Supplementary Material

A Large Dimensional Analysis of Least Squares Support Vector Machines

APPENDIX A

PROOF OF THEOREM 1

Our key interest here is on the decision function of LS-SVM: \( g(x) = \alpha^T k(x) + b \) with \((\alpha, b)\) given by

\[
\begin{align*}
\alpha &= S^{-1} \left( I_n - \frac{1}{n} S^{-1} I_n \right) y \\
b &= \frac{1}{n} S^{-1} y
\end{align*}
\]

and \( S^{-1} = \left( K + \frac{\alpha}{n} I_n \right)^{-1} \).

Before going into the detailed proof, as we will frequently deal with random variables evolving as \( n, p \) grow large, we shall use the extension of the \( O(\cdot) \) notation introduced in [20]: for a random variable \( x \equiv x_n \) and \( u_n \geq 0 \), we write \( x = O(u_n) \) if for any \( \eta > 0 \) and \( D > 0 \), we have \( n^D P(x \geq n^\eta u_n) \rightarrow 0 \). Note that under Assumption 1 it is equivalent to use either \( O(u_n) \) or \( O(u_p) \) since \( n, p \) scales linearly. In the following we shall use constantly \( O(u_n) \) for simplicity.

When multidimensional objects are concerned, \( v = O(u_n) \) means the maximum entry of a vector (or a diagonal matrix) \( v \) in absolute value is of order \( O(u_n) \) and \( M = O(u_n) \) means that the operator norm of \( M \) is of order \( O(u_n) \). We refer the reader to [20] for more discussions on these practical definitions.

Under the growth rate settings of Assumption 1 from [20], the approximation of the kernel matrix \( K \) is given by

\[
K = -2 f'(\tau) \left( P \Omega^T \Omega P + A \right) + \beta I_n + O(n^{-\frac{3}{2}}) \tag{12}
\]

with \( \beta = f(0) - f(\tau) + \tau f'(\tau) \) and \( A = A_n + A_{\sqrt{\pi}} + A_1 \), \( A_n = -\frac{f'(\tau)}{2 \gamma^2} I_n I_n^T \) and \( A_{\sqrt{\pi}} A_1 \) given by [18] and [19] at the top of next page, where we denote

\[
t_a \triangleq \frac{\text{tr}(C_n - C_0^2)}{\sqrt{\beta}},
(\psi)^2 \triangleq [(\psi_1)^2, \ldots, (\psi_n)^2]^T.
\]

We start with the term \( S^{-1} \). The terms of leading order in \( K \), i.e., \(-2 f'(\tau) A_n \) and \( \frac{\alpha}{n} I_n \), are both of operator norm \( O(n) \). Therefore a Taylor expansion can be performed as

\[
S^{-1} = \left( K + \frac{n}{\gamma} I_n \right)^{-1} = \frac{1}{n} \left[ L^{-1} - \frac{2 f'(\tau)}{n} \right]
\]

\[
= \left( A_{\sqrt{\pi}} A_1 + P \Omega^T \Omega P \right) + \frac{\beta I_n}{n} + O(n^{-\frac{3}{2}})
+ \frac{L}{n} + \frac{2 f'(\tau)}{n^2} \left( L + \left( Q - \frac{\beta}{n^2} I_n \right) L + O(n^{-\frac{3}{2}}) \right)
\]

with \( L = \left( f(\tau) I_n I_n^T + I_n \right)^{-1} \) of order \( O(1) \) and \( Q = \frac{2 f'(\tau)}{n^2} \left( A_1 + P \Omega^T \Omega P \right) \).

With the Sherman-Morrison formula we are able to compute explicitly \( L \) as

\[
L = \left( f(\tau) \frac{1}{n} I_n I_n^T + \frac{1}{\gamma} I_n \right)^{-1} = \gamma \left( I_n - \frac{f(\tau)}{1 + \gamma f(\tau)} \frac{1}{n} I_n \right)
= \frac{1}{1 + \gamma f(\tau)} I_n + \frac{2 f'(\tau)}{1 + \gamma f(\tau)} P = O(1). \tag{13}
\]

Writing \( L \) as a linear combination of \( I_n \) and \( P \) is useful when computing \( S^{-1} \), or \( S^{-1} I_n \), because by the definition of \( P = I_n - \frac{1}{n} I_n \), we have \( I_n P = P I_n = 0 \).

