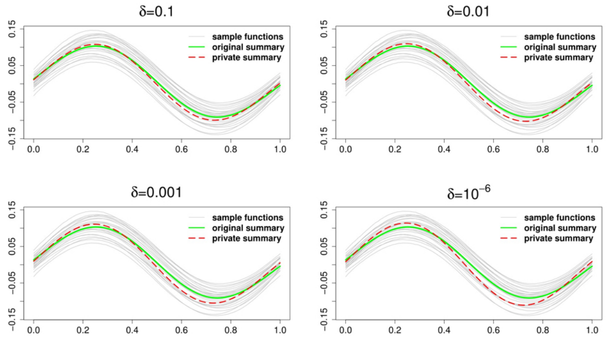

Figure 8. Original and private RKHS smoothing mean with Gaussian Kernel (C_1) for different values of privacy level parameter δ when $\epsilon = 1$

Scenario 6: Varying sample size N

In Figure 9 we vary the sample size from 5 to 100. The results are very similar to changing δ and ϵ, as the sample size does not influence the smoothness of the curves (in terms of derivatives), but the accuracy of the estimate (green curve) gets much better and so does the utility of the privacy enhanced version.

Scenario 7: Different underlying mean function μ

Lastly, in Figure 10 we consider three additional mean functions. Overall, the actual function being estimated does not influence the utility of the privacy enhanced version, only

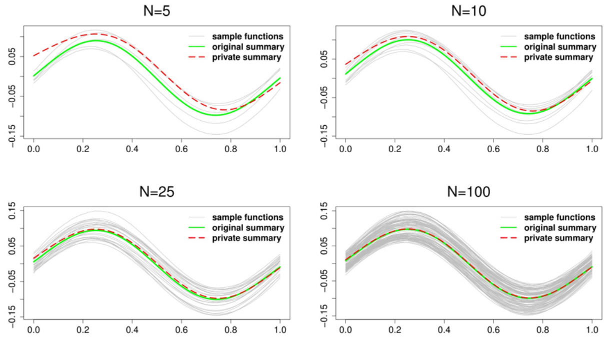

Figure 9. Original and private RKHS smoothing mean with Gaussian Kernel (C_1) for different sample sizes N

the accuracy of the original estimate. This is because the noise to be added is computed from the different smoothing parameters as well as the range of the L^2 norm of the data, not the underlying estimate itself.

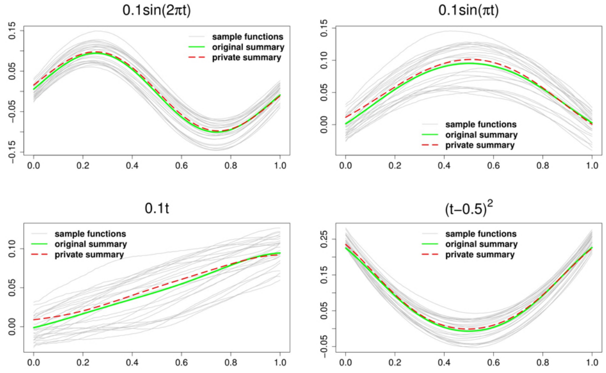

Figure 10. Original and private RKHS smoothing mean with Gaussian Kernel (C_1) for different initial mean functions μ

C. Extension of Diffusion Tensor Imaging

In this section our aim is to see the privacy enhanced RKHS smoothing estimate of the mean function discussed in Section 6 for C_3 and C_4 based on the optimal parameters in Table 1 for CV. The results are given in Figures 11 and 12 for Matern and Exponential kernels, respectively. In each case, we see that the utility of the privacy enhanced versions increases as ϕ increases, however, the largest values of ϕ produce estimates that are over smoothed. Here Table 2 represents the optimal parameters to generate privacy enhanced estimates for PCV.

		Exp. Kernel	Mat $3/2$ Kernel	Gau. Kernel
	ϕ	optimum ρ	optimum ρ	optimum ρ
1	0.0001	0.25	0.10	0.01
2	0.001	0.20	0.15	0.01
3	0.01	0.30	0.15	0.03
4	0.03	0.80	0.30	0.05

Table 1. Optimum kernel range parameters ρ for different kernels with using CV for each fixed penalty parameter ϕ in DTI dataset

Kernel	range ϕ	range ρ	optimum ϕ	optimum ρ
C_1	$[10^{-4}, 0.1]$	$[0.01, 0.1]$	0.005	0.030
C_3	$[10^{-4}, 0.1]$	$[0.05, 0.5]$	0.005	0.250
C_4	$[10^{-4}, 0.1]$	$[0.2, 1]$	0.010	0.466

Table 2. Optimum penalty and range parameters (ϕ, ρ) for different kernels with PCV in CCA application.

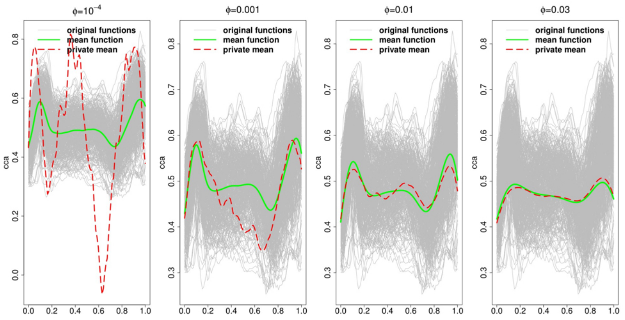

Figure 11. Mean estimate for CCA and its private release using Matérn$(3/2)$ kernel (C_3) with CV.

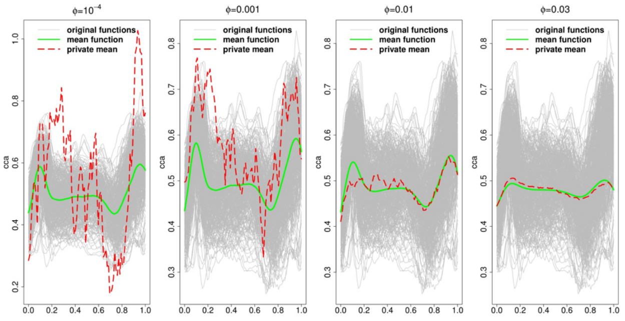

Figure 12. Mean estimate for cca and its private release using Exponential kernel (C_4) with CV.