We shall start with the term \( 1_n^T S^{-1} \), since it is the basis of several other terms appearing in \( \alpha \) and \( b \).

\[
1_n^T S^{-1} = \frac{\gamma}{1 + \gamma f(\tau)} \left[ I_n + \frac{2 f'(\tau)}{n^2} A_{\sqrt{\pi}} L + \left( Q - \frac{\beta}{n^2} I_n \right) L \right]
+ O(n^{-\frac{3}{2}})
\]

since \( 1_n^T L = \frac{\gamma}{1 + \gamma f(\tau)} 1_n^T \).

With \( 1_n^T S^{-1} \) at hand, we next obtain,

\[
1_n^T S^{-1} y = \frac{\gamma}{1 + \gamma f(\tau)} \left[ c_2 - c_1 + \frac{2 f'(\tau)}{n^2} 1_n^T A_{\sqrt{\pi}} L y + \left( Q - \frac{\beta}{n^2} I_n \right) L y \right] + O(n^{-\frac{3}{2}})
+ \frac{1}{n} \left( Q - \frac{\beta}{n^2} I_n \right) L y \right] + O(n^{-\frac{3}{2}}) \tag{14}
\]

\[
1_n^T S^{-1} l_n = \frac{\gamma}{1 + \gamma f(\tau)} \left[ \frac{1}{n} + \frac{2 f'(\tau)}{n^2} \frac{1}{1 + \gamma f(\tau)} \right] + O(n^{-\frac{3}{2}})
+ \frac{\gamma}{1 + \gamma f(\tau)} \frac{1}{n} \left( Q - \frac{\beta}{n^2} I_n \right) L y \right] + O(n^{-\frac{3}{2}}). \tag{15}
\]

The inverse of \( 1_n^T S^{-1} l_n \) can consequently be computed using a Taylor expansion around its leading order, allowing an error term of \( O(n^{-\frac{3}{2}}) \) as

\[
1_n^T S^{-1} l_n = \frac{1}{n} + \frac{\gamma f(\tau)}{n} \left[ \frac{1}{n} - \frac{2 f'(\tau)}{n^2} \frac{1}{1 + \gamma f(\tau)} \right]
+ \frac{\gamma}{1 + \gamma f(\tau)} \frac{1}{n} \left( Q - \frac{\beta}{n^2} I_n \right) L y \right] + O(n^{-\frac{3}{2}}). \tag{15}
\]
\[
A_{\sqrt{n}} = \frac{-1}{2} \left[ \psi \mathbf{1}_n^T + \mathbf{1}_n \psi^T + \left\{ t_a \frac{\mathbf{1}_{n_a}}{\sqrt{p}} \right\}_{a=1}^2 \mathbf{1}_n^T + \mathbf{1}_n \left\{ t_b \frac{1}{\sqrt{p}} \right\}_{b=1}^2 \right]^{n-1} \]

\[
A_1 = \frac{-1}{2} \left\{ \| \mu_a - \mu_b \|^2 \frac{\mathbf{1}_{n_a}}{p} \right\}_{a,b=1}^2 + 2 \left\{ \frac{\Omega \mathbf{P}_{a}}{\sqrt{p}} (\mu_a - \mu_b) \frac{\mathbf{1}_{n_b}}{p} \right\}_{b=1}^2 - 2 \left\{ \frac{\mathbf{1}_{n_a} (\mu_a - \mu_b)^T \Omega \mathbf{P}_{b}}{\sqrt{p}} \right\}_{a,b=1}^2 \]

\[
- \frac{f''(\tau)}{4f'(\tau)} \left[ (\psi^T) + \mathbf{1}_n \frac{(\psi^T)^2}{n} + \left\{ t_a \frac{\mathbf{1}_{n_a}}{p} \right\}_{a=1}^2 \mathbf{1}_n^T + \mathbf{1}_n \left\{ t_b \frac{1}{\sqrt{p}} \right\}_{b=1}^2 + 2 \left\{ t_a t_b - \frac{\mathbf{1}_{n_a} \mathbf{1}_{n_b}}{p} \right\}_{a,b=1}^2 + 2D \left\{ \left( t_a \mathbf{1}_{n_a} \right)^2 - 1 \right\}_{a=1}^2 \psi \frac{\mathbf{1}_n^T}{\sqrt{p}} \right]^{n-1} \]

\[
\mathbf{k}(x) = f'(\tau) \left\{ \frac{\| \mu_a - \mu_b \|^2}{p} \right\}_{b=1}^2 - \frac{2}{\sqrt{p}} \left\{ \left( \left( t_a + t_b \right)^2 - 1 \right) \right\}_{b=1}^2 + 2D \left\{ \left( t_a + t_b \right)^2 - 1 \right\}_{b=1}^2 \psi \frac{\mathbf{1}_n^T}{\sqrt{p}} \right]^{n-1} \]

\[
= \frac{4}{p^2} \text{tr}(C_a C_b) \mathbf{1}_n^2 \left[ Q - \frac{\beta}{n} \mathbf{1}_n \right] \]

\[
\text{At this point, for } \alpha = S^{-1} \left( I_n - \frac{1}{n} \mathbf{1}_n^T S^{-1} \right) y, \text{ we have}
\]

\[
\alpha = S^{-1} \left[ I_n - \frac{2\gamma f'(\tau)}{n^2} \mathbf{1}_n \mathbf{1}_n^T A_{\sqrt{n}} \right] y + O(n^{-\frac{3}{2}}).
\]

Here again, we use \( \mathbf{1}_n^T \mathbf{L} = \frac{1}{n+\gamma f(\tau)} \mathbf{1}_n^T \) and \( \mathbf{L} - \frac{1}{n+\gamma f(\tau)} \mathbf{1}_n^T = \gamma \mathbf{P} \), to eventually get

\[
\alpha = \frac{\gamma}{n} \frac{\mathbf{L} \mathbf{P}}{\gamma} + \frac{\gamma^2}{n} \left( Q - \frac{\beta}{n} \mathbf{1}_n \right) \mathbf{P} \mathbf{y} + O(n^{-\frac{3}{2}}).
\]

\[
\text{Note here the absence of a term of order } O(n^{-3/2}) \text{ in the expression of } \alpha \text{ since } \mathbf{P} \mathbf{A}_{\sqrt{n}} \mathbf{P} = 0 \text{ from (13).}
\]

We shall now work on the vector \( \mathbf{k}(x) \) for a new datum \( x \), following the same analysis as in (20) for the kernel matrix \( \mathbf{K} \), assuming that \( x \sim \mathcal{N}(\mu_x, \mathbf{C}_x) \) and recalling the random variables definitions,

\[
\omega_x \triangleq (x - \mu_x) / \sqrt{p}, \quad \psi_x \triangleq ||\omega_x||^2 - E[\omega_x||^2]
\]
we show that the \( j \)-th entry of \( k(x) \) can be written as

\[
[k(x)]_j = f(\tau) + f'(\tau) \left[ \frac{t_o + t_b}{\sqrt{p}} + \psi_x + \psi_j - 2(\omega_x)^T \omega_j \right] + \frac{\| \mu_b - \mu_a \|^2}{p} + \frac{2}{\sqrt{p}} (\mu_b - \mu_a)^T (\omega_j - \omega_x) + f''(\tau) \left[ \left( \frac{t_o + t_b}{\sqrt{p}} + \psi_j + \psi_x \right)^2 + \frac{4}{p^2} \text{tr} C_a C_b \right] + O(n^{-\frac{3}{2}}).
\]

Combining (21) and (22), we deduce

\[
\alpha^T k(x) = \frac{2\gamma}{\sqrt{n}} c_1 c_2 f''(\tau) (t_2 - t_1) + \frac{\gamma}{n} Y^T P \tilde{k}(x) + \frac{\gamma f''(\tau)}{n} Y^T P (\psi - 2P \Omega^T \omega_x) + O(n^{-\frac{3}{2}})
\]

with \( \tilde{k}(x) \) given in (20).

At this point, note that the term of order \( O(n^{-\frac{3}{2}}) \) in the final object \( g(x) = \alpha^T k(x) + b \) disappears because in both (17) and (23) the term of order \( O(n^{-\frac{3}{2}}) \) is \( \frac{2\gamma}{\sqrt{n}} c_1 c_2 f''(\tau) (t_2 - t_1) \) but of opposite signs. Also, we see that the leading term \( c_2 - c_1 \) in \( b \) will remain in \( g(x) \) as stated in Remark 2.

The development of \( y^T P k(x) \) induces many simplifications, since i) \( P \mathbb{1}_n = 0 \) and ii) random variables as \( \omega_x \) and \( \psi \) in \( k(x) \), once multiplied by \( y^T P \), thanks to probabilistic averaging of independent zero-mean terms, are of smaller order and thus become negligible. We thus get

\[
\frac{\gamma}{n} y^T P \tilde{k}(x) = 2\gamma c_1 c_2 f''(\tau) \left[ \frac{\| \mu_x - \mu_a \|^2 - \| \mu_1 - \mu_a \|^2}{p} \right] - 2(\omega_x)^T \frac{\mu_a - \mu_1}{\sqrt{p}} \left[ \frac{t_o}{\sqrt{p}} + \psi_x \right] + \frac{f''(\tau)}{2n} y^T P (\psi - 2P \Omega^T \omega_x)^2 + \gamma c_1 c_2 f''(\tau) \left[ \left( \frac{t_o}{\sqrt{p}} + \psi_x \right)^2 + \frac{2(t_2 - t_1)}{\sqrt{p}} + \frac{t_2^2 - t_1^2}{p} + \frac{4}{p^2} \text{tr} C_a C_2 - C_a C_1 \right] + O(n^{-\frac{3}{2}}).
\]

This result, together with (23), completes the analysis of the term \( \alpha^T k(x) \). Combining (23)–(24) with (17) we conclude the proof of Theorem 1.

## APPENDIX B

### PROOF OF THEOREM 2

This section is dedicated to the proof of the central limit theorem for

\[\hat{g}(x) = c_2 - c_1 + \gamma (\mathbb{I} + c_x \mathcal{D})\]

with the shortcut \( c_x = -2c_1 c_2 \) for \( x \in C_1 \) and \( c_x = 2c_1^2 c_2 \) for \( x \in C_2 \), and \( \mathbb{I}, \mathcal{D} \) as defined in 7 and 9.

Our objective is to show that for \( \alpha \in \{1, 2\} \), \( n(\hat{g}(X) - G_a) \Rightarrow 0 \) with

\[G_a \sim \mathcal{N}(E_a, \text{Var}_a)\]

where \( E_a \) and \( \text{Var}_a \) are given in Theorem 2. We recall that \( x = \mu_a + \sqrt{\mu} \omega_x \) with \( \omega_x \sim \mathcal{N}(0, \mathcal{C}_a/p) \).

Letting \( z_x \) such that \( \omega_x = \mathcal{C}^{1/2}_a z_x / \sqrt{p} \), we have \( z_x \sim \mathcal{N}(0, \mathcal{I}_n) \) and we can rewrite \( \hat{g}(x) \) in the following quadratic form (of \( z_x \) as)

\[\hat{g}(x) = z_x^T A z_x + z_x^T b + c\]

with

\[A = 2\gamma c_1 c_2 f''(\tau) \frac{\text{tr} (C_2 - C_1)}{p} \mathcal{C}_a,\]

\[b = -2\gamma f''(\tau) \left( \mathcal{C}_a \right)^2 \frac{\text{tr} (C_2 - C_1)}{\sqrt{p}} \mathcal{C}_a - 4c_1 c_2 \gamma f''(\tau) \left( \mathcal{C}_a \right)^2 \frac{(\mu_2 - \mu_1)}{p} \mathcal{C}_a,\]

\[c = c_2 - c_1 + \gamma c_x \mathcal{D} - 2\gamma c_1 c_2 f''(\tau) \frac{\text{tr} (C_2 - C_1)}{p} \mathcal{C}_a.\]

Since \( z_x \) is (standard) Gaussian and has the same distribution as \( U z_x \) for any orthogonal matrix \( U \) (i.e., such that \( U^T U = U U^T = \mathcal{I}_n \)), we choose \( U \) that diagonalize \( A \) such that \( A = U A U^T \), with \( A \) diagonal so that \( \hat{g}(x) \) and \( \tilde{g}(x) \) have the same distribution where

\[\hat{g}(x) = z_x^T A z_x + z_x^T b + c = \sum_{i=1}^{n} \left( z_i^2 \lambda_i + z_i \tilde{b}_i + \frac{c}{n} \right)\]

and \( \tilde{b} = U^T b \), \( \lambda_i \) the diagonal elements of \( A \) and \( z_i \) the elements of \( z_x \).

Conditioning on \( \Omega \), we thus result in the sum of independent but not identically distributed random variables

\[r_i = z_i^2 \lambda_i + z_i \tilde{b}_i + \frac{c}{n}\]

We then resort to the Lyapunov CLT [3, Theorem 27.3].

We begin by estimating the expectation and the variance

\[
\text{E}[r_i | \Omega] = \lambda_i + \frac{c}{n}
\]

\[
\text{Var}[r_i | \Omega] = \sigma_i^2 = 2\lambda_i^2 + \tilde{b}_i^2
\]

of \( r_i \), so that

\[
\sum_{i=1}^{n} \text{E}[r_i | \Omega] = c_2 - c_1 + \gamma c_x \mathcal{D} = E_a
\]

\[
s_n^2 = \sum_{i=1}^{n} \sigma_i^2 = 2 \text{tr}(A^2) + b^T b
\]

\[
= 8\gamma^2 c_1 c_2 \left( f''(\tau) \right)^2 \left( \frac{\text{tr} (C_2 - C_1)}{p} \right)^2 \frac{\text{tr} \mathcal{C}_a^2}{p^2}
\]

\[
+ 4\gamma^2 \left( \frac{f''(\tau)}{n} \right)^2 \mathcal{Y}^T \mathcal{P} \mathcal{Q}^T \frac{\mathcal{C}_a}{p} \Omega \mathcal{P} \mathcal{Y}
\]

\[
+ \frac{16\gamma^2 c_1^2 c_2^2 (f''(\tau))^2}{n} (\mu_2 - \mu_1)^T \frac{\mathcal{C}_a}{p} (\mu_2 - \mu_1)
\]

\[+ O(n^{-\frac{5}{2}})\]

We shall rewrite \( \Omega \) into two blocks as:

\[
\Omega = \left[ \begin{array}{c}
\mathcal{C}_a \frac{1}{\sqrt{p}} \mathcal{Z}_1, \\
\frac{(\mathcal{C}_a)^2}{\sqrt{p}} \mathcal{Z}_2
\end{array} \right]
\]

where \( \mathcal{Z}_1 \in \mathbb{R}^{p \times n_1} \) and \( \mathcal{Z}_2 \in \mathbb{R}^{p \times n_2} \) with i.i.d. Gaussian entries with zero mean and unit variance. Then

\[
\frac{\mathcal{C}_a}{p} \mathcal{Y}^T \mathcal{C}_a = \frac{1}{p^2} \left[ \mathcal{Z}_1^T (\mathcal{C}_a^2)^{1/2} \mathcal{C}_a (\mathcal{C}_a^2)^{1/2} \mathcal{Z}_1 + \mathcal{Z}_2^T (\mathcal{C}_a^2)^{1/2} \mathcal{C}_a (\mathcal{C}_a^2)^{1/2} \mathcal{Z}_2 \right]
\]
and with $P_y = y - (c_2 - c_1)1_n$, we deduce

$$y^\top P \Omega^\top \frac{C_a}{p} \Omega P y = \frac{4}{p^2} \left( c_1^2 \mathbb{E}_1^T (C_1)^\frac{1}{2} C_a (C_1)^\frac{1}{2} Z_1 n_1 - 2c_1c_2 \mathbb{E}_1^T (C_1)^\frac{1}{2} C_a (C_2)^\frac{1}{2} Z_1 n_2 + c_1^2 \mathbb{E}_1^T (C_2)^\frac{1}{2} C_a (C_2)^\frac{1}{2} Z_2 1_{n_2} \right).$$

Since $Z_i 1_{n_i} \sim N(0, n_i 1_{n_i})$, by applying the trace lemma [1, Lemma B.26] we get

$$y^\top P \Omega^\top \frac{C_a}{p} \Omega P y = \frac{4nc_1^2 c_2^2}{p^2} \left( \frac{\text{tr} C_1 C_a}{c_1} + \frac{\text{tr} C_2 C_a}{c_2} \right) \xrightarrow{a.s.} 0.$$  \hspace{1cm} (25)

Consider now the events

$$E = \left\{ \left| y^\top P \Omega^\top \frac{C_a}{p} \Omega P y - \rho \right| < \epsilon \right\}$$

$$\bar{E} = \left\{ \left| y^\top P \Omega^\top \frac{C_a}{p} \Omega P y - \rho \right| \geq \epsilon \right\}$$

for any fixed $\epsilon$ with $\rho = \frac{4nc_1^2 c_2^2}{p^2} \left( \frac{\text{tr} C_1 C_a}{c_1} + \frac{\text{tr} C_2 C_a}{c_2} \right)$ and write

$$\mathbb{E} \left[ \exp \left( iun \frac{\tilde{g}(x) - E_a}{s_n} \right) \right] = \mathbb{E} \left[ \exp \left( iun \frac{\tilde{g}(x) - E_a}{s_n} \right) \right] P(E) + \mathbb{E} \left[ \exp \left( iun \frac{\tilde{g}(x) - E_a}{s_n} \right) \right] P(\bar{E}) \hspace{1cm} (26)$$

We start with the variable $\tilde{g}(x)|E$ and check that Lyapunov’s condition for $\bar{r}_i = r_i - \mathbb{E}[r_i]$, conditioning on $E$,

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{i=1}^{n} \mathbb{E}[|\bar{r}_i|^4] = 0$$

holds by rewriting

$$\lim_{n \to \infty} \frac{1}{s_n^4} \sum_{i=1}^{n} \mathbb{E}[|\bar{r}_i|^4] = \lim_{n \to \infty} \frac{1}{s_n^4} \sum_{i=1}^{n} 60\lambda_i^4 + 12\lambda_i^2 \beta_i^2 + 3\beta_i^4 = 0$$

since both $\lambda_i$ and $\beta_i$ are of order $O(n^{-3/2})$.

As a consequence of the above, we have the CLT for the random variable $\tilde{g}(x)|E$, thus

$$\mathbb{E} \left[ \exp \left( iun \frac{\tilde{g}(x) - E_a}{s_n} \right) \right] \to \exp(-\frac{u^2}{2}).$$

Next, we see that the second term in (26) goes to zero because $|\mathbb{E}[\exp \left( iun \frac{\tilde{g}(x) - E_a}{s_n} \right) |E]| \leq 1$ and $P(\bar{E}) \to 0$ from (25) and we eventually deduce

$$\mathbb{E} \left[ \exp \left( iun \frac{\tilde{g}(x) - E_a}{s_n} \right) \right] \to \exp(-\frac{u^2}{2}).$$

With the help of Lévy’s continuity theorem, we thus prove the CLT of the variable $n \frac{\tilde{g}(x) - E_a}{s_n}$, since $s_n^2 \to \text{Var}_a$, with Slutsky’s theorem, we have the CLT for $\frac{\tilde{g}(x) - E_a}{\sqrt{\text{Var}_a}}$ (thus for $n \frac{\tilde{g}(x) - E_a}{\sqrt{\text{Var}_a}}$), and eventually for $n \frac{\tilde{g}(x) - E_a}{\sqrt{\text{Var}_a}}$ by Theorem 1, which completes the proof.